

[4]

TESTS FOR DETECTING A SHIFT IN THE MEAN OF HYDROLOGICAL TIME SERIES

T.A. BUISHAND

Royal Netherlands Meteorological Institute (K.N.M.I.), 3730 AE De Bilt (The Netherlands)

(Received October 5, 1983; accepted for publication January 3, 1984)

ABSTRACT

Buishand, T.A., 1984. Tests for detecting a shift in the mean of hydrological time series. *J. Hydrol.*, 73: 51–69.

A practical problem in time-series analysis of hydrological and meteorological data is to find statistical techniques for testing for an abrupt change in the mean at an unknown time. Suitable techniques have been developed in the situation of a single time series, $\{y_i\}$. Attention is paid to the likelihood ratio statistic V and to a Bayesian statistic U .

Complications arise if we want to test for a shift in the mean using a regression on a second correlated sequence, $\{x_i\}$, because the critical values of the test statistic in general depend on the configuration of the x_i 's. For the statistic U , this problem can be solved using techniques similar to those for testing for serial correlation in least-squares regression. A so-called bounds test can be performed on the least-squares residuals. Unfortunately, there is quite a large possibility that this test is inconclusive. As an alternative to the bounds test, the statistic U can be applied to transformed residuals. The tests are illustrated with runoff and precipitation data for the Colorado River Basin, U.S.A., and the River Thames at Teddington, U.K.

INTRODUCTION

A practical problem in analysing time series of meteorological and hydrological data is that such time series are not always homogeneous. Sometimes there are abrupt changes in the mean, for example, due to a station relocation or the use of another measuring device. Therefore, several techniques have been developed for testing homogeneity.

The technique to be used depends on prior knowledge about systematic changes. For instance, traditional analysis of variance techniques can be applied, if it is known or suspected when the changes occurred. On the other hand, there is no well-accepted statistical method for testing shifts in the mean occurring at unknown times. In fact, the simple case that there is only, at most, one unknown change-point has not yet been solved satisfactorily.

In the literature (Hawkins, 1977; Worsley, 1979; Buishand, 1982),

attention has been paid to the following model with a single shift in the mean:

$$y_i = \begin{cases} \mu + \epsilon_i, & i = 1, \dots, m \\ \mu + \Delta + \epsilon_i, & i = m + 1, \dots, n \end{cases} \quad (1)$$

where the ϵ_i 's are independent random normal variables with zero mean and common unknown variance σ^2 ; the change-point m and the parameters μ and Δ are also unknown. Such a shift in the mean may occur, for instance, in water-level records due to a change in the gauge datum or in rainfall records due to a station relocation. Statistical methods have been developed to test the null hypothesis $\Delta = 0$ against the alternative hypothesis $\Delta \neq 0$.

A natural extension of eq. 1 is the following regression model:

$$y_i = \begin{cases} a + bx_i + \epsilon_i, & i = 1, \dots, m \\ a + \Delta + bx_i + \epsilon_i, & i = m + 1, \dots, n \end{cases} \quad (2)$$

Here the ϵ_i 's have the same meaning as in eq. 1; the x_i 's are fixed non-random variables. For instance, one might investigate a possible shift in mean runoff by regressing streamflow y_i on (effective) precipitation x_i . Although precipitation is a random quantity, eq. 2 may still be useful if we consider the distribution of streamflow given precipitation. Testing $\Delta = 0$ against $\Delta \neq 0$ presents some problems because, unlike the situation of known m , the distribution of the test statistics depends on the configuration of the x_i 's.

In this paper, methods for testing $\Delta = 0$ against $\Delta \neq 0$ are presented. We start with a review of techniques for the simple model of eq. 1. Comments are made on the likelihood ratio test and a Bayesian statistic. For the latter the critical values are derived. Thereafter, testing for a systematic change in a linear regression model (eq. 2) is discussed. The emphasis is on the application of the Bayesian statistic for testing $\Delta = 0$ against $\Delta \neq 0$ in eq. 1 to regression residuals. The techniques are illustrated with two examples at the end of the paper.

TESTING FOR A SINGLE SHIFT IN THE MEAN

A number of tests for a shift in the mean at an unknown time (eq. 1) have been compared by Buishand (1982). The test statistics can usually be written in terms of the adjusted partial sums or cumulative deviations from the mean:

$$S_0^* = 0; \quad S_k^* = \sum_{i=1}^k (y_i - \bar{y}), \quad k = 1, \dots, n \quad (3)$$

in which \bar{y} stands for $S_n^* = 0$.

For the model of eq.

$$E(S_k^*) = \begin{cases} -k(n-m) \\ -m(n-k) \end{cases}$$

and

$$\text{var}(S_k^*) = k(n-k)n^{-1}$$

So the mean is zero for $k = m$ and negative for $\Delta > 0$ and positive for $\Delta < 0$. For a random sequence, $\{y_k\}$ especially for k in the middle of the cumulative deviations. On the other hand, statistical tests based on S_k^* have optimal properties.

First, we consider the likelihood ratio test (LRT) (Lad 1979). The test statistic

$$V = \max_{1 \leq k \leq n-1} |S_k^*|$$

where D_y denotes the standard deviation of y_i .

$$D_y^2 = n^{-1} \sum_{i=1}^n (y_i - \bar{y})^2$$

Large values of the test statistic V indicate that in eq. 6, the S_k^* are significantly different from zero, proportional to their standard deviation.

Let K denote the critical value of V at the maximum. The statistic V is distributed at point m . The distribution of V is obtained from a recurrence relation that is uniform over the indices k . The distribution is maximal at the end-point $k = m$.

The preference of K over other tests for departures from normality is based on the rejection of the null hypothesis by restricting the maximum value of V to $3 \leq k \leq n-3$.

Statistical inference based on Bayesian procedures. A test was derived by Gardner (1975).

in which \bar{y} stands for the average of y_1, y_2, \dots, y_n . It is obvious that $S_n^* = 0$.

For the model of eq. 1 a straightforward calculation gives:

$$(1) \quad E(S_k^*) = \begin{cases} -k(n-m)n^{-1}\Delta, & k = 0, \dots, m \\ -m(n-k)n^{-1}\Delta, & k = m+1, \dots, n \end{cases} \quad (4)$$

and

$$\text{var}(S_k^*) = k(n-k)n^{-1}\sigma^2, \quad k = 0, \dots, n \quad (5)$$

So the mean is zero for a homogeneous record ($\Delta = 0$), positive for $\Delta < 0$ and negative for $\Delta > 0$. The variance is maximal if $k = n/2$. Even for a pure random sequence, $\{y_k\}$, the values of S_k^* can differ considerably from zero, especially for k in the neighbourhood of $n/2$. For this reason, the use of cumulative deviations is sometimes discouraged (W.M.O., 1966). On the other hand, statistical theory indicates that tests on cumulative deviations have optimal properties in situations of abrupt shifts in the mean.

First, we consider the likelihood ratio test (Hawkins, 1977; Worsley, 1979). The test statistic can be written as:

$$(2) \quad V = \max_{1 \leq k \leq n-1} \left\{ |S_k^*| / [D_y \{k(n-k)\}^{1/2}] \right\} \quad (6)$$

where D_y denotes the sample standard deviation:

$$D_y^2 = n^{-1} \sum_{i=1}^n (y_i - \bar{y})^2 \quad (7)$$

Large values of the statistic V lead to rejection of the null hypothesis. Note that in eq. 6, the S_k^* 's are weighted by a factor which is inversely proportional to their standard deviation.

Let K denote the index for which the right-hand side of eq. 6 reaches its maximum. The statistic K is the maximum-likelihood estimate of the change-point m . The distribution of K under the null hypothesis ($\Delta = 0$) can be obtained from a recurrence relation (Hawkins, 1977). This distribution is not uniform over the indices $1, 2, \dots, n-1$. Hawkins showed that $\text{Pr}(K = k)$ is maximal at the end-points and minimal for k in the neighbourhood of $n/2$.

The preference of K for the end-points makes the statistic V sensitive to departures from normality (a single outlier at $k = 1$ or $k = n-1$ may lead to rejection of the null hypothesis). A more robust statistic can be obtained by restricting the maximisation over a subset of the indices, for instance $3 \leq k \leq n-3$.

(3) Statistical inference about a shift in the mean can also be made by Bayesian procedures. A two-sided Bayesian test for use where σ^2 is known was derived by Gardner (1969). Where σ^2 is unknown, it seems natural to

TABLE I

Critical values and bounds of the statistic U for testing $\Delta = 0$ against $\Delta \neq 0$ in eq. 1 (m unknown)

Sample size, n	Lower bound, U_{\min}	Significance level α			Upper bound, U_{\max}
		0.10	0.05	0.01	
10	0.023	0.333	0.416	0.574	0.929
20	0.012	0.340	0.440	0.659	1.934
30	0.008	0.343	0.447	0.688	2.944
40	0.006	0.344	0.451	0.702	3.956
50	0.005	0.345	0.453	0.710	4.968
100	0.002	0.346	0.457	0.727	10.033
∞	0	0.347	0.461	0.743	∞

replace σ^2 in the statistical expression by the sample variance. This leads to the following statistic (Buishand, 1982):

$$U = [n(n + 1)]^{-1} \sum_{k=1}^{n-1} (S_k^*/D_y)^2 \tag{8}$$

provided we have a uniform prior distribution for the position of the change-point m .

Comparing the statistics given by eqs. 6 and 8, it is seen that in the latter less weight is given to the end-points. As a consequence, the statistic V is superior to U for a shift in the mean near the beginning and the end of the sequence, whereas the opposite holds for changes in the middle of the sequence (Buishand, 1982). Further, the statistic U is rather robust against departures from normality.

Critical values of the statistic U are given in Buishand (1982), obtained by a Monte Carlo method. It is, however, not too difficult to derive the distribution of U by numerical integration, as shown in Appendix A. This appendix also shows that U has a lower and an upper bound. The bounds and critical values are given in Table I.

EXTENSION TO THE REGRESSION MODEL

In applied regression analysis it is often recommended that the validity of model assumptions be checked by making plots of the least-squares residuals:

$$e_i = y_i - \hat{a} - \hat{b}x_i, \quad i = 1, \dots, n \tag{9}$$

Here \hat{a} and \hat{b} stand for the ordinary least-squares estimates:

$$\hat{a} = \bar{y} - \hat{b}\bar{x}; \quad \hat{b} = D_{xy}/D_x^2 \tag{10}$$

where

$$D_x^2 = n^{-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

For the e_i 's, v

$$S'_0 = 0; \quad S'_k = \sum_{i=1}^k e_i$$

Since the sum of squares of the residuals changes have occurred, called residual-normality, in the hydrometeorology, not provide an adequate

Unfortunately, the statistics which depend on the e_i 's. In this section, the application of the values of the statistics

The likelihood ratio

The likelihood ratio in eq. 2, has been shown to be equivalent to a test

$$V' = \max_{1 \leq k \leq n} \frac{S'_k}{k}$$

where \bar{x}_k is the

$$D_e^2 = n^{-1} \sum_{i=1}^n e_i^2$$

The weighting function is proportional to the distance between eqs. 6 and 8 in the denominator, which is weakly related to

The exact distribution depends on the parameters, approximate test statistics discussed by Wo

where

$$D_x^2 = n^{-1} \sum_{i=1}^n (x_i - \bar{x})^2; \quad D_{xy} = n^{-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \quad (11)$$

For the e_i 's, we can define the partial sums:

$$S'_0 = 0; \quad S'_k = \sum_{i=1}^k e_i, \quad k = 1, \dots, n \quad (12)$$

Since the sum of the e_i 's is equal to zero, $S'_n = 0$. To see whether systematic changes have occurred, a graph of the S'_k 's against k can be made (the so-called residual-mass curve). Although this technique has been recommended in the hydrometeorological literature (Searcy and Hardison, 1960), it does not provide an objective test-criterion.

Unfortunately statistical arguments do not in general lead to test statistics which depend only on the partial sums of the least-squares residuals. Further, the tests in the previous section cannot be directly applied to the e_i 's. In this section, first we briefly discuss some complications that arise in the application of the likelihood ratio test and then consider the critical values of the statistic U in regression situations.

The likelihood ratio test

The likelihood ratio test for testing $\Delta = 0$ against $\Delta \neq 0$, with m unknown in eq. 2, has been derived by Maronna and Yohai (1978). Their test is equivalent to a test based on the statistic (Potter, 1981; Worsley, 1983):

$$V' = \max_{1 \leq k \leq n-1} \left\{ |S'_k| / [D_e \{k(n-k) - k^2(\bar{x}_k - \bar{x})^2 / D_x^2\}^{1/2}] \right\} \quad (13)$$

where \bar{x}_k is the average of x_1, \dots, x_k , and:

$$D_e^2 = n^{-1} \sum_{i=1}^n e_i^2 \quad (14)$$

The weighting factor of the partial sums S'_k in eq. 13 is inversely proportional to their standard deviation (see Appendix B). From a comparison between eqs. 6 and 13, it is seen the latter has an extra term, $k^2(\bar{x}_k - \bar{x})^2 / D_x^2$, in the denominator. As a consequence, the likelihood ratio statistic is only weakly related to the residual-mass curve.

The exact distribution of the statistic V' is rather intractable. Moreover, it depends on the x_i 's so that a single table of critical values is not possible. An approximate test, based on an improved Bonferroni inequality, has been discussed by Worsley (1983).

Maronna and Yohai (1978) also considered random x_i 's. But, then the distribution of the statistic V' depends on (unknown) parameters of the distribution of the x_i 's, which makes its use unattractive.

The application of the test statistic U

As an alternative to the likelihood ratio test, we may apply the statistic U to the least-squares residuals:

$$U = [n(n + 1)]^{-1} \sum_{k=1}^{n-1} (S'_k/D_e)^2 \tag{15}$$

It is, however, not appropriate to use the critical values from Table I, because the e_i 's are neither independent nor identically distributed. Although, in this case also, the distribution of U depends on the configuration of the x_i 's, it is possible to give bounds for the critical values (see Appendix A). The test for $\Delta = 0$ against $\Delta \neq 0$ in eq. 2 then takes the form:

$$\begin{aligned} U \leq u_L, & \quad \text{the null hypothesis is not rejected} \\ u_L < U < u_U, & \quad \text{the test is inconclusive} \\ U \geq u_U, & \quad \text{the null hypothesis is rejected} \end{aligned} \tag{16}$$

Table II gives the bounds u_L and u_U as a function of the sample size n . The values for the upper bound, u_U , are slightly larger than the corresponding values in Table I; the difference decreases with n . On the other hand, the values of the lower bound, u_L , differ considerably from those in Table I and there is also a marked difference between u_L and u_U . As a consequence, we cannot neglect the possibility that the test is inconclusive.

When the bounds test is inconclusive, the exact critical level for the x_i 's under consideration can be calculated (see Appendix A). This requires many

TABLE II

Lower and upper bounds of the critical values of the statistic U for testing $\Delta = 0$ against $\Delta \neq 0$ in eq. 2 (m unknown)

Sample size, n	$\alpha = 0.10$		$\alpha = 0.05$		$\alpha = 0.01$	
	u_L	u_U	u_L	u_U	u_L	u_U
10	0.114	0.369	0.134	0.458	0.170	0.621
20	0.116	0.358	0.140	0.462	0.193	0.691
30	0.116	0.354	0.142	0.462	0.201	0.710
40	0.116	0.353	0.143	0.462	0.204	0.719
50	0.116	0.352	0.143	0.462	0.206	0.724
100	0.117	0.349	0.144	0.462	0.211	0.734

additional computation applications. A section in the following section

THE USE OF RECURSIVE

Instead of applying the test, we make a transformation of the data to make a transformed regression distributed under the null hypothesis. Methods of transformation for homogeneity testing (Brown et al., 1975)

Suppose that the regression coefficients are unknown. The regression coefficient estimates:

$$\hat{a}_j = \bar{y}_j - \hat{b}_j \bar{x}_j;$$

where \bar{x}_j and \bar{y}_j are the sample means of the regression, we get

$$\hat{y}_{j+1} = \bar{y}_j + \hat{b}_j(\bar{x}_{j+1} - \bar{x}_j)$$

We now look at the residuals

$$u_{j+1} = y_{j+1} - \hat{y}_{j+1} = y_{j+1} - \bar{y}_j - \hat{b}_j(\bar{x}_{j+1} - \bar{x}_j) \quad j = 2, \dots$$

It can be shown that the residuals are independent, and

$$\begin{aligned} \text{var } u_{j+1} &= \text{var } y_{j+1} - \text{var } \bar{y}_j - \text{var } \hat{b}_j(\bar{x}_{j+1} - \bar{x}_j) \\ &= \sigma^2 \left[\frac{1}{n} - \frac{1}{n} - \frac{1}{n} \right] \\ &= \sigma^2 \left[\frac{1}{n} - \frac{1}{n} - \frac{1}{n} \right] \quad j = 2, \dots \end{aligned}$$

The recursive residuals

$$u_j^* = u_j \left[\frac{j}{j-1} - \frac{1}{j-1} \right] \quad j = 3, \dots$$

Obviously, under the null hypothesis

additional computations, which may be a serious limitation for practical applications. A simple alternative to the bounds test is discussed in the following section.

THE USE OF RECURSIVE RESIDUALS

Instead of applying a test to the raw least-squares residuals we can first make a transformation to residuals which are independent and identically distributed under the null hypothesis ($\Delta = 0$ in eq. 2). There are several methods of transforming the e_i 's to independent random variables. In testing for homogeneity, it is preferable to use the so-called recursive residuals (Brown et al., 1975) which are obtained as follows.

Suppose that only $j < n$ pairs $(x_1, y_1), \dots, (x_j, y_j)$ are used to estimate the regression coefficients a and b . For $j \geq 2$, we may write for the least-squares estimates:

$$\hat{a}_j = \bar{y}_j - \hat{b}_j \bar{x}_j; \quad \hat{b}_j = \left[\sum_{i=1}^j (x_i - \bar{x}_j)(y_i - \bar{y}_j) \right] / \left[\sum_{i=1}^j (x_i - \bar{x}_j)^2 \right] \quad (17)$$

where \bar{x}_j and \bar{y}_j are the averages over the first j observations. From the fitted regression, we get the following prediction of y_{j+1} :

$$\hat{y}_{j+1} = \bar{y}_j + \hat{b}_j(x_{j+1} - \bar{x}_j), \quad j = 2, \dots, n-1 \quad (18)$$

We now look at the error term:

$$u_{j+1} = y_{j+1} - \hat{y}_{j+1} = y_{j+1} - \bar{y}_j - \hat{b}_j(x_{j+1} - \bar{x}_j), \quad j = 2, \dots, n-1 \quad (19)$$

It can be shown (Brown et al., 1975) that under the null hypothesis the u_j 's are independent, with mean zero and variance given by:

$$\begin{aligned} \text{var } u_{j+1} &= \text{var } y_{j+1} + \text{var } \bar{y}_j + (x_{j+1} - \bar{x}_j)^2 \text{var } \hat{b}_j \\ &= \sigma^2 \left[[(j+1)/j] + (x_{j+1} - \bar{x}_j)^2 / \left\{ \sum_{i=1}^j (x_i - \bar{x}_j)^2 \right\} \right], \\ & \quad j = 2, \dots, n-1 \end{aligned} \quad (20)$$

The recursive residuals are defined by:

$$u_j^* = u_j \left[[j/(j-1)] + (x_j - \bar{x}_{j-1})^2 / \left\{ \sum_{i=1}^{j-1} (x_i - \bar{x}_{j-1})^2 \right\} \right]^{-1/2}, \quad j = 3, \dots, n \quad (21)$$

Obviously, under the null hypothesis, the u_j^* 's are independent random

, then the
ers of the

ie statistic

(15)

Table I,
distributed.
configur-
alues (see
he form:

(16)

ize n . The
esponding
hand, the
able I and
ence, we

r the x_i 's
ires many

= 0 against

u_U

0.621
0.691
0.710
0.719
0.724
0.734

variables with mean zero and constant variance σ^2 . The term "recursive residual" stems from the fact that the estimates \hat{a}_j , \hat{b}_j and their covariances can be obtained recursively. There is, however, not so much need to use recurrence relations here, because we have explicit expressions for \hat{a}_j , \hat{b}_j and var u_j which can easily be updated. It can be shown further, that the u_j^* 's are obtained from the e_j 's by a linear transformation (Philips and Harvey, 1974).

The statistic U can be applied to the recursive residuals. Since, under the null hypothesis, these residuals are independent and identically distributed, the critical values in Table I can be used to look for evidence of departures from homogeneity. If there is a single shift in the mean ($\Delta \neq 0$ in eq. 2), then the residuals u_3^*, \dots, u_m^* are not affected, whereas, for the other residuals the mean will differ from zero. For $j > m$, the mean of the u_j^* 's is not a constant, but there is, nevertheless, an abrupt change at $j = m + 1$. Therefore, we may expect that the statistic U still has reasonable power when applied to the u_j^* 's. As an alternative to U we can use the statistic V given by eq. 6.

The recursive residual, u_j^* , was obtained by making a prediction of y_j from a regression fitted to the first $(j - 1)$ pairs (x_i, y_i) . It is, of course, also possible to base the prediction on the last $(n - j)$ pairs. We then have the following estimates of the regression coefficients:

$$\hat{a}_j' = \bar{y}_j' - \hat{b}_j' \bar{x}_j'; \quad \hat{b}_j' = \left[\sum_{i=j+1}^n (x_i - \bar{x}_j')(y_i - \bar{y}_j') \right] / \left[\sum_{i=j+1}^n (x_i - \bar{x}_j')^2 \right],$$

$$j = 1, \dots, n - 2 \quad (22)$$

where \bar{x}_j' and \bar{y}_j' are the averages over the last $(n - j)$ observations.

We can now define a sequence of error terms:

$$v_j = y_j - \bar{y}_j' - b_j'(x_j - \bar{x}_j'), \quad j = 1, \dots, n - 2 \quad (23)$$

giving the following set of recursive residuals:

$$v_j^* = v_j \left[\{(n - j + 1)/(n - j)\} + (x_j - \bar{x}_j')^2 / \left\{ \sum_{i=j+1}^n (x_i - \bar{x}_j')^2 \right\} \right]^{-1/2},$$

$$j = 1, \dots, n - 2 \quad (24)$$

Under the null hypothesis, the v_j^* 's are also independent normal variables with mean zero and constant variance σ^2 , but they differ from the u_j^* 's. This leads to different values for the test statistics.

Rescaled adjusted partial sums

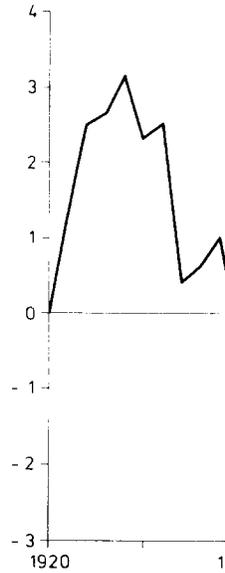


Fig. 1. Residual-mass Arizona. The term "re"

EXAMPLES

Annual precipitation Canyon, Arizona, U

As a first example annual data for the For the 26-y. period runoff on effective tation plus 0.3 times rescaled adjusted residual-mass curve by Searcy and Har a statistical test of 0.086, which is sm significance level o of systematic chang

January precipitation Teddington, U.K.

The second exam 80-y. period 1884-

m "recursive
 or covariances
 need to use
 for \hat{a}_j , \hat{b}_j and
 t the u_j^* 's are
 arvey, 1974).
 ce, under the
 tributed, the
 artures from
 2), then the
 residuals the
 t a constant,
 fore, we may
 plied to the
 p. 6.
 n of y_j from
 course, also
 en have the

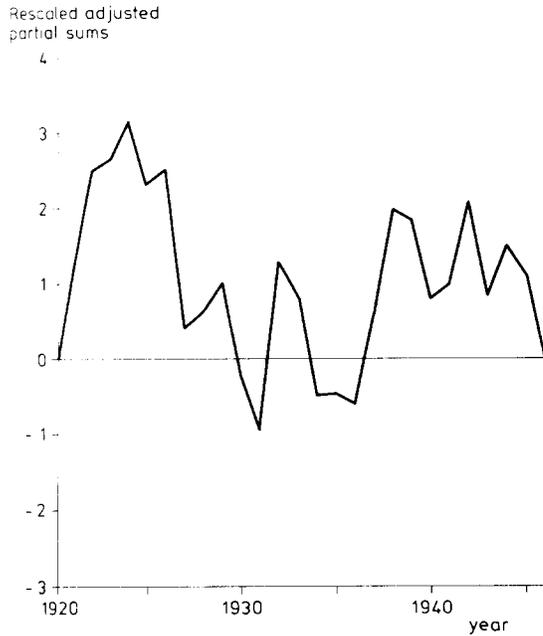


Fig. 1. Residual-mass curve of annual runoff for the Colorado River near Grand Canyon, Arizona. The term "rescaled" refers to division by the residual standard deviation, D_e .

EXAMPLES

Annual precipitation and runoff for the River Colorado Basin near Grand Canyon, Arizona, U.S.A.

As a first example to illustrate the methods described above, we consider annual data for the Colorado Basin taken from Searcy and Hardison (1960). For the 26-y. period 1921–1946, a linear regression was obtained of annual runoff on effective precipitation (being 0.7 times the current year's precipitation plus 0.3 times the previous year's precipitation). Fig. 1 gives the rescaled adjusted partial sums (S'_k/D_e) of the least-squares residuals. This residual-mass curve does not reveal a particular trend, which was also noticed by Searcy and Hardison (1960). To support this graphical result, we perform a statistical test on the S'_k 's. The realization of the statistic U in eq. 15 is 0.086, which is smaller than the lower bound u_L in Table II if we choose a significance level of 5%. So we conclude that there is no statistical evidence of systematic changes.

January precipitation and runoff for the River Thames Basin above Teddington, U.K.

The second example refers to data from the River Thames Basin for the 80-y. period 1884–1963, taken from Cooper and Clarke (1980). A linear

$$-\bar{x}_j)^2 \Big],$$

(22)

(23)

(24)

al variables
 n the u_j^* 's.



Fig. 2. Residual-mass curve of January runoff for the River Thames at Teddington. The term "rescaled" refers to division by the residual standard deviation, D_e .

regression was obtained of January streamflow on January precipitation. The plot of the partial sums of the least-squares residuals is given in Fig. 2. There are some indications of a shift in the mean, e.g., there is an apparent break in the slope of the curve about 1910. The realization of the statistic U in eq. 15 is 0.182. Tested at the 5% level, this value is just between the bounds u_L and u_U in Table II. This means that the test is inconclusive. We therefore have to consider the exact distribution of U under the null hypothesis. Some numerical work gives $\Pr(U > 0.182) = 0.30$, from which it is concluded that the null hypothesis is not rejected at the 5% level.

As an alternative to the bounds test, a test is performed on the adjusted partial sums, S_k^* , of the recursive residuals. For both the residuals u_j^* and v_j^* , the plot of the adjusted partial sums is given in Fig. 3. The curve for the v_j^* 's resembles that of the least-squares residuals in Fig. 2. The u_j^* 's show a slightly different plot, especially at the beginning of the sequence. This somewhat different behaviour can readily be explained from the residual-mass curve in Fig. 2. For instance, the almost linear decrease of this curve during the first 25 years does not appear in the plot for the u_j^* 's because there is a constant regression relation over this period during which there is no systematic deviation of the u_j^* 's from zero.

The realization of the test statistic U is 0.306 for the u_j^* 's and 0.245 for the v_j^* 's. From Table I (with $n = 78$), it is seen that these values are not

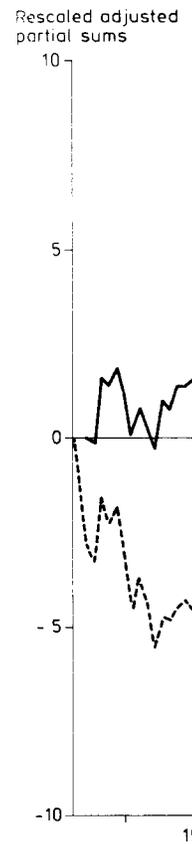


Fig. 3. Rescaled adjusted partial sums for the River Thames at Teddington. The dashed line represents the u_j^* 's. The solid line represents the v_j^* 's. The standard deviation of the residuals is D_e .

significant at the 5% level. There is no statistical evidence of a shift in the mean.

CONCLUDING REMARKS

In the beginning of the testing an abrupt change in the mean of the other systematic changes was observed.

It was also seen that the residuals were not adequate in many situations. The following form:

$$y_i = x_i - r_i$$

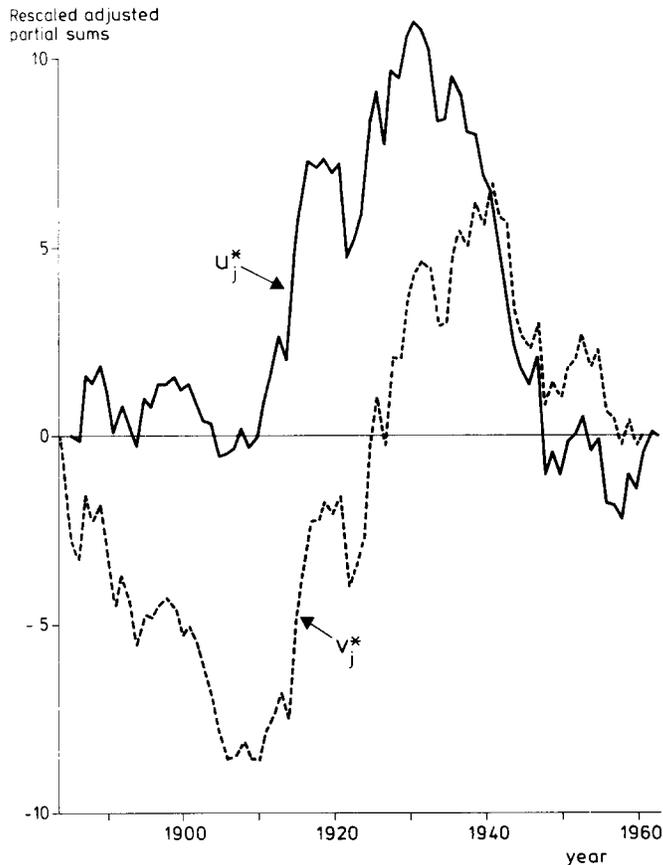


Fig. 3. Rescaled adjusted partial sums of recursive residuals (u_j^* and v_j^*) of January runoff for the River Thames at Teddington. The term "rescaled" refers to division by the sample standard deviation of the recursive residuals.

significant at the 5% level. Therefore we may conclude that there is no statistical evidence of systematic changes.

CONCLUDING REMARKS

In the beginning of this paper some statistical methods were reviewed for testing an abrupt change in the mean of a single time series (eq. 1). For other systematic changes, the test statistics may still have reasonable power.

It was also seen that extension to a regression relation (eq. 2) gave rise to a number of complications. Fortunately, the simple model of eq. 1 is adequate in many situations. The sequence (y_i) in this equation can take the following form:

$$y_i = x_i - r_i \quad (25)$$

or

$$y_i = \ln x_i - \ln r_i \quad (26)$$

where the x_i 's refer to data for the station under consideration and the r_i 's refer to a homogeneous record of a nearby station or to a regional average. For instance, for precipitation records, the statistic U can be applied to differences in annual amounts (eq. 25) or to differences in their logarithms (eq. 26). For such data it is, in general, unnecessary to use a regression relation, which is sometimes done (Maronna and Yohai, 1978; Potter, 1981).

The regression of streamflow data on precipitation data gives a variance reduction which improves the power of statistical tests. For annual runoff and effective precipitation of the Colorado Basin, the estimated correlation coefficient $\hat{\rho}$ is 0.84, which reduces the variance by a factor of $1 - \hat{\rho}^2 = 0.29$. For the data from the Thames Basin, a value of 0.62 was found for $\hat{\rho}$, which gives a much smaller variance reduction, ($1 - \hat{\rho}^2 = 0.62$).

Another reason for using a regression relation between streamflow and precipitation is to eliminate serial correlation effects. For instance, for the annual runoff data from the Colorado River there is some evidence of serial correlation. The estimate of the first-order serial correlation coefficient is 0.31. By taking a combination of the current year's precipitation and the previous year's precipitation as an explanatory variable, a sequence of error terms may be obtained which are nearly independent and the techniques described in this paper can still be applied.

APPENDIX A — THE DISTRIBUTION OF THE STATISTIC U UNDER THE NULL HYPOTHESIS

The derivation of distributional properties of the statistic U is similar to that of the Durbin—Watson statistic for testing serial correlation in least-squares regression. In this appendix the main lines are indicated; more details can be found in the classical paper of Durbin and Watson (1950) or in Hannan (1970).

We start with a short discussion on the distribution of quadratic forms in normal variables. Then the critical values and the bounds U_{\min} and U_{\max} in Table I are derived. Finally, we examine the distribution of the statistic U when it is applied to least-squares residuals.

The distribution of quadratic forms of normal variables

A quadratic form is defined by:

$$Q = u^T A u = \sum_{i=1}^n \sum_{j=1}^n (a_{ij} u_i u_j) \quad (A-1)$$

where $u = (u_1, \dots, u_n)$ is a column vector and A is a symmetric matrix; the a_{ij} 's are the elements of A .

A well-known result is that if A is a symmetric matrix, A , can be diagonalized by an orthogonal matrix, H , such that

$$H^T A H = \Lambda = \text{diag}(\lambda_1, \dots, \lambda_m, 0, \dots, 0)$$

The term "orthogonal" means that $H^{-1} = H^T$. The columns of H are the eigenvectors of A and the λ_i 's are the eigenvalues of A .

$$Q = v^T H^T A H v = v^T \Lambda v$$

where $\lambda_1, \dots, \lambda_m$ are the eigenvalues of A .

We now consider the distribution of Q when the u_i 's are independent normal variables, so that Q in eq. A-1 is a quadratic form in independent χ^2 -variates (Kendall, 1945).

The numerical evaluation of the distribution of Q is given by Imhof (1961). If the eigenvalues of A are

$$\Pr(Q < q) = \frac{1}{2} - \pi^{-1} \int_0^{\infty} \left[\sum_{i=1}^m \lambda_i w^2 \right]^{-1/2} \exp\left\{-\frac{1}{2} \sum_{i=1}^m \lambda_i w^2\right\} dw$$

where

$$\theta(w) = \frac{1}{2} \sum_{i=1}^m \arctan(\lambda_i w^2)$$

and

$$\rho(w) = \prod_{i=1}^m (1 + \lambda_i^2 w^2)^{-1/2}$$

The integral in eq. A-4 is over a finite range $0 \leq w \leq S$, where S is the upper bound of Q .

$$\lim_{w \rightarrow 0} \left\{ \frac{[\sin \theta(w)]}{[w \rho(w)]} \right\}$$

The upper bound, S , should be chosen such that $\Pr(Q > S) < \epsilon$. This can be achieved by

$$S^{m/2} = 2 / \left[\pi m t_s \prod_{i=1}^m (1 + \lambda_i^2 t_s^2)^{-1/2} \right]$$

where $u = (u_1, \dots, u_n)^T$ denotes an $n \times 1$ column vector and A an $n \times n$ symmetric matrix; the symbol T refers to transposition.

A well-known result from matrix algebra is that every real symmetric matrix, A , can be diagonalized by an orthogonal transformation, H :

$$H^T A H = \Lambda = \text{diag}(\lambda_1, \dots, \lambda_n) \quad (\text{A-2})$$

The term "orthogonal" implies that H is such that $H^T H$ is the $n \times n$ identity matrix I_n . The columns of H are orthonormal eigenvectors of A ; the λ_i 's are the eigenvalues of A . Let $u = H v$, then eq. A-1 becomes:

$$Q = v^T H^T A H v = v^T \Lambda v = \sum_{i=1}^m (\lambda_i v_i^2) \quad (\text{A-3})$$

where $\lambda_1, \dots, \lambda_m$ are the non-zero eigenvalues of A ($m \leq n$).

We now consider the case that the elements of u are independent standard normal variables. Then the v_i 's are also independent standard normal variables, so that Q in eq. A-3 is distributed as a linear combination of independent χ^2 -variates (Kendall and Stuart, 1969, Ch. 15).

The numerical evaluation of the distribution of Q has been discussed by Imhof (1961). If the eigenvalues are all distinct, $\Pr(Q < q)$ follows from:

$$\Pr(Q < q) = \frac{1}{2} - \pi^{-1} \int_0^{\infty} \{ \sin \theta(w) \} / \{ w \rho(w) \} dw \quad (\text{A-4})$$

where

$$\theta(w) = \frac{1}{2} \sum_{i=1}^m \arctan(\lambda_i w) - \frac{1}{2} q w \quad (\text{A-5})$$

and

$$\rho(w) = \prod_{i=1}^m (1 + \lambda_i^2 w^2)^{1/4} \quad (\text{A-6})$$

The integral in eq. A-4 can be calculated by the trapezoidal rule over the finite range $0 \leq w \leq S$, making use of the fact that:

$$\lim_{w \rightarrow 0} \left\{ [\sin \theta(w)] / [w \rho(w)] \right\} = \frac{1}{2} \sum_{i=1}^m \lambda_i - \frac{1}{2} q \quad (\text{A-7})$$

The upper bound, S , should be chosen such that the truncation error is small. This can be achieved by using the following relation:

$$S^{m/2} = 2 / \left[\pi m t_S \prod_{i=1}^m (|\lambda_i|)^{1/2} \right] \quad (\text{A-8})$$

in which t_s denotes an upper bound for the truncation error. In this study t_s has been set equal to 0.00001.

The derivation of the critical values in Table I

Note that under the null hypothesis ($\Delta = 0$ in eq. 1), the distribution of the statistic U does not depend on μ and σ . Therefore, for ease of notation, $\mu = 0$ and $\sigma = 1$ in this appendix.

To derive the distribution of U , the following quadratic forms are defined in the y_i 's:

$$Q_1 = (n+1)^{-1} \sum_{k=1}^{n-1} \{S_k^*\}^2 \quad (\text{A-9})$$

and

$$Q_2 = nD_y^2 \quad (\text{A-10})$$

We first consider the second quadratic form Q_2 , which in matrix notation reads:

$$Q_2 = y^T M y \quad (\text{A-11})$$

with

$$M = I_n - n^{-1} s s^T \quad (\text{A-12})$$

where $y = (y_1, \dots, y_n)^T$ and s is the $n \times 1$ vector of 1's. The matrix M has $(n-1)$ non-zero eigenvalues which are all equal to 1. Application of eq. A-3 gives the well-known result that Q_2 is a χ^2 -variate with $(n-1)$ degrees of freedom.

To obtain an expression for the quadratic form Q_1 in matrix notation, write for the k th adjusted partial sum S_k^* in eq. 3:

$$S_k^* = \tilde{y}^T i_k \quad (\text{A-13})$$

where $\tilde{y} = (y_1 - \bar{y}, \dots, y_n - \bar{y})^T = y - \bar{y}s$ and i_k is the vector with the first k elements equal to one and the other elements equal to zero. Substitution of eq. A-13 in eq. A-9 gives:

$$\begin{aligned} Q_1 &= (n+1)^{-1} \sum_{k=1}^{n-1} (\tilde{y}^T i_k)^2 = (n+1)^{-1} \tilde{y}^T \left(\sum_{k=1}^{n-1} i_k i_k^T \right) \tilde{y} \\ &= \tilde{y}^T B \tilde{y} \end{aligned} \quad (\text{A-14})$$

where B is the $n \times n$ symmetric matrix with entries $b_{ij} = b_{ji} = (n-i)/(n+1)$, $i \leq j$. Since $\tilde{y} = My$:

$$Q_1 = y^T C y \quad (\text{A-15})$$

with $C = M^T B M$. The

$$C = (n+1)^{-1} \sum_{k=1}^{n-1} c_k$$

with

$$c_k^T = i_k^T M = n^{-1} (n -$$

It was demonstrated
($n-1$) positive eigen

$$v_k = 1/[4(n+1) \cos$$

The corresponding e
 $C p_i = v_i p_i$. Note tha
decreasing for $0 < x$
associated eigenvalue

The matrices M an
the same orthogonal
Durbin and Watson (
noting that the eigen
the orthogonal transf

$$Q_1 = y^T C y = z^T H$$

$$Q_2 = y^T M y = z^T H$$

in which the z_i 's a
 Q_1/Q_2 :

$$U = \left[\sum_{i=1}^{n-1} (v_i z_i^2) \right] /$$

and thus:

$$\Pr(U < u) = \Pr \left\{ \sum_{i=1}^{n-1} \right.$$

This probability can
and $\lambda_i = v_i - u$. The
 $\Pr(U < u)$ for differ
table are the same
(Buishand, 1982).

In this study t_s

with $C = M^T B M$. The matrix C has the form (Gardner, 1969):

$$C = (n+1)^{-1} \sum_{k=1}^{n-1} c_k c_k^T$$

with

$$c_k^T = i_k^T M = n^{-1} (\underbrace{n-k, \dots, n-k}_k, \underbrace{-k, \dots, -k}_{n-k})$$

distribution of
ase of notation,

rms are defined

It was demonstrated further in Gardner (1969), that the matrix C has $(n-1)$ positive eigenvalues:

$$(A-9) \quad v_k = 1/[4(n+1) \cos^2 \{k\pi/(2n)\}], \quad k = 1, \dots, n-1 \quad (A-16)$$

The corresponding eigenvectors are denoted by p_1, p_2, \dots, p_{n-1} , i.e., $Cp_i = v_i p_i$. Note that $v_1 < v_2 < \dots < v_{n-1}$, since $\cos^2 x$ is monotonically decreasing for $0 < x < \frac{1}{2}\pi$. The vector s is also an eigenvector of C with associated eigenvalue 0.

(A-10)

matrix notation

The matrices M and C have the property that they can be diagonalized by the same orthogonal matrix, H . A rather complicated proof of this is given in Durbin and Watson (1950); the result is, however, immediately obtained by noting that the eigenvectors of C are also eigenvectors of M . Application of the orthogonal transformation $y = Hz$ gives, for quadratic forms Q_1 and Q_2 :

(A-11)

(A-12)

matrix M has
tion of eq. A-3
-1) degrees of

$$Q_1 = y^T C y = z^T H^T C H z = \sum_{i=1}^{n-1} (v_i z_i^2) \quad (A-17)$$

matrix notation,

$$Q_2 = y^T M y = z^T H^T M H z = \sum_{i=1}^{n-1} z_i^2 \quad (A-18)$$

in which the z_i 's are independent standard normal variates. Since $U = Q_1/Q_2$:

(A-13)

ctor with the
zero. Substi-

$$U = \left[\sum_{i=1}^{n-1} (v_i z_i^2) \right] / \left[\sum_{i=1}^{n-1} z_i^2 \right] \quad (A-19)$$

and thus:

$$\Pr(U < u) = \Pr \left\{ \sum_{i=1}^{n-1} (v_i z_i^2) < u \sum_{i=1}^{n-1} z_i^2 \right\} = \Pr \left\{ \sum_{i=1}^{n-1} (v_i - u) z_i^2 < 0 \right\} \quad (A-20)$$

(A-14)

$b_{ji} = (n-i)/$

This probability can be computed from eqs. A-4–A-6 with $m = n-1$, $q = 0$ and $\lambda_i = v_i - u$. The critical values in Table I are obtained by calculating $\Pr(U < u)$ for different values of u . The asymptotic values ($n \rightarrow \infty$) in this table are the same as those of Smirnov's ω^2 for testing goodness-of-fit (Buishand, 1982).

(A-15)

The bounds U_{\min} and U_{\max}

Because $v_1 < v_2 < \dots < v_{n-1}$:

$$v_1 \sum_{i=1}^{n-1} z_i^2 \leq \sum_{i=1}^{n-1} (v_i z_i^2) \leq v_{n-1} \sum_{i=1}^{n-1} z_i^2 \quad (\text{A-21})$$

Dividing each term in this inequality by $\sum_{i=1}^{n-1} z_i^2$ we obtain the following bounds for the statistic U in eq. A-19:

$$U_{\min} = v_1 \quad \text{and} \quad U_{\max} = v_{n-1} \quad (\text{A-22})$$

The lower bound, U_{\min} , is attained if the vector y coincides with the eigenvector p_1 of C , whereas $U = U_{\max}$ if y coincides with p_{n-1} .

From eq. A-16 it follows that, as $n \rightarrow \infty$:

$$U_{\min} = 1/[4(n+1) \cos^2 \{\pi/(2n)\}] = [4(n+1)]^{-1} + 0(n^{-3}) \quad (\text{A-23})$$

and

$$U_{\max} = 1/[4(n+1) \cos^2 \{\pi(n-1)/(2n)\}] = (n-1)\pi^{-2} + 0(n^{-1}) \quad (\text{A-24})$$

These approximations are already quite reasonable for $n = 10$.

The distribution of U for least-squares residuals

When the statistic U is applied to least-squares residuals, either an exact test or a bounds test can be made. To examine these tests, the vector $e = (e_1, \dots, e_n)^T$ of least-squares residuals is introduced and eq. 15 is written in the form:

$$U = (e^T B e)/(e^T e) = (e^T M^T B M e)/(e^T e) = (e^T C e)/(e^T e) \quad (\text{A-25})$$

Here use has been made of the fact that $e = M \bar{e}$ (because $\bar{e} = 0$).

Let $\epsilon = (\epsilon_1, \dots, \epsilon_n)^T$ denote the vector of the disturbance terms in eq. 2. Standard regression theory gives the following relation between e and ϵ :

$$e = N \epsilon \quad (\text{A-26})$$

where

$$N = I_n - n^{-1} s s^T - \tilde{x} \tilde{x}^T / (\tilde{x}^T \tilde{x}) \quad (\text{A-27})$$

with $\tilde{x} = (x_1 - \bar{x}, \dots, x_n - \bar{x})^T$. The matrix N projects each vector in the n -dimensional Euclidean space onto the orthogonal complement of s and \tilde{x} . For such a projection matrix it may be verified that $N = N^T = N^2$. Therefore, substitution of eq. A-26 in eq. A-25 gives:

$$U = (\epsilon^T N^T C N \epsilon) / (\epsilon^T N^T N \epsilon) = (\epsilon^T N C N \epsilon) / (\epsilon^T N \epsilon) \quad (\text{A-28})$$

Both the matrices N and $N C N (= N B N)$ have two zero and $(n-2)$ positive

eigenvalues. The positive known result that the degrees of freedom. The $v_1^*, v_2^*, \dots, v_{n-2}^*$ (v_k^* is Again it is readily v alized by the same o eq. A-19, application o

$$U = \left[\sum_{i=1}^{n-2} (v_i^* z_i^2) \right] /$$

Through the matrix $N x_i$'s.

To perform an exa methods, and then Pr(For the eigenvalues

$$v_i \leq v_i^* \leq v_{i+1}, \quad i$$

from which it follow variables:

$$U_L = \left[\sum_{i=1}^{n-2} (v_i z_i^2) \right] /$$

and

$$U_U = \left[\sum_{i=1}^{n-2} (v_{i+1} z_i^2) \right]$$

where the z_i 's are in eigenvalues of the ma by Durbin and Watson

The random variab vector p_{n-1} of C ; th Neither U_L nor U_U butions in Table II are

From eq. A-23, it $n \rightarrow \infty$. As a consequ distribution as the st eigenvalue, v_{n-1} , ren therefore the lower b if $n \rightarrow \infty$. This is in c lower bound tends to

eigenvalues. The positive eigenvalues of N are all equal to 1, giving the well-known result that the denominator in eq. A-28 is a χ^2 -variate with $(n-2)$ degrees of freedom. The ordered positive eigenvalues of NCN are denoted by $v_1^*, v_2^*, \dots, v_{n-2}^*$ (v_k^* is the k th smallest positive eigenvalue).

Again it is readily verified that the matrices N and NCN can be diagonalized by the same orthogonal matrix L (Hannan, 1970). Analogous to eq. A-19, application of the orthogonal transformation $\epsilon = Lz$ results in:

$$U = \left[\sum_{i=1}^{n-2} (v_i^* z_i^2) \right] / \left[\sum_{i=1}^{n-2} z_i^2 \right] \quad (\text{A-29})$$

Through the matrix N , the v_i^* 's and the distribution of U depend on the x_i 's.

To perform an exact test, first the v_i^* 's must be calculated by numerical methods, and then $\Pr(U < u)$ can be obtained from Imhof's algorithm.

For the eigenvalues v_i^* it can be shown that:

$$v_i \leq v_i^* \leq v_{i+1}, \quad i = 1, \dots, n-2 \quad (\text{A-30})$$

from which it follows that the statistic U is bounded by the two random variables:

$$U_L = \left[\sum_{i=1}^{n-2} (v_i z_i^2) \right] / \left[\sum_{i=1}^{n-2} z_i^2 \right] \quad (\text{A-31})$$

and

$$U_U = \left[\sum_{i=1}^{n-2} (v_{i+1} z_i^2) \right] / \left[\sum_{i=1}^{n-2} z_i^2 \right] \quad (\text{A-32})$$

where the z_i 's are independent standard normal variables; the v_i 's are the eigenvalues of the matrix C given by eq. A-16. This property was discovered by Durbin and Watson (1950); a short proof can be found in Hannan (1970).

The random variables U_L and U are identical if \tilde{x} coincides with the eigenvector p_{n-1} of C ; the upper bound U_U is attained if \tilde{x} coincides with p_1 . Neither U_L nor U_U depend on the x_i 's. The critical points of their distributions in Table II are obtained in the same way as those in Table I.

From eq. A-23, it is seen that the smallest eigenvalue, v_1 , tends to zero if $n \rightarrow \infty$. As a consequence, the random variable U_U has the same asymptotic distribution as the statistic U in eq. A-19. On the other hand, the largest eigenvalue, v_{n-1} , remains large compared with the other eigenvalues, and therefore the lower bound, U_L , does not tend to the limiting value in Table I if $n \rightarrow \infty$. This is in contrast with the Durbin-Watson statistic, for which the lower bound tends to the upper bound for large n .

APPENDIX B — THE VARIANCE OF PARTIAL SUMS OF LEAST-SQUARES RESIDUALS

For the k th partial sum of the least-squares residuals, we may write, analogous to eq. A-13:

$$S'_k = e^T i_k \quad (\text{B-1})$$

where $e = (e_1, \dots, e_n)^T$. Substitution of eq. A-26 gives:

$$S'_k = \epsilon^T N^T i_k \quad (\text{B-2})$$

Eq. A-26 holds only under the null hypothesis. However, if $\Delta \neq 0$, then the right-hand side of eq. B-2 remains valid for the reduced partial sums $S'_k - E(S'_k)$. Therefore, independent of Δ , the variance is:

$$\text{var } S'_k = E(\epsilon^T N^T i_k)^2 = E(i_k^T N \epsilon \epsilon^T N^T i_k) = \sigma^2 i_k^T N i_k \quad (\text{B-3})$$

because $E(\epsilon \epsilon^T) = \sigma^2 I_n$. Substitution of eq. A-27 gives:

$$\begin{aligned} \text{var } S'_k &= \sigma^2 [i_k^T I_n i_k - n^{-1} i_k^T s s^T i_k - (i_k^T \tilde{x})(\tilde{x}^T i_k) / (\tilde{x}^T \tilde{x})] \\ &= \sigma^2 \left[k - n^{-1} k^2 - \frac{\left\{ \sum_{i=1}^k (x_i - \bar{x}) \right\}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right] \\ &= \sigma^2 [k(n-k) - k^2(\bar{x}_k - \bar{x})^2 / D_x^2] / n \end{aligned} \quad (\text{B-4})$$

The variance does not depend on the regression coefficients a , b and Δ , but it is influenced by the configuration of the x_i 's.

ACKNOWLEDGEMENTS

The author is most grateful to A. Otten (Agricultural University of Wageningen) and T.A.B. Snijders (University of Groningen) for their valuable comments on an earlier effort.

REFERENCES

- Brown, R.L., Durbin, J. and Evans, J.M., 1975. Testing for the constancy of regression relationships over time (with Discussion). *J.R. Stat. Soc., Ser. B*, 37: 149–192.
- Buishand, T.A., 1982. Some methods for testing the homogeneity of rainfall records. *J. Hydrol.*, 58: 11–27.
- Cooper, D.M. and Clarke, R.T., 1980. Distribution-free methods for estimating flood and streamflow exceedance probabilities by correlation. *Water Resour. Res.*, 16: 1121–1127.
- Durbin, J. and Watson, G.S., 1950. Testing for serial correlation in least squares regression, I. *Biometrika*, 37: 409–428.
- Gardner Jr., L.A., 1969. On detecting changes in the mean of normal variates. *Ann. Math. Stat.*, 40: 116–126.
- Hannan, E.J., 1970. *Multiple Time Series*. Wiley, New York, N.Y., 536 pp.

- Hawkins, D.M., 1977. *Te*
Stat. Assoc., 72: 180–
- Imhof, J.P., 1961. *Comp*
Biometrika, 48: 419–
- Kendall, M.G. and Stuart
bution Theory. Charle
- Maronna, R. and Yohai,
change in mean. *J. Am*
- Philips, G.D.A. and Harv
analysis. *J. Am. Stat. A*
- Potter, K.W., 1981. *Illust*
series. *Mon. Weather R*
- Searcy, J.K. and Hardisc
Part I. *General Surface*
31–59.
- W.M.O. (World Meteorolo
Org., Geneva, Tech. No
- Worsley, K.J., 1979. *On*
lations. *J. Am. Stat. A*
- Worsley, K.J., 1983. *Te*
35–42.

ARES

may write,

(B-1)

, then the
partial sums

(B-3)

(B-4)

and Δ , but

ersity of
valuable

regression
2.
records. J.

flood and
Res., 16:

egression,

in. Math.

- Hawkins, D.M., 1977. Testing a sequence of observations for a shift in location. *J. Am. Stat. Assoc.*, 72: 180-186.
- Imhof, J.P., 1961. Computing the distribution of quadratic forms in normal variables. *Biometrika*, 48: 419-426.
- Kendall, M.G. and Stuart, A.D., 1969. *The Advanced Theory of Statistics. Vol. 1. Distribution Theory.* Charles Griffin, London, 3rd ed., 439 pp.
- Maronna, R. and Yohai, V.J., 1978. A bivariate test for the detection of a systematic change in mean. *J. Am. Stat. Assoc.*, 73: 640-645.
- Philips, G.D.A. and Harvey, A.C., 1974. A simple test for serial correlation in regression analysis. *J. Am. Stat. Assoc.*, 69: 935-939.
- Potter, K.W., 1981. Illustration of a new test for detecting a shift in mean in precipitation series. *Mon. Weather Rev.*, 109: 2040-2045.
- Searcy, J.K. and Hardison, C.H., 1960. Double-mass curves. In: *Manual of Hydrology, Part I. General Surface Water Techniques.* U.S. Geol. Surv., Water-Supply Pap., 1541-B: 31-59.
- W.M.O. (World Meteorological Organization), 1966. *Climatic change.* World Meteorol. Org., Geneva, Tech. Note 79.
- Worsley, K.J., 1979. On the likelihood ratio test for a shift in location of normal populations. *J. Am. Stat. Assoc.*, 74: 365-367.
- Worsley, K.J., 1983. Testing for a two-phase multiple regression. *Technometrics*, 25: 35-42.