Biometrika (1987), **74**, 1, *pp.* 33–43
Printed in Great Britain

Hypothesis testing when a nuisance parameter is present only under the alternative

BY ROBERT B. DAVIES
Applied Mathematics Division, DSIR, Wellington, New Zealand

SUMMARY

We wish to test a simple hypothesis against a family of alternatives indexed by a one-dimensional parameter, θ. We use a test derived from the corresponding family of test statistics appropriate for the case when θ is given. Davies (1977) introduced this problem when these test statistics had normal distributions. The present paper considers the case when their distribution is chi-squared. The results are applied to the detection of a discrete frequency component of unknown frequency in a time series. In addition quick methods for finding approximate significance probabilities are given for both the normal and chi-squared cases and applied to the two-phase regression problem in the normal case.

Some key words: Chi-squared process; Frequency component; Hypothesis test; Maximum; Nuisance parameter; Quick test; Two-phase regression; Time series; Upcrossing.

1. INTRODUCTION

We wish to test a hypothesis in the presence of a nuisance parameter, θ, which enters the model only under the alternative. In other words, θ is meaningless under the null hypothesis. Traditional methods for deriving hypothesis tests do not work in this situation.

Example 1. Suppose $X_1, \ldots, X_n$ are independent normal random variables with constant known variance, or else n is very large so the variance can be estimated. Suppose under the hypothesis there is a constant linear trend, but under the alternative the linear trend changes at some unknown point, θ, but with the composite regression line remaining continuous. That is there is no jump at the point θ but there is a jump in the derivative. This is the two-phase problem of Hinkley (1969) and the continuous case of Hawkins (1980). See Worsley (1983) for a discussion of the discontinuous case and other references. Naturally, the parameter θ is meaningless under the hypothesis of no change in slope.

Example 2. This is the same as Example 1, except that we observe several series subject to linear trend and the alternative is that at least one changes its trend at some unknown time, θ.

Example 3. Here $X_1, \ldots, X_n$ are as in Example 1, except that the expectations are given by

$$E(X_j) = \xi_1 \sin(j\theta) + \xi_2 \cos(j\theta). \qquad (1{\cdot}1)$$

The hypothesis is $\xi_1 = \xi_2 = 0$ and the alternative is that at least one of ξ_1 and ξ_2 is nonzero. This corresponds to a discrete frequency component at an unknown frequency, θ. Again θ is meaningless under the hypothesis.

Other examples are given by Davies (1977). Because θ cannot be estimated under the hypothesis, traditional large sample theory is not applicable. However, if θ were known it would be easy to find an appropriate test. Suppose $S(\theta)$ is the appropriate test statistic, with large values corresponding to the alternative being true. Then the test statistic we suggest for the case when θ is unknown is

$$M = \sup \{S(\theta): \mathscr{L} \leq \theta \leq \mathscr{U}\}, \tag{1·2}$$

where $[\mathscr{L}, \mathscr{U}]$ is the range of possible values of θ. The test would be to reject the hypothesis for large values of M. Of course, $S(\theta)$ has to be normalized in some way, for example by having mean and variance independent of θ under the null hypothesis. The problem is to find the significance probability of the resulting test. Davies (1977) studied the case when $S(\theta)$ had a normal distribution for each value of θ and gave a sharp bound on the significance probability. This bound was calculated from the autocorrelation function of $S(\theta)$.

One problem with this procedure is that of finding the autocorrelation function. It would not usually be realistic to do this in a 'once-off' problem, particularly where the test is only part of the overall analysis. In §2 we describe a very simple way of finding an approximation to the significance probability from the graph of $S(\theta)$ itself.

In § 3 we consider the case when $S(\theta)$ has a chi-squared distribution rather than a normal distribution. This arises when testing the hypothesis that a vector is zero against the alternative that at least one component is nonzero. Examples 2 and 3 are instances. We derive a bound corresponding to that found by Davies (1977) for the normal case and an approximation corresponding to that found in § 2. Some properties of the chi-squared process used in this section are derived in Appendix 1.

An important time series example of the chi-squared case is Example 3. The appropriate test is introduced in § 4. The derivation of the main formula is given in Appendix 2.

Our significance levels are either bounds or approximations to bounds and it is important to know how accurate they are. A series of computer simulations for Examples 1 and 3 are reported in § 5. In most instances our methods perform very well.

2. Quick calculation of significance: Normal case

We test the hypothesis $\xi = 0$ against the alternative $\xi > 0$ in the presence of a nuisance parameter $\theta \in [\mathscr{L}, \mathscr{U}]$ which enters the model only when $\xi > 0$. Suppose that an appropriate test, if θ was known would be to reject the hypothesis for large values of $S(\theta)$ where, for each θ, $S(\theta)$ has a standard normal distribution under the hypothesis. We suppose further that $S(\theta)$ is continuous on $[\mathscr{L}, \mathscr{U}]$ with a continuous derivative except possibly for a finite number of jumps in the derivative, and forms a Gaussian process. Davies (1977) recommends rejecting the hypothesis for large values of (1·2) and provides the bound

$$\mathrm{pr}\{\sup S(\theta) > c: \mathscr{L} \leq \theta \leq \mathscr{U}\} \leq \Phi(-c) + \exp(-\tfrac{1}{2}c^2) \int_{\mathscr{L}}^{\mathscr{U}} \{-\rho_{11}(\theta)\}^{\frac{1}{2}}\, d\theta/(2\pi), \tag{2·1}$$

where Φ denotes the cumulative normal distribution function,

$$\rho_{11}(\theta) = [\partial^2 \rho(\phi, \theta)/\partial\phi^2]_{\phi=\theta}, \quad \rho(\phi, \theta) = \mathrm{corr}\{S(\phi), S(\theta)\}.$$

The continuity condition given here is weaker than the one originally given by Davies (1977); see Marcus (1977) and Sharpe (1978).

In some cases it will be reasonable to calculate (2·1) analytically. In others, one would prefer a quick rule for deciding whether the result is significant. Let $T(\theta)=\partial S(\theta)/\partial\theta$. Then $\operatorname{var}\{T(\theta)\}=-\rho_{11}(\theta)$, so that $\{-\rho_{11}(\theta)\}^{\frac{1}{2}}=(\frac{1}{2}\pi)^{\frac{1}{2}}E\{|T(\theta)|\}$. Our proposal is to estimate

$$\int_{\mathscr{L}}^{\mathscr{U}} E\{|T(\theta)|\}\,d\theta \tag{2·2}$$

from the total variation

$$V=\int_{\mathscr{L}}^{\mathscr{U}} |T(\theta)|\,d\theta=|S(\theta_1)-S(\mathscr{L})|+|S(\theta_2)-S(\theta_1)|+\ldots+|S(\mathscr{U})-S(\theta_n)|, \tag{2·3}$$

where $\theta_1,\ldots,\theta_n$ are the successive turning points of $S(\theta)$. We presume n will be finite in practical problems.

Hence if M denotes the maximum of $S(\theta)$, then our estimate of the significance probability is

$$\Phi(-M)+V\exp(-\tfrac{1}{2}M^2)/(8\pi)^{\frac{1}{2}}. \tag{2·4}$$

Naturally, (2·4) is only approximate, but one would expect it to be much better than just $\Phi(-M)$. The second term in (2·4) will be important when $S(\theta)$ scans across a range of widely differing hypotheses and then values of $T(\theta)$ might tend to be independent for separated values of θ. In this case we would expect the law of large numbers to apply so that V would give a good estimate of (2·2). Simulations in § 5 bear this out. Formula (2·4) is for the one-sided case; for the two-sided case let M denote the maximum of $|S(\theta)|$ and then double (2·4).

3. Chi-squared case

We suppose that for each value of $\theta\in[\mathscr{L},\mathscr{U}]$, the test statistic appropriate for that value of θ is of the form

$$S(\theta)=Z_1^2(\theta)+\ldots+Z_s^2(\theta),$$

where the $\{Z_i(\theta)\}$ are continuous with continuous first derivatives, except possibly for a finite number of jumps in the derivatives, and form a vector Gaussian process. Suppose further that under the hypothesis the $Z_i(\theta)$ have zero expectations and for each θ the random variables $Z_1(\theta),\ldots,Z_s(\theta)$ are independent with unit variance. We will say that $S(\theta)$ is a chi-squared process. This is a generalization of the definition of Sharp (1978). In vector notation $S(\theta)=\|Z\|^2$, where $Z(\theta)$ is the column vector composed of the $Z_i(\theta)$. Here and elsewhere $\|Z\|=(Z'Z)^{\frac{1}{2}}$ denotes the L_2 norm of a vector, Z, and Z' denotes the transpose of a vector or matrix. Let $Y(\theta)=\partial Z(\theta)/\partial\theta$. Then $Y(\theta)$ and $Z(\theta)$ are jointly normally distributed. Suppose that

$$\operatorname{var}\begin{pmatrix}Z(\theta)\\ Y(\theta)\end{pmatrix}=\begin{bmatrix}I & A(\theta)\\ A'(\theta) & B(\theta)\end{bmatrix}. \tag{3·1}$$

Let $\lambda_1(\theta),\ldots,\lambda_s(\theta)$ be the eigenvalues of $B(\theta)-A'(\theta)A(\theta)$ and $\eta(\theta)$ be independent centred normal random variables with variances given by the $\lambda_i(\theta)$.

Then, according to Corollary A·1 in Appendix 1,

$$\operatorname{pr}[\sup\{S(\theta):\mathscr{L}\leq\theta\leq\mathscr{U}\}>u]\leq\operatorname{pr}(\chi_s^2>u)+\int_{\mathscr{L}}^{\mathscr{U}}\psi(\theta)\,d\theta, \tag{3·2}$$

where

$$\psi(\theta) = E(\|\eta(\theta)\|)u^{\frac{1}{2}(s-1)} e^{-\frac{1}{2}u} \pi^{-\frac{1}{2}} 2^{-\frac{1}{2}s}/\Gamma(\tfrac{1}{2}s+\tfrac{1}{2}) \qquad (3\cdot3)$$

and χ_s^2 denotes a chi-squared random variable with s degrees of freedom.

Sharpe (1978) shows that the number of high level upcrossings in the stationary independent case is approximately Poisson and we would expect the same to be true here and consequently the bound (3·2) should be sharp. An alternative definition of $\lambda_1(\theta), \ldots, \lambda_s(\theta)$ is given in Appendix 1.

To find the approximate form of the significance probability, let

$$V = \int_{\mathscr{L}}^{\mathscr{U}} |\partial S^{\frac{1}{2}}(\theta)/\partial\theta|\, d\theta = |S^{\frac{1}{2}}(\theta_1) - S^{\frac{1}{2}}(\mathscr{L})| + |S^{\frac{1}{2}}(\theta_2) - S^{\frac{1}{2}}(\theta_1)| + \ldots + |S^{\frac{1}{2}}(\mathscr{U}) - S^{\frac{1}{2}}(\theta_n)|,$$

where $\theta_1, \ldots, \theta_n$ are the turning points of $S^{\frac{1}{2}}(\theta)$. Let M denote the maximum of $S(\theta)$. Then it follows from Theorem A·2 that the estimate of the significance level corresponding to (2·4) is

$$\operatorname{pr}(\chi_s^2 > M) + VM^{\frac{1}{2}(s-1)} e^{-\frac{1}{2}M} 2^{-\frac{1}{2}s}/\Gamma(\tfrac{1}{2}s). \qquad (3\cdot4)$$

Alternatively, J. R. Harvey in a North Carolina State University Ph.D. thesis gives a number of analytic and approximate expressions for $E(\|\eta\|)$. One of these reduces to

$$E(\|\eta\|) = (2\pi)^{-\frac{1}{2}} \int_0^\infty \left\{1 - \prod_{j=1}^{s} (1+\lambda_j t)^{-\frac{1}{2}}\right\} t^{-3/2}\, dt.$$

Look at two special cases. Suppose all the λ_i are all equal with common value λ. Then

$$E(\|\eta\|) = (2\lambda)^{\frac{1}{2}}\Gamma(\tfrac{1}{2}s+\tfrac{1}{2})/\Gamma(\tfrac{1}{2}s),$$

in agreement with Sharpe's (1978, formula (3·2)). If $s=2$ then, again following Harvey,

$$E(\|\eta\|) = (2\lambda_1/\pi)^{\frac{1}{2}}\mathscr{E}(1-\lambda_2/\lambda_1), \qquad (3\cdot5)$$

where $\lambda_1 \geq \lambda_2$ and $\mathscr{E}$ denotes a complete elliptic integral of the second kind (Abramowitz & Stegun, 1972, formula (17.3.3)).

This ends our derivation of the extensions of the results of Davies (1977) and § 2 for the chi-squared situation. We now carry out some matrix manipulation which simplifies the evaluation of the eigenvalues $\lambda_1(\theta), \ldots, \lambda_s(\theta)$ for a class of problems which includes the time series example of § 4.

Suppose we observe $X_1, \ldots, X_n$ denoted collectively by the column vector X, where the X_i are independently normally distributed with unit variances and $E(X) = W'(\theta)\xi$ and where $W(\theta)$ is an $s \times n$ matrix of rank s, and ξ is an s-dimensional vector of unknown parameters. If θ were known, the most stringent test for testing $\xi = 0$ against the alternative $\xi \neq 0$ rejects the hypothesis for large values of

$$S(\theta) = X'Q(\theta)X, \qquad (3\cdot6)$$

where $Q(\theta) = W'(\theta)\{W(\theta)W(\theta)'\}^{-1}W(\theta)$. Write $Q(\theta) = U'(\theta)U(\theta)$, where $U(\theta)$ is an $s \times n$ matrix with $U(\theta)U'(\theta) = I$. Then setting $Z(\theta) = U(\theta)X$ puts the problem in the form already discussed and identifies $S(\theta)$ as being a chi-squared process under the null hypothesis. Then $Y(\theta) = \{\partial U(\theta)/\partial\theta\}X$ and so $A = U(\partial U'/\partial\theta)$, $B = (\partial U/\partial\theta)(\partial U'/\partial\theta)$, where for simplicity we are dropping references to θ.

Write $F \equiv G$ if matrices F and G have the same nonzero eigenvalues. Note that $FG \equiv GF$ if both products are defined. Also note that A defined in (3·1) is skew-symmetric. We need to find the eigenvalues of

$$\begin{aligned} B - A'A &= B + A^2 + (A')^2 + AA' \\ &\equiv U'\{(\partial U/\partial\theta)(\partial U'/\partial\theta) + U(\partial U'/\partial\theta)U(\partial U'/\partial\theta) \\ &\quad + (\partial U/\partial\theta)U'(\partial U/\partial\theta)U' + U(\partial U'/\partial\theta)(\partial U/\partial\theta)U'\}U \\ &= Q(\partial Q/\partial\theta)^2 Q. \end{aligned} \tag{3·7}$$

Let $R(\theta) = W'(\theta)\{W(\theta)W'(\theta)\}^{-1}\{\partial W(\theta)/\partial\theta\}$. Then

$$Q(\partial Q/\partial\theta)^2 Q = R(I-Q)R \tag{3·8}$$

$$\begin{aligned} &= W'(WW')^{-1}\partial W/\partial\theta\{I - W'(WW')^{-1}W\}(\partial W'/\partial\theta)(WW')^{-1}W \\ &\equiv (WW')^{-\frac{1}{2}}\{(\partial W/\partial\theta)(\partial W'/\partial\theta) \\ &\quad - (\partial W/\partial\theta)W'(WW')^{-1}W(\partial W'/\partial\theta)\}(WW')^{-\frac{1}{2}}. \end{aligned} \tag{3·9}$$

Thus $\lambda_1(\theta), \ldots, \lambda_s(\theta)$ can be found as the nonzero eigenvalues of (3·7), (3·8) or (3·9). Formula (3·9) is particularly convenient if

$$WW', \quad (\partial W/\partial\theta)(\partial W'/\partial\theta), \quad (\partial W/\partial\theta)W' \tag{3·10}$$

are all diagonal because then (3·9) reduces to

$$(WW')^{-1}\{(\partial W/\partial\theta)(\partial W'/\partial\theta)\} - (WW')^{-2}\{(\partial W/\partial\theta)W'\}^2. \tag{3·11}$$

4. Detection of a discrete frequency component

We observe $X = (X_1, \ldots, X_n)'$, where the X_j are independently normally distributed with unit variances and with

$$E(X_j) = \xi_1 \sin\{(j - \tfrac{1}{2}n - \tfrac{1}{2})\theta\} + \xi_2 \cos\{(j - \tfrac{1}{2}n - \tfrac{1}{2})\theta\}. \tag{4·1}$$

That is, $X_1, \ldots, X_n$ is a sequence of independent standard normal variables on to which has been superimposed a cyclic effect with period $2\pi/\theta$. Formula (4·1) is just a change of parameterization of (1·1) to simplify the calculation. Now suppose we wish to test the hypothesis, $\xi_1 = \xi_2 = 0$, that is, there is no frequency component, against the alternative that at least one of ξ_1 and ξ_2 is nonzero. Traditionally (Hannan, 1960, pp. 76–83), this problem has been handled by looking at only the values of θ of the form $2\pi k/n$, for $k = 1, \ldots, [\frac{1}{2}n]$. For these values of θ the corresponding values of (3·6) will be independent and so significance levels can be calculated. However a loss of power occurs if the true value falls between these values. We should emphasise that we are concerned with discrete frequency components. The method considered here has little relevance to the problem of detecting peaks in the frequency spectrum which have bandwidth greater than $2\pi/n$ cycles per sampling interval.

Now apply the theory of the previous section. The matrices (3·10) are derived in Appendix 2 and all turn out to be diagonal so (3·11) is applicable. Applying (3·6) and (4·1) we find

$$S(\theta) = \left[\sum_{j=1}^{n} X_j \sin\{(j - \tfrac{1}{2}n - \tfrac{1}{2})\theta\}\right]^2 \Big/ v_1 + \left[\sum_{j=1}^{n} X_j \cos\{(j - \tfrac{1}{2}n - \tfrac{1}{2})\theta\}\right]^2 \Big/ v_2, \tag{4·2}$$

where $v_1 = \frac{1}{2}n - \frac{1}{2}\sin(n\theta)/\sin(\theta)$, $v_2 = \frac{1}{2}n + \frac{1}{2}\sin(n\theta)/\sin(\theta)$.

For the moment suppose $0<\mathscr{L}\leqslant\theta\leqslant\mathscr{U}<\pi$, so that $S(\theta)$ is defined. Then it is shown in Appendix 2 that the eigenvalues λ_1 and λ_2 can be expressed

$$(n^2-1)/(3G)-n^2/4+(1-F^2/G^2)/(4\sin^2\theta), \tag{4·3}$$

where $F=\cos(n\theta)-\sigma\cos\theta$, $G=1-\sigma\sin(n\theta)/(n\sin\theta)$, $\sigma=+1$ to give λ_1 and -1 to give λ_2.

For each θ the value of $E(\|\eta\|)=a(\theta)$, say, in (3·3) can be found from (3·5) and so the bound (3·2) is equal to

$$\int_{\mathscr{L}}^{\mathscr{U}} a(\theta)\,d\theta\, u^{\frac{1}{2}}\, e^{-\frac{1}{2}u}/\pi+e^{-\frac{1}{2}u}. \tag{4·4}$$

In fact, (4·3) tends to zero as θ tends to 0 or π, so, provided (4·2) and (4·3) are defined by continuity at $\theta=0$ or π, we can allow θ to take on any value in the range $[0,\pi]$. For n large and not near 0 or π, formula (4·3) can be approximated by $n^2/12$ leading to $a(\theta)$ in (4·4) being approximated by $n(\pi/24)^{\frac{1}{2}}$, which in turn leads to (4·4) being approximated by

$$nu^{\frac{1}{2}}\, e^{-\frac{1}{2}u}(\mathscr{U}-\mathscr{L})/(24\pi)^{\frac{1}{2}}+e^{-\frac{1}{2}u}. \tag{4·5}$$

The approximation of $a(\theta)$ by $n(\pi/24)^{\frac{1}{2}}$ turns out to be adequate if $n>4$ and $2\pi/n<\theta<\pi-2\pi/n$. This approximation is also good for moderate and large values of n, when θ covers the whole range $0\leqslant\theta\leqslant\pi$. In Table 1 the values of

$$J=\int_0^{\pi} a(\theta)\,d\theta/\pi$$

and its approximation $n(\pi/24)^{\frac{1}{2}}$ are given for various values of n. The approximation is especially good if the approximate value is reduced by $\frac{1}{2}$. Presumably this adjustment is appropriate because of errors in the approximation at the ends of the range of integration.

Table 1. *Exact and approximate values for discrete frequency testing*

n	J	$n(\pi/24)^{\frac{1}{2}}$	n	J	$n(\pi/24)^{\frac{1}{2}}$
5	1·26	1·81	40	13·96	14·47
10	3·09	3·62	50	17·58	18·09
15	4·91	5·43	60	21·20	21·71
20	6·72	7·23	80	28·44	28·94
25	8·53	9·05	100	35·68	36·18
30	10·34	10·85			

Simulations described in § 5 show that the bound (4·4) and also the approximate bound based on (3·4) give very good approximations to the actual significance probabilities.

Formula (4·5) can be compared with the formula one obtains when one considers the test based on the maximum of the $S_k=S(2\pi k/n)$, where k is an integer. Assuming $\mathscr{U}$ and $\mathscr{L}$ to be of the form $(2k+1)\pi/n$, we have

$$\operatorname{pr}\{\sup(S_k\colon \mathscr{L}\leqslant 2\pi k/n\leqslant\mathscr{U})>u\}=1-(1-e^{-\frac{1}{2}u})^{\frac{1}{2}n(\mathscr{U}-\mathscr{L})/\pi}$$
$$\sim\tfrac{1}{2}n\,e^{-\frac{1}{2}u}(\mathscr{U}-\mathscr{L})/\pi. \tag{4·6}$$

In fact, there is relatively little difference between the values of u required to give the same value to formulae (4·5) and (4·6) when n is large and thus there is only a small loss of sensitivity when the test based on the maximum of the S_k is replaced by the test introduced in this section and θ is of the form $2\pi k/n$. On the other hand, suppose θ is of the form $2\pi(k+\frac{1}{2})/n$ and $\xi^2=\xi_1^2+\xi_2^2=O(n^{-1}\log n)$ in (4·1). Then $\sup_\theta E\{S(\theta)\}=n\xi^2/4+o(n)$, whereas $\sup_k E(S_k)=n\xi^2/\pi^2+o(n)$. Thus for large n, our test can provide substantial improvement in sensitivity over the test based on S_k. The relative efficiency is $\pi^2/4$. For smaller n one might expect the difference to be smaller and this is confirmed by the simulations in § 5.

In practice, of course, one would need to normalize the time series $X_1,\ldots,X_n$ by subtracting the sample mean and dividing by the sample standard deviation. It may also be necessary to compensate for serial correlation by fitting simple autoregressive or moving average processes. Heuristic arguments suggest that, for large n, the preceding results should still be applicable to the normalized series. Our simulations show, however, that this is so only for larger n and the results are less applicable for small n.

5. Simulations

We use simulation to investigate the formulae introduced in this paper for the univariate continuous two-phase problem, Example 1, and the discrete frequency problem, Example 3. We begin with the two-phase problem.

Observe $X_1,\ldots,X_n$, a sequence of independent normal random variables with unit variance and expectations given by

$$E(X_i)=\begin{cases} a+bt_i & (t_i<\theta),\\ a+bt_i+\xi(t_i-\theta) & (t_i\geqslant\theta),\end{cases}$$

where t_i denotes the time at which the ith measurement was made and θ the unknown time at which the change in slope occurred. We suppose t_i to be centred so that $\Sigma\, t_i=0$. We test the hypothesis that $\xi=0$ against the alternative that $\xi\neq 0$. Using $C(\alpha)$ principles one can find an appropriate test statistic for the case when θ is known: $S(\theta)=\Sigma_1\,(X_i-\hat{a}-\hat{b}t_i)(t_i-\theta)/V^{\frac{1}{2}}$, where

$$\hat{a}=\sum X_i/n,\quad \hat{b}=\sum t_iX_i/s_0,\quad V=s_1s_2/s_0+s_3s_4/n,$$
$$s_0=\sum t_i^2,\quad s_1=\sum\nolimits_1 t_i(t_i-\theta),\quad s_2=\sum\nolimits_2 t_i(t_i-\theta),\quad s_3=\sum\nolimits_1 (t_i-\theta),\quad s_4=\sum\nolimits_2 (t_i-\theta),$$

Σ_1 is over $t_i>\theta$ and Σ_2 is over $t_i<\theta$.

Under the hypothesis of no change of slope, $\xi=0$, $S(\theta)$ has a standard normal distribution. If ξ is nonzero, $|S(\theta)|$ will tend to be large. Since $S(\theta)$ is a linear function of the X_i, it forms a Gaussian process when considered as a whole. The function $S(\theta)$ is continuous, and continuously differentiable at all but a finite number of points. This is all that is required for the results given in § 2 to hold. Rather than calculating ρ_{11} in (2·1) analytically, the values of V in formula (2·3) obtained in a preliminary simulation have been averaged and used in (2·4). A sample size of 1000 was used.

In our simulation, $n=20$, $t_i=-9{\cdot}5,-8{\cdot}5,\ldots,9{\cdot}5$, $\mathscr{L}=-8$, $\mathscr{U}=8$. For evaluating M and V, the function $S(\theta)$ was evaluated at intervals of 0·2. In each case, two-sided tests were used. The first simulation has $\xi=0$ and a sample size of 1000. The results are given in the first part of Table 2. The 'accurate test' is the one just described using a preliminary simulation. That is, V in (2·4) is replaced by the average value from the preliminary simulation. The 'approximate test' uses the two-sided version of (2·4) directly. The 'naive test' uses the critical points for a normal distribution and ignores the fact that we are

Table 2. *Simulation of test for change in slope*; *N, sample size*

	Nominal significance level (%)				
	20	10	5	2	1
Per cent significant under null hypothesis, $N = 1000$					
Accurate test	18·8	9·8	5·0	1·5	0·9
Approximate test	18·5	9·1	4·4	1·3	0·7
Naive test	58·2	33·2	20·8	9·5	5·2
Per cent significant under alternative, $N = 100$					
Accurate test	58	45	33	20	12
Approximate test	56	43	35	19	10

scanning across a range of values of θ. The accurate and approximate tests give very satisfactory results whereas, not surprisingly, the naive test would be very misleading. In the second part of the table the corresponding results are given from a simulation with sample size of 100 and a value of ξ chosen to give a moderate power. This simulation was carried out to verify that the approximate test does not lead to a significant power loss.

Now examine the discrete frequency problem. The results of our first simulation are given in Table 3. In this and the other time series simulations, $\mathscr{L} = 0$, $\mathscr{U} = \pi$. For this simulation $n = 16$ and $S(\theta)$ was evaluated at intervals of $\pi/128$. Sample size was 4000. A fast Fourier transformation provided an effective way of carrying out these calculations. Critical points were found using formula (3·5). The table gives the observed per cent of significant results for the exact bound, the approximate bound, and for the test based on the S_k. Similar simulations were carried out with $n = 4$ and $n = 64$ with equally good results except that when $n = 4$ there was a modest drop in the number of significances for the approximate test.

In Table 4 we give a series of simulations with $n = 64$ and $S(\theta)$ evaluated at intervals of $\pi/256$. The first two simulations examine the performance of the tests under the alternative. The value of ξ was chosen to give moderate power. In the first simulation the discrete frequency is of the form $2\pi k/n$ and in the second it is of the form $2\pi(k+\frac{1}{2})/n$. As expected, the test based on the S_k has a minor advantage in the first simulation and a modest loss of power in the second.

We have also simulated the tests when the series is normalized by subtracting the sample mean and dividing by the sample standard deviation. In addition we applied them to the series after prewhitening by fitting an autoregressive process. In both cases the number of significances reported was substantially reduced for $n = 64$ but the effect was minor for $n = 256$. In the final simulation in Table 4 we report the results of a

Table 3. *Simulation of test for discrete frequency components*; *series length* = 16

	Nominal significance level (%)				
	20	10	5	2	1
Critical level, u	8·85	10·39	11·90	13·88	15·36
Per cent significant under null hypothesis, $N = 4000$					
Accurate test	17·2	9·0	4·7	2·1	1·1
Approximate test	15·6	7·5	3·8	1·6	0·9
Discrete test	18·0	8·7	4·4	1·7	0·9

Table 4. *Simulation of the test for discrete frequency components; series length* = 64

	Nominal significance level (%)				
	20	10	5	2	1
Critical level, u	11·97	13·47	14·96	16·91	18·38
Per cent significant under alternative, $N = 200$					
Accurate test	74	66	54	47	37
Approximate test	73	65	54	45	37
Discrete test	81	71	63	50	43
Per cent significant under alternative, $N = 200$					
Accurate test	69	59	50	37	30
Approximate test	68	60	49	35	30
Discrete test	55	36	27	20	15
Per cent significant under hypothesis (exponential case): $N = 500$					
Accurate test	17	11	7	3	2
Approximate test	16	10	7	3	2
Discrete test	19	12	7	3	2

simulation in which the time series consisted of independent centred exponential random variables. These results indicate that our tests are moderately robust against nonnormality.

Acknowledgements

Some of this work was carried out while the author was visiting the Statistical Laboratory of the University of California at Berkeley. It was partially supported by the National Institute of Health and the Office of Naval Research.

Appendix 1

The upcrossings of a chi-squared process

The notation and conditions are those of § 3.

THEOREM A·1. *The expected number of upcrossings of the level u by the process $S(\theta)$ over the range $\theta \in [\mathscr{L}, \mathscr{U}]$ is*

$$\int_{\mathscr{L}}^{\mathscr{U}} \psi(\theta)\, d\theta, \tag{A·1}$$

where ψ is as defined in (3·3).

COROLLARY A·1. *The probability of the maximum of a chi-squared process exceeding a given level is subject to the bound given in* (3·2).

Proof. For simplicity drop explicit references to θ. First note that A defined in (3·1) is skew-symmetric since

$$A + A' = \lim_{\Delta \to 0} E[Z(\theta)\{Z(\theta+\Delta) - Z(\theta)\}' + \{Z(\theta+\Delta) - Z(\theta)\}Z'(\theta+\Delta)]/\Delta = 0.$$

Let $f(.)$ denote the density of $S(\theta)$, a chi-squared variable with s degrees of freedom. Let $T = \partial S/\partial\theta = 2Z'Y$. The theory of § 5 of Marcus (1977) can be extended to include the chi-squared process and it follows that upcrossing formulae apply to the present problem for $u > 0$. Following Sharpe (1978), the expected number of upcrossings of the level u is given by (A·1) with

$$\psi = E\{T 1_{T>0} | S = u\} f(u). \tag{A·2}$$

We must show ψ is as given by (3·2). Let $\eta = Y - A'Z$. Then the covariance matrix of $(Z, \eta)'$ is diag $(I, B - A'A)$. In view of the skew-symmetry of A, we have $T = 2Z'\eta$. Choose U orthogonal and Λ diagonal so that $\Lambda = U(B - A'A)U'$. Premultiplication of Z and η by U will not affect S and T, so we can assume without loss of generality, that the covariance matrix of $(Z, \eta)'$ is diag (I, Λ).

The definition of η is now in conformity with that in § 3 if $\lambda_1, \ldots, \lambda_s$ are the elements of Λ. From (A·2)

$$\psi = 2E\{E(Z'\eta 1_{Z'\eta>0} | \|Z\|, \eta) | \|Z\|^2 = u\} f(u).$$

But

$$E(Z'\eta 1_{Z'\eta>0} | \|Z\|, \eta) = c_\eta \|Z\|$$

for some c_η which may depend on η, but not on $\|Z\|$ since Z and η are independent and the conditional distribution of Z given $\|Z\|$ is just the uniform distribution on the surface of the sphere of radius $\|Z\|$. To evaluate c_η take expectations:

$$E(Z'\eta 1_{Z'\eta>0} | \eta) = c_\eta E(\|Z\|). \tag{A·3}$$

On the left-hand side rotate Z so that Z_1 lies in the direction of η. Then (A·3) is equal to

$$E(Z_1 1_{Z_1>0}) \|\eta\| = (2\pi)^{-\frac{1}{2}} \|\eta\|.$$

Combining these results and noting that $f(u) = u^{\frac{1}{2}s-1} e^{-\frac{1}{2}u} 2^{-\frac{1}{2}s} / \Gamma(\frac{1}{2}s)$ and $E(\|Z\|) = 2^{\frac{1}{2}}\Gamma(\frac{1}{2}s + \frac{1}{2})/\Gamma(\frac{1}{2}s)$, we obtain (3·2) as required.

The corollary is proved in the same way as formula (3·6) of Davies (1977). This completes the proofs. □

At this point we note an alternative formula for $\Lambda(\theta)$. Let

$$R_{\theta,\Delta} = E\{Z(\theta + \Delta) Z'(\theta)\}. \tag{A·4}$$

Then

$$R_{\theta,\Delta} - I = E[\{Z(\theta + \Delta) - Z(\theta)\} Z'(\theta)] = \Delta A'(\theta) + o(\Delta),$$

$$2I - R_{\theta,\Delta} - R'_{\theta,\Delta} = E[\{Z(\theta + \Delta) - Z(\theta)\}\{Z(\theta + \Delta) - Z(\theta)\}'] = \Delta^2 B(\theta) + o(\Delta^2).$$

Let $D_{\theta,\Delta} = R_{\theta,\Delta} - R'_{\theta,\Delta}$. Now

$$\begin{aligned} R_{\theta,\Delta} \exp(-\tfrac{1}{2} D_{\theta,\Delta}) &= \{I + \tfrac{1}{2} D_{\theta,\Delta} + \tfrac{1}{2}(R_{\theta,\Delta} + R'_{\theta,\Delta} - 2I)\}(I - \tfrac{1}{2} D_{\theta,\Delta} + D^2_{\theta,\Delta}/8) + o(\Delta^2) \\ &= I - \tfrac{1}{2}\Delta^2 (B - A'A) + o(\Delta^2). \end{aligned}$$

But $\exp(-\frac{1}{2} D_{\theta,\Delta})$ is orthogonal. Hence the singular values of (A·4) are of the form $I - \frac{1}{2}\Delta^2 \Lambda(\theta) + o(\Delta^2)$. This provides a way of finding the eigenvalues directly from the autocovariance function of $Z(\theta)$ and enables a direct comparison to be made with the formulae in § 2.

Now look at the formula for $E\{\|\eta(\theta)\|\}$ in terms of the derivative of $S(\theta)$.

THEOREM A·2. *We have that*

$$E(\|\eta\|) = E|\partial S(\theta)^{\frac{1}{2}}/\partial\theta| \pi^{\frac{1}{2}} \Gamma(\tfrac{1}{2}s + \tfrac{1}{2})/\Gamma(\tfrac{1}{2}s).$$

Proof. From the proof of Theorem A·1, $\partial S(\theta)^{\frac{1}{2}}/\partial\theta = Z'\eta/\|Z\|$. Taking expectations

$$E|\partial S(\theta)^{\frac{1}{2}}/\partial\theta| = E\{E(|Z'\eta| \, | \, \|Z\|, \eta)/\|Z\|\} = 2c_\eta,$$

and the result follows from the proof of Theorem A·1. □

Appendix 2

Derivation of formula (4·3)

Using the notation of § 3 and letting $m=\frac{1}{2}(n+1)$,

$$W=\begin{bmatrix}\sin\{(1-m)\theta\} & \sin\{(2-m)\theta\} & \ldots & \sin\{(n-m)\theta\}\\ \cos\{(1-m)\theta\} & \cos\{(2-m)\theta\} & \ldots & \cos\{(n-m)\theta\}\end{bmatrix}.$$

All the off-diagonal elements of the matrices (3·10) are zero. The diagonal terms of WW' can be expressed as

$$\sum_{k=1}^{n}\sin^2\{(k-m)\theta\},\quad \sum_{k=1}^{n}\cos^2\{(k-m)\theta\}, \tag{A·5}$$

which are equal to $\frac{1}{2}nG$, with G as in (4·3) with $\sigma=+1$ for the first term and -1 for the second. The diagonal terms of $(dW/d\theta)W'$ can be found by differentiating (A·5). Differentiating again and noting that $\Sigma\,(k-m)^2=n(n^2-1)/12$, where the sum is over $k=1,\ldots,n$, enables one to find the diagonal elements of $(dW/d\theta)(dW'/d\theta)$. Then (4·3) can be derived from (3·11).

References

Abramowitz, M. & Stegun, I. A. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* New York: Dover.

Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **64**, 247–54.

Hannan, E. J. (1960). *Time Series Analysis.* London: Methuen.

Hawkins, D. M. (1980). A note on continuous and discontinuous segmented regressions. *Technometrics* **22**, 443–4.

Hinkley, D. V. (1969). Inference about the intersection in two-phase regression. *Biometrika* **56**, 495–504.

Marcus, M. B. (1977). Level crossings of a stochastic process with absolutely continuous sample paths. *Ann. Prob.* **5**, 52–71.

Sharpe, K. (1978). Some properties of the crossings process generated by a stationary chi-squared process. *Adv. Appl. Prob.* **10**, 373-91.

Worsley, K. J. (1983). Testing for a two-phase multiple regression. *Technometrics* **25**, 35–42.

[*Received November* 1984. *Revised April* 1986]