Hypothesis testing when a nuisance parameter is present only under the alternative - linear model case

By ROBERT B. DAVIES

Statistics Research Associates Limited

PO Box 12 649, Thorndon, Wellington, New Zealand

robert@statsresearch.co.nz

Summary

The results of Davies (1977, 1987) are extended to a linear model situation with unknown residual variance.

Some key words: Change point; F-process; Frequency component; Hypothesis test; Nuisance parameter; t-process; Two-phase regression; Up-crossing.

1. Introduction

Davies (1977, 1987) introduced a method for testing a hypothesis in the presence of a nuisance parameter, θ, which enters into the model only under the alternative. Examples 1 and 3 in the second paper were concerned with detecting a discrete frequency component and detecting a change in slope. Both are of the form in which we observe $Y_1, \ldots, Y_n$, denoted by a column vector, Y, which are independently normally distributed with variance σ^2 and mean

$$E(Y) = X\gamma + W(\theta)\xi, \tag{1}$$

where X is an $n \times s$ matrix and $W(\theta)$ is an $n \times p$ matrix function of the scalar nuisance parameter $\theta \in [\mathcal{L}, \mathcal{U}]$. The column vectors γ and ξ are unknown parameters. The objective is to test the hypothesis $\xi = 0$ against the alternative that at least one element of ξ is nonzero. If $p = 1$ we may be interested in testing $\xi = 0$ against the one-sided alternative $\xi > 0$. If $\xi = 0$ then θ does not enter into the model, so the model is of the general form considered by Davies (1977, 1987).

This paper extends the results of Davies (1977, 1987) to the situation when the residual variance σ^2 is unknown and n is too small for it to be estimated accurately. This requires us, in particular, to replace the Gaussian and χ^2 processes by t and F processes.

The main results are given in §2, the application to the examples in §3, and the derivation of the results in the Appendix. The notation is similar to that in Davies (1987). In particular, Y' denotes the transpose of a vector Y and $\|Y\|$ its L_2 norm.

2. Main Results

Assume that $q = n - s - p > 0$ and that the matrix $(X, W(\theta))$ is of full rank. For each θ carry out a QR factorisation of $(X, W(\theta))$ (Golub & Van Loan, 1996, p. 223) or some other orthogonalisation procedure so that

$$\begin{pmatrix} S \\ P(\theta) \\ Q(\theta) \end{pmatrix} \begin{pmatrix} X & W(\theta) \end{pmatrix} = \begin{pmatrix} G & H(\theta) \\ 0 & J(\theta) \\ 0 & 0 \end{pmatrix}, \tag{2}$$

where $(S', P(\theta)', Q(\theta)')'$ is an orthogonal matrix and G, $H(\theta)$ and $J(\theta)$ are $s \times s$, $s \times p$ and $p \times p$ matrices. The $s \times n$ matrix S is independent of θ whereas the $p \times n$ and $q \times n$ matrices $P(\theta)$ and $Q(\theta)$ will usually depend on θ. The QR factorisation (2) provides a convenient way of calculating the orthogonal components, SY and $Z(\theta) = P(\theta)Y$, of the data vector Y corresponding to X and $W(\theta)$ and the component corresponding to the residuals $R(\theta) = Q(\theta)Y$. Then, for each θ,

$$F(\theta) = \frac{\|Z(\theta)\|^2 / p}{\|R(\theta)\|^2 / q} \tag{3}$$

will have an F distribution with p and q degrees of freedom when $\xi = 0$. If θ were known then the usual test of the hypothesis $\xi = 0$ would reject the hypothesis when (3) was large. Suppose the elements of $P(\theta)$ and $Q(\theta)$ are continuous functions of $\theta \in [\mathcal{L}, \mathcal{U}]$ and have continuous derivatives except possibly for a finite number of jumps. Then we will say that (3) is an F process with p and q degrees of freedom. If $p = 1$ then we will say

$$t(\theta) = \frac{Z(\theta)}{\|R(\theta)\| / q^{\frac{1}{2}}} \tag{4}$$

is a t process with q degrees of freedom. Here, $Z(\theta)$ is regarded as a scalar. If θ were known then the usual test of the hypothesis $\xi = 0$ against the one-sided alternative $\xi > 0$ would reject the hypothesis when (4) was large. This is assuming that we have chosen the signs of the elements of $P(\theta)$ so that large values of $Z(\theta)$ correspond to large values of ξ.

When θ was not known, the approach of Davies (1977, 1987) would be to use the maximum with respect to θ of either (3) or (4) as the test statistic. The remainder of this section is concerned with calculating approximations to the distributions of these maxima. As in Davies (1977, 1987), these approximations are based on the expected number of upcrossings by (3) or (4), regarded as a stochastic process indexed by θ, of the critical value.

Define an analogue of the Beta distribution,

$$\beta(\theta) = \|Z(\theta)\|^2 / D^2, \tag{5}$$

where $D^2 = \|Z(\theta)\|^2 + \|R(\theta)\|^2 = \|Y\|^2 - \|SY\|^2$. Hence $F(\theta) = q\beta(\theta)/[p\{1 - \beta(\theta)\}]$.

Let $A(\theta) = P(\theta)\{\partial P'(\theta)/\partial\theta\}$, $B(\theta) = \{\partial P(\theta)/\partial\theta\}\{\partial P'(\theta)/\partial\theta\}$. Let $\eta(\theta)$ be a column vector whose elements are centred normal random variables with variance-covariance matrix given by $B(\theta) - A'(\theta)A(\theta)$.

The main result, derived in the Appendix, is that

$$\mathrm{pr}[\sup_{\mathcal{L}\le\theta\le\mathcal{U}}\{\beta(\theta)\} > u] \le 1 - I_u(\frac{1}{2}p, \frac{1}{2}q) + \int_{\mathcal{L}}^{\mathcal{U}} \psi(u,\theta)d\theta, \tag{6}$$

where I_u denotes the incomplete Beta function (Abramowitz & Stegun, 1972, formula 26.5.1) and

$$\psi(u,\theta) = \frac{u^{(p-1)/2}(1-u)^{(q-1)/2}\Gamma(\frac{1}{2}p+\frac{1}{2}q)E\{\|\eta(\theta)\|\}}{(2\pi)^{\frac{1}{2}}\Gamma(\frac{1}{2}p+\frac{1}{2})\Gamma(\frac{1}{2}q+\frac{1}{2})}. \tag{7}$$

Then the analogue of (3·2, 3·3) of Davies (1987) is

$$\mathrm{pr}[\sup_{\mathcal{L}\le\theta\le\mathcal{U}}\{F(\theta)\} > u] \le \mathrm{pr}(F_{p,q} > u) + \int_{\mathcal{L}}^{\mathcal{U}} \psi\{up/(q+up),\theta\}d\theta, \tag{8}$$

where $F_{p,q}$ denotes an F random variable with p and q degrees of freedom. When $p = 1$, the analogue of formula (3·7) of Davies (1977) is

$$\mathrm{pr}[\sup_{\mathcal{L}\le\theta\le\mathcal{U}}\{t(\theta)\} > u] \le \mathrm{pr}(t_q > u) + \frac{1}{2}\int_{\mathcal{L}}^{\mathcal{U}} \psi\{u^2/(q+u^2),\theta\}d\theta, \tag{9}$$

where t_q denotes a t random variable with q degrees of freedom. For the reasons noted in Davies (1977, 1987) we expect these bounds to be very tight in most situations for probabilities less than 0.2. They will, therefore, be good for evaluating statistical significance.

To evaluate (6), (8) or (9) we need a way of calculating

$$\int_{\mathcal{L}}^{\mathcal{U}} E\{\|\eta(\theta)\|\}d\theta. \tag{10}$$

Let $\lambda_1(\theta), \ldots, \lambda_p(\theta)$ be the eigenvalues of $B(\theta) - A'(\theta)A(\theta)$. Davies (1987) gives formulae for calculating $E\{\|\eta(\theta)\|\}$ from $\lambda_1(\theta), \ldots, \lambda_p(\theta)$. For example, if $p = 1$, then $E\{\|\eta(\theta)\|\} = (2\lambda_1/\pi)^{\frac{1}{2}}$. If $p = 2$ and $\lambda_1 \ge \lambda_2$ then $E\{\|\eta(\theta)\|\} = (2\lambda_1/\pi)^{\frac{1}{2}}\mathcal{E}(1 - \lambda_2/\lambda_1)$, where $\mathcal{E}$ denotes the complete elliptic integral of the second kind (Abramowitz & Stegun, 1972, formula (17.3.3)).

It follows from Theorem A3 in the Appendix that these eigenvalues are the squares of the singular values of

$$Q(\theta)\frac{\partial W(\theta)}{\partial\theta}J(\theta)^{-1}, \tag{11}$$

where W is defined in (1) and J and Q in (2). Formula (11) can be calculated as part of the QR factorisation and then the integration (10) carried out numerically. Alternatively, one can use the nonzero eigenvalues of $L(\partial W/\partial\theta)M(\partial W'/\partial\theta)L$, where $L = Q'Q = I - (X,W)\{(X,W)'(X,W)\}^{-1}(X,W)'$ and $M = (J'J)^{-1}$ is the lower right hand $p \times p$ submatrix of $\{(X,W)'(X,W)\}^{-1}$. This completes the discussion of the exact bounds on the probability distributions of the maxima of (3) and (4).

Formulae (2·4) and (3·4) of Davies (1987) introduced an approximate method for finding (10). A corresponding method could be based on the total variation of some function of $\beta(\theta)$. Theorem A4 suggests that an appropriate function is the arcsine square-root transformation. Let $\mathcal{V}$ denote the total variation of $\arcsin\{\beta(\theta)^{\frac{1}{2}}\}$; that is

$$\mathcal{V} = \sum_{i=1}^{m} \left| \arcsin\{\beta(\theta_i)^{\frac{1}{2}}\} - \arcsin\{\beta(\theta_{i-1})^{\frac{1}{2}}\} \right| = \int_{\mathcal{L}}^{\mathcal{U}} \left| \partial \arcsin\{\beta(\theta)^{\frac{1}{2}}\}/\partial\theta \right| d\theta,$$

where $\theta_1, \ldots, \theta_{m-1}$ denote the successive turning points of $\beta(\theta)$, with $\theta_0 = \mathcal{L}$ and $\theta_m = \mathcal{U}$.

Corresponding expressions for (3) and (4) are the total variation of respectively $\arctan[\{pF(\theta)/q\}^{\frac{1}{2}}]$ and $\arctan\{q^{-\frac{1}{2}}t(\theta)\}$. Then, according to formula (A5),

$$\int_{\mathcal{L}}^{\mathcal{U}} E\{\|\eta(\theta)\|\} d\theta = \frac{(2\pi)^{\frac{1}{2}}\Gamma(\frac{1}{2}p+\frac{1}{2})\Gamma(\frac{1}{2}q+\frac{1}{2})}{\Gamma(\frac{1}{2}p)\Gamma(\frac{1}{2}q)} E(\mathcal{V}).$$

This suggests

$$\frac{(2\pi)^{\frac{1}{2}}\Gamma(\frac{1}{2}p+\frac{1}{2})\Gamma(\frac{1}{2}q+\frac{1}{2})}{\Gamma(\frac{1}{2}p)\Gamma(\frac{1}{2}q)} \mathcal{V}$$

as an estimate of (10), leading to the following approximate significance probability corresponding to (8):

$$\operatorname{pr}(F_{p,n-p} > \mathcal{M}) + \frac{u^{(p-1)/2}(1-u)^{(q-1)/2}\Gamma(\frac{1}{2}p+\frac{1}{2}q)}{\Gamma(\frac{1}{2}p)\Gamma(\frac{1}{2}q)} \mathcal{V}, \tag{12}$$

where $u = \mathcal{M}p/(q + \mathcal{M}p)$ and $\mathcal{M}$ is the maximum of the F process. If $p = 1$ and we want a one-sided test the corresponding term for (9) is

$$\operatorname{pr}(t_q > \mathcal{M}) + \frac{(1-u)^{(q-1)/2}\Gamma(\frac{1}{2}q+\frac{1}{2})}{2\pi^{\frac{1}{2}}\Gamma(\frac{1}{2}q)} \mathcal{V},$$

where $u = \mathcal{M}^2/(q + \mathcal{M}^2)$ and $\mathcal{M}$ is the maximum of the t process.

3. Examples

I applied the method to Examples 1 and 3 in Davies (1987), allowing a nonzero mean in Example 3. However, in contrast to Davies (1987), I calculated the integral (10) numerically using (11) rather than analytically. Simulation revealed generally good agreement with the theoretical values for both the more accurate (8) and approximate (12) calculations of the levels of significance for 20%, 5%, 1% and 0.2% levels with $10 \leq n \leq 200$ and a variety of values of $\mathcal{L}$ and $\mathcal{U}$. For small values of n the approximate test became slightly conservative but would still be usable. In contrast, the formulae of Davies (1977, 1987) did not give satisfactory results when an estimated value of σ was used for values of $n \leq 50$ but were reasonably satisfactory for $n \geq 100$ in the case of Example 1. In the case of Example 3 the results were unsatisfactory for $n \leq 200$.

Table 1 shows the actual significance levels found by simulation for the discrete frequency test, Example 3, in Davies (1987) for various values of sample size, n, and nominal significance level. The number of simulations was 100,000. The approximate test uses formula (12) and the accurate test uses formula (8). Both of these tests show a satisfactory agreement with the nominal significance level. The chi-squared versions use formula (3·2) of Davies (1987). The A version divides each $S(\theta)$ value by the variance of the raw data and the B version by the residual mean square after the frequency component corresponding to θ has been fitted. Neither is satisfactory, particularly for small values of the significance level even when $n = 200$.

Table 1. *Actual significance levels for discrete frequency test*

n	nominal significance level	20	5	1	0.2
10	approximate test	19.5	4.4	0.80	0.13
	accurate test	19.7	4.8	0.95	0.17
	chi-sq test - A	3.2	0.0	0.00	0.00
	chi-sq test - B	66.5	40.3	24.56	15.87
50	approximate test	19.4	5.0	0.93	0.18
	accurate test	19.2	4.9	0.92	0.18
	chi-sq test - A	13.0	2.2	0.21	0.01
	chi-sq test - B	33.5	12.3	3.82	1.21
200	approximate test	18.5	4.8	0.96	0.17
	accurate test	18.3	4.7	0.94	0.17
	chi-sq test - A	15.6	3.6	0.60	0.09
	chi-sq test - B	23.1	6.8	1.58	0.34

Appendix

Derivation of main results

Assume that $P(\theta)$ and $Q(\theta)$ defined in §2 are continuous with continuous derivatives except possibly for a finite number of jumps in the derivative at fixed values of θ. For simplicity assume that $\sigma = 1$ and drop explicit references to θ. Let $\eta = \partial Z/\partial\theta - A'Z$ and $T = \partial\beta/\partial\theta = 2Z'(\partial Z/\partial\theta)/D^2 = 2Z'\eta/D^2$ since $A = P(\partial P/\partial\theta)'$ is skew-symmetric.

Theorem A1. *The expected number of up-crossings of the level u by the process $\beta(\theta)$ defined in formula (5) over the range $[\mathcal{L},\mathcal{U}]$ is*

$$\int_{\mathcal{L}}^{\mathcal{U}} \psi(u,\theta)d\theta, \tag{A1}$$

where ψ is as in (7).

Proof. The proof is similar to that of Davies (1987, Theorem A·1). Following Sharpe (1978, §3), the expected number of up-crossings of the level u is given by (A1) with

$$\psi(u) = E(T1_{T>0}|\beta = u)f(u),$$

where $f(u)$ is the density function of β. That is

$$f(u) = \frac{u^{p/2-1}(1-u)^{q/2-1}\Gamma(\frac{1}{2}p+\frac{1}{2}q)}{\Gamma(\frac{1}{2}p)\Gamma(\frac{1}{2}q)}.$$

The variance matrix of $(Z', \eta')'$ is $\mathrm{diag}(I, B - A'A)$, so that the definition of η conforms to that given in §2. Also η is independent of Z and so is a linear function of $R = QY$. Hence,

$$\begin{aligned}
\psi(u) &= 2E\left\{\frac{E(Z'\eta 1_{Z'\eta>0}|R, \|Z\|)}{D^2} | \, \|Z\|^2 = uD^2\right\} f(u) \\
&= 2E\left(\frac{c_\eta \|Z\|}{D^2} | \, \|Z\|^2 = uD^2\right) f(u) \\
&= 2u^{\frac{1}{2}}(1-u)^{\frac{1}{2}} E\left(\frac{c_\eta}{\|R\|} | \, \|Z\|^2 = uD^2\right) f(u),
\end{aligned}$$

where $c_\eta = E(Z'\eta 1_{Z'\eta>0}|R, \|Z\|)/\|Z\| = \frac{1}{2}\|\eta\| \pi^{-\frac{1}{2}}\Gamma(\frac{1}{2}p)/\Gamma(\frac{1}{2}p+\frac{1}{2})$ as in Davies (1987). Now η is a linear function of R and so $\eta/\|R\|$ is independent of $\|R\|$ by Basu's lemma (Lehmann, 1983, p. 46). R is independent of Z and so $\eta/\|R\|$ is independent of $\|R\|$ and $\|Z\|$ and hence also of D. Hence,

$$\psi(u) = 2u^{\frac{1}{2}}(1-u)^{\frac{1}{2}} E(c_\eta/\|R\|) f(u) = 2u^{\frac{1}{2}}(1-u)^{\frac{1}{2}} E(c_\eta) f(u)/E(\|R\|).$$

Also $E(\|R\|) = 2^{\frac{1}{2}}\Gamma(\frac{1}{2}q+\frac{1}{2})/\Gamma(\frac{1}{2}q)$ and the result follows. □

Corollary A1. *A bound on the probability* $\mathrm{pr}[\{\sup \beta(\theta)\} > u]$ *is given by (6).*

Proof. This follows by an argument similar to that in Davies (1977). □

Worsley (1994) considers versions of t and F processes and has a theorem related to our Theorem A1, for multi-dimensional θ. A generalisation of Worsley's result would be the way to proceed if we needed to extend our results to allow for multi-dimensional θ.

Theorem A2. *The nonzero eigenvalues of $B - A'A$ are the squares of the nonzero singular values of (11).*

Proof. Note that $P'P + Q'Q + S'S = I$. Also, $QP' = 0$ so that $\partial Q/\partial\theta P' + Q\partial P'/\partial\theta = 0$. Similarly $S\partial P'/\partial\theta = 0$ and $\partial Q/\partial\theta S' = 0$. Then

$$B - A'A = \frac{\partial P}{\partial\theta}(I - P'P)\frac{\partial P'}{\partial\theta} = \frac{\partial P}{\partial\theta}(Q'Q + S'S)\frac{\partial P'}{\partial\theta} = \frac{\partial P}{\partial\theta}(Q'Q)\frac{\partial P'}{\partial\theta}.$$

Differentiate $QW = 0$ to find $\partial Q/\partial\theta(P'P + Q'Q + S'S)W + Q\partial W/\partial\theta = 0$. Hence $Q\partial P'/\partial\theta J = Q\partial W/\partial\theta$ and the result follows. □

Now consider the quick way of estimating (10). This is based on the total variation, $\mathcal{V}$, of an monotonically increasing function, m, of $\beta(\theta)$. Thus

$$\mathcal{V} = \int_{\mathcal{L}}^{\mathcal{U}} |\partial m(\beta)/\partial\theta| \, d\theta = \int_{\mathcal{L}}^{\mathcal{U}} |m_1(\beta)T| \, d\theta,$$

where m_1 denotes the first derivative of m.

Theorem A3. *An unbiased estimator of (10) is given by*

$$\frac{\pi^{\frac{1}{2}}\Gamma(\frac{1}{2}p+\frac{1}{2})\Gamma(\frac{1}{2}q+\frac{1}{2})}{2^{\frac{1}{2}}E(K)\Gamma(\frac{1}{2}p)\Gamma(\frac{1}{2}q)}\mathcal{V}, \tag{A2}$$

where

$$K = m_1(\|Z\|^2/D^2)\,\|R\|\,\|Z\|/D^2.$$

Proof.

$$\begin{aligned}
E(|m_1(\beta)T|) &= 2E\{m_1(\|Z\|^2/D^2)\,|Z'\eta|/D^2\} \\
&= 2E\{m_1(\|Z\|^2/D^2)E(|Z'\eta|\,|R,\|Z\|)/D^2\} \\
&= 4E\{m_1(\|Z\|^2/D^2)c_\eta\,\|Z\|/D^2\} \\
&= \frac{2^{\frac{1}{2}}E(K)E(\|\eta\|)\Gamma(\frac{1}{2}p)\Gamma(\frac{1}{2}q)}{\pi^{\frac{1}{2}}\Gamma(\frac{1}{2}p+\frac{1}{2})\Gamma(\frac{1}{2}q+\frac{1}{2})},
\end{aligned} \tag{A3}$$

since $E(|Z'\eta|\,|R,\|Z\|) = 2c_\eta\,\|Z\|$, the ratio $\eta/\|R\|$ is independent of $(\|R\|, Z)$ and $D^2 = \|R\|^2 + \|Z\|^2$. Integrate over θ and the result follows. □

We would like to choose the function m to minimise the variance (A2). It does not seem likely that there would be a simple formula for this variance. However, we can calculate the variance of

$$\frac{\pi^{\frac{1}{2}}\Gamma(\frac{1}{2}p+\frac{1}{2})\Gamma(\frac{1}{2}q+\frac{1}{2})}{2^{\frac{1}{2}}E(K)\Gamma(\frac{1}{2}p)\Gamma(\frac{1}{2}q)}\,|m_1(\beta)T|, \tag{A4}$$

which leads to (A2) when integrated over θ. Since the correlation structure of (A4) is unlikely to be sensitive to the choice on m, we might expect minimising the variance of (A4) to come close to minimising the variance of (A2).

Theorem A4. *The variance of (A4) is minimised by choosing* $m(x) = \arcsin(x^{\frac{1}{2}})$. *With this choice of* m

$$E(|m_1(\beta)T|) = \frac{\Gamma(\frac{1}{2}p)\Gamma(\frac{1}{2}q)}{(2\pi)^{\frac{1}{2}}\Gamma(\frac{1}{2}p+\frac{1}{2})\Gamma(\frac{1}{2}q+\frac{1}{2})}E(\|\eta\|). \tag{A5}$$

Proof. Using an argument similar to that used in Theorem A3, we have

$$\begin{aligned}
E\{m_1(\beta)T\}^2 &= 4E[m_1(\|Z\|^2/D^2)^2E\{(Z'\eta)^2|R,\|Z\|\}/D^4] \\
&= 4E(K^2)E(\|\eta\|^2)/(pq),
\end{aligned} \tag{A6}$$

since $E\{(Z'\eta)^2|R,\|Z\|\} = E(Z_1^2/\|Z\|^2)\,\|\eta\|^2\,\|Z\|^2 = \|\eta\|^2\,\|Z\|^2/p$. To minimise the variance of (A4) it is sufficient to minimise the ratio of (A6) to the square of (A3). This can be done by choosing $\mathrm{var}(K) = 0$; for example, we can take $K = \frac{1}{2}$. This implies that $m(x) = \arcsin(x^{\frac{1}{2}})$. Formula (A5) follows from (A3). □

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.

Davies, R.B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **74**, 33-43.

Lehmann, E.L. (1983). *Theory of Point Estimation.* New York: Wiley.

Golub, G.H. & Van Loan, C.F. (1996). *Matrix Computations*, 3rd ed. Baltimore: Johns Hopkins University Press.

Sharpe, K. (1978). Some properties of the crossings process generated by a stationary chi-squared process. *Adv. Appl. Prob.* **10**, 373-391.

Worsley, K.J. (1994). Local maxima and the expected Euler characteristic of excursion sets of χ^2, F and t fields. *Adv. Appl. Prob.* **26**, 13-42.