Testing a Model Parameter When Another is Unidentified Under the Null

DENIS CONNIFFE*
The Economic and Social Research Institute

Abstract: Some standard test procedures, such as Score and Likelihood Ratio, replace nuisance parameters by their maximum likelihood estimates under the null hypothesis about the parameter of interest. In some models, however, a nuisance parameter is not identified under the null, so that these approaches need modification. By taking a mathematically tractable case, this paper illustrates the issues that arise and the solutions that have been proposed in the literature. The rival tests are compared in terms of power and robustness to misspecification.

I INTRODUCTION

Until relatively recently the topic of hypothesis testing when a "nuisance" parameter (that is, a parameter not assigned a value by the hypothesis under test) is unidentifiable under the null was a rather esoteric sideline in mathematical statistics, with Davies (1977) the best known reference. Engle (1984) did discuss how the problem can arise in econometrics and Watson and Engle (1985) employed the Davies approach to test for a time varying regression coefficient. Godfrey (1988) continued the discussion, applying the Davies method to some cases and suggesting another approach. In the 1990s interest in the topic has escalated in the econometrics literature with contributions including Bera and Higgins (1992); King and Shively (1993); Andrews and Ploberger (1994); Bera and Ra (1995); Hansen (1996); Bera, Ra and Sarkar (1997); and Conniffe (1998).

The topic, when treated in a general way, can lead to demanding mathematics and sometimes to expressions that are not amenable to algebraic analysis, but only to computationally intensive numerical methods. Yet it is important for any user of this developing methodology to understand the issues and assumptions involved and these do not differ fundamentally between a mathematically easy case and a very difficult one. Most of this paper will concentrate on a simple model that adequately demonstrates the important points and the key differences in approaches.

II NUISANCE PARAMETER UNIDENTIFIABLE UNDER THE NULL

The situation is easiest understood by comparing two simple models

$$y_i = \theta + \phi x_i + e_i \tag{1}$$

$$y_i = \theta x_i^\phi + e_i \tag{2}$$

where the e's are assumed independently normally distributed with variance σ^2 and the null hypothesis is $\theta = 0$. When the null is true model (1) becomes $y_i = \phi x_i + e_i$ and the nuisance parameter can be estimated (probably by maximum likelihood) "under the null", with the estimator denoted $\tilde{\phi}$. But model (2) becomes $y_i = e_i$, which does not contain ϕ and so $\tilde{\phi}$ does not exist. The model

$$y_i = \theta x_{1i} + \theta \phi x_{2i} + e_i \tag{3}$$

is another example and will be examined in detail in this paper. Of course, models with more than one nuisance parameter unidentifiable are easily visualised, for example, the Cobb-Douglas with additive disturbance term

$$y_i = \theta x_{1i}^\phi x_{2i}^\psi + e_i,$$

where neither of the nuisance parameters are identifiable under the null.

The Testing Problem

The three best known approaches for constructing a statistical test are probably the Score (Lagrange Multiplier), the Likelihood Ratio and the Wald procedures. All involve the likelihood function and are closely related. The Score test criterion for testing $\theta = \theta_t$, given a nuisance parameter ϕ is:

$$\frac{\partial l(\theta_t, \tilde{\phi})}{\partial \theta} \tilde{V}^{-1} \frac{\partial l(\theta_t, \tilde{\phi})}{\partial \theta}, \tag{4}$$

where

$$l = \log L(\theta, \phi, y), \tag{5}$$

and

$$V = \text{Var} \left\{ \frac{\partial l(\theta_t, \tilde{\phi})}{\partial \theta} \right\} = I_{\theta\theta}(1 - \rho^2),$$

where

$$\begin{bmatrix} I_{\theta\theta} & I_{\theta\phi} \\ I_{\theta\phi} & I_{\phi\phi} \end{bmatrix} = -E \begin{bmatrix} \dfrac{\partial^2 l}{\partial \theta^2} & \dfrac{\partial^2 l}{\partial \theta \partial \phi} \\ \dfrac{\partial^2 l}{\partial \theta \partial \phi} & \dfrac{\partial^2 l}{\partial \phi^2} \end{bmatrix}, \tag{6}$$

with $\rho^2 = I^2_{\theta\phi} / I_{\theta\theta} I_{\phi\phi}$. Both $I_{\theta\theta}$ and ρ^2 can be functions of ϕ and $\tilde{V}$ replaces ϕ by $\tilde{\phi}$. But, of course, this approach requires $\tilde{\phi}$.

So does the Likelihood Ratio criterion,

$$-2[l(\theta_t, \tilde{\phi}, y) - l(\hat{\theta}, \hat{\phi}, y)],$$

where $\hat{\theta}$ and $\hat{\phi}$ are the unrestricted maximum likelihood estimates, given by the solutions of

$$\frac{\partial l(\theta, \phi)}{\partial \theta} = 0 \quad \text{and} \quad \frac{\partial l(\theta, \phi)}{\partial \phi} = 0.$$

The Wald criterion, which compares $\hat{\theta} - \theta_t$ to its standard error does not require $\tilde{\phi}$. It may require $\hat{\phi}$, as a step in obtaining the corresponding $\hat{\theta}$ and in estimating its variance, but $\hat{\phi}$ can be estimated. The criterion's null distribution may be complicated by effects associated with non-identification of $\tilde{\phi}$, but its relative neglect in the literature may be unjustified and it will be returned to.

III THE ALTERNATIVES IN THE LITERATURE

In the regular case when ϕ is estimable under the null the expectation of the Score

$$\frac{\partial l(\theta, \phi)}{\partial \theta} \tag{7}$$

is usually zero only when $\theta = \theta_t$ *and* ϕ is at its true value, although the expectation is asymptotically zero when $\tilde{\phi}$ replaces ϕ. This is central to the validity of the Score test. However, Davies (1977, 1987) observed that when ϕ is unidentifiable under the null, the expectation of (7) is zero at $\theta = \theta_t$, for arbitrary *constant* ϕ. The variance of (7) is then $I_{\theta\theta}$ and if the y's are independent, so that $l(\theta, \phi) = \sum f(\theta, \phi, y_i)$, the central limit ensures that

$$\frac{\partial l(\theta_t, \phi)}{\partial \theta} V^{-1} \frac{\partial l(\theta_t, \phi)}{\partial \theta} \tag{8}$$

is asymptotically χ_1^2. Davies considered conducting a (possibly infinite) series of tests over the whole possible range of values of ϕ, which is equivalent to basing the test on the maximum of (8) with respect to ϕ. The null distribution of this maximum is no longer χ_1^2, however. Davies proceeded by treating (8) as a random function, or stochastic process, of ϕ, which could be assumed Gaussian for large sample size and considering the probability of the process crossing a barrier. Very similar approaches have been used in recent years in obtaining asymptotic critical values for unit root tests and tests of structural change. However, except for special cases such as model (3), the task of obtaining critical values can be analytically intractable. Davies employed bounds to significance levels, while Hansen (1996) used transformation and simulation to estimate critical values.[1]

Other authors have suggested alternatives to the Davies procedure. Godfrey (1988) suggested choice of some (constant) ϕ, so (8) is asymptotically χ_1^2 under the null, but the power of the test depends on choice of ϕ. Bera, Ra and Sarkar (1997) have reviewed examples where this approach has appeared in the literature, including some that amount to the choice of $\phi = 0$. Conniffe (1998) argued that although $\tilde{\phi}$ is unobtainable, $\hat{\phi}$, the maximum likelihood estimator under the alternative $\theta \neq \theta_t$, is estimable and substituted it into (7). This could have been done even if ϕ was identifiable under the null and then the modified score test would be obtained by adjusting the variance of (7) to allow for the estimation of ϕ. Since

$$l(\theta, \hat{\phi}) = l(\theta, \phi) + (\hat{\phi} - \phi)\frac{\partial l}{\partial \phi} + \tfrac{1}{2}(\hat{\phi} - \phi)^2 \frac{\partial^2 l}{\partial \phi^2} + ...,$$

where ϕ now means the true value,

---

1.   A referee has remarked that that the Davies test can be bootstrapped fairly easily in many cases.

$$\text{Var}\left\{\frac{\partial l(\theta_t, \hat{\phi})}{\partial\theta}\right\} \approx \text{Var}\left\{\frac{\partial l(\theta_t, \phi)}{\partial\theta} + (\hat{\phi} - \phi)\frac{\partial^2 l(\theta_t, \phi)}{\partial\theta\partial\phi}\right\}$$

Now approximating the second derivative by $-I_{\theta\phi}$ and noting that the covariance of

$$\frac{\partial l(\theta_t, \phi)}{\partial\theta} \quad \text{and} \quad \hat{\phi} - \phi$$

is asymptotically zero, the variance is approximately

$$I_{\theta\theta}\left(1 + \frac{\rho^2}{1-\rho^2}\right) = \frac{I_{\theta\theta}}{1-\rho^2} = \hat{V}, \text{ say,}$$

and the test criterion is

$$\frac{\partial l(\theta_t, \hat{\phi})}{\partial\theta}\hat{V}^{-1}\frac{\partial l(\theta_t, \hat{\phi})}{\partial\theta}. \tag{9}$$

The test criterion (9) differs from (4), except when ρ^2 is zero, because $\hat{\phi}$ rather than $\tilde{\phi}$ is inserted in (7) and the variance modification involves division by $1-\rho^2$ rather than multiplication by it. When ϕ is unidentifiable under the null, (4) cannot be calculated , but (9) can and it will be applied to model (3) in the next section. For a fairly general regression model with some parameters identifiable under the null and some not, Conniffe (1998) compared the appropriate generalisation of (9) to the Davies test statistic and shows its advantages, in at least some circumstances, of robustness, power and even simplicity. It was also remarked that the generalisation of (9) is often of Wald test form. However, comparisons in general cases are not mathematically trivial, but the key differences between test procedures can be illustrated through analysis of model (3), which permits simple algebraic analysis that can be seen as intuitively plausible.

IV THE SIMPLE EXPOSITORY MODEL

Model (3) has already been employed by Godfrey (1988) to illustrate the Davies test and his own idea of assuming some constant ϕ. Here the illustration will be extended to the modified score test (9) and then the three tests will be compared. The log likelihood is:

$$l = -\frac{n}{2}\log \pi\sigma^2 - \frac{1}{2\sigma^2}\Sigma(y_i - \theta x_{1i} - \theta\phi x_{2i})^2 .$$

So

$$\frac{\partial l}{\partial \theta} = \frac{1}{\sigma^2}\Sigma(y_i - \theta x_{1i} - \theta\phi x_{2i})(x_{1i} + \phi_{2i}),$$

which at $\theta = 0$ is

$$\frac{1}{\sigma^2}\Sigma y_i(x_{1i} + \phi x_{2i}). \tag{10}$$

Note that (10), the Score at $\theta = 0$ contains ϕ even though it is not identifiable under the null, since the model is then just $y_i = e_i$, but that the expectation of (10) is zero whatever the value of ϕ. The σ^2 parameter is estimable under the null, of course.

The variance of (10) is

$$\frac{1}{\sigma^2}\Sigma(x_{1i} + \phi x_{2i})^2 . \tag{11}$$

So for any constant ϕ

$$\frac{\left[\Sigma y_i(x_{1i} + \phi x_{2i})\right]^2}{\sigma^2 \Sigma(x_{1i} + \phi x_{2i})^2} \tag{12}$$

would (when σ^2 is replaced by its estimate under the null $\Sigma y^2/n$) be distributed as asymptotically χ_1^2 (actually $F_{1,n}$ in finite samples). The simple test suggested by Godfrey just makes an arbitrary *single* choice of ϕ in (12). The Davies test criterion maximises (12) with respect to ϕ (assuming no limitations on its range) and this works out to be

$$\frac{1}{\sigma^2}(b_1^*\Sigma x_1 y + b_2^*\Sigma x_2 y), \tag{13}$$

where b_1^* and b_2^* are the OLS estimators of coefficients in a regression of y on x_1 and x_2 and so (13) is asymptotically χ_2^2 under the null, or F with 2 and n-2 degrees of freedom in finite samples. This is intuitively plausible, since the choice of models is between $y_i = e_i$ and $y = b_1 x_1 + b_2 x_2 + e$, with $b_1 = \theta$ and $b_2 = \theta\phi$.

By obtaining the second derivatives of the likelihood and taking expectations

$$\rho^2 = \frac{\left[\sum x_{2i}(x_{1i} + \phi x_{2i})\right]^2}{\sum x_{2i}^2 \sum (x_{1i} + \phi x_{2i})^2}.$$

Substituting $\hat{\phi}$ into (10), squaring and dividing by its variance, which is (11) divided by $1-\rho^2$ (again inserting $\hat{\phi}$ where required) gives the Score test statistic:

$$\frac{\left[\sum y_i(x_{1i} + \hat{\phi} x_{2i})\right]^2}{\sigma^2 \sum (x_{1i} + \hat{\phi} x_{2i})^2} \left(1 - \frac{\left[\sum x_{2i}(x_{1i} + \hat{\phi} x_{2i})\right]^2}{\sum x_{2i}^2 \sum (x_{1i} + \hat{\phi} x_{2i})^2}\right). \qquad (14)$$

From $y = b_1 x_1 + b_2 x_2 + e$, it is clear that the MLE's of the coefficients are b_1^* and b_2^* and so the MLE of ϕ is b_2^* / b_1^*, which in this model is also the value that maximises (12). Then (14) simplifies greatly to:

$$\frac{1}{\sigma^2}\left[b_1^* \sum x_1 y + b_2^* \sum x_2 y - \frac{(\sum x_2 y)^2}{\sum x_2^2}\right].$$

This is the regression sum of squares for fitting both x_1 and x_2 minus the regression sum of squares for fitting x_2 alone which, of course, is the sum of squares for testing x_1 given x_2 fitted and equals

$$\frac{(b_1^*)^2}{\text{var}(b_1^*)}, \qquad (15)$$

the square of the usual "t" value. Its null distribution is asymptotically χ_1^2, or $F_{1,n-2}$, in finite samples. Since the MLE of θ is b_1^*, (15) is also the Wald test, which is not surprising given the equivalence of Wald and Score tests in standard linear regression. A point worth noting is that (15) is a valid test for θ ($=b_1$) = 0, whether $b_2 (= \theta\phi) = 0$ or not.

V COMPARING TESTS

The three possible tests are given by (12), (13) and (15). The first assumes a value for ϕ and is $F_{1,n-1}$ if some assumptions hold. The second avoids the issue of the value of ϕ by testing over its full range and is $F_{2,n-2}$, again if assumptions are correct. The third estimates ϕ by its unrestricted MLE and is $F_{1,n-2}$ and does not require one key assumption of the other two tests.

Robustness to Model Specification

Implicit in model (3)

$$y_i = \theta x_{1i} + \theta\phi x_{2i} + e_i,$$

is the idea that the coefficients of both x_1 and x_2 become zero at $\theta = 0$. But it is clear that *any* two variable regression model

$$y_i = \theta x_{1i} + \zeta x_{2i} + e_i,$$

where ζ is not zero if θ is zero, can be written in the form of model (3) by taking $\phi = \zeta / \theta$. As $\theta \to 0$ ϕ becomes ever larger, so that the true model with $\theta = 0$ remains $y_i = \zeta x_{2i} + e_i$. Perhaps economic theory strongly suggests that the model is properly formulated in the sense that $\zeta = 0$ when $\theta = 0$, but model misspecification is a constant danger in econometrics . If ζ is non-zero when $\theta = 0$, both the test statistics (12) and (13) are non-central chi-squared with the non-centrality parameter a function of ζ and the Type 1 error probability tending to unity for large n. The Score test (15), however, still has its correct significance level, because, unlike (13), the sum of squares corresponding to x_2 has been subtracted out. Conniffe (1998) shows that this property of the Score test extends to more general models and that the Score test statistic can be seen as a correction to the Davies statistic to ensure robustness. A situation where (12) or (13) were highly significant, but (15) not nearly significant, would strongly suggest misspecification. If (12) or (13) were just significant (at 5 per cent, say) and (15) a little short of significance, it might be plausible to take the specification as correct and attribute the discrepancy to greater power for tests (12) and (13). That, however, raises the question of how test powers compare when the specification is correct.

Relative Power of Tests

Test powers depend on the true values of parameters and on the variation in x_1 and x_2 and the correlation between them, but these factors are encapsulated by the non-centrality parameters. For the F test based on (12), with 1 and n-1 degrees of freedom, the non-centrality parameter is:

$$\frac{\theta^2 [\sum (x_1 + \phi_a x_2)(x_1 + \phi x_2)]^2}{\sigma^2 \sum (x_1 + \phi_a x_2)^2},$$

where ϕ_α is the value substituted for the unknown true ϕ. The non-centrality parameter of the F test based on (13), with 2 and n-2 degrees of freedom, is

$$\frac{\theta^2}{\sigma^2}[\sum x_1^2 + 2\phi\sum x_1 x_2 + \phi^2 \sum x_2^2],$$

while that for the Score F test based on (15), with 1 and n-2 degrees of freedom, is:

$$\frac{\theta^2}{\sigma^2}\left[\sum x_1^2 - \frac{(\sum x_1 x_2)^2}{\sum x_2^2}\right] = \frac{\theta^2}{\sigma^2}\sum x_1^2 (1-r^2),$$

where $r = \sum x_1 x_2 / \sqrt{\sum x_1^2 \sum x_2^2}$. For convenience, assume $\sum x_1 x_2 = 0$ and rescale the x's so that $\sum x_1^2 = \sum x_2^2 = n\sigma^2$. Then the non-centrality parameters become

$$n\theta^2 \frac{(1+\phi\phi_a)^2}{1+\phi_a^2}, \quad n\theta^2(1+\phi^2) \text{ and } n\theta^2$$

respectively. The comparison between (12) and (15) is simplest, because both are single degree of freedom (for numerators) tests so the larger non-centrality parameter means the most powerful test (neglecting n-1 v n-2 in denominators). Test (12) is more powerful than test (15) if the ratio

$$\frac{(1+\phi\phi_a)^2}{1+\phi_a^2} > 1.$$

It clearly will be if the guess is correct, although there will be little to gain if ϕ is small. If $\phi = .1$ and the guess correct the ratio is 1.01. If the guess is wrong, say .5, (12) could be considerably less powerful than (15) because the ratio is .8. On the other hand, if $\phi = .5$ and the guess is right, (12) is better since the ratio is 1.25. If the guess is 1.0, the ratio is 1.125, so (12) still has some advantage, but if it is 2.0, the ratio is .8 and so (15) is better.

Comparisons of tests based on (13) and (15) require actual evaluations of power because degrees of freedom differ as well as non-centrality parameters. The choice of θ is important for comparisons. There is little point in comparing tests at large values of θ, because both powers will be almost unity, or at very small values, when both powers are near zero. The tests should be compared at intermediate levels of power, where there is scope for appreciable differences between them. For a "t" test the 5 per cent critical value is about 2 and the probability that a (symmetrically distributed) statistic exceeds its own mean is .5, so taking $\theta = 2/\sqrt{n}$ makes the power of the Score test for a 5 per cent significance level approximately .5, since the F ratio is just the square of a "t" statistic. Table 1 compares test powers.

Table 1: *Powers of Score and Davies Tests for* $\alpha=.05$

$F_{1,n-2}$ (Score test)		$F_{2,n-2}$						
		$\phi^2 = 0$	.05	.1	.2	.3	.4	.5
$\theta^2=.2$ (n=20)	.474	.359	.375	.390	.422	.452	.481	.510
$\theta^2=.08$ (n=50)	.500	.393	.411	.428	.461	.494	.525	.556
$\theta^2=.04$ (n=100)	.500	.403	.420	.438	.472	.506	.538	.568

The Score test is more powerful than the Davies test, especially for low degrees of freedom, unless values of ϕ are much larger than θ. However, the case of $\sum x_1 x_2 = 0$ is rather favourable to the score test. For $\sum x_1 x_2 \neq 0$, the power of the Score test decreases as r^2 increases from zero. The power of the rival test also falls if r is negative (assuming ϕ positive) and the pattern of Table 1 persists. Indeed, for n = 20, r = −.5, the Score test is also more powerful at $\phi^2 = .4$ and .5 . However, for positive r, the value of ϕ^2 below which the Score test is more powerful, decreases rapidly with r. So for r = .3 the Score test becomes less powerful at $\phi^2 = .1$ and at r = .5 it is less powerful even at $\phi^2 = .05$. There is no uniformly most powerful test and the best choice depends on the circumstances.

What is clear is that even if the model (3) is known to be perfectly specified, the Score test (15) is sometimes preferable to (13) on power grounds. As regards Godfrey's test (12), it has already been seen that it can be better or worse than the Score test , depending on whether a good or bad guess is made for ϕ. Without prior knowledge of some sort, guessing seems a dubious way to proceed. If there is any possibility that model (3) is incorrectly specified in the sense that $\zeta(=\theta\phi)$ is non-zero under the null, preference must swing decisively to the score test, since it is then the only valid test. Overall, the Score test seems to be the preferable procedure.

These remarks apply to more general models, although mathematical difficulties complicate comparisons. The null distributions need not be χ^2 asymptotically or F in finite samples and can involve mixtures of distributions, even in only moderately complex models, although the generalisation of (15) can be much simpler than that of (13) in this regard (Conniffe, 1998). The virtue of model (3) in this paper has been its tractability. It does, however, have credibility in suggesting the wider scope of the findings, because a class of non-linear models can be approximated by model (3).

$$y_i = \theta g(\phi, x_i) + e_i \approx \theta g(\hat{\phi}, x_i) + \theta(\phi - \hat{\phi})g'(\hat{\phi}, x_i) + e_i$$
$$\approx \theta(\hat{g}_i - \hat{\phi}\hat{g}_i') + \theta\phi\hat{g}_i' + e_i,$$

or

$$y_i \approx \theta z_{1i} + \zeta z_{2i} + e_i,$$

where prime denotes derivative, $\zeta = \theta \phi$ and z_{1i} and z_{2i} are not functions of parameters.

REFERENCES

ANDREWS, D.W.K., and W. PLOBERGER, 1994. "Optimal Tests when a Nuisance Parameter is Present only Under the Alternative", *Econometrica*, Vol. 62, pp. 1383-1414.

BERA, A.K., and M.L. HIGGINS, 1992. "A Test for Conditional Heteroscedasticity in Time Series Models", *Journal of Time Series Analysis*, Vol. 13, pp. 501-519.

BERA, A.K., and S. RA, 1995. "A Test for the Presence of Conditional Heteroscedasticity Within Arch-M Framework", *Econometrics Reviews*, Vol. 14, pp. 473-485.

BERA, A.K., S. RA, and N. SARKAR, 1997. "Hypothesis Testing for Some Nonregular Cases in Econometrics", in S. Chakravarty, D. Coondoa and R. Mukerjee (eds.), *Quantitative Economics — Theory and Practice; Essays in Honour of Professor N. Bhattacharya*, New Delhi: Allied Publishers, pp. 324-356.

CONNIFFE, D., 1998. "Score Tests when a Nuisance Parameter is Unidentified Under the Null Hypothesis", *Journal of Statistical Planning and Inference*, in press.

DAVIES, R.B., 1977. "Hypothesis Testing when a Nuisance Parameter is Present Only Under the Alternative", *Biometrika* , Vol. 64, pp. 247-254.

DAVIES, R.B., 1987. "Hypothesis Testing when a Nuisance Parameter is Present Only Under the Alternative", *Biometrika*, Vol. 74, pp. 33-43.

ENGLE, R.F., 1984. "Wald, Likelihood Ratio and Lagrange Multiplier Tests in Econometrics", in Z. Griliches and M.D. Intriligator (eds.), *Handbook of Econometrics* Vol. 2, New York: North-Holland, pp. 776-826.

GODFREY, L.G., 1988. *Misspecification Tests in Econometrics: The Lagrange Multiplier Principle and Other Approaches*. Cambridge: Cambridge University Press.

HANSEN, B.E., 1996. "Inference when a Nuisance Parameter is not Identified Under the Null Hypothesis", *Econometrica* Vol. 64, pp. 413-430.

KING, M.L., and T.S. SHIVELY, 1993. "Locally Optimal Testing when a Nuisance Parameter is Present Only Under the Alternative", *The Review of Economics and Statistics*, Vol. 75, pp. 1-7.

WATSON, M.W., and R.F. ENGLE, 1985. "Testing for Regression Coefficient Stability with a Stationary AR(1) Alternative", *Review of Economics and Statistics*, Vol. 67, pp. 341-346.