SERIE RESEARCHMEMORANDA

COMPUTING WALD CRITERIA FOR NESTED HYPOTHESES

D.A. Kodde and F.C. Palm

Researchmemorandum 1985-16

VRIJE UNIVERSITEIT
EKONOMISCHE FAKULTEIT
AMSTERDAM

COMPUTING WALD CRITERIA FOR NESTED HYPOTHESES

D.A. Kodde* and F.C. Palm**

Comments welcome
March 1985

We present a numerically convenient procedure for computing Wald criteria for nested hypotheses. Similar to Szroeter's (1983) generalized Wald test, the suggested procedure does not require explicit derivation of the restrictions implied by the null hypothesis and hence its use might eliminate an intricate step in testing linear and nonlinear hypotheses.
We show that the traditional Wald test, Szroeter's (1983) generalized Wald test and our procedure are asymptotically equivalent under H_0. A class of nonlinear transformations of the restrictions for which the Wald statistic is asymptotically invariant is discussed. Finally, we illustrate the use of our procedure for testing the common factor restrictions in a dynamic regression model.

* Katholieke Universiteit, Postbus 9108, 6500 HK Nijmegen.

** Vrije Universiteit, Postbus 7161, 1007 MC Amsterdam.

The research reported in this paper was initiated when the first author was affiliated with the Vrije Universiteit. A computer program which implements the general procedure has been written in FORTRAN and is available on request. An earlier version of the paper was presented at the ESEM 1983 in Pisa.

1. Introduction

The Wald test (see Wald (1943)) is a very useful tool in empirical econometrics. For computational convenience, a Wald test will be preferred to a likelihood ratio test or a score test, when estimates of the unrestricted parameters can be easily obtained. For instance, this is frequently the case in a specification analysis or 'top-down'approach to model-building, where a fairly general model is taken as the maintained hypothesis throughout the modeling process. Also, a Wald test can be used when consistent but not fully efficient parameter estimates are available whose asymptotic distribution is known (see e.g. Stroud (1971)).

In this paper, we present a procedure for the computation of the Wald criteria when testing nested hypotheses. The suggested procedure does not require explicit derivation of the restrictions implied by the null hypothesis and hence its use might eliminate an intricate step in testing linear and nonlinear nested hypotheses. We show that the traditional Wald test, which can be computed if the restrictions are expressed explicitly, Szroeter's (1983) generalized Wald method and our procedure asymptotically yield the same value for the statistic under the null hypothesis. In practice it is often possible to express the restrictions under the null hypothesis in various forms. For the three statistics, we discuss a general class of nonlinear transformations of the restrictions, which yield the same value for the Wald statistic in large samples.

The plan of the paper is as follows. In section 2, we present our procedure for testing nested hypotheses. For the ease of reference, we briefly describe the generalized Wald test and we introduce some basic notation. The asymptotic equivalence of the three statistics is established in section 3. Then, a class of nonlinear transformations of the restrictions for which the Wald statistic is invariant, is discussed. In section 4, we consider the implications of a lack of global identification of the model under the null hypothesis for our procedure and the generalized Wald method. Section 5 contains an example which illustrates how the Wald statistic can be computed in a fairly straightforward way for common factor restrictions in a dynamic regression model. Finally, in section 6 we briefly present some conclusions.

2. Wald criteria for nested hypotheses

In this section, we present a procedure to compute Wald criteria, which does not require explicit derivation of the restrictions implied by the null hypothesis. For the ease of reference, we also give the generalized Wald test proposed by Szroeter (1983) in our notation.

Let us assume that we have a model defined in terms of n parameters forming a vector θ, and that $\hat{\theta}$ is some consistent asymptotically normally distributed estimate of θ_0 such that $\sqrt{T}(\hat{\theta} - \theta_0)$, with T being the sample size and θ_0 being the true value of θ, has a covariance matrix Ω_θ which can be consistently estimated by $\hat{\Omega}_\theta$. A nested null hypothesis H_0 implies a set of constraints on θ

$$h(\theta) = 0 \ , \tag{2.1}$$

which form a vector of r independent, continuously differentiable functions. Under the alternative hypothesis , the equality in (2.1) does not hold true.
The Wald statistic for testing the set of restrictions is

$$W = T\, h(\hat{\theta})' \hat{\Omega}_h^{-1} h(\hat{\theta}) \ , \tag{2.2}$$

where

$$\hat{\Omega}_h = \left(\frac{\partial h}{\partial \theta'}\right) \hat{\Omega}_\theta \left(\frac{\partial h}{\partial \theta'}\right)' \ , \tag{2.3}$$

with $\frac{\partial h}{\partial \theta'}$ denoting the first derivative matrix of h with respect to θ which we evaluate at $\hat{\theta}$.

On the null hypothesis that all the constraints (2.1) are satisfied, W is χ^2-distributed in large samples with r degrees of freedom, provided that plim $\hat{\Omega}_h$ is nonsingular and that $\frac{\partial h}{\partial \theta'}$ is a continuous function of θ at the true parameter value θ_0. In the sequel, we denote the first and second partial derivatives of y with respect to a vector x' by $D_x y$, with y being a scalar or a vector, and by $D^2_{xx} y$ respectively, when y is a scalar. Finally, 'o_p' denotes the order in probability.

For a given set of restrictions, the Wald statistic is easily computed. Explicit derivation of the restrictions, however, can be

tedious and intricate. The method we propose here simplifies explicit formulation of the restrictions. We show how $h(\hat{\theta})$ and $D_\theta h$ can be determined by implicitly using the restrictions. Once $h(\hat{\theta})$ and $D_\theta h$ have been computed, the Wald statistic (2.2) can be directly obtained.
In empirical work, the restrictions implied by H_0 are usually given in the form of

$$f(\beta,\theta) = 0 \quad , \tag{2.4}$$

where β is a vector of m parameters of the restricted model, f is a continuously differentiable mapping from an $m+n$ dimensional space into an $m+r$ dimensional one. The $m+r$ relations in (2.4) are implicit if H_0 is true. From the system in (2.4), we now choose m equations, $f_1(\beta,\theta) = 0$, such that β can be solved explicitly as a function of θ , that is $\beta = \beta(\theta)$. This solution is substituted in the r remaining relations that we denote by $f_2(\beta,\theta) = 0$ to give

$$h(\theta) = f_2(\beta(\theta),\theta) = 0 \ . \tag{2.5}$$

As indicated above, we only need the restrictions and the corresponding partial derivatives both evaluated at $\hat{\theta}$ to compute the Wald statistic.
First, we determine $h(\hat{\theta})$ along the lines just described, which means we solve $f_1(\beta,\hat{\theta}) = 0$ for β to get $\hat{\beta}$ and substitute it into (2.5).
Next, we obtain an expression for the partial derivatives evaluated at $\hat{\theta}$.
For the sake of simplicity, we define the following matrices $D_\beta f = F$, $D_\theta f = Q$, $D_\beta f_i = F_i$, $D_\theta f_i = Q_i$, $i = 1,2$, where the arguments β and θ have been deleted. When we evaluate these matrices at $\hat{\theta}$ and $\beta(\hat{\theta})$, we use the notation $\hat{F}$, $\hat{Q}$, $\hat{F}_i$ and $\hat{Q}_i$ respectively. Assuming that f_1 has been chosen such that F_1 is continuous and nonsingular at (β_0,θ_0) , we have as a result from the implicit function theorem (see e.g. Rudin (1976)) that the solution of (2.5) is continuous and differentiable in θ with first derivative given by

$$D_\theta \beta(\theta) = - F_1^{-1} Q_1 \ . \tag{2.6}$$

If the matrix F_1 is nonsingular at (β_0,θ_0), there exists only one solution to $f_1(\beta,\theta) = 0$ in some neighborhood of (β_0,θ_0).
Applying the chain-rule of differentiation to (2.5) and using expression (2.6), the partial derivatives of h become

$$D_\theta h = - F_2 F_1^{-1} Q_1 + Q_2 = H Q \quad , \tag{2.7a}$$

with $H = [-F_2 F_1^{-1} \vdots I_r]$. As a result of the implicit function theorem, $D_\theta h$ is continuous in θ at θ_0 .
When we evaluate (2.7a) at a consistent estimate of θ, we get (see e.g. Billingsley (1968)) under H_0

$$\hat{D}_\theta h = HQ + o_p(1) \quad , \tag{2.7b}$$

with H and Q being evaluated at (β_0, θ_0).
Formulae (2.5) and (2.7) are suited for various kinds of nested hypotheses. However, quite often the set of restrictions (2.4) has the special form, $f(\beta) - \theta = 0$, so that expression (2.7a) can be simplified. For instance, the constraints implied by the common factor structure (e.g. Sargan (1977), 1980a)), the polynomial distributed lags (e.g. Almon (1965) and Sargan (1980b)) and the rational expectations restrictions on the reduced form of a simultaneous equation model (e.g. Hoffman and Schmidt (1981)) are of this special form. For this form of the implicit relations, $Q = -I_n$, so that we obtain:

$$\begin{aligned} h(\hat{\theta}) &= f_2(\hat{\beta}) - \hat{\theta}_2 \\ \text{and} \qquad & \qquad (2.8) \\ D_\theta h &= - H \quad , \end{aligned}$$

with $\hat{\theta}_2$ being the appropriate subvector of $\hat{\theta}$.
A procedure for computing Wald tests for different kinds of nested hypotheses consists in:

1) choosing a set of m equations f_1, solving them for $\hat{\beta}$ for a given $\hat{\theta}$ and substituting $\hat{\beta}$ in f_2 to obtain $h(\hat{\theta}) = f_2(\beta(\hat{\theta}), \hat{\theta})$,
2) computing the matrices F_i and Q_i , $i = 1, 2$, to obtain $\hat{D}_\theta h$ in (2.7b), and
3) calculating the value of W in (2.2).

To conclude, the approach yields a convenient procedure to compute Wald criteria. It also accommodates sequential testing (see e.g. Smith (1983)), when f_2 is successively extended, given the choice of f_1 and the parametrization θ , β .

The generalized Wald test proposed by Szroeter (1983) for the set of restrictions (2.4) can be obtained as follows. Given $\hat{\theta}$, a consistent estimator β is found by minimizing

$$f(\beta, \hat{\theta})' S \, f(\beta, \hat{\theta}) \tag{2.9}$$

with respect to β, where S is a positive semi-definite symmetric matrix such that F'S F has rank r. The requirement that rank F'S F = r is a generalization of Szroeter (1983) since he chooses a matrix S with rank m+r . Notice that the estimate which minimizes (2.9) is the asymptotic least squares estimate (see Gouriéroux et al. (1983)). Applying the implicit function theorem to the first order conditions for a minimum, $F'S\, f(\beta,\hat{\theta}) = 0$, we get

$$\hat{\beta} - \beta_0 = P\, Q(\hat{\theta} - \theta_0) + o_p(T^{-1/2}) \quad , \tag{2.10}$$

with $P = -(F'S\ F)^{-1}\ F'S$. The mean value theorem applied to f at the true parameters yields

$$\begin{aligned} f(\hat{\beta},\hat{\theta}) &= [I + \tilde{\tilde{F}}\tilde{P}]\ \tilde{Q}\ (\hat{\theta} - \theta_0) \quad , \\ &= [I + FP]\ Q\ (\hat{\theta} - \theta_0) + o_p(T^{-1/2}) \quad , \end{aligned} \tag{2.11}$$

where a tilde '~' denotes evaluation at a suitable point between $(\beta(\hat{\theta}),\hat{\theta})$ and (β_0,θ_0).
The generalized Wald test is now given by

$$W_g = T\, f(\hat{\beta},\hat{\theta})'\hat{\Omega}^{-}\, f(\hat{\beta},\hat{\theta}) \quad , \tag{2.12}$$

where $\hat{\Omega}$ denotes the matrix $\Omega = (I+FP)Q\ \Omega_\theta\ Q'(I+FP)'$ evaluated at $(\beta(\hat{\theta}),\hat{\theta})$. As a result of the continuity of the derivatives of f and of Slutsky's theorem, $\hat{\Omega}$ is $o_p(1)$, and (2.12) can be expressed as

$$W_g = T\, f(\hat{\beta},\hat{\theta})'\Omega^{-}\, f(\hat{\beta},\hat{\theta}) + o_p(1) \quad . \tag{2.13}$$

To finish this section, we comment on the implementation of Szroeter's (1983) procedure in practice.
When $S = [\hat{Q}\ \hat{\Omega}_\theta \hat{Q}]^{-1}$, the asymptotic covariance matrix of $f(\hat{\beta},\hat{\theta})$ in (2.11) is

$$[Q\Omega_\theta Q' - F\{F'(Q\Omega_\theta Q')^{-1}F\}^{-1}\ F'] \quad , \tag{2.14}$$

and S is a g-inverse of this covariance matrix evaluated at $(\hat{\beta},\hat{\theta})$, so that the generalized Wald test (2.12) becomes

$$W_g = T\, f(\hat{\beta},\hat{\theta})'[\hat{Q}\ \hat{\Omega}_\theta \hat{Q}']^{-1} f(\hat{\beta},\hat{\theta}) = T\, f(\hat{\beta},\hat{\theta})'S\ f(\hat{\beta},\hat{\theta}) \quad . \tag{2.15}$$

W_g is Szroeter's objective function (2.9) evaluated at the

minimum for β and multiplied by T. Expression (2.15) gives an alternative way of computing Wald crieteria. Notice, however, that Q may depend on β so that a consistent estimate of β is required for obtaining S in (2.15).

Of course if the restrictions are of the form $f(\beta) - \theta = 0$, $Q = -I_n$ and the computation of the Wald test using (2.15) is straightforward. To summarize the practical implications, Szroeter's procedure requires computing the global minimum of (2.9), whereas our procedure requires obtaining the solutions of $f_1(\beta,\hat{\theta}) = 0$ and checking whether they satisfy $f_2(\beta,\hat{\theta}) = 0$. Of course our procedure stops as soon as H_0 is not rejected for a given solution. Notice that solving $f_1(\beta,\hat{\theta}) = 0$ corresponds to minimizing (2.9) for diagonal S with a one on the diagonal when the corresponding equation of f is included in f_1 and zero otherwise.

3. Asymptotic equivalence relationships

In this section, we investigate whether the value of the Wald statistic is affected by choosing alternative formulations for the constraints. We give a general class of nonlinear transformations of the restrictions for which the value of the traditional and generalized Wald statistics is asymptotically invariant under H_0. Furthermore, we consider the influence of the choice of f_1 and f_2 on the Wald test. Finally, we show that our procedure is asymptotically equivalent with the traditional and the generalized Wald tests.

3.1 Transforming the restrictions

Consider the case where the set of restrictions $h(\theta) = 0$ is such that Ω_h is nonsingular. As can be seen from (2.2) and (2.3), an alternative formulation of the restrictions say $g(\theta) = 0$, for which there exists a nonsingular matrix A such that $D_\theta = AD_\theta h$ will asymptotically yield the same value for the Wald statistic, both under H_0 and under a sequence of local alternative hypotheses. This result, which we call the equivalence condition of the partial derivatives, directly follows from the lemma of Holly and Monfort (1982), that we give in appendix I. That the identity for the Wald statistic usually does not hold true when there exists no matrix A that transforms $D_\theta h$ into $D_\theta g$ can be seen by showing that the plim of the difference between the two Wald statistics is nonzero.
Given the set of restrictions $h(\theta) = 0$, we consider a transformation $g(h(\theta),\theta)$, with $g(h(\theta),\theta) = 0$ if and only if $h(\theta) = 0$, g having continuous first and second derivatives and $D_y\, g(y,\theta)$ being nonsingular at $(0,\theta_0)$. Then, h and g yield the same value for W in large samples. This result follows from the equivalence condition of the partial derivatives. The matrices of partial derivatives of h and g with respect to θ are given by

$$D_\theta h(\theta) \quad \text{and} \quad D_y\, g(y,\theta)\, D_\theta y + D_\theta\, g(y,\theta)\,. \tag{3.1}$$

But on H_0, as a result of Slutsky's theorem, we have

$$\operatorname{plim} D_\theta g(y,\hat{\theta}) = \operatorname{plim} D_\theta\, g(0,\hat{\theta}) = D_\theta\, g(0,\theta_0) = 0, \tag{3.2}$$

where $\hat{\theta}$ is a consistent estimate of θ. The second term of the derivative of g with respect to θ in (3.1) vanishes in large samples and we obtain the asymptotic invariance of the Wald

statistic with respect to transformations of the type $g(h(\theta),\theta)$.

Next, we consider some equivalence properties of the generalized Wald test. First, Szroeter (1983) shows that the asymptotic local power of his test does not depend on the particular choice of S. The asymptotic efficiency of $\hat{\beta}$, however, depends on S. In fact $S = [\hat{Q}\ \hat{\Omega}_\theta\ \hat{Q}']^{-1}$ maximizes the asymptotic efficiency of $\hat{\beta}$, which then is an optimal asymptotic least squares estimate. Second, we consider general transformations of $f(\beta,\theta) = 0$ which take the form $g(f(\beta,\theta), \beta, \theta)$, with

$$g\ (f\ (\beta,\theta),\ \beta,\ \theta) = 0 \tag{3.3}$$

if and only if $f(\beta,\theta) = 0$. Furthermore, g has continuous first and second derivatives and $D_y\ g(y,\beta,\theta)$ is nonsingular at $(0,\beta_0,\theta_0)$. Again, we will show that in large samples f and g yield the same value for the generalized Wald test. Without loss of generality, we only consider the case where the optimal weighting matrix S is chosen. When g is evaluated at the optimal asymptotic least squares estimator $\beta(\hat{\theta})$, the matrix of partial derivatives of g with respect to θ is given by

$$D_y g(y,\beta,\theta)[FD_\theta\ \beta + Q] + D_\beta g(y,\beta,\theta)D_\theta\beta + D_\theta g(y,\beta,\theta). \tag{3.4}$$

But on H_0, as a result of Slutsky's theorem and similar to the analysis in (3.2), the second and third term of (3.4) converge to zero, when evaluated at a consistent estimate θ. In addition, the difference between $D_\theta\beta$ based on f and g respectively, vanishes in large samples (see also Gouriéroux et al. (1983)). Therefore,

$$[D_y\ g(\hat{y},\hat{\beta},\hat{\theta})]^{-1} D_\theta g(f(\hat{\beta},\hat{\theta}),\hat{\beta},\hat{\theta}) = [I + F\ P]Q + o_p(1)\ , \tag{3.5}$$

and the lemma by Holly and Monfort (1982) establishes the asymptotic invariance of the generalized Wald test for transformations of the type mentioned above.

3.2 The choice of f_1

Next, we analyze the consequences of the partition of f into f_1 and f_2 for the value of the Wald statistic. Without loss of generality, we only consider two alternative choices for f_1 and f_2. To do so, we partition the system of constraints into four subsets, which consist of k, m-k, k and r-k relations respectively

$$
\begin{aligned}
f_1^* (\beta,\theta) &= 0 \\
f_2^* (\beta,\theta) &= 0 \\
f_3^* (\beta,\theta) &= 0 \\
f_4^* (\beta,\theta) &= 0
\end{aligned}
\tag{3.6}
$$

To simplify the notation, we delete the arguments β and θ and we denote the subset of restrictions $\begin{pmatrix} f_i^* \\ f_j^* \end{pmatrix}$ by f_{i+j}^* and its partial derivatives with respect to β and θ by F_{i+j} and Q_{i+j} respectively.

As our choice of $f_1 = 0$, we use the sets $f_{1+2}^* = 0$ and $f_{2+3}^* = 0$ respectively to derive a solution for β. Using the result in (2.7a), the partial derivatives can be written as

$$D_\theta h_1 = [- F_{3+4} F_{1+2}^{-1} Q_{1+2} + Q_{3+4}] \tag{3.7}$$

and

$$D_\theta h_2 = [- F_{1+4} F_{2+3}^{-1} Q_{2+3} + Q_{1+4}] \quad , \tag{3.8}$$

where the subscript $i = 1,2$ indicates the choice of f_1.

The value of the Wald statistic will asymptotically not be affected by the choice of f_1, if there exists a nonsingular matrix A such that the partial derivatives in (3.7) and (3.8) satisfy the equivalence condition, $D_\theta h_2 = A D_\theta h_1$. A nonsingular matrix that gives the desired result is

$$\underset{r\times r}{A} = \left[- F_{1+4} B_2 \;\vdots\; \begin{matrix} 0_{k\ r-k} \\ \vdots \\ I_{r-k} \end{matrix}\right] \quad , \tag{3.9}$$

where $0_{k\ r-k}$ is a zero-matrix of order $k \times (r-k)$ and B_2 consists of the last k columns of the matrix

$$\underset{m\times m}{[B_1 \;\vdots\; B_2]} = [F_{2+3}]^{-1} \quad . \tag{3.10}$$

After premultiplication of (3.7) by (3.9), we get an expression that is identical with (3.8) (the details of the derivation are given in appendix II). The choice of a subset of restrictions f_1 does not affect the value of the Wald statistic, provided f_1 is such that its solution $\hat{\beta}$ converges to β and the matrix of partial derivatives is continuous at the true parameter values. Similar to our analysis in section 3.1, we can also show that transformations of the implicit functions asymptotically have no effect on the value of the Wald test in this case.

3.3 Equivalence of the traditional and the generalized Wald tests

We show that the traditional Wald test and the generalized Wald test yield the same value in large samples. Therefore, it is not necessary to derive the restrictions $h(\theta) = 0$ explicitly. It is sufficient to directly apply the procedure put forward in section 2 or Szroeter's (1983) generalized Wald method to $f(\beta,\theta) = 0$. From (2.7), we obtain that

$$h(\hat{\theta}) = H\,Q\,(\hat{\theta} - \theta_0) + o_p(T^{-1/2}) \quad . \tag{3.11}$$

The traditional Wald test and our procedure (2.2) can then be written as

$$W = T(\hat{\theta} - \theta_0)'Q'H'[HQ\Omega_\theta Q'H']^{-1} HQ(\hat{\theta} - \theta_0) + o_p(1) \quad . \tag{3.12}$$

Since $HF = 0$, from (2.11) one obtains that

$$H\,f(\hat{\beta},\hat{\theta}) = H\,Q\,(\hat{\theta} - \theta_0) + o_p(T^{-1/2}) =$$

$$= h\,(\hat{\theta}) + o_p(T^{-1/2}) \quad , \tag{3.13}$$

which establishes, using Holly and Monfort's lemma (see appendix I), the asymptotic equivalence of the generalized Wald test, the traditional Wald test and our approach, as H has full rank so that rank $(H) =$ rank $(H\Omega H')$. When $f(\beta,\theta) = 0$ is linear in β and θ, the three criteria are also equivalent in finite samples.

4. Multiple solutions for β under H_0

We consider the case where $f(\beta,\theta) = 0$, can have multiple solutions for β.

First, the subset $f_1(\beta,\theta) = 0$ we choose, possibly has multiple solutions. However, not every solution of $f_1(\beta,\theta) = 0$ will also satisfy the remaining implicit relations. As the sample size T increases, the Wald statistic tends to infinity for those solutions for which $f_2(\beta,\theta) \neq 0$.

Second, the complete system $f(\beta,\theta) = 0$ can admit several solutions for β. We assume that the set of restrictions can be expressed in the form $f(\beta) - \theta = 0$ and that each solution for β is locally identified. Under these assumptions, the various forms of the Wald test asymptotically yield the same result for each solution β.

The traditional Wald test (2.2) is used to test the restrictions $h(\theta) = 0$. These restrictions are expressed in terms of the parameters θ only, which are uniquely identified. Therefore, this statistic is not affected by the presence of multiple solutions for the implicit parameters β. For an example, we refer to section 5.

To test $f(\beta) - \theta = 0$, the generalized Wald statistic equals

$$W_g = \min_{\beta} T(f(\beta) - \hat{\theta})' \, \Omega_{\theta}^{-1} \, (f(\beta) - \hat{\theta}) \quad . \tag{4.1}$$

Let β^* denote the value of β which minimizes expression (4.1) and let θ^* be given by $\theta^* = f(\beta^*)$. Then we get

$$W_g = T(\theta^* - \hat{\theta})' \, \Omega_{\theta}^{-1} \, (\theta^* - \hat{\theta}) \quad . \tag{4.2}$$

Now with multiple solutions to $f(\beta) = \theta^*$, we obtain the same value of W_g for each solution.

In section 3.3, we have shown that the asymptotic equivalence of the three Wald criteria hinges upon the fact that $HF = 0$. In the presence of multiple solutions, this condition is satisfied too. To show this directly, we use $h(\theta) = 0$ and $f(\beta) = \theta$. By differentiating $h(\theta)$ with respect to β and applying the chain rule, we find

$$0 = D_{\beta} \, h(\theta) = D_{\theta} \, h(\theta) \, D_{\beta} \, f(\beta) = HF \quad , \tag{4.3}$$

which yields the desired result. The three statistics are asymptotically equivalent in case of multiple solutions for β.
It is interesting to note that the Lagrange multiplier test, the likelihood ratio test and the Wald test also asymptotically yield the same value under H_0 in case maximum likelihood estimates of θ are used, even if β in $f(\beta) - \theta = 0$ is not globally identified.

The practical implication of the existence of multiple solutions for $f_1(\beta,\theta) = 0$ is that one can only reject H_0 if for each solution of f_1 the Wald statistic is significantly different from zero. In other words, once we have a solution β to $f_1(\beta,\hat{\theta}) = 0$ for which the test is not significant, we conclude that the null hypothesis is not rejected. Therefore, one will preferably choose f_1 such that its solutions can be easily obtained. For example, if there are at least m linear restrictions in f, one may want to select f_1 as a linear system in β (one has to make sure that is has a unique solution). The occurrence of multiple solutions will be illustrated by an example of common factor restrictions in section 5.

5. An example: common factor restrictions

Common factor restrictions, which are widely used in dynamic econometric models, can easily be tested using the methods presented in section 2. The main reason for which we discuss the common factor approach here is to show how multiple solutions for the subset of nonlinear restrictions f_1 arise and how alternative formulations for the restrictions imply the same asymptotic values for the Wald statistic under H_0.

Sargan (1980a) presents a method for testing common factor restrictions in a dynamic single equation model. His method is based on a condition on the determinant of a given matrix. Sargan (1977) generalizes the method to vector dynamic models. Mizon and Hendry (1980) give an application of Sargan's (1980a) method. A single regression equation with common factors can be written as

$$\phi(L)\,\alpha(L)\,y_t = \sum_{i=1}^{k} \phi(L)\,\gamma_i(L)\,x_{it} + \varepsilon_t \quad , \qquad (5.1)$$

where y_t is the endogenous variable ε_t is a white noise error term with zero mean and constant variance σ^2 and independent of the exogenous variable $x_{it'}$, for all t and t' and $i = 1,\ldots,k$. The polynomials $\phi(L)$, $\alpha(L)$ and $\gamma_i(L)$, $i = 1,\ldots,k$, have degree p, r_0 and r_i respectively. The roots of $\phi(L)\alpha(L)$ lie outside the unit circle. The model (5.1) arises as a special case of the dynamic regression model

$$\theta_0(L)\, y_t = \sum_{i=1}^{k} \theta_i(L)\, x_{it} + \varepsilon_t \quad , \tag{5.2}$$

when $\theta_0(L) = \phi(L)\alpha(L)$ and $\theta_i(L) = \phi(L)\gamma_i(L)$, $i = 1,\ldots,k$. The number of parameters in (5.1) and (5.2) is $m = p + \sum_{i=0}^{k} r_i + k$ and $n = (1+k)p + \sum_{i=0}^{k} r_i + k$ respectively, so that the common factor structure in (5.1) leads to pk restrictions by equating the corresponding coefficients in (5.1) and (5.2). The restrictions are of the form $f(\beta) - \theta = 0$ and the computation of the Wald test is straightforward in this case.

For a given choice of f_1, there might exist two or more solutions, not all of them yielding the same asymptotic value for the Wald statistic under H_0. However, all solutions to f yield the same value of W asymptotically. A simple example given by Mizon and Hendry (1980) is illuminating in this respect. They consider a special case of models (5.1) and (5.2) written as

$$y_t = (\phi+\alpha)y_{t-1} - \phi\alpha\, y_{t-2} + \gamma_0 x_t + (\gamma_1 - \phi\gamma_0)x_{t-1} - \phi\gamma_1 x_{t-2} + \varepsilon_t$$

with $k = p = r_0 = r_1 = 1$, $\phi(L) = 1-\phi L$, $\alpha(L) = 1-\alpha L$, $\gamma_1(L) = \gamma_0 + \gamma_1 L$, and $y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + \theta_3 x_t + \theta_4 x_{t-1} + \theta_5 x_{t-2} + \varepsilon_t$.

When H_0 is true, we have the following set of implicit relations between $\beta = (\phi\,,\alpha\,,\gamma_0\,,\gamma_1)'$ and $\theta = (\theta_1,\ldots,\theta_5)'$

$$\begin{aligned}
f_1(\beta,\theta) = 0 : \quad \phi + \alpha - \theta_1 &= 0 \\
-\phi\alpha \quad - \theta_2 &= 0 \\
\gamma_0 - \theta_3 &= 0 \\
\gamma_1 - \phi\,\gamma_0 - \theta_4 &= 0 \\
f_2(\beta,\theta) = 0 : \quad -\phi\,\gamma_1 - \theta_5 &= 0
\end{aligned} \tag{5.3}$$

When $\theta_1^2 + 4\theta_2 > 0$, $f_1 = 0$ has two real solutions. However, if H_0 is true, only one of these solutions also satisfies $f_2 = 0$, except when there exists a functional relationship on β, namely $\gamma_0\alpha = -\gamma_1$, in which case both solutions satisfy $f_2 = 0$ and the model has two common factors. The requirement that $(1 - \theta_1 L - \theta_2 L^2) = 0$ and $(1 - \alpha L)(1 - \phi L) = 0$ have their roots outside the unit circle does not resolve the problem of multiple solutions. For instance, for $\theta' = (.5\ ,\ .2\ ,\ 1\ ,\ 5\ ,\ 1\ )$, the characteristic roots of the unrestricted model and the restricted model lie inside the unit circle, whereas (5.3) still has two solutions.

The Wald statistic can be computed for both solutions using the formulae in (2.8). The partial derivatives are then given by

$$D_\theta\, h\Big|_{\theta=\hat{\theta}} = \left(\frac{\gamma_1\phi + \gamma_0\phi^2}{\alpha - \phi}\ ,\ \frac{\gamma_1 + \gamma_0\phi}{\alpha - \phi}\ ,-\phi^2\ ,\ -\phi\ ,\ -1\right)\Big|_{\beta=\hat{\beta}} . \tag{5.4}$$

Computation of the Wald test when (2.8) is evaluated in a solution of $f_1 = 0$ that also satisfies $f_2 = 0$ asymptotically yields the value of the test statistic that ought to be used in testing. The value of the Wald statistic for the second solution of $f_1 = 0$ will tend to infinity as plim $h(\hat{\theta})$ = constant $\neq 0$ and plim $\hat{\Omega}_h$ is a constant matrix.
In small samples, we may not be able to discriminate between these values, but in large samples we can.

Mizon and Hendry (1980) derive the restrictions on θ implied by (5.3) explicitly. They find

$$\theta_5 + \phi\theta_4 + \phi^2\theta_3 = 0 \quad \text{and} \quad \phi = \frac{\theta_1\theta_5 - \theta_2\theta_4}{\theta_2\theta_3 + \theta_5} . \tag{5.5}$$

If the implicit relations (5.3) are substituted in (5.5) , it is obvious that the restriction on θ implied by (5.5) must be valid under H_0 . However, the formulation of the restriction in (5.5) is not unique. After some transformation of (5.3) , we also find

$$\theta_5 + \phi\theta_4 + \phi^2\theta_3 = 0 \quad \text{and} \quad \phi = \frac{-\theta_2\theta_3 - \theta_5}{\theta_1\theta_3 + \theta_4} \tag{5.6}$$

as a restriction. According to Sargan (1980a), common factor restrictions emerge from conditions on the rank of a certain matrix Ψ. For the problem at hand,

$$\text{rank } (\Psi) = \text{rank} \begin{bmatrix} -1 & \theta_1 & \theta_2 & 0 \\ \theta_3 & \theta_4 & \theta_5 & 0 \\ 0 & -1 & \theta_1 & \theta_2 \\ 0 & \theta_3 & \theta_4 & \theta_5 \end{bmatrix} = 3$$

gives the restriction as can be verified by substituting (5.3).
The rank condition yields the determinantal condition

$$\theta_5^2 + 2\theta_2\theta_3\theta_5 + \theta_1\theta_4\theta_5 + \theta_1^2\theta_3\theta_5 + \theta_2^2\theta_3^2 - \theta_2\theta_4^2 - \theta_1\theta_2\theta_3\theta_4 = 0 \,, \qquad (5.7)$$

which is equivalent to the relationship obtained from (5.5) or (5.6) after eliminating ϕ. This result shows the equivalence between the Mizon-Hendry approach and the Sargan procedure. The equivalence with our procedure and the generalized Wald test can be shown along the lines of section 3.3 as (5.7) is equivalent to $f(\beta(\hat{\theta}), \hat{\theta})$ and for (5.3), $D_\theta h = -H$ which is orthogonal to F.

If $\gamma_1 + \alpha\gamma_0 = 0$, the matrix Ψ has rank 2 when H_0 is true. Sequential testing for the presence of two common factor polynomials can be performed along the lines proposed by Sargan (1980a) by first testing for rank $(\Psi) = 3$ and subsequently for rank $(\Psi) = 2$. Alternatively, in our method we could extend f_2 in (5.3) by adding the restriction $\gamma_1 + \alpha\gamma_0 = 0$.

7. Some concluding remarks

In this paper, we presented a general procedure for computing Wald criteria to test linear and nonlinear nested hypotheses. The procedure can also be applied when the restrictions are in implicit form, as is often the case in econometric modeling. Along with Szroeter's (1983) generalized Wald test, the proposed procedure is expected to save the investigator from the time-consuming activity of expressing the restrictions in explicit form.

We gave a class of nonlinear transformations of the restrictions, that we want to test, for which the various Wald criteria are asymptotically invariant. We discussed the properties of the proposed procedure. In particular, we showed the asymptotic equivalence between the proposed procedure, the traditional Wald test and the generalized Wald test. The problem of multiple solutions to a set of nonlinear constraints on the parameters under H_0 has been discussed. Some of the problems which may arise when testing nonlinear constraints have been illustrated using a dynamic regression model with common factor restrictions. Additional applications including the test of overidentifying restrictions in a simultaneous equations model are given in Kodde and Palm (1982).

Finally we like to point out that the Wald encompassing test for testing nonnested hypotheses (see e.g. Mizon and Richard (1982) can beobtained using the procedure proposed here after some modifications. Also, β can be efficiently estimated by asymptotic nonlinear least squares applied to the 'asymptotic' model $f(\beta,\theta) = 0$ along the lines proposed by Gouriéroux et al. (1983), provided an efficient estimate of θ is available.

Appendix I

For the ease of reference, we give lemma 2 obtained by Holly and Monfort (1982):

Lemma: Let V be a p-dimensional random vector such that Variance (V) = Ω is of rank r ($\leq$ p) and $EV = \mu \in R(\Omega)$, the range of Ω .
Let $Z = AV$ where A is a non-random matrix. Then, $Z'(A\Omega A')^{-}Z = V'\Omega^{-}V$ with probability one (for any choice of the generalized inverse $(A\Omega A')^{-}$ and Ω^{-}) if, and only if, rank $(A\Omega A')$ = rank (Ω) .

For the proof, see Holly and Monfort (1982).

Appendix II

In this appendix, we show that

$$A\,[-F_{3+4}\,F_{1+2}^{-1}\,Q_{1+2} + Q_{3+4}] = [-F_{1+4}\,F_{2+3}^{-1}\,Q_{2+3} + Q_{1+4}], \qquad (A.1)$$

where $A = [-F_{1+4}\,B_2 \vdots {}^{0}{}_{k\,r-k}\ \vdots\ I_{r-k}]$ is defined in (3.9) and B_2 is given in (3.10) and the formulae are evaluated at $(\beta,\theta) = (\hat{\beta}_1,\hat{\theta})$.
The matrix multiplication in the 1.h.s. of (A.1) gives

$$\left[F_{1+4}\,B_2\,F_3^* + \begin{pmatrix} 0_{km} \\ -F_4^* \end{pmatrix}\right]\,F_{1+2}^{-1}\,Q_{1+2} + \left[-F_{1+4}\,B_2\,Q_3^* + \begin{pmatrix} 0_{kn} \\ -Q_4^* \end{pmatrix}\right] . \qquad (A.2)$$

From the definition (3.10) we have the following identity

$$B_2\,F_3^* = I_m - B_1\,F_2^* \quad ,$$

which we substitute into the first term of (A.2) to yield, after some algebraic transformations,

$$\left[\begin{bmatrix} I_k \vdots 0_{k\,m-k} \\ F_4^*\,F_{1+2}^{-1} \end{bmatrix} - F_{1+4}\,B_1\,(0_{m-k\,k} \vdots I_{m-k}) + \right.$$

$$\left. + \begin{bmatrix} 0_{km} \\ -F_4^*\,F_{1+2}^{-1} \end{bmatrix}\right] Q_{1+2} - F_{1+4}\,B_2\,Q_3^* + \begin{bmatrix} 0_{k\,n} \\ Q_4^* \end{bmatrix} . \qquad (A.3)$$

Expression (A.3) is equivalent to:

$$\begin{bmatrix} Q_1^* \\ 0_{r-k\,n} \end{bmatrix} - F_{1+4}\,B_1(0_{m-k\,n} + Q_2^*) - F_{1+4}\,B_2\,Q_3^* + \begin{bmatrix} 0_{kn} \\ Q_4^* \end{bmatrix} . \qquad (A.4)$$

Using (3.10) in (A.4), we find the desired result:

$$-F_{1+4}\,F_{2+3}^{-1}\,Q_{2+3} + Q_{1+4} \quad .$$

References

Almon, S. (1965): "The distributed lag between capital appropriations and expenditures", Econometrica, 30, 178-196.

Billingsley, P. (1968): Convergence of Probability Measures, New York, John Wiley.

Gouriéroux, Ch., Monfort, A., and A. Trognon (1983): "Asymptotic least squares", paper presented at the ESEM 1983 in Pisa.

Holly, A., and A. Monfort (1982): "Some useful equivalence properties of Hausman's test", Cahiers de recherches économiques, Université de Lausanne.

Hoffman, D.L., and P. Schmidt (1981): "Testing the restrictions implied by the rational expectations hypothesis", Journal of Econometrics, 15, 265-287.

Kodde, D.A., and F.C. Palm (1982): "Computing Wald criteria for nested hypotheses, with econometric applications", Vrije Universiteit, Economische Faculteit, Researchmemorandum 1982-27.

Mizon, G.E., and D.F. Hendry (1980): "An empirical application and Monte Carlo analysis of tests of dynamic specification", The Review of Economic Studies, 47, 21-46.

Mizon, G.E., and J.F. Richard (1982): "The encompassing principle and its application to non-nested hypotheses", paper presented at the ESEM in Dublin, September 1982.

Rudin, W. (1976): Principles of Mathematical Analysis, 3rd edit., Tokyo, McGraw-Hill.

Sargan, J.D. (1977): "Dynamic specification for models with autoregressive errors. Vector autoregressive case", Discussion paper, London School of Economics.

Sargan, J.D. (1980a): "Some tests of dynamic specification for a single equation", Econometrica, 48, 879-898.

Sargan, J.D. (1980b): "The consumer price equation in the postwar British economy: An exercise in equation specification testing", The Review of Economic Studies, 47, 113-136.

Smith, R.J. (1983): "Alternative asymptotically optimal tests in econometrics", Queen's University, Kingston, Discussion Paper 544.

Stroud, T.W.F. (1971): "On obtaining large-sample tests from asymptotically normal estimators", The Annals of Mathematical Statistics, 42, 1412-1424.

Szroeter, J. (1983): "Generalized Wald methods for testing nonlinear implicit and overidentifying restrictions", Econometrica, 51, 335-353.

Wald, A. (1943): "Tests of statistical hypotheses concerning several parameters when the number of observations is large", Transactions of the American Mathematical Society, 54, 426-482.