Computing Wald criteria for nested hypotheses
F.C. Palm and D.A. Kodde

Received: March 1, 1988; revised version: June 16, 1988

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.

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 when 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 in explicit form, Szroeter's (1983) generalized Wald method and our procedure asymptotically yield the same value for the statistic under the null hypothesis. 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 Szroeter's (1983) 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

Let us assume that we have a model defined in terms of \( n \) parameters forming a vector \( \theta \), and that \( \hat{\theta} \) is some consistent asymptotically normally distributed estimate of \( \theta_0 \) such that \( \sqrt{T}(\hat{\theta} - \theta_0) \), with \( T \) being the sample size and \( \theta_0 \) being the true value of \( \theta \), has a covariance matrix \( \Omega_\theta \) which can be consistently estimated by \( \hat{\Omega}_\theta \). A nested null hypothesis \( H_0 \) implies a set of constraints on \( \theta \).
\[ h(\theta) = 0, \tag{2.1} \]

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

\[ W = T h(\hat{\theta})' \hat{\Sigma}^{-1}_h h(\hat{\theta}), \tag{2.2} \]

where

\[ \hat{\Sigma}_h = (Dg_h) \hat{\Sigma}_\theta (Dg_h)', \tag{2.3} \]

with \( Dg_h \) denoting the first derivative matrix of \( h \) with respect to \( \theta \) which we evaluate at \( \hat{\theta} \). In the sequel, we denote the first and second partial derivatives of \( y \) with respect to a vector \( x' \) by \( D_{xy} \), with \( y \) being a scalar or a vector, and by \( D^2_{xx}y \) respectively, when \( y \) is a scalar.

On the null hypothesis that all the constraints (2.1) are satisfied, \( W \) is \( \chi^2 \)-distributed in large samples with \( r \) degrees of freedom, provided that \( \text{plim} \hat{\Sigma}_h \) is nonsingular and that \( Dg_h \) is a continuous function of \( \theta \) at the true parameter value \( \theta_0 \).

When the restrictions are given in the form (2.1), the Wald statistic is easily computed. Derivation of the restrictions in the form (2.1), however, can be tedious and intricate. We propose a method that simplifies explicit formulation of the restrictions and we show how \( h(\hat{\theta}) \) and \( Dg_h \) can be determined by implicitly using the restrictions. In empirical work, the restrictions implied by \( H_0 \) are usually given in the "mixed" form (see e.g. Gouriéroux and Monfort (1987)) of
\[ f(\theta, \beta) = 0. \quad (2.4) \]

where \( \beta \) 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. Under \( H_0 \), \( \theta_0 \) satisfies the implicit restrictions (2.4) and it does so for a unique value \( \beta_0 \) of \( \beta \) (in the interior of the parameter space for \( \beta \)). The matrices \( D_\theta f \) and \( D_\beta f \) are assumed to have rank \( m \) and \( m+r \) respectively \( (m+r \leq n) \).

From the system in (2.4), we now choose \( m \) equations, \( f_1(\theta, \beta) = 0 \), such that \( \beta \) can be solved explicitly as a function of \( \theta \), that is \( \beta = \hat{\beta}(\theta) \). When locally no solution exists to \( f_1(\theta, \beta) = 0 \), our result still holds true asymptotically if \( \hat{\beta} \) converges in probability to \( \theta_0 \), because we assume that \( f(\theta, \beta_0) = 0 \) has a solution. This solution is substituted in the \( r \) remaining relations that we denote by \( f_2(\theta, \beta) = 0 \) to give

\[ h(\theta) = f_2(\beta(\theta), \theta) = 0. \quad (2.5) \]

Next, we obtain an expression for the partial derivatives. For the sake of simplicity, we define the following matrices \( D_\theta f = F_1 \), \( D_\beta f = Q_1 \), \( D_\theta f_1 = F_1 \), \( D_\beta f_1 = Q_1 \), \( i = 1, 2 \), where the arguments \( \theta \) and \( \beta \) have been deleted. When we evaluate these matrices at \( \hat{\beta} \) and \( \beta(\hat{\theta}) \), we use the notation \( \hat{F}_i \), \( \hat{Q}_i \), \( \hat{F}_i \) and \( \hat{Q}_i \) respectively. Assuming that \( f_1 \) has been chosen such that \( \hat{F}_i \) is continuous and nonsingular at \( (\beta_0, \theta_0) \), 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 \( \theta \) with first derivative given by

\[ D_\theta \hat{f}(\theta) = -F_1^{-1}Q_1. \quad (2.6) \]

If the matrix \( F_1 \) is nonsingular at \( (\beta_0, \theta_0) \), there exists only one solution to \( f_1(\beta, \theta) = 0 \) in some neighborhood of \( (\beta_0, \theta_0) \).
Applying the chain-rule of differentiation to (2.5) and using expression (2.6), the partial derivatives of \( h \) become

\[
D_\theta h = -F_2 F_1^{-1} Q_1 + Q_2 = HQ,
\]

(2.7a)

with \( H = [-F_2 F_1^{-1} I_r] \). As a result of the implicit function theorem, \( D_\theta h \)

is continuous in \( \theta \) at \( \theta_0 \).

When we evaluate (2.7a) at a consistent estimate of \( \theta \), we get (see e.g. Billingsley (1968)) under \( H_0 \)

\[
\hat{D}_\theta h = HQ + o_p(1),
\]

(2.7b)

with \( H \) and \( Q \) being evaluated at \((\theta_0, \theta_0)\) and "\( o_p \)" denoting the order of probability. 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, \( f(\theta) - \theta = 0 \), 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 (1960b)) 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

\[
h(\theta) = f_2(\theta) - \theta_2 \text{ and } D_\theta h = -H,
\]

(2.8)

with \( \theta_2 \) being the appropriate subvector of \( \theta \).

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 \( \hat{\theta} \) for a given \( \theta \) and substituting \( \hat{\theta} \) in \( f_2 \) to obtain \( h(\hat{\theta}) = f_2(\theta(\hat{\theta}), \hat{\theta}) \),

(2) computing the matrices \( F_1 \) and \( Q_1, i = 1, 2 \), to obtain \( \hat{D}_{\theta} h \) in (2.7b),

and (3) calculating the value of \( W \) in (2.2). In the incidental case where
Dgh in (2.7a) does not have full rank r, a consistent estimate of the
generalized inverse of \( \Omega_h \) in (2.3) has to be substituted into (2.2).
The approach yields a convenient procedure to compute Wald criteria.
It also accommodates sequential testing when \( f_2 \) is successively
extended, given the choice of \( f_1 \) and the parametrization \( \theta, \tilde{\theta} \).
The generalized Wald test proposed by Sroeter (1983) for the set of
restrictions (2.4) can be obtained as follows. Given \( \hat{\theta} \), a consistent
estimator \( \tilde{\theta} \) is found by minimizing

\[

f(\tilde{\theta}, \hat{\theta}) = f(\tilde{\theta}, \hat{\theta}) \quad (2.9)
\]

with respect to \( \hat{\theta} \), where \( S \) is a postive semi-definite symmetric matrix
such that \( F' S F \) has rank \( r \). The requirement that rank \( F' S F = r \) is a
generalization of Sroeter (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. (1985) and Kodde, Palm and
Pfann (1987)). Applying the implicit function theorem to the first order
conditions for a minimum, \( F' S f(\tilde{\theta}, \hat{\theta}) = 0 \), we get

\[

\hat{\theta} - \theta_0 = PQ(\tilde{\theta} - \theta_0) + \theta_0 (I - H)
\quad (2.10)
\]

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

\[

f(\hat{\theta}, \tilde{\theta}) = [I + \tilde{\theta} FP] Q(\tilde{\theta} - \theta_0),
\]

\[

= [I + \tilde{\theta} FP] Q(\tilde{\theta} - \theta_0) + \theta_0 (I - H)
\quad (2.11)
\]

where a tilde """" denotes evaluation at a suitable point between
\( (\tilde{\theta}(\hat{\theta}), \hat{\theta}) \) and \( (\tilde{\theta}_0, \theta_0) \).

The generalized Wald test is now given by

\[

W_g = T f(\tilde{\theta}, \hat{\theta})' \tilde{\theta} - f(\tilde{\theta}, \hat{\theta}).
\quad (2.12)
\]
where $\hat{\Omega}$ denotes the matrix $\Omega = (I + FP)Q \Omega Q' (I + FP)'$ evaluated at $(\hat{\beta}, \hat{\theta})$. 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

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

(2.13)

Some comments on the implementation of Szroeter's (1983) procedure are in order.

When $S = [\hat{\Theta} \hat{R} \hat{G}]^{-1}$, the asymptotic covariance matrix of $f(\hat{\beta}, \hat{\theta})$ in (2.11) is

$$[Q \Omega Q' - F' (Q \Omega Q')^{-1} F]^{-1},$$  

(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

$$W_g = T f(\hat{\beta}, \hat{\theta})' [\hat{\Theta} \hat{R} \hat{G}]^{-1} f(\hat{\beta}, \hat{\theta}) = T f(\hat{\beta}, \hat{\theta})' S f(\hat{\beta}, \hat{\theta}).$$  

(2.15)

$W_g$ is Szroeter's objective function (2.9) evaluated at the minimum for $\beta$ and multiplied by $T$. Expression (2.15) gives an alternative way of computing Wald criteria. Notice, however, that $Q$ may depend on $\beta$ so that a consistent estimate of $\beta$ is required for obtaining $S$ in (2.15).

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 \( A_h \) 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 \( Dg = ADgH \) 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 (1985), 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 \( DgH \) into \( Dg \) 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, \( Dg(\theta, \theta) \) being nonsingular and \( Dgg(\theta, \theta) \) being zero 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 par-
tial derivatives. The matrices of partial derivatives of \( h \) and \( g \) with respect to \( \theta \) are given by

\[
Dg(h(\theta)) \text{ and } Dg(y,\theta)Dg(y) + Dg(y,\theta).
\]  

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

\[
\lim Dg(\hat{y},\hat{\theta}) = \lim Dg(0,\tilde{\theta}) = Dg(0,\theta_0) = 0,
\]

where \( \hat{\theta} \) is a consistent estimate of \( \theta \) and \( \hat{y} = h(\hat{\theta}) \). The second term of the derivative of \( g \) with respect to \( \theta \) 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, Sproeter (1983) shows that the asymptotic local power of his test does not depend on the particular choice of \( S \). The asymptotic efficiency of \( \hat{\theta} \), however, depends on \( S \). In fact \( S = [\hat{Q} \hat{N}_g \hat{Q}']^{-1} \) maximizes the asymptotic efficiency of \( \hat{\theta} \), which then is an optimal asymptotic least squares estimate.

Second, we consider general transformations of \( f(\theta,\theta) = 0 \) which take the form \( g(f(\theta,\theta),\theta,\theta) \), with

\[
g(f(\theta,\theta),\theta,\theta) = 0
\]

if and only if \( f(\theta,\theta) = 0 \). Furthermore, \( g \) has continuous first and second derivatives, \( Dy(g(y,\beta,\theta)) \) is nonsingular, \( Dg(g(y,\beta,\theta)) = 0 \) and \( Dg(y,\beta,\theta) = 0 \) at \((0,\theta_0,\theta_0)\). 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 \( \hat{\theta} \), the matrix of partial derivatives of \( g \) with respect to \( \theta \) is given by
\[ D_y g(y, \theta, \delta) [F D \theta + Q] + D_g g(y, \theta, \delta) D \theta + D_g g(y, \theta, \delta). \]  

(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 \( \hat{\delta} \). In addition, the difference between \( D_g \theta \) based on \( f \) and \( g \) respectively, vanishes in large samples (see also Gouriéroux et al. (1985)). Therefore,

\[ [D_y g(y, \hat{\delta}, \hat{\theta})]^{-1} D_g g(f(\hat{\delta}, \hat{\theta}), \hat{\theta}, \hat{\delta}) = [I + F P] Q + o_p(1), \]  

(3.5)

and the lemma by Holly and Monfort (1985) 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 \( f_2 \). We partition the system of constraints into four subsets, which consist of \( k, m-k, k \) and \( r-k \) relations respectively.

\[ f_1^k(\theta, \beta) = 0, \quad i = 1, \ldots, 4. \]  

(3.6)

To simplify the notation, we delete the arguments \( \beta \) and \( \theta \) and we denote the subset of restrictions \( f_i^k \) and \( f_j^k \) by \( f_{i+j}^k \) and its partial derivatives with respect to \( \beta \) and \( \theta \) by \( F_{i+j}^k \) and \( Q_{i+j}^k \) respectively.

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

\[ \text{D}_{\phi} = \begin{bmatrix} -F_{3+4} & F_{1+2}^{-1} & Q_{1+2} + Q_{3+4} \end{bmatrix} \]  \hspace{1cm} (3.7) \]

and

\[ \text{D}_{\psi} = \begin{bmatrix} -F_{1+4} & F_{2+3}^{-1} & Q_{2+3} + Q_{1+4} \end{bmatrix}, \]  \hspace{1cm} (3.8) \]

where the subscript \( i = 1, 2 \) indicates the choice of \( f_i \).

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, \( \text{D}_{\phi} = A \text{D}_{\psi} \). A nonsingular matrix that gives the desired result is

\[ A = \begin{bmatrix} -F_{1+4} & B_2 \\ F_{1+4} & B_2 \end{bmatrix}, \]  \hspace{1cm} (3.9) \]

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

\[ \begin{bmatrix} B_1 & B_2 \end{bmatrix} = [F_{2+3}]^{-1}. \]  \hspace{1cm} (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 \( \beta \) 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. From (2.7), we obtain that

\[ h(\hat{\theta}) = HQ(\hat{\theta} - \theta_0) + o_p(T^{-\frac{1}{2}}). \]  

(3.11)

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

\[ W = T(\hat{\theta} - \theta_0)'Q' \left[HQ_0Q'H'\right]^{-1}HQ(\hat{\theta} - \theta_0) + o_p(1). \]  

(3.12)

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

\[ HF(\hat{\theta}, \hat{\theta}) = HQ(\hat{\theta} - \theta_0) + o_p(T^{-\frac{1}{2}}) = h(\hat{\theta}) + o_p(T^{-\frac{1}{2}}), \]  

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 \( \text{rank}(H) = \text{rank}(H'H') \).

When \( f(\theta, \theta) = 0 \) is linear in \( \theta \) and \( \theta \), the three criteria are also equivalent in finite samples.

4. MULTIPLE SOLUTIONS FOR \( \theta \) UNDER \( H_0 \)

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

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

Second, the complete system \( f(\theta, \theta) = 0 \) can admit several solutions for \( \theta \).

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

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

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

\[
W_g = \min_{\beta} T(f(\beta) - \theta)\Omega^{-1}_g(f(\beta) - \theta). \tag{4.1}
\]

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

\[
W_g = T(\beta^* - \theta)\Omega^{-1}_g(\beta^* - \theta). \tag{4.2}
\]

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 \( \theta \) and applying the chain rule, we find

\[
0 = Dh(\theta) = Dh(\theta)Dgf(\beta) = HF, \tag{4.3}
\]

which yields the desired result. The three statistics are asymptotically equivalent in case of multiple solutions for \( \beta \).
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 $\theta$ are used, even if $\beta$ 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 $\beta$ to $f_1(\beta, \theta) = 0$ 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 $\beta$ (one has to make sure that it 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 regression models with autocorrelated disturbances 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 (1960) 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

$$\phi(L)\alpha(L)y_t = \sum_{i=1}^{k} \phi(L)\gamma_i(L)x_{it} + \epsilon_t,$$  \hspace{1cm} (5.1)

where $y_t$ is the endogenous variable, $\epsilon_t$ is a white noise error term with zero mean and constant variance $\sigma^2$ 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 $r_1$ 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

$$\theta_0(L)y_t = \sum_{i=1}^{k} \theta_i(L)x_{it} + \epsilon_t,$$ \hspace{1cm} (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 on the parameters of (5.2). The restrictions are of the form $f(\theta) - \theta = 0$ 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

\[ y_t = (\phi \alpha) y_{t-1} - \phi \alpha y_{t-2} + \gamma_0 x_t + (\gamma_1 - \phi \gamma_0) x_{t-1} - \phi \gamma_1 x_{t-2} + \epsilon_t \]

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 \( y_t = \beta_1 y_{t-1} + \beta_2 y_{t-2} + \beta_3 x_t + \beta_4 x_{t-1} + \beta_5 x_{t-2} + \epsilon_t \).

When \( H_0 \) is true, we have the following set of implicit relations between \( \beta = (\phi, \alpha, \gamma_0, \gamma_1)' \) and \( \theta = (\beta_1, \ldots, \beta_5)' \):

\[
\begin{align*}
 f_1(\beta, \theta) &= 0 : \quad \phi + \alpha - \beta_1 = 0 \\
 & \quad -\phi \alpha \quad - \beta_2 = 0 \\
 & \quad \gamma_0 - \beta_3 = 0 \\
 & \quad \gamma_1 - \phi \gamma_0 - \beta_4 = 0
\end{align*}
\]

\[
 f_2(\beta, \theta) = 0 : \quad -\phi \gamma_1 - \beta_5 = 0.
\]

(5.3)

When \( \beta_1^2 + 4\beta_2 > 0 \), \( f_1 = 0 \) 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 \( \beta \), 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 \( (1-\beta_1 L-\beta_2 L^2) = 0 \) 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
\[
D_{\theta h} = \left( \frac{Y_1 \phi + Y_0 \phi^2}{\alpha - \phi}, \frac{Y_1 + Y_0 \phi}{\alpha - \phi} \right), \left\{ \phi^2, \phi, -1 \right\}
\] (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 \( \text{plim} \ h(\hat{\theta}) = \text{constant} \neq 0 \) and \( \text{plim} \ \hat{h}_h \) is a constant matrix.

In small samples, we may not be able to discriminate between these values, but in large samples we can.

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

\[
\theta_5 + \phi \theta_4 + \phi^2 \theta_3 = 0 \quad \text{and} \quad \phi = \frac{\theta_1 \theta_5 - \theta_2 \theta_4}{\theta_2 \theta_3 + \theta_5}.
\] (5.5)

If the implicit relations (5.3) are substituted in (5.5), it is obvious that the restriction on \( \theta \) 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

\[
\theta_5 + \phi \theta_4 + \phi^2 \theta_3 = 0 \quad \text{and} \quad \phi = \frac{-\theta_2 \theta_3 - \theta_5}{\theta_1 \theta_3 + \theta_4}
\] (5.6)

as a restriction. According to Sargan (1980a), common factor restrictions emerge from conditions on the rank of a certain matrix \( \Psi \). For the problem at hand,
\[
\begin{pmatrix}
-1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix}
\]

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

\[b_5^2 + 2b_2b_3b_4 + b_1b_4b_5 + b_2^2b_5 + b_2b_4^2 - b_2b_4^2 - b_1b_2b_3b_4 = 0, \quad (5.7)\]

which is equivalent to the relationship obtained from (5.5) or (5.6) after eliminating \(b\). 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 \(f(\beta(\hat{\beta}), \hat{\delta}) = 0\) and for (5.3), \(D_{gh} = -H\) which is orthogonal to \(F\).

If \(\gamma_1 + a\gamma_0 = 0\), the matrix \(\psi\) 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 \(\gamma_1 + a\gamma_0 = 0\).

6. 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 avoids expressing the restrictions
in explicit form, which can be intricate and time consuming. We gave a class of nonlinear transformations of the restrictions to be tested, 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 Hg 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. Finally, as mentioned in section 2, additional applications include the test of overidentifying restrictions and the rational expectations constraints in a simultaneous equations model and polynomial distributed lags. Also, $\beta$ can be efficiently estimated by asymptotic nonlinear least squares applied to the "asymptotic" model $f(\beta, \theta) = 0$ provided a consistent estimate of $\theta$ is available.

**APPENDIX I**

For the ease of reference, we give lemma 2 obtained by Holly and Monfort (1985).

**Lemma**: Let $V$ be a $p$-dimensional random vector such that $\text{Variance}(V) = \Omega$ is of rank $r (\leq p)$ and $EV = \mu \in R(\Omega)$, the range of $\Omega$. Let $Z = AV$ where $A$ is a non-random matrix. Then, $Z'(A\Omega A')^{-1}Z = V'\Omega^{-1}V$ with probability one (for any choice of the generalized inverse $(A\Omega A')^{-1}$ and $\Omega^{-1}$) if, and only if, $\text{rank}(A\Omega A') = \text{rank}(\Omega)$.

For the proof, see Holly and Monfort (1985).
APPENDIX II

In this appendix, we show that

$$A[-F_{3+4}F_{4+2}^{-1}Q_{1+2} + Q_3 + 4] = [-F_1 + 4 F_{2+3}^{-1} Q_{2+3} + 4_1 + 4].$$  (A.1)

where A is defined in (3.9) and B_2 is given in (3.10) and the formulae are evaluated at (\hat{\theta}, \hat{\theta}).

The matrix multiplication in the l.h.s. of (A.1) gives

$$[F_1 + 4 B_2 - F_{4+2}^{-1} + \begin{pmatrix} O_k^\text{m} & m \end{pmatrix}] F_{4+2}^{-1} + \begin{pmatrix} \text{I}_m - B_1 F_{4+2}^{-1} \end{pmatrix}.  \quad (A.2)$$

From the definition (3.10) we have the following identity

$$B_2 F_5^\star = \text{I}_m - B_1 F_2^\star,$$

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

$$\begin{bmatrix} I_k & O_k & m-k \\ \text{I}_m - F_{4+2}^{-1} & \text{I}_m - k \end{bmatrix} - F_1 + 4 B_1 (O_{m-k} k) +$$

$$+ \begin{pmatrix} O_k & m \\ -F_{4+2}^{-1} \end{pmatrix} Q_{1+2} - F_1 + 4 B_2 Q_3^* + \begin{pmatrix} O_k & n \\ Q_4^* \end{pmatrix}. \quad (A.3)$$

Expression (A.3) is equivalent to

$$\begin{bmatrix} O_k^* \\ 0 \end{bmatrix} = -F_1 + 4 B_1 (O_{m-k} n + Q^*) - F_1 + 4 B_2 Q_3^* + \begin{pmatrix} O_k & n \end{pmatrix}.$$  \quad (A.4)

Using (3.10) in (A.4), we find the desired result $- F_{1+4} F_{2+3}^{-1} Q_{2+3} + Q_{2+3} + Q_{1+4}$. 

REFERENCES


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.

D.A. Kodde
Catholic University of Nijmegen
Economic Institute
P.O. Box 9108
6500 HK Nijmegen
The Netherlands

F.C. Palm
University of Limburg
Department of Economics
P.O. Box 616
6200 MD Maastricht
The Netherlands