Computing Wald Criteria for Nested Hypotheses with Econometric Applications

D.A. Kodde
F.C. Palm

Research Memorandum 1982-27  December 1982
Computing Wald Criteria for Nested Hypotheses with Econometric Applications

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

December 1982
comments welcome

In this paper, we present a general 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 a time-consuming step in testing linear and nonlinear nested hypotheses. We also indicate how the procedure can be used to get restricted parameter estimates. Next, the properties of the general procedure are discussed. Finally, three econometric applications illustrate how the Wald statistic can be computed in a fairly straightforward way.

* Katholieke Universiteit, Nijmegen.
** Vrije Universiteit, Amsterdam, 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.
1. **Introduction**

In recent years, the Wald test has received increasing attention in the econometric literature. This trend will probably last for some time as the Wald test (see Wald (1943)) has proved to be a very useful tool in empirical econometrics and economic statistics. Among the areas of application, we like to mention the specification analysis or the 'top-down' approach to model-building, where a fairly general model is taken as the maintained hypothesis throughout the modeling process. Restrictions on the maintained model lead to a nested model and can be tested by means of a Wald test, which only requires estimates of the parameters of the unrestricted model. 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.

Another important area of application is the population-sample decomposition approach that is becoming popular in econometrics. Again, at the start, one assumes that an estimate of some population moment can be computed from a given sample. For a nonstandard sample, such as e.g. observations with measurement errors, incomplete sampling, presence of selectivity bias, the standard moment estimators will not be consistent. However, after an appropriate transformation of the moment estimator, consistency (and asymptotic normality) may be achieved. In a second stage, restrictions implied by the process for the population can be checked using a Wald test. Although the adjusted moment estimator will usually not be fully efficient, the population-sample decomposition is certainly attractive from a computational point of view and it is expected to be more robust than a joint analysis of a model for the population allowing for deficiencies of sampling.

In this paper, we present a general 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 a time-consuming step in testing linear and nonlinear nested hypotheses. In section 2, we present the procedure along with some
basic notation. We also indicate how it can be used to get restricted parameter estimates. The properties of the general procedure are discussed in section 3. Section 4 contains several examples to illustrate how the Wald statistic can be computed in a fairly straightforward way. Finally, in section 5 we briefly present some conclusions.

2. A general procedure for computing Wald criteria

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 \( \Theta \) such that \( \sqrt{T}(\hat{\Theta} - \Theta) \), with \( T \) being the sample size, has a covariance matrix which can be consistently estimated by \( \hat{\Sigma}_n \). A nested null hypothesis \( H_0 \) implies a set of implicit constraints on \( \Theta \)

\[
\mathbf{h}(\Theta) = \mathbf{0}, \tag{2.1}
\]

which form a vector of \( r \) independent, continuously differentiable functions. Under the alternative hypothesis \( H_a \), the equality in (2.1) does not hold true.

The Wald statistic for testing the set of implicit restrictions is

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

where

\[
\hat{\Sigma}_n = \left( \frac{\partial \mathbf{h}}{\partial \Theta}' \right) \left( \frac{\partial \mathbf{h}}{\partial \Theta} \right)' \tag{2.3}
\]

with \( \frac{\partial \mathbf{h}}{\partial \Theta} \) denoting the first derivative matrix of \( \mathbf{h} \) with respect to \( \Theta \) evaluated at \( \hat{\Theta} \).

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 \( \operatorname{plim} \hat{\Sigma}_n \) is nonsingular and that \( \frac{\partial \mathbf{h}}{\partial \Theta} \) is a continuous function of \( \Theta \) at the true parameter value \( \Theta_0 \). 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 \( D_{xx'y} \) respectively, when \( y \) is a scalar.
For a given set of restrictions, the Wald statistic is easily computed. Explicit derivation of the restrictions however, can be tedious and time-consuming. The method we propose here simplifies explicit formulation of the restrictions. We show how $h(\theta)$ and $D_\theta h$ can be determined by using the restrictions implicitly. Once $h(\theta)$ 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

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

where $\beta$ is a vector of $m$ parameters of the restricted model, $f$ is a continuously differentiable mapping from a $m + n$ dimensional space into a $m + r$ dimensional one. In section 4, some illustrative examples are given for various forms of $f(\beta, \theta) = 0$. 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 $\beta$ can be solved explicitly as a function of $\theta$, 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

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

As indicated above, we only need the restrictions and the corresponding partial derivatives both evaluated at $\hat{\theta}$ to compute the Wald statistic. Now we show how these magnitudes can be derived form (2.4). First, we determine $h(\hat{\theta})$ along the lines just described, where an estimate $\hat{\theta}$ is substituted for $\theta$, which means we solve $f_1(\beta, \hat{\theta}) = 0$ for $\beta$ to get $\hat{\beta}$. Next, we obtain an expression for the partial derivatives evaluated at $\hat{\theta}$. Assuming that $f_1$ has been chosen such that $D_\beta f_1(\beta, \theta)$ is non-singular at $(\hat{\beta}, \hat{\theta})$, 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 at $(\hat{\beta}, \hat{\theta})$ given by

$$D_\theta \beta(\theta) = - \left[ D_\beta f_1(\beta, \theta) \right]^{-1} D_\theta f_1(\beta, \theta) \bigg|_{(\hat{\beta}, \hat{\theta})}. \quad (2.6)$$
The matrix $D_{\hat{\beta}} f_1$ is nonsingular if only one solution of $f_1(\beta, \hat{\theta}) = 0$ exists in some neighborhood of $(\hat{\beta}, \hat{\theta})$.

Using the chain rule of differentiation and expression (2.6), the partial derivations of $h$ at $\hat{\theta}$ become

$$D_0 h = D_{\hat{\theta}} f_2,$$

$$= \left[ - D_{\hat{\beta}} f_2 (D_{\hat{\beta}} f_1)^{-1} D_{\hat{\theta}} f_1 + D_{\hat{\theta}} f_2 \right] (\hat{\beta}, \hat{\theta}).$$

(2.7)

For the sake of simplicity, we delete the arguments $\beta$ and $\theta$.

Formulae (2.5) and (2.7) are suited for all kind of nested hypotheses. However, quite often the set of restrictions (2.4) has the special form, $f(\beta) - \theta = 0$, so that expression (2.7) 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. This list of examples is not exhaustive but it contains some major areas of application for the Wald test. For this special form of the implicit relations, we obtain

$$h(\hat{\theta}) = f_2(\hat{\theta}) - \hat{\theta}_2$$

and

$$D_{\hat{\theta}} h = \left[ - D_{\hat{\beta}} f_2 (D_{\hat{\beta}} f_1)^{-1} \hat{\theta}_1 - I_2 \right]_{\hat{\beta}},$$

(2.8)

with $\hat{\theta}_2$ being the appropriate subvector of $\hat{\theta}$.

In this context, it should be noted that the system (2.4) can be used to get an efficient estimate of $\beta$, when $\beta$ is the complete set of parameters of the restricted model. Under $H_0$, the log-likelihood function denoted by $L$ can be expressed in terms of the parameters $\beta$, i.e. $L(\Theta(\beta))$, provided the restrictions in (2.4) are or can be written in the special form $f(\beta) - \theta = 0$. Using the chain rule of differentiation, the first order conditions for a maximum can be expressed as

$$(D_{\hat{\beta}} \Theta)' D_{\hat{\theta}} L \bigg|_{\hat{\beta}} = 0 = (D_{\hat{\beta}} \Theta)' D_{\theta} L \bigg|_{\beta, \hat{\theta}(\hat{\beta})} (D_{\hat{\theta}} \Theta)' D_{\hat{\theta}} L (D_{\hat{\beta}} \Theta) \bigg|_{\beta, \hat{\theta}(\hat{\beta})} (\hat{\beta} - \beta)$$

(2.9)
where \( \hat{\beta}, \hat{\theta}, \) and \( \hat{\beta}_{\text{ML}} \) denote respectively a consistent initial estimator, a first order efficient two-step and the efficient maximum likelihood estimator of \( \beta \). When \( D_{\theta} f \) is nonsingular, the matrix of partial derivatives

\[
D_{\beta \theta} = -[D_{\theta} f]^{-1} D_{\beta f}
\]  

(2.10)

obtained from (2.4) can be substituted into (2.9) to yield the following expression for \( \hat{\beta} \)

\[
\hat{\beta} = \beta + \left[ (D_{\beta f})'(D_{\theta f})^{-1} D_{\theta \theta} L (D_{\theta f})^{-1} D_{\beta f} \right]^{-1} (D_{\beta f})'(D_{\theta f})^{-1} D_{\theta \theta} L (D_{\theta f})^{-1} D_{\theta \theta} (2.11)
\]

The large sample covariance matrix for \( \sqrt{T}(\hat{\beta} - \beta_0) \), with \( \beta_0 \) being the true parameter value is given by

\[
T \left[ (D_{\beta f})'(D_{\theta f})^{-1} D_{\theta \theta} L (D_{\theta f})^{-1} D_{\beta f} \right]^{-1}
\]  

(2.12)

Two remarks can be made. First, this estimation procedure cannot be used when the restrictions are of the implicit form (2.1) and \( \theta \) is the parameter of interest. The Lagrange multiplier approach (see e.g. Silvey (1959)) is suited for efficient estimation under implicit restrictions. Second, when \( D_{\theta \theta}^2 L \) can be rearranged to become a block diagonal matrix with the first diagonal block being of order \( s \geq m \) and with the corresponding matrix \( D_{\beta \theta} \) having zero's in the last \( n - s \) rows, then (2.12) can be expressed as a block diagonal matrix, so that the last \( n - s \) elements of \( D_{\theta} L \) are not needed to efficiently estimate \( \beta \).

To conclude, besides yielding a convenient procedure to compute Wald criteria and restricted parameter estimates, the approach also accommodates sequential testing, which is accomplished by successively extending the set of restrictions \( f_2 \) for a given choice of \( f_1 \). For most problems, the formulation of the restrictions in (2.5) and the choice of \( f_1 \) and \( f_2 \) in the procedure described in this section are not unique. The implication of this choice for the value of the Wald statistic will be analyzed in the next section.
3. The Wald statistic and alternative formulations of the restrictions

In this section we investigate whether the value of the Wald statistic is affected by choosing an alternative formulation for the constraints. We give a class of transformations of the restrictions, which do not affect the value of the Wald statistic in large samples given that $H_0$ is true. Furthermore, we consider the impact of multiple solutions for $f_1(\beta,\theta) = 0$ and the influence of the choice of $f_1$ and $f_2$ on the Wald test.

In the present context, the following result by Holly and Monfort (1982) (see lemma 2) will be very useful:

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

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

Consider now the case where the set of restrictions $h(\theta) = 0$ is such that $h_0$ is nonsingular. Any alternative formulation of the restrictions say $g(\theta) = 0$, for which there exists a nonsingular matrix $A$ such that $D_\theta g = AD_\theta h$ (We call this the equivalence condition of the partial derivatives), will asymptotically yield the same value for the Wald statistic, both under $H_0$ and under a sequence of local alternative hypotheses. 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.

The problem we now consider is the existence of multiple solutions for a given choice of $f_1(\beta,\theta) = 0$. If $H_0$ is true, the data generating process can be characterized by just one point in the parameter space for $\beta$, defined as the solution of $f(\beta,\theta_0) = 0$, where $\theta_0$ is the true value of $\theta$. Otherwise, the parameter $\beta$ is not identified under $H_0$. However, not every solution of $f_1(\beta,\theta_0) = 0$ will also satisfy the remaining implicit relations. Usually, only one - occasionally several - of the solutions for $f_1(\beta,\theta_0) = 0$ also satisfies $f_2(\beta,\theta_0) = 0$.
As the sample size $T$ increases, the Wald statistic tends to infinity for those solutions for which $f_2(\beta, \theta_0) \neq 0$. If there exist two or more solutions for $f_1(\beta, \theta_0) = 0$, which also satisfy $f_2(\beta, \theta_0) = 0$, their Wald statistics will usually not be identical, as the equivalence condition for the partial derivatives need not to be fulfilled. 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. This point will be illustrated by the example of common factor restrictions in section 4. However, this problem is only relevant if we cannot find a set of restrictions $f_1(\beta, \theta_0) = 0$ so that just one solution exists.

Next, we investigate the consequences of the choice of $f_1$ 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,

$$f_1^*(\beta, \theta) = 0 \quad (3.1)$$

$$f_2^*(\beta, \theta) = 0 \quad (3.2)$$

$$f_3^*(\beta, \theta) = 0 \quad (3.3)$$

$$f_4^*(\beta, \theta) = 0 \quad (3.4)$$

To simplify the notation, we delete the arguments $\beta$ and $\theta$ and we denote the subset of restrictions $\left( f_1^*, f_2^* \right)$ by $f_{i+1}^*$. As our first choice of $f_1 = 0$, we use the set $f_1^*, f_2^*$ to derive a solution for $\beta, \hat{\beta}_1$. A second solution $\hat{\beta}_2$ is derived from $f_{2+3}^* = 0$. Using the result in (2.7), the partial derivatives can be written as (the subscript $i = 1, 2$ indicates the choice of $f_1$)

$$D_\theta h_1 = \left[ -D_\beta f_{3+4}^* (D_\beta f_{1+2}^*)^{-1} D_\theta f_{1+2}^* + D_\theta f_{3+4}^* \right]_{(\hat{\beta}_1, \hat{\theta})}$$

and

$$D_\theta h_2 = \left[ -D_\beta f_{1+4}^* (D_\beta f_{2+3}^*)^{-1} D_\theta f_{2+3}^* + D_\theta f_{1+4}^* \right]_{(\hat{\beta}_2, \hat{\theta})}$$
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 (3.2) and (3.3) satisfy the equivalence condition, \( D_{\theta} h_2 = A D_{\theta} h_1 \). A nonsingular matrix that gives the desired result is

\[
A = \begin{bmatrix}
-D_{\beta} f^*_{1+4} & B_2 \\
-0_{k \times (r-k)} & I_{r-k}
\end{bmatrix},
\]

(3.4)

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

\[
[B_1 : B_2] = \left(D_{\beta} f^*_{2+3} \right)^{-1}. (\hat{\beta}_1, \theta)
\]

(3.5)

After premultiplication of (3.2) by (3.4), we get an expression that is identical with (3.3) except that it is evaluated at \((\hat{\beta}_1, \theta)\) (the details of the derivation are given in an appendix). The choice of a subset of restrictions \( f_1 \) does not affect the value of the Wald statistic, provided \( \text{plim} (\hat{\beta}_1 - \hat{\beta}_2) = 0 \) and the matrices of partial derivatives are continuous at the true parameter values.

Under \( H_0 \), \( \text{plim} (\hat{\beta}_1 - \hat{\beta}_2) = 0 \) if there exists just one solution of the implicit functions for both choices of \( f_1 \). In the presence of multiple solutions, there will be a combination of these solutions such that \( \text{plim} (\hat{\beta}_1 - \hat{\beta}_2) = 0 \), but not every combination will necessarily have this property. Therefore, we have to conclude that different choices of \( f_1 \) may not yield the same value for \( W \) if there is a choice of \( f_1 \), for which multiple solutions exist.

Finally, we consider the case where the set of implicit restrictions \( h(\theta) = 0 \) is given. We prove that \( h(\theta) = 0 \) and \( g(h(\theta), \theta) = 0 \), \( g \) being continuous at the true parameter values, yield the same value for \( W \) in large samples, if \( g(0, \theta) = 0 \) and \( D_y g(y, \theta) \) is nonsingular. The matrices of partial derivatives of \( h \) and \( g \) with respect to \( \theta \) are given by

\[
D_{\theta} h(\theta) \bigg|_{\theta = \hat{\theta}} \quad \text{and} \quad \left[ D_y g(y, \theta) D_{\theta} y + D_{\theta} g(y, \theta) \right]_{(y, \theta) = (h(\hat{\theta}), \hat{\theta})}.
\]

(3.6)
Premultiplication of the second expression by \( [D_y g(y, \Theta)]^{-1} \)
evaluated at \((h(\Theta), \Theta)\) yields

\[ D_\Theta h(\Theta) \bigg|_{\Theta=\hat{\Theta}} + \left[ D_y g(y, \Theta) \right]^{-1} D_\Theta g(y, \Theta) \bigg|_{(y, \Theta)=(h(\Theta), \Theta)}. \tag{3.7} \]

But on \( H_0 \), \( \text{plim} D_\Theta g(y, \hat{\Theta}) = \text{plim} D_\Theta g(0, \hat{\Theta}) = 0 \); the second term in (3.7) vanishes in large samples and we obtain the asymptotic invariance of the Wald statistics with respect to transformations of the type \( g(h(\Theta), \Theta) \).

In section 4, we present some selected examples which illustrate the theoretical results given in sections 2 and 3.

4. Some econometric applications

The purpose of this section is to illustrate the wide range of applications of the Wald test. Thereby, we pay attention to the problems discussed in the preceding sections.

The list of examples considered is not exhaustive, but it should give the reader a fairly good indication of the usefulness in empirical econometrics of the Wald procedure proposed in the paper. Each of the subsections can be read separately.

4.1 Overidentifying exclusion restrictions in the linear simultaneous equation model

The structural form of a simultaneous equation model (SEM) is useful for, among other things, generating restrictions on the data-generating process (DGP). Byron (1974) proposed a Wald test for overidentifying restrictions on a single structural equation and on a structural system. His test can be implemented straightforwardly using the procedure described in section 2.

To illustrate the implementation of the test, we consider overidentifying restrictions on a single - say the first - structural equation. As we ignore other constraints on the system, we can limit ourselves to the set of reduced form equations

\[ (y_1, Y_1) = X \Pi_1 + V_1, \tag{4.1} \]

\[ T_x g_1 \quad T_x k x g_1 T_x g_1 \]
for the \( g_1 \) endogenous variables included in the first-structural equation

\[
y_1 + Y_1 \gamma_1 + X_1 \delta_1 = u_1 . \tag{4.2}
\]

When the exclusion restrictions on (4.2) are true, the matrix \( \Pi_i \)
partitioned appropriately satisfies the relations

\[
\Pi_i \begin{pmatrix} 1 \\ Y_1 \end{pmatrix} = \begin{pmatrix} \pi_{11} & \Gamma_{11} \\ \pi_{21} & \Pi_{21} \end{pmatrix} \begin{pmatrix} 1 \\ \gamma_1 \end{pmatrix} = \begin{pmatrix} \delta_1 \\ \Theta \end{pmatrix} , \tag{4.3}
\]

\[
(k_1 + k - k_i) \times g_j (g_j \times 1)
\]

where \( k_i \) is the number of predetermined variables included in (4.2).

The overidentifying restrictions can be written as

\[
f(\beta, \Theta) = 0 = (\pi_{21} \Pi_{21}) \begin{pmatrix} 1 \\ Y_1 \end{pmatrix} , \tag{4.4}
\]

\[
(g_j - 1 + r) \times 1
\]

with \( \beta = \gamma_1 \) and \( \Theta = \text{vec} (\Pi_i) = \pi \), \( r = k - k_j - g_j + 1 \) (assumed to be strictly positive). Notice that \( \delta_1 \) is a vector of nuisance parameters in the present case. In order to get a solution for \( \gamma_1 \), we partition \( (\pi_{21} \Pi_{21}) \) as \( \begin{pmatrix} \pi_{21} & \Pi_{21} \end{pmatrix} \), where the number of rows of the blocks is \( (g_j - 1) \) and \( r \) respectively, and \( \Pi_{21} \) is assumed to be nonsingular. Solving \( f_1(\gamma_1, \pi) = \pi_{21} + \Pi_{21} \gamma_1 = 0 \) for \( \gamma_1 \) gives \( \gamma_1 = -\Pi_{21}^{-1} \Pi_{21} \gamma_1 \), so that from \( f_2(\gamma_1, \pi) = \pi_{21} + \Pi_{21} \gamma_1 = 0 \), we have

\[
h(\Theta) = h(\pi) = \pi_{21} - \Pi_{21}^{-1} \Pi_{21} \pi_{21} = 0 . \tag{4.5}
\]

Notice that for a given choice of \( f \), there is only one solution for \( \gamma_1 \), so that under \( H_0 \), the test statistic will always give the same value in large samples. We use (2.7) to obtain the matrix of partial derivatives \( D_\Theta h \).

For this purpose, we need the following results

\[
D_{\gamma_1} f_1 = \Pi_{21} , \quad D_{\gamma_1} f_2 = \Pi_{21} ,
\]

\[
D_\pi f_1 = (1 \gamma_1') \otimes [0 (g_j - 1) \times k_j]^{-1} (g_j - 1) \times r]
\]

\[
D_\pi f_2 = (1 \gamma_1') \otimes [0 (k_j + g_j - 1) \times r] , \tag{4.6}
\]
where $\otimes$ denotes the Kronecker product, $A \otimes B = [a_{ij}B]$. We substitute the expressions (4.6) in (2.7) and obtain after some transformations the partial derivatives

$$D^h = (1 Y_1^t) \otimes [0_{r \times k_1} - H_2 \bar{H}_2^{-1} I_r].$$

The results (4.5) and (4.7) evaluated at the OLS estimate of $II_1$, $\hat{f}_1 = (X'X)^{-1} X'(y Y_1)$, and the variance of vec $(\hat{f}_1) = \Omega$, $\sum_{y} \otimes (X'X)^{-1}$, with $\sum_{y}$ being a consistent estimate of the covariance matrix of the reduced form disturbances in (4.1), can be substituted into (2.2) to yield the value of the Wald statistic for overidentifying restrictions. A Wald test for the restrictions on a complete SEM can be derived along the same lines.

4.2 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 different 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 SEM's. Mizon and Hendry (1980) give an application of Sargan's (1980a) method. A single regression equation with common factors can be written as

$$y_t = \sum_{i=1}^{k} \phi_i(L)x_{it} + \varepsilon_t,$$

where $y_t$ is the endogenous variable, $\varepsilon_t$ 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 $i=1,...,k$. The polynomials $\phi(L)$ and $\phi_i(L)$, $i=0,...,k$, have degree $p$ and $r_i$ respectively. The roots of $\phi(L)$ $\phi_0(L)$ lie outside the unit circle. The model (4.8) is a special case of

$$y_t = \sum_{i=1}^{k} \theta_i(L)x_{it} + \varepsilon_t,$$
where \( \Theta_i(L) \), \( i = 0, \ldots, k \), are polynomials in \( L \) of degree \( p + r_i \), and the roots of \( \Theta_0(L) \) lie outside the unit circle. The number of parameters in (4.8) and (4.9) is \( m = p + \sum_{i=0}^{k} r_i \) and \( n = (1+k)p + \sum_{i=0}^{k} r_i + k \) respectively, so that the common factor structure in (4.8) leads to \( pk \) restrictions on the parameters in (4.9).

For the procedure presented in section 2, we determine the implicit restrictions by equating the corresponding coefficients in (4.8) and (4.9). A subset of \( m \) restrictions forms the system \( f_1(\beta, \theta) = 0 \) and is used to obtain a solution for \( \beta \), the set of parameters in (4.8). The formulae given in (2.8) are then used to compute the Wald statistic, because the implicit restrictions are of the form \( f(\beta) - \Theta = 0 \).

Computation of the Wald test is straightforward in this case.

However, 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 \( H_0 \). A simple example given by Mizon and Hendry (1980) is illuminating in this respect. They consider a special case of models (4.8) and (4.9) written as

\[
y_t = (\gamma + \alpha)y_{t-1} - \phi_2 y_{t-2} + \gamma_0 x_t + (\gamma_1 - \phi_0) x_{t-1} - \phi_1 x_{t-2} + \varepsilon_t
\]

with \( k = p = r_0 = r_1 = 1 \),

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 \( \Theta = (\theta_1, \ldots, \theta_5)' \)

\[
f_1(\beta, \Theta) = 0 : \begin{align*}
\phi + \alpha - \Theta_1 &= 0 \\
-\phi_2 - \Theta_2 &= 0 \\
\gamma_0 - \Theta_3 &= 0 \\
\gamma_1 - \phi \gamma_0 - \Theta_4 &= 0
\end{align*}
\]

\[
f_2(\beta, \Theta) = 0 : \begin{align*}
-\phi_1 - \Theta_5 &= 0
\end{align*}
\]  

(4.10)

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 \( f_2 = 0 \). Notice that if both solutions satisfy \( f_2 = 0 \), there exists a functional relationship on \( \beta \), namely \( \gamma_0 \alpha = \gamma_1 \). The requirement that
(1 - \theta_1 L - \theta_2 L^2) = 0 \quad \text{and} \quad (1 - \alpha L)(1 - \phi L) = 0 \quad \text{have their roots outside the unit circle does not resolve the problem of multiple solutions. For instance, for } \theta' = (.5, .2, 1., 5, 1.), \text{ the characteristic roots of the unrestricted model and the restricted model lie inside the unit circle, whereas } (4.10) \text{ 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

\[ D_{\theta} h \bigg|_{\theta = \hat{\theta}} = \left( \frac{\gamma_1 \phi + \gamma_0 \phi^2}{\alpha - \phi}, \frac{\gamma_1 + \gamma_0 \phi}{\alpha - \phi}, -\phi^2, -\phi, -1 \right) \bigg|_{\phi = \hat{\phi}}. \]  

(4.11)

Computation of the Wald test when (2.8) is evaluated in a solution of \( f_1 = 0 \) that also satisfies \( f_2 = 0 \) asymptotically will yield the value for 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 \( \text{plim} \, \hat{\theta}_h \) 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 \( \theta \) implied by (4.10) explicitly. They find

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

(4.12)

If the implicit relations (4.10) are substituted in (4.12), it is obvious that the restriction on \( \theta \) implied by (4.12) must be valid under \( H_0 \). However, the formulation of the restriction in (4.12) is not unique. After some transformation of (4.10), we also find

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

(4.13)

as a restriction. There does not exist an equivalence between the partial derivatives of (4.12) and (4.13), as can be seen by determining \( D_{\theta} h \) for both of them and substituting the implicit relations in (4.10). According to the results of section 3, the two associated Wald statistics will not be equivalent asymptotically. The problem caused by multiple solutions is present whether we choose an explicit or an implicit form for the restrictions. In empirical work, one will have to compute the different solutions for the set of restrictions. The null hypothesis \( H_0 \)
is not rejected once we have found a solution for which the test is in favor of $H_0$. It can only be rejected when for all solutions for the restrictions, $H_0$ has to be rejected.

4.3 Testing for LISREL

In this section, we consider an example of the linear structural relations (LISREL) which have been widely used in economics and other social sciences and we show how the Wald test applies in this case. As the LISREL-model linearly relates observed variables to some unobservable quantities, it can be used to generate restrictions on the second moments of the observed variables. In the economic time series literature, LISREL-models have been postulated for second-order stationary processes (see e.g. among others Geweke and Singleton (1981)) and tested by means of frequency domain methods. In addition to the requirement of weak stationarity, we assume that the observed variables are generated by a finite order vector autoregressive model. This assumption which is frequently made in applied work simplifies the presentation. We show how covariance structures for a multivariate dynamic model can be tested in the time domain.

We assume now that aggregate expenditures on good $i$ in constant prices $c_{it}$, $i = 1, \ldots, N$, and aggregate personal disposable income in constant prices $y_t$, are (after appropriate transformation) generated by a weakly stationary $p$-th order autoregressive model.

$$
\begin{bmatrix}
c_{1t} \\
c_{2t} \\
\vdots \\
c_{Nt} \\
y_t
\end{bmatrix}
= x_t = A(L) x_t + u_t, \quad (4.14)
$$

where $A(L) = \sum_{j=1}^{p} A_j L^j$ is a $(N+1)$ order matrix polynomial lag operator of degree $p$ with constant coefficients, $u_t$ is a vector white noise with zero mean and nonsingular covariance matrix $\Omega$. The roots of the determinant $|I - A(L)|$ are assumed to lie outside the unit circle.

Now we consider a simplified version of a model recently used by Geweke and Singleton (1981).
\[ c_{it} = \delta_i(L) z_t + \varepsilon_{it} \]
\[ y_t = z_t + v_t \quad (4.15) \]

where \( z_t \) denotes the unobserved permanent income at time \( t \) in constant prices, \( \varepsilon_{it} \) and \( v_t \) are disturbances that can be interpreted as transitory consumption and income respectively. We assume that \( \varepsilon_{it} \) and \( v_t \) have zero mean, constant variances \( \sigma_{\varepsilon_i}^2 \) and \( \sigma_v^2 \), zero autocorrelations and zero cross-correlations at all leads and lags and that they are independent of \( z_t \), for all \( t \) and \( t' \). The \( \delta_i(L) \)'s are one-sided scalar polynomials of degree \( d_i \) in \( L \). For the sake of simplicity, we assume that all variables in (4.14) and (4.15) have mean zero. By postulating model (4.15), we introduce an unobservable variable \( z_t \), but more importantly, we are able to formulate restrictions on the parameters of (4.14). Imposing these restrictions, (4.14) leads to a nested null hypothesis which can be tested against the more general model (4.14). At this point, several comments have to be made:

- Model (4.15) is an example of confirmatory dynamic factor analysis, of which the static version (see e.g. Jöreskog (1969)) is readily obtained by appropriate specialization of (4.15).
- The variable \( z_t \) in (4.15) can also be interpreted as an unobserved component (see e.g. Nerlove et al. (1979)), for which a process can be specified and combined with (4.15) to yield the joint process of \( y_t \) (instead of assuming (4.14)).
- The disturbance \( v_t \) in (4.15) can be interpreted as an error of measurement. If the second moments for \( v_t \) are known, the estimated second moments for \( y_t \) can be corrected for the effect of \( v_t \) to yield consistent (and asymptotically normally distributed) estimates of the corresponding moments for \( z_t \). Then, the set of restrictions implied by the first part of (4.15) can be tested using a Wald criterion. This approach is an example of the population-sample decomposition, (see e.g. van Praag (1982), for more examples), where the sample moments are adjusted for deficiencies in the observations before being used to test restrictions on the population parameters. The estimates of these parameters will not always be efficient and the corresponding Wald test is not necessarily most powerful (for local alternatives). However, it may be computationally more convenient and more robust than a likelihood ratio test or a Wald test based on maximum likelihood estimates for a complete model taking account of the sample deficiencies. Nevertheless, the adjustment of the estimated moments for the nonstandard
sampling requires a priori knowledge, that may not be available in practice.

To illustrate the use of a Wald test, we consider a special case of (4.14) and (4.15), with $N=2$, $p=1$, $d_1=d_2=0$. For (4.14), we have

$$\Gamma_0 = A_1 \Gamma_{-1} + \Omega$$
$$\Gamma_1 = A_1 \Gamma_0$$

(4.16)

where $\Gamma_1 = \sum \chi \chi_{i-1}' = \Gamma_{-1}'$.

From (4.15), we get

$$\Gamma_0 - \begin{bmatrix} \delta_{10} \\ \delta_{20} \\ 1 \end{bmatrix} (Y_{z10} Y_{z20} Y_{z30}) - \begin{bmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{33} \end{bmatrix} = 0$$

and

$$\Gamma_1 = \begin{bmatrix} \delta_{10} \\ \delta_{20} \\ 1 \end{bmatrix} (Y_{z11} Y_{z21} Y_{z31}) = 0$$

(4.17)

with $\gamma_{zji} = \sum \chi \chi_{i-1}$. The system (4.17) implies 18 relations between the 15 parameters of $\Gamma_0$ and $\Gamma_1$ and the 11 parameters in $\beta = (\delta_{10}, \delta_{20}, Y_{z10}, Y_{z20}, Y_{z30}, Y_{z11}, Y_{z21}, Y_{z31}, \sigma_{11}, \sigma_{22}, \sigma_{33})'$. Alternatively, we can relate the parameters of $\beta$ to those of $\Omega$ and $A_1$ by substituting (4.17) for $\Gamma_0$ and $\Gamma_1$ in (4.16). The two corresponding Wald statistics have the same value in large samples because of the one-to-one relationship between $(\Gamma_0, \Gamma_1)$ and $(A_1, \Omega)$. Through vectorizing the set of relations (4.17), we get the restrictions in the form of (2.4). The choice of $f_1$ is fairly obvious in this case. Provided estimates for $\theta = \text{vec} (\Gamma_0^*, \Gamma_1^*)$ or $\theta = \text{vec} (\Omega^* A_1^*)$, where * indicates that only the freely varying parameters are included in $\theta$, are available and their joint (asymptotic) distribution is known, the Wald statistic can be computed as outlined in section 2. As (4.17) implies 7 implicit restrictions on $\theta$, under $H_0$, the Wald statistic is chi-square distributed with 7 degrees of freedom. When we use OLS estimates for $A_1$ and $\Omega$, the Wald test will be equivalent in large samples to a likelihood ratio test conditional on initial values. Notice finally that when $\sigma_{33}$ is known a priori, $Y_{z30}$ can be
consistently estimated by \( \sum_{t} z_{t}^{2} - \sigma_{zz} \), and as all remaining second moments of \( z_{t} \) appearing in (4.17) can be consistently estimated by the corresponding sample moments for \( y_{t} \), the population-sample decomposition can be straightforwardly applied.

From the discussion in this subsection, it should be clear that the Wald test can be useful for analyzing covariance structures in static and dynamic models, when an initial model such as e.g. (4.14) has been specified.

5. 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. The proposed procedure is expected to save the investigator from the time-consuming activity of expressing the restrictions in explicit form. We also investigated the consequences of the choice of a particular form of the restrictions for the value of the Wald statistic. The troublesome problem of multiple solutions for a subset of nonlinear constraints used to compute an initial value for \( \beta \) has also been analyzed. In particular, it has been shown that for some solutions, the Wald statistic will tend to infinity.

Three econometric applications for the Wald test have been presented in order to illustrate the possibilities for using Wald criteria and the problems that might arise. Several important applications have not been discussed. Among them, we mention the set of nonlinear constraints implied by the rational expectation hypothesis in a SEM (see e.g. Hoffman and Schmidt (1981) and Revankar (1980)), the Hausman specification test (see e.g. Hausman (1978), Holly (1982)), the Almon polynomial lag constraint (see e.g. Sargan (1980b)). Finally, the Wald encompassing test can be used to test nonnested hypotheses (see e.g. Mizon and Richard (1982)). Its implementation however, requires some modifications of the general procedure presented here.
Appendix

In this appendix, we show that

$$A \begin{bmatrix} -D_B f_3^{*+4} (D_B f_{1+2}^*)^{-1} D_\theta f_{1+2}^* + D_\theta f_3^{*+4} \\ \vdots \\ -D_B f_3^{*+4} (D_B f_{2+3}^*)^{-1} D_\theta f_{2+3}^* + D_\theta f_3^{*+4} \end{bmatrix} =$$

$$\begin{bmatrix} -D_B f_3^{*+4} (D_B f_{1+2}^*)^{-1} D_\theta f_{1+2}^* + D_\theta f_3^{*+4} \end{bmatrix} , \quad (A.1)$$

where $A = \begin{bmatrix} -D_B f_3^{*+4} (D_B f_{1+2}^*)^{-1} D_\theta f_{1+2}^* + D_\theta f_3^{*+4} \\ \vdots \\ -D_B f_3^{*+4} (D_B f_{2+3}^*)^{-1} D_\theta f_{2+3}^* + D_\theta f_3^{*+4} \end{bmatrix}$ is defined in (3.4) and $B_2$ is given in (3.5) and the formulae are evaluated at $(\beta, \theta) = (\hat{\beta}, \hat{\theta})$. The matrix multiplication in the l.h.s. of (A.1) gives:

$$\begin{bmatrix} D_\beta f_1^{*+4} B_2 D_\beta f_3^* \end{bmatrix} \begin{bmatrix} 0_{k \times m} \\ -D_B f_{1+2}^* \end{bmatrix} D_\theta f_{1+2}^* +$$

$$\begin{bmatrix} D_\beta f_1^{*+4} B_2 D_\theta f_3^* \end{bmatrix} \begin{bmatrix} 0_{k \times m} \\ -D_B f_{1+2}^* \end{bmatrix} D_\theta f_{1+2}^* +$$

$$\begin{bmatrix} D_\beta f_1^{*+4} B_2 D_\theta f_3^* \end{bmatrix} \begin{bmatrix} 0_{k \times n} \\ -D_B f_{1+2}^* \end{bmatrix} D_\theta f_{1+2}^* + \begin{bmatrix} 0_{k \times n} \\ -D_B f_{1+2}^* \end{bmatrix} D_\theta f_{1+2}^* +$$

From the definition (3.5), we have the following identity

$$B_2 D_\beta f_3^* = I_m - B_1 D_\beta f_2^* ,$$

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

$$\begin{bmatrix} I_k \\ 0_{k \times m-k} \\ D_\beta f_{1+4} (D_B f_{1+2}^*)^{-1} \end{bmatrix} - D_\beta f_{1+4} B_1 (0_{m-k \times k} \vdots I_{m-k}) +$$

$$\begin{bmatrix} 0_{k \times m} \\ D_\beta f_{1+4} (D_B f_{1+2}^*)^{-1} \end{bmatrix} D_\theta f_{1+2}^* - D_\beta f_{1+4} B_2 D_\theta f_3^* + \begin{bmatrix} 0_{k \times n} \\ D_\theta f_4^* \end{bmatrix} . \quad (A.3)$$

Expression (A.3) is equivalent to

$$\begin{bmatrix} D_\theta f_1^* \\ 0_{r-k \times n} \end{bmatrix} - D_\beta f_{1+4} B_1 \begin{bmatrix} 0_{m-k \times n} + D_\theta f_2^* \end{bmatrix} - D_\beta f_{1+4} B_2 D_\theta f_3^* +$$

$$\begin{bmatrix} 0_{k \times n} \\ D_\theta f_4^* \end{bmatrix} . \quad (A.4)$$
Using (3.5) in (A.4), we find the desired result

\[ - D_\beta f^*_{1+4} (D_\beta f^*_{2+3})^{-1} D_\alpha f^*_{2+3} + D_\alpha f^*_{1+4} = 0. \]
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.