Statistische Hefte 28, 263–270 (1987)

Statistische Hefte
Statistical Papers

Testing the stability of a linear dynamic model

David A. Kodde and Franz C. Palm

Received: Feb. 2, 1987; revised version: May 29, 1987

In this paper we present a Wald or distance test for testing the stability of a linear dynamic model. Stability requires that all latent roots of the system simultaneously satisfy inequality restrictions. Unlike previous tests proposed in the literature our procedure is capable of testing the restrictions simultaneously. Therefore, the test asymptotically has the correct size. The procedure can be applied in practice if stability is not a requirement for identification of the dynamic model.

I. INTRODUCTION

Linear dynamic models are widely used in applied econometrics. The conditions for the stability of linear dynamic models with constant coefficients have been derived in the literature. For continuous time linear models, the requirement for stability is that the real part of the roots of the characteristic equation is negative. To assure the stability of discrete time dynamic systems, the roots of the characteristic equation must have modulus smaller than one. We refer the reader to e.g. Aoki (1976) for a discussion of the stability conditions.
The stability condition is testable if it implies overidentifying restrictions on the parameters of the model. In many contributions to the time series literature, the stability condition is part of the identification requirements (see e.g. Hannan (1969, 1971) and Hatanaka (1975)). However, Deistler (1976) gives conditions for the identification of dynamic models which do not require stability. In particular he assumes that the initial values of the process are zero. A similar result holds when we estimate a dynamic model conditional on initial values. In these cases, the stability condition becomes testable.

To check the stability of a linear system, it has been suggested in the literature to compute the characteristic roots and to test whether the dominant root satisfies the stability condition. For instance, Theil and

Boot (1962) derive the dominant root and its large sample variance for the estimated characteristic equation of Klein's model I and test whether the dominant root differs significantly from 1 in absolute value.
This procedure can lead to incorrect inference as the number of estimated roots which are outside the stability region is stochastic. The stability requirement implies inequality restrictions on all the characteristic roots. These inequality restrictions have to be jointly tested. Therefore, a test based on the dominant root only is not sufficient. As an alternative one might consider testing each root separately and to control the overall level of significance by using the Bonferroni inequality (see e.g. Savin (1980, 1983)). However, since the individual latent roots will be correlated, the size of the test is not correct asymptotically. As the null hypothesis of stability corresponds to inequality constraints, standard tests based on quadratic forms of the estimated characteristic roots are not χ^2-distributed in large samples. Under the null hypothesis, their asymptotic distribution is a mixture of χ^2-distributions (see e.g. Gouriéroux et al. (1980, 1982)).

In the main section of this paper, we present a Wald test which can be used in a fairly straightforward way to test the stability of a linear dynamic model. We show how the test can be implemented in applied work and we give its asymptotic distribution under the null hypothesis. Our Wald test solves the problems mentioned above. The concluding section comments on other related applications of the Wald test and on the relationship between the Wald test and other test criteria.

II. A WALD TEST FOR THE STABILITY CONDITION

Consider the following first order k dimensional dynamic system for x

$$x_t = b + A(\theta)x_{t-1} + \varepsilon_t, \tag{2.1 A}$$

when t is a discrete time parameter, and for continuous time

$$x(t) = b + B(\theta)x(t) + \varepsilon(t). \tag{2.1 B}$$

The elements of the $k \times k$ matrices $A(\theta)$ and $B(\theta)$ are functions of the vector of parameters θ, b is a vector of intercepts and ε is a disturbance term. Notice that the system is written in reduced form. The elements of θ correspond to the underlying structural parameters. When the underlying structural form is an n-th order system, it has to be converted into a first order system to get (2.1). We assume that the stability condition is testable.

For the discrete time model the stability requires the eigenvalues of $A(\theta)$, λ_i, to satisfy the restrictions

$$\lambda_i \lambda_i^* < 1, \ i = 1, \ldots, p, \tag{2.2}$$

where λ_i^* is the complex conjugate of λ_i and $p = k - r$, with r being the number of pairs of complex roots. The real part of the eigenvalues of $B(\theta)$ has to be negative to assure the stability of the continuous time model, or alternatively

$$\lambda_i + \lambda_i^* < 0, \ i = 1, \ldots, p. \tag{2.3}$$

In both cases, the set of inequality constraints can be straightforwardly expressed as nonlinear functions of θ,

$$H_0 : h(\theta) > 0, \tag{2.4}$$

where the dimension of $h(\theta)$ equals p. Under the alternative hypothesis, the eigenvalues are unrestricted, which we express as

$$H_1 : h(\theta) \not> 0. \tag{2.5}$$

To introduce a Wald or distance test, we assume that θ can be consistently estimated by $\bar{\theta}$ such that the asymptotic distribution is given by

$$T^{\frac{1}{2}}(\bar{\theta} - \theta_0) \underset{A}{\sim} N(0,\Omega), \tag{2.6}$$

where θ_0 is the true value of θ and Ω can be consistently estimated by $\bar{\Omega}$, T denotes the sample size. We transform the functions $h(\theta)$ into a new parameter vector γ defined as

$$\gamma = T^{\frac{1}{2}}\ h(\theta) \text{ and } \bar{\gamma} = T^{\frac{1}{2}}\ h(\bar{\theta}). \tag{2.7}$$

The large sample covariance matrix of $\bar{\gamma}$ is given by

$$\Sigma = (\partial h/\partial\theta')\ \Omega\ (\partial h'/\partial\theta). \tag{2.8}$$

The covariance matrix Σ can be consistently estimated by $\bar{\Sigma}$ evaluating expression (2.8) at $\bar{\theta}$ and $\bar{\Omega}$.
We define the distance of γ from $\bar{\gamma}$ in the metric Σ as

$$\|\bar{\gamma} - \gamma\| = (\bar{\gamma} - \gamma)'\ \Sigma^{-1}\ (\bar{\gamma} - \gamma). \tag{2.9}$$

Let $\tilde{\gamma}$ be the *minimum* distance estimator satisfying the restrictions under $H_0 : \gamma > 0$. The distance between the restricted estimator and unrestricted estimator, $\bar{\gamma}$ (which is the minimum distance estimator under $H_1 : \gamma \not> 0$), is given by

$$D = \inf_{\gamma > 0} \|\bar{\gamma} - \gamma\| = \|\bar{\gamma} - \tilde{\gamma}\|. \tag{2.10}$$

The statistic D in (2.10) defines the Wald or distance test for $H_0 : h(\theta_0) > 0$ against $H_1 : h(\theta_0) \not> 0$. Under H_0 the large sample distribution of D in (2.10) is given by

$$\sup_{\gamma > 0} \Pr(D > c \mid \Sigma) = \sum_{i=0}^{p} \Pr(\chi^2(p-i) > c)\ w(p,i,\Sigma), \tag{2.11}$$

where $w(p,i,\Sigma)$ denotes the probability that precisely i of the p components of γ are strictly positive. For more details and for the derivation, we refer to Kodde and Palm (1986) and to Perlman (1969) and the references therein. Kodde and Palm (1986) give lower and upper bounds for the critical value level of the distance test which can often be used instead of computing the weights w needed to get (2.11).
In order to implement the Wald test, we have to derive Σ, for which we need

$\partial h/\partial\theta'$. Contributions to the literature on the asymptotic variance of latent roots are Oberhofer and Kmenta (1973), Neudecker and van de Panne (1966) and Schmidt (1974). The partial derivative of h with respect to θ' is given by

$$\partial h/\partial\theta' = (\partial h/\partial \text{ vec}'A)(\partial \text{ vec}A/\partial\theta') \tag{2.12}$$

where the p x 1 vector h has an typical element $1 - \lambda_i\lambda_i^*$. For the continuous time model, we have

$$\partial h/\partial\theta' = (\partial h/\partial \text{ vec}'B)\ (\partial \text{ vec}B/\partial\theta') \tag{2.13}$$

with the typical element of h being $-(\lambda_i + \lambda_i^*)$. The partial derivative of h with respect to vec'A is given by

$$\begin{aligned}\partial(1 - \lambda_i\lambda_i^*)/\partial \text{ vec}'A &= -2\ \lambda_i(q_i' \otimes w_i) \text{ if } \lambda_i \text{ is real} \\ &= -\lambda_i^*(q_i' \otimes w_i) - \lambda_i((q_i^*)' \otimes w_i^*) \text{ if } \lambda_i \text{ is complex}\end{aligned} \tag{2.14}$$

with q_i and q_i^* being the column eigenvectors associated with λ_i and λ_i^* respectively and w_i and w_i^* the column latent vectors of λ_i and λ_i^*. For (2.13), we have

$$\begin{aligned}\partial(-\lambda_i - \lambda_i^*)/\partial \text{ vec}'B &= -2(q_i' \otimes w_i) \text{ if } \lambda_i \text{ is real} \\ &= -(q_i' \otimes w_i) - ((q_i^*)' \otimes w_i^*) \text{ if } \lambda_i \text{ is complex.}\end{aligned} \tag{2.15}$$

The vectors q_i, w_i and their complex conjugates are associated with λ_i and λ_i^* but now correspond to the matrix B. For the derivation, we refer the reader to the appendix. The derivative of vec A or vec B with respect to θ' can be derived as soon as the structure of these matrices is given.

To compute the Wald statistic, the following steps have to be carried out :

1) Given the estimate $\bar\theta$, compute A (or B) by $A(\bar\theta)$ (or $B(\bar\theta)$) and compute $1 - \bar\lambda_i\bar\lambda_i^*$ (or $-\bar\lambda_i - \bar\lambda_i^*$).
2) Choose all real eigenvalues and one out of each pair of complex eigenvalues.

3) Compute D in (2.10) with Σ being the covariance matrix of the eigenvalues that have been included in D. Σ is computed using the results in (2.12) - (2.15).
4) The large sample distribution of D is given in (2.11). Reject the null hypothesis if D exceeds the critical value c in (2.11) or the upper bound in Kodde and Palm (1986) for a given significance level.

III. CONCLUDING REMARKS

To conclude, we presented a procedure for testing the stability of a linear dynamic model with discrete or continuous time parameter. The procedure can be implemented in a fairly straightforward way provided the stability condition is testable and a consistent asymptotically normally distributed estimate of the parameters in the model is available. The test can be applied to models with unrestricted as well as restricted reduced form. The latter case encompasses stability tests for rational expectations models and induced dynamic optimal control models with a quadratic penalty function. When asymptotically efficient parameters estimates are available, the Wald test presented in section II is asymptotically equivalent to a likelihood ratio (LR) test and a Kuhn-Tucker (KT) test, in which case the test criteria satisfy the following inequalities : KT $\leqq$ LR $\leqq$ D. Notice that as $h(\theta)$ is a nonlinear function of θ, only local properties of the test have been derived, that is the distribution of the test statistic has been given for values of θ in a neighborhood of θ_0. The fact that Wald criteria are based on an unrestricted estimate of θ should ensure that in large sample under the null hypothesis the test statistic W will indeed be evaluated at a value of θ close to θ_0. This may not be the case for the criteria LR of KT. Finally, we like to note that the testing procedure presented in section II can be easily modified to test other inequality restrictions such as the invertibility condition in moving average models (provided it is testable). With respect to further extensions, it is worthwile to mention that Farebrother (1986) obtains the inequality restrictions implied by the stability condition when the order of the characteristic equation does not exceed 4 but without knowledge about the number of complex roots. To test these restrictions, Farebrother (1986) proposes to use the distance test discussed above.

APPENDIX

In this appendix we derive the results presented in formulae (2.14) and (2.15). First we observe that

$$\partial(1-\lambda_i\lambda_i^*)/\partial \text{ vec}'A = -\lambda_i^* \, \partial\lambda_i/\partial \text{ vec}'A - \lambda_i \partial\lambda_i^*/\partial \text{ vec}'A \tag{A.1}$$

and

$$\partial(-\lambda_i-\lambda_i^*)/\partial \text{ vec}'B = -\partial\lambda_i/\partial \text{ vec}'B - \partial\lambda_i^*/\partial \text{ vec}'B. \tag{A.2}$$

Formulae (A.1) and (A.2) depend on the sensitivity of a latent root with respect to a real matrix. Let C be a square real matrix and μ the latent root under consideration, where w (dimension $1 \times k$) and q (dimension $k \times 1$) are the corresponding latent row and latent column vectors. We have

$$Cq = \mu q, \; wC = \mu w \text{ and } wq = 1. \tag{A.3}$$

Therefore in differentials

$$(dC)q + C(dq) - \mu(dq) = (d\mu)q. \tag{A.4}$$

Premultiplying (A.4) with w and using (A.3) gives $w(dC)q = d\mu$, so that

$$\partial\mu/\partial \text{ vec}'C = (q' \otimes w). \tag{A.5}$$

Result (A.5) can directly be applied to (A.1) and (A.2) if we substitute the triples (λ_i, q_i, w_i) and $(\lambda_i^*, q_i^*, w_i^*)$ for (μ, q, w). The results in (2.13) and (2.14) emerge when we use $\lambda_i = \lambda_i^*$ if λ_i is real. Furthermore it is easy to check that the expressions in (2.14) and (2.15) are real since q_i^* and w_i^* are the complex conjugates of respectively q_i and w_i.

REFERENCES

Aoki, M. (1976), Optimal Control and System Theory in Dynamic Economic Analysis, Amsterdam, North-Holland Publ. Co.

Deistler, M. (1976), "The identifiability of linear econometric models with autocorrelated errors", International Economic Review, 17, 26-46.

Farebrother, R.W. (1986), "Independent conditions for the stability of a dynamic linear model", University of Manchester, Research paper.

Gouriéroux, Ch., Holly, A., and Monfort, A. (1980), "Kuhn-Tucker, likelihood ratio and Wald tests for nonlinear models with inequality constraints on the parameters", Cambridge, Harvard University, Discussion paper 770.

Gouriéroux, Ch., Holly, A., and Monfort, A. (1982), "Likelihood ratio test, Wald test and Kuhn-Tucker test in linear models with inequality constraints on the regression parameters, Econometrica, 50, 63-80.

Hannan, E.J. (1969), "The identification of vector mixed autoregressive moving average systems", Biometrika, 56, 223-225.

Hannan, E.J. (1971), "The identification problem for multiple equation systems with moving average errors", Econometrica, 39, 751-765.

Hatanaka, M. (1975), "On the global identification of the dynamic simultaneous equations model with stationary disturbances", International Economic Review, 16, 545-554.

Kodde, D.A., and F.C. Palm (1986), "Wald criteria for jointly testing equality and inequality restrictions", Econometrica, 54, 1243-1248.

Neudecker, H., and C. van de Panne (1966), "Note on the asymptotic standard errors of latent roots of econometric equation systems", Review of the International Statistical Institute, 34, 43-47.

Oberhofer, W., and J. Kmenta (1973), "Estimation of standard errors of the characteristic roots of a dynamic econometric model", Econometrica, 41, 171-177.

Perlman, M.D. (1969), "One sided problems in multivariate analysis", Annals of Mathematical Statistics, 40, 549-567.

Savin, N.E. (1980), "The Bonferroni and the Scheffé multiple comparison procedures", Review of Economic Studies, 47, 255-273.

Savin, N.E. (1983), "Multiple hypothesis testing", In Griliches, Z. and M.P. Intriligator, Handbook of Econometrics, I, North-Holland Publ. Co., Amsterdam.

Schmidt, P. (1974), "The algebraic equivalence of the Oberhofer-Kmenta and Theil-Boot formulae for the asymptotic variance of a characteristic root of a dynamic econometric model", Econometrica, 42, 591-592.

Theil, H., and J.C.G. Boot (1962), "The final form of econometric equation systems", Review of the International Statistical Institute, 30, 136-152.

Kodde D.A.
Department of Economics
Catholic University of Nijmegen
Thomas van Aquinostraat 6
6525 EM Nijmegen
The Netherlands

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

NCC/IBL aanvraagbon A07885850X 10062005

Materiaal	: Abx	PPN: 108880508	
Titel	: Statistical papers		
	:		
Auteur	:		
Deel/Supplem.	:		
Corporatie	:	Externe database:	
Jaar/Editie	: 1988	Extern nummer	:
Uitgave	: Berlin	Springer	
Serie/Sectie	:		
Pag-ISSN/ISBN	:	0932-5026	

Plaatscode : 108880508 ; T 02383 ; ; 1988 V29 - 1990 V36

Jaar	: 1988-00-00	Datum indienen	: 09-06-2005 13:41
Volume	: 29	Datum plaatsing	: 09-06-2005 13:41
Aflevering	:	Datum rappel	: 07-07-2005
Leenvorm	: KOPIE	Particulier	: N
Leveringswijze	: P	Geplaatst bij	: 0017
Cooperatiecode	: B	Indiener	: 0036/9999
Aanvrager	: 0036/9015 U UKB		
Eindgebruiker	: UM55046	Aanvraagident.	:
Auteur	: Kodde, Palm	Aanvragerident.	: MW. Y.A.C. PEETERS
Artikel	: Computing wald criteria for nested hypotheses.		
Bladzijden	: 169-190		
Bron	:		
Opmerking	: Email heeft voorkeur.		

Indiener	: 0036/9999	Stuur rekening	: N
Aanvrager	: 0036/9015	Eindgebruiker	: UM55046
Aanvragerident.	: MW. Y.A.C. PEETE	Aanvraagident.	:
Client	:		

Afleveradres Post Mw. Y.A.C. Peeters
Universiteit Maastricht

UB/ISD, Postbus 616 (GL17)
6200 MD MAASTRICHT

NL

Fax
Email Y.Peeters@ub.unimaas.nl
Ftp
Ariel
Telefoon

Factuuradres Clearing House

[1] origineel gestuurd	[4] nog niet aanwezig	[7] uitgeleend
[2] kopie gestuurd	[5] niet aanwezig	[8] wordt niet uitgeleend
[3] overige	[6] niet beschikbaar	[9] bibliografisch onjuist
		[0] bij de binder

Aantal eenheden :
Aanvraagnummer : A07885850X

Statistical Papers 29, 169–190 (1988)

Statistical Papers
Statistische Hefte

Articles

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 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

$$\hat{\Omega}_h = (D_\theta h)\ \hat{\Omega}_\theta\ (D_\theta h)', \tag{2.3}$$

with $D_\theta h$ denoting the first derivative matrix of h with respect to θ' 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 $D_x y$, 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 χ^2-distributed in large samples with r degrees of freedom, provided that plim $\hat{\Omega}_h$ is nonsingular and that $D_\theta h$ is a continuous function of θ at the true parameter value θ_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 $D_\theta h$ 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

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

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. Under H_0, θ_0 satisfies the implicit restrictions (2.4) and it does so for a unique value β_0 of β (in the interior of the parameter space for β). The matrices $D_\beta f$ and $D_\theta f$ 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, $f_1(\beta,\theta) = 0$, such that β can be solved explicitly as a function of θ, that is $\beta = \beta(\theta)$. When locally no solution exists to $f_1(\beta,\hat{\theta}) = 0$, our result still holds true asymptotically if $\hat{\theta}$ converges in probability to θ_0, because we assume that $f(\beta,\theta_0) = 0$ has a solution. This solution is substituted in the r remaining relations that we denote by $f_2(\beta,\theta) = 0$ to give

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

Next, we obtain an expression for the partial derivatives. 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

$$D_\theta h = -F_2F_1^{-1}Q_1 + Q_2 = HQ, \tag{2.7a}$$

with $H = [-F_2F_1^{-1}\ I_r]$. 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

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

with H and Q being evaluated at (β_0,θ_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, $f(\beta) - \theta = 0$, 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

$$h(\theta) = f_2(\beta) - \theta_2 \text{ and } D_\theta h = -H, \tag{2.8}$$

with θ_2 being the appropriate subvector of θ.

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). In the incidental case where

$D_\theta h$ in (2.7a) does not have full rank r, a consistent estimate of the generalized inverse of Ω_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 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 $\hat{\beta}$ is found by minimizing

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

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

$$\hat{\beta} - \beta_0 = PQ(\hat{\theta} - \theta_0) + o_p(T^{-½}), \tag{2.10}$$

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

$$\begin{aligned} f(\hat{\beta},\hat{\theta}) &= [I + \tilde{F}\tilde{P}]\tilde{Q}(\hat{\theta} - \theta_0), \\ &= [I + FP]Q(\hat{\theta} - \theta_0) + o_p(T^{-½}), \end{aligned} \tag{2.11}$$

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

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

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

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

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

When $S = [\hat{Q}\ \hat{\Omega}_\theta\ \hat{Q}']^{-1}$, the asymptotic covariance matrix of $f(\hat{\beta},\hat{\theta})$ in (2.11) is

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

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

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

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 criteria. Notice, however, that Q may depend on β so that a consistent estimate of β 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 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 $D_\theta g = AD_\theta h$ 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 (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 $D_\theta h$ 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, $D_y g(y,\theta)$ being nonsingular and $D_\theta g(y,\theta)$ 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 partial derivatives. The matrices of partial derivatives of h and g with respect to θ are given by

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

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

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

where $\hat{\theta}$ is a consistent estimate of θ and $\hat{y} = h(\hat{\theta})$. 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, $D_y g(y,\beta,\theta)$ is nonsingular, $D_\beta g(y,\beta,\theta) = 0$ and $D_\theta g(y,\beta,\theta) = 0$ 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 $\hat{\theta}$. In addition, the difference between $D_\theta\beta$ based on f and g respectively, vanishes in large samples (see also Gouriéroux et al. (1985)).
Therefore,

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

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

$$f_i^*(\beta,\theta) = 0, \quad i = 1,...4. \tag{3.6}$$

To simplify the notation, we delete the arguments β and θ and we denote the subset of restrictions f_i^* and f_j^* 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

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

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, $D_\theta h_2 = AD_\theta h_1$. A nonsingular matrix that gives the desired result is

$$\underset{r \times r}{A} = \begin{bmatrix} -F_{1+4}B_2 & \begin{matrix} O_{k\ r-k} \\ I_{r-k} \end{matrix} \end{bmatrix}, \tag{3.9}$$

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

$$[B_1\ \ B_2] = [F_{2+3}]^{-1}. \tag{3.10}$$

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. From (2.7), we obtain that

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

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

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

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

$$\begin{aligned} Hf(\hat{\beta},\hat{\theta}) &= HQ(\hat{\theta} - \theta_0) + o_p(T^{-1/2}) = \\ &= h(\hat{\theta}) + o_p(T^{-1/2}), \end{aligned} \tag{3.13}$$

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

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

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

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

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

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

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 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 (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

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

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

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

when $\theta_0(L) = \phi(L)\alpha(L)$ and $\theta_i(L) = \phi(L)\gamma_i(L)$, $i = 1,\ldots,k$. The number of 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 $f(\beta)-\theta = 0$ 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

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

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 $y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + \theta_3 x_t + \theta_4 x_{t-1} + \theta_5 x_{t-2} + \epsilon_t$.

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

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

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

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

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

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

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

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

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

$$\theta_5^2 + 2\theta_2\theta_3\theta_5 + \theta_1\theta_4\theta_5 + \theta_1^2\theta_3\theta_5 + \theta_2^2\theta_3^2 - \theta_2\theta_4^2 - \theta_1\theta_2\theta_3\theta_4 = 0, \tag{5.7}$$

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 $f(\beta(\hat{\theta}),\hat{\theta}) = 0$ 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 (ψ) = 3 and subsequently for rank (ψ) = 2. Alternatively, in our method we could extend f_2 in (5.3) by adding the restriction $\gamma_1 + \alpha\gamma_0 = 0$.

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 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. 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, β can be efficiently estimated by asymptotic nonlinear least squares applied to the "asymptotic" model $f(\beta,\theta) = 0$ provided a consistent estimate of θ 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 Variance $(V) = \Omega$ 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, $\mathrm{rank}(A\Omega A') = \mathrm{rank}(\Omega)$.

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

APPENDIX II

In this appendix, we show that

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

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

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

$$[F_{1+4}B_2F_3^* + \binom{0_{k\,m}}{-F_4^*}]\, F_{1+2}^{-1}Q_{1+2} + [-F_{1+4}B_2Q_3^* + \binom{0_{k\,n}}{-Q_4^*}] \,. \tag{A.2}$$

From the definition (3.10) we have the following identity

$$B_2F_3^* = I_m - B_1F_2^*,$$

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

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

$$\left. + \begin{bmatrix} 0_{k\,m} \\ -F_4^*F_{1+2}^{-1} \end{bmatrix}\right] Q_{1+2} - F_{1+4}B_2Q_3^* + \begin{bmatrix} 0_{k\,n} \\ Q_4^* \end{bmatrix} \,. \tag{A.3}$$

Expression (A.3) is equivalent to

$$\begin{bmatrix} Q_1^* \\ 0_{r-k\,n} \end{bmatrix} - F_{1+4}B_1(0_{m-k\,n} + Q^*) - F_{1+4}B_2Q_3^* + \begin{bmatrix} 0_{k\,n} \\ Q_4^* \end{bmatrix} \,. \tag{A.4}$$

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

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, C., Monfort, A., and A. Trognon (1985) : "Moindres carrés asymptotiques, Annales de l'INSEE, 58, 91-122.

Gouriéroux, C., and A. Monfort (1987) : "A general framework for testing a null hypothesis in a 'mixed' form", Institut National de Statistique et des Etudes Economiques, Paris, Document de travail 8711.

Holly, A., and A. Monfort (1985) : "Some useful equivalence properties of Hausman's test", Cahiers de recherches économiques, 8505, 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., Palm F.C. and G.A. Pfann (1987) : "Asymptotic least squares estimation : efficiency considerations", Research-memorandum, University of Limburg.

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.

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.

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.

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