COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 801

ON THE FORMULATION OF WALD TESTS OF NONLINEAR RESTRICTIONS

Peter C.B. Phillips & Joon Y. Park

September 1986

ON THE FORMULATION OF WALD TESTS OF NONLINEAR RESTRICTIONS*

by

P. C. B. Phillips and Joon Y. Park

Cowles Foundation for Research in Economics
Yale University

0. ABSTRACT

This paper utilizes asymptotic expansions to investigate alternative forms of the Wald test of nonlinear restrictions. Some formulae for the asymptotic expansion of the distribution of the Wald statistic are provided for a general case. When specialized to the simple cases that have been studied recently in the literature, these formulae are found to explain rather well the discrepancies in sampling behavior that have been observed by other authors. It is further shown how the correction delivered by the Edgeworth expansion may be used to find transformations of the restrictions which accelerate convergence to the asymptotic distribution.

August, 1986

*Our thanks go to Glena Ames for her skill and effort in typing the manuscript of this paper and to the NSF for research support under Grant Number SES 8519595. This paper was completed while Joon Y. Park was an Alfred P. Sloan Doctoral Dissertation Fellow.

1. INTRODUCTION

The numerical value of the Wald test depends not only on the restriction to be tested but also on its algebraic formulation. Under general conditions Wald statistics which are based upon different but algebraically equivalent forms all have the same asymptotic distribution. However, numerical outcomes of the tests and their finite sample distributions can be substantially different for different forms of the same restrictions. Thus, the adequacy of the usual asymptotic χ^2 approximation may also vary substantially as we change the algebraic form of the restriction.

Gregory and Veall (1985) recently studied this phenomenon by simulation. They observed that the distributions of alternative Wald statistics for a simple nonlinear restriction can be widely divergent in small samples; and they conclude that the algebraic form of the restrictions to be tested is likely to be important in many different empirical applications of the Wald test. Lafontaine and White (1986) go further and argue that it is even possible to obtain any value of the Wald statistic by suitably reformulating the restriction. Breusch and Schmidt (1985) make a similar point in a closely related paper.

The purpose of the present paper is to study this phenomenon by direct analytical methods. Our approach is to develop an asymptotic expansion of the distribution of the Wald statistic in a form that is sufficiently general to permit different formulations of the restrictions. Higher order terms in the expansion then provide a mechanism by which deviations from the common asymptotic theory may be measured for alternative forms of the Wald test. In this sense the asymptotic expansion provides more complete distributional information than crude asymptotic theory and, in general, leads to distinct higher order terms for different algebraic representations of the restrictions. As is common in finite sample theory, these terms involve parametric dependencies, so that the adequacy of the asymptotics in each case will depend on the relevant region of the parameter space. This information can be used to improve decision making since it provides advice concerning the behavior of alternative forms of the Wald test in different regions of the parameter space.

Our results indicate that terms in the asymptotic expansion up to $O(T^{-1})$ where T is the sample size provide enough additional information to capture the main distributional effects that are incurred by using alternative forms of the Wald test. The correction terms may be used to determine which version of a Wald test has a sampling distribution that is more closely approximated by the asymptotic. In simple cases the correction terms themselves suggest transformations of the restrictions which will improve the asymptotic approximation by eliminating the correction to a certain order and thereby accelerate convergence to the asymptotic distribution. An example is provided in Section 3. This idea is inspired by earlier work by Phillips (1979) and Konishi (1981) on the use of Edgeworth expansions to determine the form of normalizing transformations.

The paper is organized as follows. Section 2 develops formulae for the asymptotic expansion of the distribution of the Wald test in a general setting. These formulae are specialized in Section 3 to study the examples given recently by other authors and to examine transformations which accelerate convergence. Some brief conclusions are given in Section 4. Proofs are provided in Appendix A and Appendix B details some useful additional formulae for the Edgeworth expansion of the Wald test.

2. EDGEWORTH EXPANSION IN A GENERAL CASE

We start by assuming that the data generating mechanism depends on a set of parameters represented by the $p \times 1$ vector β whose true value we denote by β^0. The hypothesis to be tested takes the form

$$H_0 : g(\beta^0) = 0 \quad (1)$$

where $g : R^p \to R^r$, $r \leq p$, is a vector valued function that is continuously differentiable at least to the third order. Given a sample of size T, let $\hat{\beta}$ denote an estimator of β^0 and define $q = \hat{\beta} - \beta^0$ and $\bar{q} = \sqrt{T}q$.

Under general conditions we can expect that

$$\bar{q} \Rightarrow N(0,\Omega) \quad (2)$$

as $T \uparrow \infty$, where the symbol " $\Rightarrow$ " represents weak convergence of the associated probability measures and Ω is some positive semi-definite matrix, possibly dependent on β.

The vector $\bar{q}$ may be regarded as being composed of low order sample moments of the underlying data or simple functions (often rational functions, as in the case of instrumental variables estimation) of such sample moments. As shown in Phillips (1977) and Sargan and Satchell (1986) the distribution of $\bar{q}$ will admit a valid asymptotic expansion under very general conditions. Details of conditions which are sufficient to ensure the existence of a valid expansion are given in these articles and a constructive process by which the expansion may be obtained is detailed in Phillips (1982). The asymptotic expansion of the density of $\bar{q}$ has the form:

$$(3) \qquad pdf(\bar{q}) = (2\pi)^{-p/2}(\det \Omega)^{-1/2}\exp\{-(1/2)\bar{q}'\Omega^{-1}\bar{q}\} \cdot [1 + \Sigma_{j=1}^{\nu-1}P_j(\bar{q})T^{-j/2}] + o(T^{-(\nu+1)/2})$$

where $P_j(\bar{q})$ is a real polynomial in the elements of $\bar{q}$ of degree $3j$ and where we have assumed that Ω is positive definite. The coefficients in the polynomials P_j in (3) are determined by the cumulants of the underlying sample moments upon which $\bar{q}$ depends and the derivatives of the functions which define that dependence, both to a sufficiently high order that is determined in turn by the order, ν , of the expansion. P_j is an odd (respectively, even) function when the index j is odd (even).

For the main purpose of this paper we shall be working at a sufficient level of generality if we require that standardizing transformations have been carried out which ensure that $\Omega = I_p$ and that, instead of (3), we have quite simply:

$$(4) \qquad \bar{q} \equiv N(0,I)$$

where the symbol " $\equiv$ " signifies equality in distribution. Our results may be extended quite simply to the fully general case of (3) by transformation and by carrying the additional terms induced by (3) in our subsequent computations of the asymptotic expansion of the Wald test. Formulae that apply in this case of full generality are given in Appendix B. They should be useful in many additional contexts.

We now define $G = \partial g/\partial\beta'$ and use carets over such functions of β to signify their evaluation at $\hat{\beta}$, so that $\hat{G} = G(\hat{\beta})$ whereas $G = G(\beta^0)$. In what follows we assume that G (respectively, $\hat{G}$) is of full rank r (with probability one). We shall also use the tensor summation convention of a repeated suffix.

The Wald statistic for testing (1) now has the form

$$(5) \qquad W = T\hat{g}'(\hat{G}\hat{G}')^{-1}\hat{g}$$

and as $T \uparrow \infty$

$$(6) \qquad W \Rightarrow \chi_r^2$$

under the stated conditions. To refine (6) we first develop the following Taylor representations of $\sqrt{T}\hat{g}$ and $\hat{\Psi}^{-1} = (\hat{G}\hat{G}')^{-1} = (\hat{\psi}^{ia})$. Thus, to $O_p(T^{-1})$ we have:

$$\sqrt{T}\hat{g}_i = g_{ij}\bar{q}_j + \frac{1}{2\sqrt{T}} g_{ijk}\bar{q}_j\bar{q}_k + \frac{1}{6T} g_{ijk\ell}\bar{q}_j\bar{q}_k\bar{q}_\ell + O_p(T^{-3/2})$$

$$\hat{\psi}^{ia} = \psi^{ia} + \frac{1}{\sqrt{T}} \psi^{ia}_m\bar{q}_m + \frac{1}{2T} \psi^{ia}_{mn}\bar{q}_m\bar{q}_n + O_p(T^{-3/2})$$

Here subscripts j, k, ℓ of g and m, n of ψ imply differentiation with respect to the corresponding components of β. Using these expansions in (5) and collecting terms up to $O_p(T^{-1})$ we obtain:

$$(7) \quad W = (g_{ij}\psi^{ia}g_{ab})\bar{q}_j\bar{q}_b + \frac{1}{\sqrt{T}}(g_{ij}\psi^{ia}_m g_{ab} + g_{ijm}\psi^{ia}g_{ab})\bar{q}_j\bar{q}_b\bar{q}_m$$
$$+ \frac{1}{T}\left[\left(\frac{1}{2}\right)g_{ij}\psi^{ia}_{mn}g_{ab}+g_{ijm}\psi^{ia}_n g_{ab}+\left(\frac{1}{4}\right)g_{ijm}\psi^{ia}g_{abn}+\left(\frac{1}{3}\right)g_{ijmn}\psi^{ia}g_{ab}\right]\bar{q}_j\bar{q}_b\bar{q}_m\bar{q}_n$$
$$+ O_p(T^{-3/2})$$
$$= \bar{q}'G'(GG')^{-1}G\bar{q} + T^{-1/2}u(\bar{q}) + T^{-1}v(\bar{q}) + O_p(T^{-3/2}) \quad , \quad \text{say.}$$

We rewrite $u(\bar{q})$ and $v(\bar{q})$ as:

$$u(\bar{q}) = \mathrm{vec}(J)'(\bar{q} \otimes \bar{q} \otimes \bar{q})$$

$$v(\bar{q}) = \mathrm{tr}\{L(\bar{q}\,\bar{q}' \otimes \bar{q}\,\bar{q}')\}$$

where J is the $p^2 \times p$ matrix which stacks the $p \times p$ matrices

$$G'(GG')^{-1}_{(i)}G + G'_{(i)}(GG')^{-1}G \quad (i = 1, \ldots, p) \tag{8}$$

$\mathrm{vec}(J)$ stacks rows of J and L is the $p^2 \times p^2$ matrix whose $(i,j)^{th}$ $p \times p$ block is given by

$$\left(\frac{1}{2}\right)G'(GG')^{-1}_{(ij)}G + G'_{(i)}(GG')^{-1}_{(j)}G + \left(\frac{1}{4}\right)G'_{(i)}(GG')^{-1}G_{(j)} + \left(\frac{1}{3}\right)G'_{(ij)}(GG')^{-1}G \tag{9}$$

$$(i, j = 1, \ldots, p) .$$

In (8) and (9) we use the notation $A_{(i)} = \partial A/\partial\beta_i$, $A_{(ij)} = \partial^2 A/\partial\beta_i\partial\beta_j$ for any matrix $A = A(\beta)$.

The characteristic function of W may now be written:

$$\begin{aligned} cf_W(t) &= (2\pi)^{-p/2}\int_{R^p} e^{itW} e^{-(1/2)\bar{q}'\bar{q}} d\bar{q} \\ &= (2\pi)^{-p/2}\int_{R^p} \exp\left\{-(1/2)\bar{q}'[I - 2itG'(GG')^{-1}G]\bar{q}\right\} \\ &\quad \cdot\left[1 + \frac{it}{\sqrt{T}}\, u(\bar{q}) + \frac{it}{T}\, v(\bar{q}) - \frac{t^2}{2T}\, u^2(\bar{q})\right] d\bar{q} + o(T^{-1}) . \end{aligned} \tag{10}$$

To compute the integral in (10) we first transform $\bar{q} \to z = R^{-1}\bar{q}$ where we define R by writing

$$S = I - 2itP_{G'} = \bar{P}_{G'} + (1-2it)P_{G'}$$

and

$$R = \bar{P}_{G'} + (1-2it)^{-1/2} P_{G'} = S^{-1/2} .$$

Here we use the notation $P_A = A(A'A)^{-1}A'$ for any matrix A of full column rank.

Noting that the Jacobian of the transformation $\bar{q} \to z$ is $\det R = (1-2it)^{-r/2}$ we find that (10) reduces to:

$$(11)\quad cf_W(t) = (1-2it)^{-r/2}\left\{1+(it)T^{-1/2}E(u)+(it)T^{-1}E(v)-\left(\frac{t^2}{2}\right)T^{-1}E(u^2)\right\} + o(T^{-1})$$

where E is the integral expectation operator with respect to the density $(2\pi)^{-p/2}\exp\{-(1/2)z'z\}$ and

$$(12)\qquad u = u(Rz) \;, \quad v = v(Rz) \;.$$

Note that the leading term in (11) is just the characteristic function of the χ^2_r distribution, corresponding to the usual first order asymptotics of W. Higher order terms in (11) are computed simply by taking expectations of polynomials in independent standard normal variates. We observe that $E(u) = 0$ since u involves only a polynomial of odd degree in z. Moreover,

$$v = \mathrm{tr}\{(R \otimes R)L(R \otimes R)(zz' \otimes zz')\}$$

and

$$u^2 = \mathrm{vec}(J)'(R \otimes R \otimes R)(zz' \otimes zz' \otimes zz')(R \otimes R \otimes R)\mathrm{vec}(J) \;.$$

Expectations of v and u^2 are now obtained through the following Lemmas.

LEMMA 2.1. *If* $z = N(0, I_p)$ *then*

(13) $$E(zz' \otimes zz') = I + K_{pp} + (\mathrm{vec}\ I)(\mathrm{vec}\ I)' ;$$

(14) $$E(zz' \otimes zz' \otimes zz') = I \otimes I \otimes I$$
$$+ (1/2)\Sigma_{ij}[(I \otimes T_{ij} \otimes T_{ij}) + (T_{ij} \otimes I \otimes T_{ij}) + (T_{ij} \otimes T_{ij} \otimes I)]$$
$$+ \Sigma_{ijk}(T_{ij} \otimes T_{ik} \otimes T_{jk}) ;$$

Alternatively,

(15) $$E(zz' \otimes zz' \otimes zz') = I + K_{pp^2} + K_{p^2p}$$
$$+ (I + K_{pp^2} + K_{p^2p})(I \otimes (\mathrm{vec}\ I)(\mathrm{vec}\ I)')(I + K_{pp^2} + K_{p^2p})$$
$$+ I \otimes K_{pp} + K_{pp} \otimes I + K_{pp^2}(I \otimes K_{pp})K_{p^2p} ;$$

where

$$T_{ij} = E_{ij} + E_{ji} \quad , \quad E_{ij} = e_i e_j'$$

and e_i *is the* i^{th} *unit vector in* R^p . K_{mn} *denotes the commutation matrix of order* $mn \times mn$.

LEMMA 2.2. *If* $x = N_p(0,V)$ *then*

(16) $$E(xx' \otimes xx') = (I + K_{pp})(V \otimes V) + (\mathrm{vec}\ V)(\mathrm{vec}\ V)'$$

(17) $$E(xx' \otimes xx' \otimes xx') = (I + K_{pp^2} + K_{p^2p})(V \otimes V \otimes V)$$
$$+ (I + K_{pp^2} + K_{p^2p})(V \otimes (\mathrm{vec}\ V)(\mathrm{vec}\ V)')(I + K_{pp^2} + K_{p^2p})$$
$$+ V \otimes \{K_{pp}(V \otimes V)\} + \{K_{pp}(V \otimes V)\} \otimes V + K_{pp^2}[V \otimes \{K_{pp}(V \otimes V)\}]K_{p^2p} .$$

LEMMA 2.3.

$$(18) \qquad E(v) = a_0 + a_1(1-2it)^{-1} + a_2(1-2it)^{-2}$$

$$(19) \qquad E(u^2) = b_0 + b_1(1-2it)^{-1} + b_2(1-2it)^{-2} + b_3(1-2it)^{-3}$$

where

$$a_i = tr(A_i) \ , \quad (i = 0,\ 1,\ 2)$$

$$A_0 = L\{(I + K_{pp})(\overline{P}_{G'} \otimes \overline{P}_{G'}) + (\mathrm{vec}\ \overline{P}_{G'})(\mathrm{vec}\ \overline{P}_{G'})\}$$

$$A_1 = L[(I + K_{pp})\{(\overline{P}_{G'} \otimes P_{G'} + P_{G'} \otimes \overline{P}_{G'}) + (\mathrm{vec}\ \overline{P}_{G'})(\mathrm{vec}\ P_{G'})'$$
$$+ (\mathrm{vec}\ P_{G'})(\mathrm{vec}\ \overline{P}_{G'})'\}]$$

$$A_2 = L\{(I + K_{pp})(P_{G'} \otimes P_{G'}) + (\mathrm{vec}\ P_{G'})(\mathrm{vec}\ P_{G'})']$$

$$b_i = (\mathrm{vec}\ J)'B_i(\mathrm{vec}\ J) \quad (i = 0,\ 1,\ 2,\ 3)$$

$$B_0 = H(\overline{P}_{G'} \otimes \overline{P}_{G'} \otimes \overline{P}_{G'}) + H(\overline{P}_{G'} \otimes (\mathrm{vec}\ \overline{P}_{G'})(\mathrm{vec}\ \overline{P}_{G'}))H$$
$$+ \overline{P}_{G'} \otimes K_{pp}(\overline{P}_{G'} \otimes \overline{P}_{G'}) + K_{pp}(\overline{P}_{G'} \otimes \overline{P}_{G'}) \otimes \overline{P}_{G'}$$
$$+ K_{pp^2}(\overline{P}_{G'} \otimes K_{pp}(\overline{P}_{G'} \otimes \overline{P}_{G'}))K_{p^2p} = C_0(\overline{P}_{G'}) \ , \quad \textit{say;}$$

$$B_1 = H(P_{G'} \otimes \bar{P}_{G'} \otimes \bar{P}_G)H$$

$$+ H(P_{G'} \otimes (\mathrm{vec}\ \bar{P}_{G'})(\mathrm{vec}\ \bar{P}_{G'})' + \bar{P}_{G'} \otimes (\mathrm{vec}\ P_{G'})(\mathrm{vec}\ \bar{P}_{G'})'$$

$$+ \bar{P}_{G'} \otimes (\mathrm{vec}\ \bar{P}_{G'})(\mathrm{vec}\ P_{G'})')H$$

$$+ P_{G'} \otimes K_{pp}(\bar{P}_{G'} \otimes \bar{P}_{G'}) + \bar{P}_{G'} \otimes K_{pp}(P_{G'} \otimes \bar{P}_{G'}) + \bar{P}_{G'} \otimes K_{pp}(\bar{P}_{G'} \otimes P_{G'})$$

$$+ K_{pp}(P_{G'} \otimes \bar{P}_{G'}) \otimes \bar{P}_{G'} + K_{pp}(\bar{P}_{G'} \otimes P_{G'}) \otimes \bar{P}_{G'} + K_{pp}(\bar{P}_{G'} \otimes \bar{P}_{G'}) \otimes P_{G'}$$

$$+ K_{pp^2}((P_{G'} \otimes K_{pp}(\bar{P}_{G'} \otimes \bar{P}_{G'})) + (\bar{P}_{G'} \otimes K_{pp}(P_{G'} \otimes \bar{P}_{G'}))$$

$$+ (\bar{P}_{G'} \otimes K_{pp}(\bar{P}_{G'} \otimes P_{G'})))K_{p^2p} = C_1(\bar{P}_{G'}, P_{G'}) \ , \quad \textit{say;}$$

$$B_2 = C_1(P_{G'}, \bar{P}_{G'})$$

$$B_3 = C_0(P_{G'})$$

and where

$$H = I + K_{pp^2} + K_{p^2p} \ .$$

Using (18) and (19) in (11) we find the following expression for the characteristic function of W :

$$\text{(20)} \quad cf_W(t) = (1-2it)^{-r/2}\Big[1 + \frac{1}{T}\Big\{\frac{1}{2}b_0(it)^2 + \Big(a_0 - \frac{1}{4}b_1\Big)(it)$$

$$+ \Big(a_1 + \frac{1}{4}b_1 - \frac{1}{4}b_2\Big)it(1-2it)^{-1} + \Big(a_2 + \frac{1}{4}b_2 - \frac{1}{4}\ b_3\Big)it(1-2it)^{-2}$$

$$+ \frac{1}{4}b_3 it(1-2it)^{-3}\Big\}\Big] + o(T^{-1}) \ .$$

Upon inversion of (20) we obtain asymptotic expansions of the density and distribution function of W up to $O(T^{-1})$. The final result is given in the following theorem.

THEOREM 2.4. *The asymptotic expansion of the distribution function of* W *up to* $O(T^{-1})$ *as* $T \uparrow \infty$ *is given by:*

$$\text{(21)} \qquad \mathrm{cdf}(w) = F_r(w) - \frac{1}{T}c(w)f_r(w) + o(T^{-1})$$

$$= F_r(w - T^{-1}c(w)) + o(T^{-1})$$

where f_r *and* F_r *denote the density and distribution function, respectively, of a* χ_r^2 *variate and where*

$$\text{(22)} \qquad c(w) = \Sigma_{n=-1}^{3}\alpha_n w^n$$

with

$$\alpha_{-1} = -b_0(r-2)/4$$

$$\alpha_0 = (4a_0 + b_0 - b_1)/4$$

$$\alpha_1 = (4a_1 + b_1 - b_2)/4r$$

$$\alpha_2 = (4a_2 + b_2 - b_3)/4r(r+2)$$

$$\alpha_3 = b_3/4r(r+2)(r+4) \ .$$

Finally, we observe that if w_α is the critical value of the χ_r^2 distribution at the level α then the corresponding critical value of cfd(w) correct to $O(T^{-1})$ is given by the solution of

$$w_\alpha^* - T^{-1}c(w_\alpha^*) = w_\alpha \ .$$

This may be approximated by

$$w_\alpha^* = w_\alpha + T^{-1}c(w_\alpha)$$

which is correct to the same accuracy of $O(T^{-1})$.

3. SPECIALIZATIONS

We shall examine various examples and Monte Carlo results that have appeared in the recent literature in the light of the previous section.

(i) Gregory and Veall (1985)

These authors study by simulation methods the behavior of alternative Wald tests of the algebraically equivalent restrictions:

(I) $\beta_1 - 1/\beta_2 = 0$

(II) $\beta_1\beta_2 - 1 = 0$

where β_1 and β_2 are coefficients in a classical linear regression model. Their evidence suggests that the Wald test based on formulation (I) performs poorly in finite samples, even in samples as large as $T = 500$, when β_2 is in the vicinity of the origin (more specifically, $\beta_2 = 0.2, 0.1$ in their experiments).

Asymptotic expansions of the distributions of these alternative Wald tests may be deduced from Theorem 4.2 by applying the formulae of the previous section. In particular, the coefficients of the correction factor $c(w)$ in (22) are displayed in Table 1 for easy comparison.

TABLE 1

COEFFICIENTS OF THE CORRECTION TERM c(w)

	(I): $\beta_1 - \frac{1}{\beta_2} = 0$	(II): $\beta_1\beta_2 - 1 = 0$
α_{-1}	0	0
α_0	0	0
α_1	$-\frac{\beta_2^6}{(1 + \beta_2^4)^3}$	$-\frac{\beta_2^6}{(1 + \beta_2^4)^3}$
α_2	$-\frac{4 + 10\beta_2^4}{\beta_2^2(1 + \beta_2^4)^3}$	$-\frac{\beta_2^2 - 4\beta_2^6 + 10\beta_2^{10}}{(1 + \beta_2^4)^3}$
α_3	$\frac{1}{\beta_2^2(1 + \beta^4)^3}$	$\frac{\beta_2^6}{(1 + \beta_2^4)^3}$

The first three coefficients α_{-1} , α_0 and α_1 agree under the null hypothesis. Discrepancies to $O(T^{-1})$ occur in α_2 and α_3 . For small β_2 both α_2 and α_3 are of $O(\beta_2^{-2})$ under (I); but, under (II), $\alpha_2 = O(\beta_2^2)$ and $\alpha_3 = O(\beta_2^6)$ as $\beta_2 \to 0$. Thus, substantial deviations from the nominal asymptotic size are to be expected for the Wald test under (I), whereas rather good approximations from first order asymptotics are to be expected under (II), at least in this region of the parameter space. For example, when $\beta_2 = 0.1$ we find that:

$$\alpha_2 = -400 \ , \quad \alpha_3 = 100 \quad \text{under (I)}$$

while

$$\alpha_2 = -0.01 \ , \quad \alpha_3 = 10^{-6} \quad \text{under (II).}$$

Even for sample sizes as large as $T = 100$ it is clear from the polynomial correction factor

$$(1/T)c(w) = (1/T)(-10^{-6}w - 400w^2 + 100w^3)$$

that the gap between the finite sample and asymptotic distributions of W is substantial. This confirms the experimental results of Gregory and Veall. Theory also strongly supports their recommendation that (II), the multiplicative form of the Wald test, is to be preferred on the basis of the accuracy of nominal size if we have reason to believe that β_2 may be in the vicinity of the origin.

(ii) Lafontaine and White (1986)

These authors consider the algebraically equivalent restrictions:

(I) $\beta = 1$

$(\text{II})_k$ $\beta^k = 1 \ , \quad k \in Z$.

In this case the coefficients in our correction factor $c(w)$ for the distribution of the Wald test based on $(\text{II})_k$ are given by:

$$\alpha_{-1} = \alpha_0 = \alpha_1 = 0$$

$$\alpha_2 = -\frac{2}{3}(k-1)(k-2)$$

$$\alpha_3 = \frac{1}{4}(k-1)^2 \ .$$

Note that all coefficients vanish when $k = 1$, as is to be expected since the Wald statistic has an exact χ_1^2 distribution in this case. Moreover,

$\alpha_2 = O(|k|^2)$, $\alpha_3 = O(|k|^2)$ as $|k| \to \infty$. Thus, the correction on the first order asymptotics becomes more substantial as $|k|$ increases. In other words, as the nonlinearity of the formulation $(II)_k$ increases the adequacy of the asymptotic χ^2 approximation deteriorates and the nominal critical values for the test that are delivered from asymptotic theory become less reliable. For large $|k|$ the deviations from the asymptotic theory may be expected to be large. This confirms the finding of Lafontaine and White.

(iii) Breusch and Schmidt (1985)

These authors give the example of the equivalent null hypotheses:

(I) $\beta = 0$

(II) $h(\beta) = 0$

where h satisfies $h(0) = 0$ and $h'(\beta) > 0$ for all β . To test these equivalent hypotheses the following two Wald statistics are considered

$$W_1 = T\bar{X}^2 \ , \quad W_2 = Th^2(\bar{X})/(h'(\bar{X}))^2$$

where $\bar{X}$ is the sample mean of a random sample of iid $N(\beta,1)$ variates. Taking h to be continuously differentiable to the third order we deduce from Theorem 2.4 after some elementary manipulations that the distribution of W_2 has the following Edgeworth expansion to $O(T^{-1})$:

$$\text{(23)} \qquad \text{cdf}(w) = F_1(w) - \frac{1}{T}c(w)f_1(w) + o(T^{-1})$$

where

$$(24) \qquad c(w) = -\frac{2}{3}\frac{h^{(3)}(0)}{h'(0)} w^2 + \frac{1}{4}\left[\frac{h^{(2)}(0)}{h'(0)}\right]^2 w^3 .$$

For certain functions h in (II) the correction term (24) will lead to substantial deviations from the first order asymptotics. Breusch and Schmidt suggest the following function to illustrate possibilities:

$$\begin{aligned} h(\beta) &= \left\{1 + e^{-b(\beta-c)}\right\}^{-1} - \left\{1 + e^{bc}\right\}^{-1} , && \beta \geq 0 \\ &= -h(-\beta) , && \beta < 0 \end{aligned}$$

where $b > 0$ and c is unrestricted. Simple calculations verify that

$$\frac{h^{(2)}(0)}{h'(0)} = -\frac{b(1 - e^{bc})}{1 + e^{bc}}$$

and

$$\frac{h^{(3)}(0)}{h'(0)} = \frac{b^2(1 - 4e^{bc} + e^{2bc})}{(1 + e^{bc})^2} .$$

These expressions can be large for large b . Thus, we find

$$\frac{h^{(2)}(0)}{h'(0)} = O(b) , \quad \frac{h^{(3)}(0)}{h'(0)} = O(b^2)$$

as $b \rightarrow \infty$. In such cases the distribution of W_2 can be expected to be poorly approximated by the asymptotic χ_1^2 .

By contrast, W_1 is quadratic in $\bar{X}^2$ and $W_1 \equiv \chi_1^2$ for all T . Thus, when h is linear the asymptotic distribution is exact.

The form of the Edgeworth expansion (23) and the correction factor (24) suggest that it may be possible to select formulations (II) which accelerate convergence to the asymptotic χ_1^2 distribution by eliminating (or, at least reducing the magnitude of) the correction. Thus, any function h for which

$$(25) \qquad h^{(2)}(0) = h^{(3)}(0) = 0$$

will be sufficient to eliminate the $O(T^{-1})$ correction term on the asymptotic and hence accelerate the convergence to an error of $O(T^{-2})$. Clearly a function h satisfying (25) is approximately linear in the vicinity of the origin, so that this approach goes some way towards suggesting the formulation (I). In fact, if h were assumed to be analytic in fixed neighborhood of the origin, we could develop a complete asymptotic series for the distribution of W_2 . This would involve an expansion of the form (23) taken to an arbitrary number of terms. Although we shall not report the details here, application of the argument given above on accelerated convergence in this case would lead to the choice of a function h for which

$$(26) \qquad h^{(j)}(0) = 0 \quad \text{all} \quad j = 2, 3, \ldots .$$

Since h is analytic and has a convergent power series representation, (26) would then imply a preferred choice of a linear function. In this case, therefore, the argument leads directly to the formulation (I) of the Wald test.

4. CONCLUDING REMARKS

Since the finite sample distribution of the Wald statistic for testing a nonlinear restriction can depend substantially on the algebraic representation of the restriction, the study of functional forms is of great importance in this context. The present paper provides a theoretical study of the problem using asymptotic expansions. The correction terms that are delivered by these expansions are shown to be extremely useful in evaluating algebraically equivalent, competing formulations of the Wald test. We have derived the explicit form of the asymptotic expansion of the distribution of the Wald statistic to $O(T^{-1})$ for a general class of parametric restrictions. The formulae obtained should be of independent interest. Moreover, upon specialization to the testing problems considered recently by other authors, these formulae explain well the discrepancies that have been observed in the finite sample behavior of alternative Wald tests. Finally, as we have seen in Section 3, Edgeworth expansions are not only useful in explaining phenomena such as the Wald test discrepancies. They may also be used to locate transformations which attenuate or even eliminate the effect of higher order terms, thereby accelerating convergence to the asymptotic distribution and helping to make it more reliable in finite samples.

APPENDIX A

Proof of Lemma 2.1. (13) and (14) are given by Magnus and Neudecker (1979, theorem 4.1). To prove (15) we note that

$$\Sigma_{ij}(E_{ij} \otimes T_{ij}) = \Sigma_{ij}(E_{ij} \otimes E_{ij}) + \Sigma_{ij}(E_{ij} \otimes E_{ij})$$

$$= K_{pp} + (\mathrm{vec}\ I)(\mathrm{vec}\ I)'$$

so that

$$\Sigma_{ij}(T_{ij} \otimes T_{ij}) = 2\{K_{pp} + (\mathrm{vec}\ I)(\mathrm{vec}\ I)'\} .$$

Also

$$\Sigma_{ij}(T_{ij} \otimes I \otimes T_{ij}) = \Sigma_{ij}K_{pp^2}(I \otimes T_{ij} \otimes T_{ij})K_{p^2p}$$

$$= 2K_{pp^2}\{I \otimes [K_{pp} + (\mathrm{vec}\ I)(\mathrm{vec}\ I)']\}K_{p^2p} .$$

Hence

$$\Sigma_{ij}\{(I \otimes T_{ij} \otimes T_{ij}) + (T_{ij} \otimes I \otimes T_{ij}) + (T_{ij} \otimes T_{ij} \otimes I)\}$$

$$\text{(A1)} \qquad = 2\{I \otimes \{K_{pp} + (\mathrm{vec}\ I)(\mathrm{vec}\ I)'\}$$

$$+ K_{pp^2}\{I \otimes [K_{pp} + (\mathrm{vec}\ I)(\mathrm{vec}\ I)']\}K_{p^2p}$$

$$+ \{K_{pp} + (\mathrm{vec}\ I)(\mathrm{vec}\ I)'\} \otimes I\} .$$

Now

$$T_{ij} \otimes T_{ik} \otimes T_{jk} = (E_{ij} + E_{ji}) \otimes (E_{ik} + E_{ki}) \otimes (E_{jk} + E_{kj})$$

and simple manipulations along the following lines yield:

$$\Sigma_{ijk}(E_{ij} \otimes E_{ik} \otimes E_{jk}) = \Sigma_{ijk}(e_i \otimes e_i)(e_j' \otimes e_k') \otimes E_{jk}$$

$$= \Sigma_{ijk}\{\text{vec}(E_{ii})(e_j' \otimes e_k')\} \otimes E_{jk}$$

$$= \{\Sigma_i \text{ vec}(E_{ii}) \otimes I_p\}\Sigma_{jk}\{(e_j' \otimes e_k') \otimes e_j e_k'\}$$

$$= \{(\text{vec } I) \otimes I_p\{\Sigma_{jk}(e_j' \otimes e_k' \otimes e_j \otimes e_k')\}$$

$$= \{(\text{vec } I) \otimes I_p\}\Sigma_{jk}\{(e_j' \otimes 1 \otimes e_j \otimes 1)(I_p \otimes e_k' \otimes 1 \otimes e_k')\}$$

$$= \{(\text{vec } I) \otimes I_p\}\{\Sigma_{jk}(e_j' \otimes e_j)(I \otimes e_k' \otimes e_k')\}$$

$$= \{(\text{vec } I) \otimes I_p\}\{(\Sigma_j E_{jj})(I \otimes \Sigma_k(e_k' \otimes e_k'))\}$$

$$= \{(\text{vec } I_p) \otimes I_p\}\{I_p \otimes (\text{vec } I_p)'\} \ .$$

Note that

$$K_{pp^2}((\text{vec } I) \otimes I)K_{1p} = K_{pp^2}((\text{vec } I) \otimes I) = I \otimes (\text{vec } I)$$

so that, using the fact that $K'_{pp^2} = K_{p^2p}$ we deduce that

$$\Sigma_{ijk}(E_{ij} \otimes E_{ik} \otimes E_{jk}) = K_{p^2p}(I \otimes (\text{vec } I)(\text{vec } I)') \ .$$

In a similar way we find that:

$$\Sigma_{ijk}(E_{ji} \otimes E_{ik} \otimes E_{jk}) = K_{pp^2}(I \otimes (\text{vec } I)(\text{vec } I)')$$

$$\Sigma_{ijk}(E_{ij} \otimes E_{ki} \otimes E_{jk}) = (I \otimes I \otimes I)K_{p^2p}$$

$$\Sigma_{ijk}(E_{ji} \otimes E_{ki} \otimes E_{jk}) = K_{pp^2}\{I \otimes (\text{vec } I)(\text{vec } I)'\}K_{pp^2}$$

$$\Sigma_{ijk}(E_{ij} \otimes E_{ik} \otimes E_{kj}) = K_{p^2p}\{I \otimes (\text{vec } I)(\text{vec } I)'\}K_{p^2p}$$

$$\Sigma_{ijk}(E_{ji} \otimes E_{ik} \otimes E_{kj}) = (I \otimes I \otimes I)K_{pp^2}$$

$$\Sigma_{ijk}(E_{ij} \otimes E_{ki} \otimes E_{kj}) = \{I \otimes (\text{vec } I)(\text{vec } I)'\}K_{p^2p}$$

$$\Sigma_{ijk}(E_{ji} \otimes E_{ki} \otimes E_{kj}) = \{I \otimes (\text{vec } I)(\text{vec } I)'\}K_{pp^2}$$

It now follows from (14), (A1) and the above that:

(A2) $$E(zz'\otimes zz'\otimes zz') = I + K_{pp^2} + K_{p^2p} + (I\otimes K_{pp}) + (K_{pp}\otimes I) + K_{pp^2}(I\otimes K_{pp})K_{p^2p}$$
$$+ \{I\otimes(\text{vec } I)(\text{vec } I)'\}(I + K_{p^2p} + K_{pp^2}) + (\text{vec } I)(\text{vec } I)' \otimes I$$
$$+ K_{pp^2}\{I\otimes(\text{vec } I)(\text{vec } I)'\}K_{p^2p} + K_{pp^2}\{I\otimes(\text{vec } I)(\text{vec } I)'\}K_{pp^2}$$
$$+ K_{p^2p}\{I\otimes(\text{vec } I)(\text{vec } I)'\}K_{p^2p} + (K_{p^2p} + K_{pp^2})\{I\otimes(\text{vec } I)(\text{vec } I)'\} .$$

Noting that

$$(\text{vec } I)(\text{vec } I)' \otimes I = K_{p^2p}(I \otimes (\text{vec } I)(\text{vec } I)')K_{pp^2}$$

we deduce that the final six terms of (A2) sum to

$$(I + K_{p^2p} + K_{pp^2})\{I \otimes (\text{vec } I)(\text{vec } I)'\}(I + K_{p^2p} + K_{pp^2})$$

leading to result (15) as stated in the lemma.

Proof of Lemma 2.2. Let $z = V^{-1/2}x \sim N_p(0,I)$. Then

$$E(xx' \otimes xx') = (V^{1/2} \otimes V^{1/2})E(zz' \otimes zz')(V^{1/2} \otimes V^{1/2})$$
$$= (I + K_{pp})(V \otimes V) + (\text{vec } V)(\text{vec } V)'$$

as required for (16). Similarly

$$E(xx'\otimes xx'\otimes xx') = (V^{1/2} \otimes V^{1/2} \otimes V^{1/2})E(zz'\otimes zz'\otimes zz')(V^{1/2} \otimes V^{1/2} \otimes V^{1/2})$$

and result (17) follows directly in view of the properties of the commutation matrices K_{pp^2} , K_{p^2p} and K_{pp} .

Proof of Lemma 2.3. We note that $\bar{q} = Rz \sim N(0, R^2) = N(0, S^{-1})$ and

$$E(v) = \text{tr}\{LE(\bar{q}\,\bar{q} \otimes \bar{q}\,\bar{q}')\}$$

(A3) $$= \text{tr}[L\{(I + K_{pp})(S^{-1} \otimes S^{-1}) + (\text{vec } S^{-1})(\text{vec } S^{-1})'\}] .$$

Now

(A4) $$S^{-1} = \bar{P}_{G'} + (1-2it)^{-1}P_{G'}$$

so that upon expansion of (A3) we obtain

$$E(v) = \text{tr}(A_0) + \text{tr}(A_1)(1-2it)^{-1} + \text{tr}(A_2)(1-2it)^{-2}$$

as required for (18). To prove (19) we write

$$E(u^2) = (\text{vec } J)'E(\bar{q}\,\bar{q}' \otimes \bar{q}\,\bar{q}' \otimes \bar{q}\,\bar{q}')(\text{vec } J)$$

(A5) $$= (\text{vec } J)'(B_0 + B_1(1-2it)^{-1} + B_2(1-2it)^{-2} + B_3(1-2it)^{-3})(\text{vec } J) .$$

Upon evaluation using (17) with $V = S^{-1}$, (A5) yields the stated formula (19).

Proof of Theorem 2.4. Term by term Fourier inversion of (20) yields

$$\text{(A6)} \quad pdf(w) = f_r(w) + \frac{1}{T}\Big\{\emptyset_0(-1)^2 f_r^{(2)}(w) + \emptyset_1(-1)f_r'(w) + \emptyset_2(-1)f_{r+2}'(w) + \emptyset_3(-1)f_{r+4}'(w) + \emptyset_4(-1)f_{r+6}'(w)\Big\} + o(T^{-1})$$

where

$$\emptyset_0 = b_0/2 \ , \quad \emptyset_1 = a_0 - b_1/4$$

$$\emptyset_2 = a_1 + b_1/4 - b_2/4 \ , \quad \emptyset_3 = a_2 + b_2/4 - b_3/4$$

$$\emptyset_4 = b_3/4 \ .$$

Upon integration of (A6) we obtain

$$\text{(A7)} \quad cdf(w) = F_r(w) + \frac{1}{T}\Big\{\emptyset_0 f_r'(w) - \emptyset_1 f_r(w) - \emptyset_2 f_{r+2}(w) - \emptyset_3 f_{r+4}(w) - \emptyset_4 f_{r+6}(w)\Big\} + o(T^{-1})$$

and in view of the relationships

$$f_r'(w) = -(1/2)f_r(w) + ((r-2)/2w)f_r(w)$$

$$f_{r+2n}(w) = 2^{-n}w^n f_r(w)/(r/2)_n$$

(A7) can be rewritten in the form

$$\mathrm{cdf}(w) = F_r(w) + \frac{1}{T}\left[\left\{\varnothing_0\left(\frac{r-2}{2w} - \frac{1}{2}\right) - \varnothing_1\right\}f_r(w)\right.$$

$$\left. - \varnothing_2 f_{r+2}(w) - \varnothing_3 f_{r+4}(w) - \varnothing_4 f_{f+6}(w)\right] + o(T^{-1})$$

$$= F_r(w) + \frac{1}{T}f_r(w)\left[\frac{\varnothing_0(r-2)}{2w} - \left(\varnothing_1 + \frac{1}{2}\varnothing_0\right) - \frac{\varnothing_2 w}{r} - \frac{\varnothing_3 w^2}{r(r+2)} - \frac{\varnothing_4 w^3}{r(r+2)(r+4)}\right] + o(T^{-1})$$

which reduces to (21) as stated upon translation of notation. Validity of the expansion as a proper asymptotic series follows from the theorem in Sargan (1980).

APPENDIX B

This appendix derives formulae for the Edgeworth expansion to $O(T^{-1})$ of the distribution of the Wald statistic (5) in full generality. Thus, in place of assumption (4) we require only that the distribution of the component variates $\bar{q}$ have a valid Edgeworth expansion of the form (3). Let $\tilde{q} = \Omega^{-1/2}\bar{q}$ and carrying terms to $O(T^{-1})$ we write the Edgeworth expansion as:

$$\mathrm{pdf}(\tilde{q}) = (2\pi)^{-p/2}\, e^{-\tilde{q}'\tilde{q}/2}\Big\{1 + T^{-1/2}[f_1'\tilde{q} + f_3'(\tilde{q}\otimes\tilde{q}\otimes\tilde{q})]$$
$$+ T^{-1}[f_0 + \mathrm{tr}(F_2\tilde{q}\tilde{q}') + \mathrm{tr}\{F_4(\tilde{q}\tilde{q}'\otimes\tilde{q}\tilde{q}')\} + \mathrm{tr}\{F_6(\tilde{q}\tilde{q}'\otimes\tilde{q}\tilde{q}'\otimes\tilde{q}\tilde{q}')\}]\Big\} + o(T^{-1}) \ .$$

The Wald statistic for testing H_0 as given by (1) is now:

$$\text{(B1)} \qquad W = T\hat{g}'(\hat{G}\hat{\Omega}\hat{G}')^{-1}\hat{g}$$

where $\hat{\Omega} = \Omega(\hat{\beta})$ is a consistent estimator of Ω and where it is assumed that $G\Omega G'$ (respectively $\hat{G}\hat{\Omega}\hat{G}'$) is nonsingular (with probability one). The function $\Omega(\ )$ representing the estimator of Ω is assumed to be continuously differentiable to the third order. We write $\hat{\Psi}^{-1} = (\hat{G}\hat{\Omega}\hat{G}')^{-1} = (\hat{\psi}^{ia})$ and employ the Taylor representations

$$\sqrt{T}\,\hat{g}_i = \tilde{g}_{ij}\tilde{q}_j + \frac{1}{2\sqrt{T}}\,\tilde{g}_{ijk}\tilde{q}_j\tilde{q}_k + \frac{1}{6T}\,\tilde{g}_{ijk\ell}\tilde{q}_j\tilde{q}_k\tilde{q}_\ell + O_p(T^{-3/2})$$

$$\hat{\psi}^{ia} = \psi^{ia} + \frac{1}{\sqrt{T}}\,\tilde{\psi}^{ia}_m\tilde{q}_m + \frac{1}{2T}\,\tilde{\psi}^{ia}_{mn}\tilde{q}_m\tilde{q}_n + O_p(T^{-3/2})$$

where

$$\tilde{g}_{ij} = g_{is}\omega_{sj} \ , \quad \tilde{g}_{ijk} = g_{ist}\omega_{sj}\omega_{tk}$$

$$\tilde{g}_{ijk\ell} = g_{istu}\omega_{sj}\omega_{tk}\omega_{u\ell} \ , \quad \tilde{\psi}^{ia}_{m} = \psi^{ia}_{s}\omega_{sm}$$

$$\tilde{\psi}^{ia}_{mn} = \psi^{ia}_{st}\omega_{sm}\omega_{tn}$$

and where $\Omega^{1/2} = (\omega_{st})$ is the symmetric positive definite square root of Ω .

In place of (7) we now obtain the expansion

$$W = \tilde{q}'\Omega^{1/2}G'(G\Omega G')^{-1}G\Omega^{1/2}\tilde{q} + T^{-1/2}u(\tilde{q}) + T^{-1}v(\tilde{q}) + O_p(T^{-3/2})$$

where

(B2) $\quad u(\tilde{q}) = \text{vec}(\tilde{J})'(\tilde{q} \otimes \tilde{q} \otimes \tilde{q})$

(B3) $\quad v(\tilde{q}) = \text{tr}\{\tilde{L}(\tilde{q}\tilde{q}' \otimes \tilde{q}\tilde{q}')\}$

Here $\text{vec}(\tilde{J})$ has $\{p^2(j-1) + p(b-1) + m\}$'th element:

$$\tilde{g}_{ij}\tilde{\psi}^{ia}_{m}\tilde{g}_{ab} + \tilde{g}_{ijm}\psi^{ia}\tilde{g}_{ab} \ ;$$

and $\tilde{L}$ is $p^2 \times p^2$ with $(j(p-1) + b, \ m(p-1) + n)$'th element:

$$(1/2)\tilde{g}_{ij}\tilde{\psi}^{ia}_{mn}\tilde{g}_{ab} - \tilde{g}_{ijm}\tilde{\psi}^{ia}_{n}\tilde{g}_{ab} + (1/4)\tilde{g}_{ijm}\psi^{ia}\tilde{g}_{abn} + (1/3)\tilde{g}_{ijmn}\psi^{ia}\tilde{g}_{ab} \ .$$

In place of (10) the characteristic function of W therefore becomes:

$$cf_W(t) = (2\pi)^{-p/2}\int_{R^p} \exp\{-(1/2)\tilde{q}'\tilde{S}\tilde{q}\}\left\{1 + \frac{it}{\sqrt{T}}u(\tilde{q}) + \frac{it}{T}v(\tilde{q}) - \frac{t^2}{2T}u^2(\tilde{q})\right\}$$

$$\cdot\Big\{1 + T^{-1/2}[f_1'\tilde{q} + f_3'(\tilde{q}\otimes\tilde{q}\otimes\tilde{q})] + T^{-1}[f_0 + tr(F_2\tilde{q}\tilde{q}')$$

$$+ tr\{F_4(\tilde{q}\tilde{q}'\otimes\tilde{q}\tilde{q}')\} + tr\{F_6(\tilde{q}\tilde{q}'\otimes\tilde{q}\tilde{q}'\otimes\tilde{q}\tilde{q}')\}]\Big\}d\tilde{q} + o(T^{-1})$$

where

$$\tilde{S} = I - 2it\tilde{G}'(\tilde{G}\tilde{G}')^{-1}\tilde{G} \ , \quad \tilde{G} = G\Omega^{1/2} \ .$$

Upon integration and simplification this reduces to:

$$cf_W(t) = (1-2it)^{-r/2}\Big\{1 + T^{-1}[f_0 + tr(F_2\tilde{S}^{-1}) + tr\{F_4E(\tilde{q}\tilde{q}'\otimes\tilde{q}\tilde{q}')\}$$

$$+ tr\{F_6E(\tilde{q}\tilde{q}'\otimes\tilde{q}\tilde{q}'\otimes\tilde{q}\tilde{q}')\}] + (it/T)\{E(v) + tr\{F_JE(\tilde{q}\tilde{q}'\otimes\tilde{q}\tilde{q}')\}$$

$$+ vec(\tilde{J})'E(\tilde{q}\tilde{q}' \otimes \tilde{q}\tilde{q}' \otimes \tilde{q}\tilde{q}')f_3\} - (t^2/2T)E(u^2)\Big\} + o(T^{-1})$$

where E denotes the expectation operator with respect to $pdf(\tilde{q}) = N(0, \tilde{S}^{-1})$ and where $vec(F_J) = f_1 \otimes vec(\tilde{J})$.

Using Lemmas 2.2 and (2.3) we find that:

$$E(\tilde{q}\tilde{q}' \otimes \tilde{q}\tilde{q}') = D_0 + D_1(1-2it)^{-1} + D_2(1-2it)^{-2}$$

$$E(\tilde{q}\tilde{q}' \otimes \tilde{q}\tilde{q}' \otimes \tilde{q}\tilde{q}') = \tilde{B}_0 + \tilde{B}_1(1-2it)^{-1} + \tilde{B}_2(1-2it)^{-2} + \tilde{B}_3(1-2it)^{-3}$$

$$E(u^2) = vec(\tilde{J})'\{\tilde{B}_0 + \tilde{B}_1(1-2it)^{-1} + \tilde{B}_2(1-2it)^{-2} + \tilde{B}_3(1-2it)^{-3}\}vec(\tilde{J})$$

$$E(v) = tr(\tilde{A}_0) + tr(\tilde{A}_1)(1-2it)^{-1} + tr(\tilde{A}_2)(1-2it)^{-2}$$

where $\{\tilde{B}_j, \tilde{A}_k;\ j = 0, 1, 2, 3;\ k = 0, 1, 2\}$ are matrices identical in form to those defined in Lemma 2.3 but with $P_{\tilde{G}'} = \tilde{G}'(\tilde{G}\tilde{G}')^{-1}\tilde{G}$ $= \Omega^{1/2}G(G\Omega G')^{-1}G\Omega^{1/2}$ and $\tilde{L}$ in place of $P_{G'}$ and L ; and where

$$D_0 = (I + K_{pp})(\bar{P}_{\tilde{G}'} \otimes \bar{P}_{\tilde{G}'}) + (\mathrm{vec}\ \bar{P}_{\tilde{G}'})(\mathrm{vec}\ \bar{P}_{\tilde{G}'})'$$

$$D_1 = (I + K_{pp})(\bar{P}_{\tilde{G}'} \otimes P_{\tilde{G}'} + P_{\tilde{G}'} \otimes \bar{P}_{\tilde{G}'}) + (\mathrm{vec}\ \bar{P}_{\tilde{G}'})(\mathrm{vec}\ P_{\tilde{G}'}) + (\mathrm{vec}\ P_{\tilde{G}'})(\mathrm{vec}\ \bar{P}_{\tilde{G}'})$$

$$D_2 = (I + K_{pp})(P_{\tilde{G}'} \otimes P_{\tilde{G}'}) + (\mathrm{vec}\ P_{\tilde{G}'})(\mathrm{vec}\ P_{\tilde{G}'}) .$$

We now obtain

$$cf_W(t) = (1-2it)^{-r/2}\Big[1 + \frac{1}{T}\,\Sigma_{j=0}^{3} c_j (1-2it)^{-j} + \frac{it}{T}\,\Sigma_{j=0}^{3} e_j (1-2it)^{-j} - \frac{t^2}{2T}\Sigma_{j=0}^{3} \tilde{b}_j (1-2it)^{-j}\Big] + o(T^{-1})$$

where

$$c_0 = f_0 + tr(F_2 \bar{P}_{\tilde{G}'}) + tr(F_4 D_0) + tr(F_6 \tilde{B}_0)$$

$$c_1 = tr(F_2 P_{\tilde{G}'}) + tr(F_4 D_1) + tr(F_6 \tilde{B}_1)$$

$$c_2 = tr(F_4 D_2) + tr(F_6 \tilde{B}_2)$$

$$c_3 = tr(F_6 \tilde{B}_3)$$

$$e_j = \tilde{a}_j + d_j\ , \quad (j = 0,\ 1,\ 2,\ 3)$$

$$d_j = (\text{vec}\ \tilde{J})'\tilde{B}_j f\ , \quad (j = 0,\ 1,\ 2,\ 3)$$

$$\tilde{a}_j = \begin{cases} tr(\tilde{A}_j) + tr(F_J D_j) & (j = 0,\ 1,\ 2) \\ 0 & (j = 3) \end{cases}$$

$$\tilde{b}_j = (\text{vec}\ \tilde{J})'\tilde{B}_j(\text{vec}\ \tilde{J})\ , \quad (j = 0,\ 1,\ 2,\ 3)\ .$$

Inversion of $cf_W(t)$ now yields:

$$pdf(w) = f_r(w) + \frac{1}{T}\Sigma_{j=0}^{3} c_j f_{r+2j}(w) + \frac{1}{T}\{\tilde{\phi}_0 f_r^{(2)}(w) - \tilde{\phi}_1 f_r'(w) - \tilde{\phi}_2 f_{r+2}'(w) - \phi_3 f_{r+4}'(w) - \phi_4 f_{r+6}'(w)\} + o(T^{-1})$$

where

$$\tilde{\phi}_0 = \tilde{b}_0/2 \ , \quad \tilde{\phi}_1 = e_0 - \tilde{b}_1/4$$

$$\tilde{\phi}_2 = e_1 + \tilde{b}_1/4 - \tilde{b}_2/4 \ , \quad \tilde{\phi}_3 = e_2 + \tilde{b}_2/4 - \tilde{b}_3/4$$

$$\tilde{\phi}_4 = e_3 + \tilde{b}_3/4 \ .$$

Upon integration we obtain:

(B4) $$\mathrm{cdf}(w) = F_r(w) + \frac{1}{T}\Sigma_{j=0}^{3} c_j F_{r+2j}(w) + \frac{1}{T}\Sigma_{j=-1}^{3}\psi_{jr} w^j f_r(w) + o(T^{-1})$$

where

$$\psi_{-1r} = \tilde{\phi}_0(r-2)/2$$

$$\psi_{0r} = -(\tilde{\phi}_1 + \tilde{\phi}_0/2)$$

$$\psi_{1r} = -\tilde{\phi}_2/r$$

$$\psi_{2r} = -\tilde{\phi}_3/r(r+2)$$

$$\psi_{3r} = -\tilde{\phi}_4/r(r+2)(r+4)$$

Formula (B4) is the Edgeworth expansion to $O(T^{-1})$ of the distribution of the Wald statistic (B1) in the general case. It should be useful in many situations besides those considered in the present paper.

REFERENCES

Breusch, T. S. and P. Schmidt (1985). "Alternative forms of the Wald test: How long is a piece of string," mimeographed, Southampton University.

Gregory, A. and M. Veall (1985). "On formulating Wald tests of nonlinear restrictions," *Econometrica*, 53, 1465-1468.

Konishi, S. (1981). "Normalizing transformations of some statistics in multivariate analysis," *Biometrika*, 68, 647-651.

Lafontaine, F. and K. J. White (1986). "Obtaining any Wald statistic you want," *Economics Letters*.

Magnus, J. R. and H. Neudecker (1979). "The commutation matrix: Some properties and applications," *Annals of Statistics*, 7, 381-394.

Phillips, P. C. B. (1977). "A general theorem in the theory of asymptotic expansions as approximations to the finite sample distributions of econometric estimators," *Econometrica*, 45, 1517-1534.

Phillips, P. C. B. (1979). "Normalizing transformations and expansions for functions of statistics," mimeographed, Birmingham University.

Phillips, P. C. B. (1982). "Small sample distribution theory in econometric models of simultaneous models of simultaneous equations," Cowles Foundation Discussion Paper No. 617, Yale University.

Sargan, J. D. (1980). "Some approximations to the distribution of econometric criteria which are asymptotically distributed as chi-squared," *Econometrica*.

Sargan, J. D. and S. E. Satchell (1986). "A theory of validity for Edgeworth expansions," *Econometrica*, 54, 189-213.