

Wald tests when restrictions are locally singular*

Jean-Marie Dufour[†] Eric Renault[‡]

Victoria Zinde-Walsh[§]

December 3, 2013

Key words: nonlinear restriction; deficient rank; singular covariance matrix; Wald test; nonstandard asymptotic theory; bound. **Journal of Economic Literature classification:** C3.

*This work was supported by the Willam Dow Chair in Political Economy (McGill University), the Bank of Canada Research Fellowship, The Toulouse School of Economics Pierre-de-Fermat Chair of Excellence, A Guggenheim Fellowship, Conrad-Adenauer Fellowship from Alexander-von-Humboldt Foundation, the Canadian Network of Centres of Excellence program on Mathematics of Information Technology and Complex Systems, the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada and the Fonds de recherche sur la société et la culture (Québec). The authors also thank the research centres CIREQ and CIRANO for providing support and meeting space for the joint work. We thank Purevdorj Tuvaandorj for very useful comments.

[†]William Dow Professor of Economics, McGill University, Centre interuniversitaire de recherche en analyse des organisations (CIRANO) and Centre interuniversitaire de recherche en économie quative (CIREQ).

[‡]Brown University

[§]McGill University and CIREQ

1 Introduction

Tests based on asymptotic distributions typically require regularity assumptions in order to be able to obtain critical values. This is the case, in particular, for Wald-type statistics based on asymptotically normal estimators. Wald-type tests are especially convenient because they allow one to test a wide array of linear and nonlinear restrictions from a single unrestricted estimator. We focus here on the problem of implementing Wald-type tests for nonlinear restrictions.

The use of the Wald statistic has been criticized because of finite sample non-invariance (Gregory and Veall (1985), Breusch and Schmidt (1988), Phillips-Park(1988), Dagenais-Dufour(1991)) and lack of robustness to identification failure (Dufour(1997,2003)). We focus here on situations where the parameter tested is typically identified under the null hypothesis, but usual rank conditions on the Jacobian matrix may fail asymptotically.

Under regularity conditions, the standard asymptotic distribution of the test statistic is chi-square with degrees-of-freedom equal to the number of restrictions. The regularity conditions involve the assumption that the restrictions are differentiable with respect to the parameters considered, with a derivative matrix which has full column rank in an open neighborhood of the true value of the parameter vector. There are many problems, however, for which this regularity condition is violated. These include, among others:

1. hypothesis tests on bilinear and multilinear forms of model coefficients

in Gouriéroux-Monfort-Renault (1988);

2. testing whether the matrix of polynomials or multilinear forms in model coefficients has than full rank or, equivalently, whether the determinant of the matrix is zero in Gouriéroux-Monfort-Renault (1993);
3. tests of Granger noncausality in VARMA models in Boudjellaba-Dufour-Roy (1992,1994);
4. tests of noncausality at various horizons in Dufour-Renault (1998), Dufour-Pelletier-Renault (2005);
5. tests for common factors in ARMA models in Gouriéroux-Monfort-Renault (1989), Galbraith-ZindeWalsh (1997);
6. test of volatility and covolatility in Gouriéroux and Jasiak (2013).

A common feature of the above problems is the fact that the estimated asymptotic covariance matrix of the relevant nonlinear functions of coefficient estimates converges to a singular matrix on a subset of the null hypothesis – so that the usual regularity condition fails – but is non-singular (with probability one) in finite samples. The estimated covariance matrix used by the Wald-type statistic is a consistent estimator of the asymptotic covariance matrix of the corresponding nonlinear form in parameter estimates, but the rank of the estimated covariance matrix does not consistently estimate the rank of the asymptotic covariance matrix (because the rank is not a continuous function). It is important to note here that this is *not an identification*

problem, so that standard criticisms of Wald-type methods in the presence of identification problems (see Dufour (1997,2003)) do not apply in this case.

If the covariance matrix estimator can be modified so that it remains consistent and its rank converges to the appropriate asymptotic rank, then the asymptotic distribution of the modified Wald-type statistic (based on a generalized inverse of the covariance matrix) remains chi-square although with a reduced degrees-of-freedom number; see Andrews (1987). For example, Lutkepohl-Burda (1997) proposed such methods based on reducing the rank of the estimated covariance matrix by either using a form of randomization or setting “small eigenvalues” to zero. Such methods, however, effectively modify the test statistic and involve arbitrary truncation parameter for which no practical guidelines are available: in finite samples, the test statistic can become as small as one wishes leading to largely arbitrary results and unlimited power reductions.

Interestingly, except for a bound given by Sargan(1980) in a special case, the asymptotic distribution of Wald-type statistics in non-regular cases has not been studied. In this paper, we undertake this task and propose solutions to the problem that do not require modifying the test statistic. More specifically, the contributions of the paper can be summarized as follows.

First, we provide examples showing that Wald statistics in such non-regular cases can have several asymptotic distributions. We also show that usual critical values based on a chi-square distribution (with degrees-of-freedom equal to the number of constraints) can both lead to under-rejections

and over-rejections depending on the form of the function studied. Indeed, the Wald statistic may diverge under null hypothesis, so that arbitrary size distortions may occur.

Second, we study the asymptotic distribution of Wald-type statistics in non-regular cases. Surprisingly, the asymptotic behavior of the Wald statistic has not been generally studied for full classes of restrictions; here we consider the class of polynomial restrictions. We show that the Wald statistic either has a non-degenerate asymptotic distribution even when the estimated covariance converges to a singular matrix, or diverges to infinity. We provide conditions for convergence and a general characterization of this distribution. We find that the test can have several different asymptotic distributions under the null hypothesis – depending on the degree of singularity as well as various nuisance parameters – which may be non-chi-square distributions.

Third, we provide bounds on the asymptotic distribution (when it exists), which turn out to be proportional to a chi-square distribution where the proportionality constant depends on the degree of singularity of the function considered. In several cases of interest, this bound yields an easily available conservative critical value. Even when the limit distribution is non-pivotal it is sometime possible to provide pivotal bounds that would yield conservative critical values.

Fourth, we propose an adaptive consistent strategy for determining whether the asymptotic distribution exists and which form it takes; this approach also permits to determine what kind of bound is valid.

The framework considered and the test statistics are defined in Section 2. A number of examples are presented in Section 3; they illustrate the properties of the Wald test in singular cases. In Section 4 we discuss some general algebraic and analytic features of matrices of polynomials and quadratic forms and derive the asymptotic distribution of the Wald statistic. Bounds are derived in Section 5. An adaptive strategy for determining the asymptotic distribution and the bounds is developed in Section 6. Proofs are presented in the Appendix.

2 Framework

We consider testing q restrictions in a situation where an asymptotically non-singular estimator $\hat{\theta}_T$ is available for a $p \times 1$ parameter of interest $\bar{\theta}$ that satisfies the restrictions; $q \leq p$.

Assumption 2.1. *The function $g(\theta) = [g_1(\theta), \dots, g_q(\theta)]'$ is a continuously differentiable function from Θ to R^q , where Θ is an open subset of R^p and $q \leq p$.*

Assumption 2.1a. *The function $g(\theta) = [g_1(\theta), \dots, g_q(\theta)]'$ is such that each $g_i(\theta)$ is a polynomial of order m in the components of θ , i.e.*

$$\begin{aligned}
 g_i(\theta) &= \sum_{k=0}^m g_{ik}(\theta), \\
 g_{ik}(\theta) &= \sum_{i_1 + \dots + i_p = k} A_{ik}(i_1, \dots, i_p) \theta_1^{i_1} \dots \theta_p^{i_p}, \quad k = 0, 1, \dots, m, \quad i = 1, \dots, q.
 \end{aligned} \tag{1}$$

where $g_{ik}(\theta)$ represents a homogeneous polynomial of order k , each coefficient $A_{ik}(i_1, \dots, i_p)$ is a constant, and m is the maximal order of a polynomial in $g(\theta)$.

Assumption 2.2. We assume that some $\bar{\theta}$ satisfies a null hypothesis of the form:

$$H_0 : g(\theta) = 0. \quad (3)$$

Assumption 2.3. Assume that $\{\hat{\theta}_T : T \geq T_0\}$ is a sequence of $p \times 1$ random vectors such that for some positive definite matrix V and a scalar rate sequence $\lambda_T \rightarrow \infty$ as $T \rightarrow \infty$ convergence in probability holds:

$$\lambda_T V^{-\frac{1}{2}} (\hat{\theta}_T - \bar{\theta}) \rightarrow_p Z, \quad (4)$$

where Z is a random $p \times 1$ vector with a known absolutely continuous probability distribution, $Q(\bar{\theta})$ on R^p .

Assumption 2.3a. In addition to Assumption 2.3 $\lambda_T = T^{\frac{1}{2}}$, Z is a Gaussian random vector.

Assumption 2.4. $\{\hat{V}_T : T \geq T_0\}$ is a sequence of $p \times p$ random matrices such that $P[\text{rank}(\hat{V}_T) = p] = 1$, for all T , and

$$\text{plim}_{T \rightarrow \infty} \hat{V}_T = V \quad (5)$$

where the probability that \hat{V}_T be positive definite is one for $T \geq T_0$ (for some $T_0 > 0$).

We define the Wald test statistic:

$$W_T = \lambda_T^2 g'(\hat{\theta}_T) \left[\frac{\partial g}{\partial \theta'}(\hat{\theta}_T) \hat{V}_T \frac{\partial g'}{\partial \theta}(\hat{\theta}_T) \right]^{-1} g(\hat{\theta}_T), \quad (6)$$

when $\lambda_T^2 = T$, this is

$$W_T = T g'(\hat{\theta}_T) \left[\frac{\partial g}{\partial \theta'}(\hat{\theta}_T) \hat{V}_T \frac{\partial g'}{\partial \theta}(\hat{\theta}_T) \right]^{-1} g(\hat{\theta}_T). \quad (7)$$

If the distribution $Q(\bar{\theta})$ has a finite variance, we can assume without loss of generality that its variance is the identity matrix.

However, when the rate of convergence λ_T is not the standard $T^{1/2}$, a factor λ_T^2 shows up instead of T .

The statistic W_T is not well defined when the estimator $\hat{\theta}_T$ falls into the set of singularity points at which $\frac{\partial g}{\partial \theta'}(\hat{\theta}_T) \hat{V}_T \frac{\partial g'}{\partial \theta}(\hat{\theta}_T)$ is non-invertible (of rank less than q). Andrews (1987) studied the case where $\left[\frac{\partial g}{\partial \theta'}(\hat{\theta}_T) \hat{V}_T \frac{\partial g'}{\partial \theta}(\hat{\theta}_T) \right]^{-1}$ is replaced by a generalized inverse (*e.g.*, the Moore-Penrose inverse) and gave conditions under which the asymptotic distribution is chi-square. The main result there is that the asymptotic distribution of W_T under H_0 is chi-square $\chi^2(r_0)$ with $r_0 = \text{rank}[\frac{\partial g}{\partial \theta'}(\theta)]$ when $\text{rank}[\frac{\partial g}{\partial \theta'}(\hat{\theta}_T)]$ converges to r_0 under H_0 . This will be the case in particular when $\frac{\partial g}{\partial \theta'}(\theta)$ has rank r_0 in some open neighborhood of $\bar{\theta}$.

Here we study situations where the matrix $\frac{\partial g}{\partial \theta'}(\hat{\theta}_T) \hat{V}_T \frac{\partial g'}{\partial \theta}(\hat{\theta}_T)$ is non-singular in finite samples (with probability 1) but may converge to a singular matrix. Under Assumption 2.4 this non-singularity is equivalent to the matrix

$G(\theta) = \frac{\partial g}{\partial \theta}(\theta)$ having full rank almost everywhere.

Assumption 2.5. *The matrix $G(\theta)$ has full row rank for almost all θ .*

3 Examples and counter-examples

Before we move to study the asymptotic distribution of $W_T = W_T(\hat{\theta}_T, \hat{V}_T)$ in general terms we provide examples which show that, indeed, the asymptotic distribution of W_T is not regular. In particular, our examples illustrate non-invariance of the asymptotic distribution of the statistic to the form of the restriction and dependence (discontinuous) of the asymptotic distribution on the parameter value, $\bar{\theta}$; we also show that the asymptotic distribution may have either thinner or thicker tails than the standard χ_q^2 distribution and can even diverge to infinity under the null.

To streamline exposition of the examples we assume that $V = I$.

The following example illustrates lack of invariance of the asymptotic distribution.

Example 3.1. *Consider two equivalent forms for the null, $g(\theta) = 0$, one is (i) $\theta = 0$, the other (ii) $\theta^2 = 0$. Of course, the asymptotic distribution for the Wald test statistic in the case (i) under Assumption 2.3a is χ_1^2 . By contrast, for (ii) the value of $W_T = T \frac{\hat{\theta}_T^4}{4\hat{\theta}_T^2}$; the limit distribution then is $\frac{1}{4}\chi_1^2$.*

Below for the multivariate $\hat{\theta}_T$ we suppress dependence of the components, $\hat{\theta}_{T1}, \dots, \hat{\theta}_{Tp}$, on T .

The next example is the one given by Andrews (1987); we develop it to

illustrate both the fact that the distribution depends on $\bar{\theta}$, and also that despite the distribution not being pivotal, the usual χ_1^2 distribution provides here a pivotal upper bound.

Example 3.2. *Consider the restriction given by $g(\theta) = \theta_1\theta_2$. In this case, $G(\theta) = [\theta_2, \theta_1]$, and the Wald statistic for testing $H_0 : \theta_1\theta_2 = 0$ takes the form:*

$$W_T = T \frac{\hat{\theta}_1^2 \hat{\theta}_2^2}{\hat{\theta}_1^2 + \hat{\theta}_2^2}.$$

If either θ_1 or θ_2 is non-zero, under H_0 the limiting distribution is χ_1^2 . If, however, $\theta_1 = \theta_2 = 0$, we have:

$$W_T \xrightarrow[T \rightarrow \infty]{p} \frac{Z_1^2 Z_2^2}{Z_1^2 + Z_2^2}. \quad (8)$$

Writing this expression as $Z_2^2 - \frac{Z_2^4}{Z_1^2 + Z_2^2}$ we see that the limit distribution in this case under Assumption 2.3a is strictly below χ_1^2 , thus it is not pivotal. However, χ_1^2 provides a conservative bound.

A more precise bound can be obtained. Write the vector (Z_1, Z_2) in polar coordinates: $(r \sin \phi, r \cos \phi)$, with $r^2 = Z_1^2 + Z_2^2, r \geq 0$ and $\phi = \arcsin \frac{Z_1}{r}$. Then the limit ratio in (8) becomes

$$\frac{1}{4} r^2 (\sin 2\phi)^2.$$

Thus the distribution of $\frac{1}{4} r^2$ provides an upper bound on the limit distribution of W_T under the most general assumptions.

If the distribution of the vector Z is spherical (that is depends on r only), then the distribution of ϕ is uniform and independent of r ; it follows that $r \sin 2\phi$ has then the same distribution as $r \sin \phi$. Indeed, conditionally on r (denoting by $F_{\cdot|r}(\cdot)$ the conditional distribution)

$$\begin{aligned} F_{\sin 2\phi|r}\left(\frac{\alpha}{r}\right) &= F_{2\phi|r}\left(\arcsin \frac{\alpha}{r}\right) = 2F_{\phi|r}\left(\frac{1}{2} \arcsin \frac{\alpha}{r}\right) \\ &= 2 \int_0^{\frac{1}{2} \arcsin(\alpha/r)} I(0 \leq \phi \leq 2\pi) \frac{1}{2\pi} d\phi = F_{\phi|r}\left(\arcsin \frac{\alpha}{r}\right). \end{aligned}$$

Then the limit of W_T is given by the distribution of $\frac{1}{4}Z_1^2$ (the same as $\frac{1}{4}Z_2^2$).

Under normality this is distributed as $\frac{1}{4}\chi_1^2$. If the distribution of Z is such that each marginal is normal but the joint is not, then $\frac{1}{4}\chi_1^2$ provides an upper bound but not necessarily the distribution.

The limit $\frac{1}{4}\chi_1^2$ distribution under normality was obtained by Glonek (1993) who also demonstrated that this asymptotic distribution does not depend on the covariance matrix V . Thus the limit distribution for test of this hypothesis for a normal Z is either χ_1^2 or $\frac{1}{4}\chi_1^2$, therefore is not pivotal. However, χ_1^2 provides a conservative bound, so that there is a pivotal upper bound.

In the above examples, standard critical values are conservative in non-regular cases. So here if we do not know whether we are in a regular case or not, usual critical values are the appropriate ones: the test never over-rejects (asymptotically) under the null hypothesis when using critical values entailed by usual regularity assumptions.

However, it is also possible that the standard limit distribution does not

hold in any part of the parameter space and using the corresponding critical values may lead to a severely oversized test.

Example 3.3. Suppose that $g(\theta) = \theta_1^2 + \dots + \theta_p^2$; then $G(\theta) = [2\theta_1, \dots, 2\theta_p]$ and

$$W_T = T \frac{\left(\sum_{i=1}^p \hat{\theta}_i^2\right)^2}{4\sum_{i=1}^p \hat{\theta}_i^2}.$$

Then the limit distribution is that of $\frac{1}{4} \|Z\|^{\frac{1}{2}}$; under normality this is $\frac{1}{4}\chi_p^2$; it is a pivotal distribution even though non-standard. If p is large enough, the χ_1^2 will not provide an upper bound.

In the case of more than one restriction in addition to all the non-standard features that can arise for a single restriction it is also possible that the test statistic diverges even under H_0 .

Example 3.4. Suppose that $q = p = 2$ and $g(\theta) = [\theta_1^2 : \theta_1\theta_2^2]'$. Then

$$G(\theta) = \begin{bmatrix} 2\theta_1 & 0 \\ \theta_2^2 & 2\theta_1\theta_2 \end{bmatrix};$$

it follows that

$$W_T = T \frac{4\hat{\theta}_1^2 + \hat{\theta}_2^2}{16}.$$

Then if (i) $\bar{\theta}_1 = \bar{\theta}_2 = 0$ the asymptotic distribution is $\frac{1}{4}Z_1^2 + \frac{1}{16}Z_2^2$ and thus under normality is a linear combination of two independent χ_1^2 and is bounded by $\frac{1}{4}\chi_2^2$. However, if (ii) $\bar{\theta}_1 = 0$, but $\bar{\theta}_2 \neq 0$ the null still holds, but as $T \rightarrow \infty$ the Wald statistic diverges to $+\infty$.

The examples show that even for the simplest restrictions the limit distribution of the Wald statistic may be quite complex and far from standard. A number of applications require the Wald test of polynomial restriction functions where singularity could not be excluded and thus the non-standard features illustrated by the simple examples above may be present.

Several applications involve test of one restriction, such as tests of determinants and other polynomial functions in coefficients in Gourieroux, Monfort, Renault (1988, 1993), Galbraith and Zinde-Walsh (1992), Gourieroux and Jasiak (2013). In tests of Granger noncausality in VARMA models by Boudjellaba, Dufour and Roy (1992,1994) several polynomial restrictions need to hold under the null, similarly in testing noncausality at various horizons in Dufour, Renault (1998) and Dufour, Pelletier and Renault (2005).

4 Limit distribution of the Wald statistic

The asymptotic behavior of the Wald statistic has not been generally examined in the literature for full functional classes of nonlinear restrictions. Here we provide a characterization of the asymptotic distribution for restrictions given by polynomial functions. We shall work under the Assumptions 2.1a, 2.2,2.3, 2.4 and 2.5.

Two approaches are possible. The Wald statistic can be represented as a ratio of two polynomial functions in random variables; such a representation implicitly incorporates the information in the polynomial restrictions.

Another approach is based on an explicit analysis of the restrictions and represents the limit distribution in a quadratic form; this representation permits simple derivation of conservative bounds. In this paper we focus on the second representation.

The first subsection gives a few general results about matrices of polynomials; the second applies them to matrices related to the Jacobian matrix of the restrictions under test. The third subsection provides the limit distribution for the Wald statistic for polynomial restrictions; this distribution is in general not pivotal and depends on $\bar{\theta}$.

4.1 Matrices of polynomials

A polynomial function is either the zero polynomial, when it is identically zero (the coefficient on every monomial term is zero), or it is non-zero a.e..

Consider a $q \times p$ matrix $G(y)$ of polynomials of variable $y \in \mathbb{R}^p$. When $q = p$, we will say that the matrix $G(y)$ is non-singular if its determinant is a non-zero polynomial. More generally, we will define the rank of the $q \times p$ matrix $G(y)$ as the largest dimension of a square non-singular submatrix. This section considers $q \times p$ matrices $G(y)$, $q \leq p$, of full row rank q (Assumption 2.5).

We first note that, for any square $q \times q$ non-singular matrix S , $SG(y)$ is also a matrix of polynomials of rank q : if $\check{G}(y)$ is a $q \times q$ submatrix of $G(y)$ with determinant $\det(\check{G}(y))$ that is a non-zero polynomial, it is also true for the submatrix $S\check{G}(y)$ of $SG(y)$.

Consider a polynomial $h(y) = \sum_{k=0}^n h_k(y)$ with homogeneous polynomial terms of order k :

$$h_k(y) = \sum_{i_1+\dots+i_p=k} h_k(i_1, \dots, i_p) y_1^{i_1} \dots y_p^{i_p}. \quad (9)$$

Denote by \bar{k}_h the lowest order of homogeneous polynomial entering into polynomial $h(y)$:

$$\bar{k}_h = \min_{0 \leq k \leq n} \{k : h_k(i_1, \dots, i_p) \neq 0 \text{ for some } i_1 + \dots + i_p = k\}. \quad (10)$$

Note that

$$\lambda^{\bar{k}_h} h(y/\lambda) = h_{\bar{k}_h}(y) + \sum \lambda^{r_l} r_l(y), \quad (11)$$

with all $r_l < 0$ and $r_l(y)$ polynomial with $\bar{k}_{r_l} > \bar{k}_h$.

Consider all possible $\tilde{G}(y)_l$, with $\tilde{G}(y)_l$ a $q \times q$ submatrix of $G(y)$; $l = 1, \dots, L$ with $L = \frac{p!}{q!(p-q)!}$.

Define

$$\bar{\alpha} = \min_l (\bar{k}_{\det(\tilde{G}(y)_l)}) \quad (12)$$

with the convention $\bar{k}_{\det(\tilde{G}(y)_l)} = +\infty$ if $\det(\tilde{G}(y)_l)$ is the zero polynomial.

Note that for some \tilde{G}_l strict inequality $\bar{k}_{\det(\tilde{G}(y)_l)} > \bar{\alpha}$ may hold as shown in the example below.

Example 4.1. $G(y) = \frac{\partial g}{\partial y'}$ for $g(y) = (y_1^2 + y_3^3, y_2^2 + y_4^3, y_1^2 + y_2^2)'$; then

$$G(y) = \begin{bmatrix} 2y_1 & 0 & 3y_3^2 & 0 \\ 0 & 2y_2 & 0 & 3y_4^2 \\ 2y_1 & 2y_2 & 0 & 0 \end{bmatrix}.$$

We have four possible $q \times q$ submatrices (with $q = 3$):

$$\tilde{G}(y)_1 = \begin{bmatrix} 2y_1 & 0 & 3y_3^2 \\ 0 & 2y_2 & 0 \\ 2y_1 & 2y_2 & 0 \end{bmatrix}, \det(\tilde{G}(y)_1) = -12y_1y_2y_3^2$$

$$\tilde{G}(y)_2 = \begin{bmatrix} 2y_1 & 0 & 0 \\ 0 & 2y_2 & 3y_4^2 \\ 2y_1 & 2y_2 & 0 \end{bmatrix}, \det(\tilde{G}(y)_2) = -12y_1y_2y_4^2$$

$$\tilde{G}(y)_3 = \begin{bmatrix} 2y_1 & 3y_3^2 & 0 \\ 0 & 0 & 3y_4^2 \\ 2y_1 & 0 & 0 \end{bmatrix}, \det(\tilde{G}(y)_3) = 18y_1y_3^2y_4^2$$

$$\tilde{G}(y)_4 = \begin{bmatrix} 0 & 3y_3^2 & 0 \\ 2y_2 & 0 & 3y_4^2 \\ 2y_2 & 0 & 0 \end{bmatrix}, \det(\tilde{G}(y)_4) = 18y_2y_3^2y_4^2$$

Hence $\bar{\alpha} = 4$ but $\det(\tilde{G}(y)_3)$ and $\det(\tilde{G}(y)_4)$ are homogeneous polynomials of degree $5 > \bar{\alpha}$.

Thus, $\bar{\alpha}$ is the smallest possible degree of an homogeneous polynomial in

the determinant of any non-singular $q \times q$ submatrix of $G(y)$. Then $\bar{\alpha} = 0$ if and only if $y = 0$ is not a root of some such determinant and $\bar{\alpha} > 0$ otherwise. In other words, $\bar{\alpha} = 0$ if and only if $G(0)$ is of full row rank.

Select some matrix $\tilde{G}(y)_l$ for which $\bar{k}_{\det(\tilde{G}(y)_l)} = \bar{\alpha}$. Note that then (11) implies that the limit:

$$\lim_{\lambda \rightarrow \infty} \lambda^{\bar{\alpha}} \det(\tilde{G}_l(y/\lambda)) \quad (13)$$

is a polynomial in y on R^p that is distinct from zero almost everywhere.

For the matrix of polynomials $G(y)$ of rank q and any non-singular $q \times q$ matrix S , for the polynomial matrix $SG(y)$ there is some $\alpha = (\alpha_1, \dots, \alpha_p)$ such that

$$\lim_{\lambda \rightarrow \infty} \text{diag}(\lambda^{\alpha_i}) SG(y/\lambda) \quad (14)$$

exists and is a finite non-zero polynomial matrix, $\bar{G}(y)$. Indeed, define $\alpha_i = \min_j \{\bar{k}_{\{SG(y)\}_{ij}}\}$, where $\{SG(y)\}_{ij}$ denotes the polynomial that is the ij -th element of the matrix $SG(y)$. From (11) existence of the limit matrix follows.

Lemma 4.1. *Suppose that there exists $a = (\alpha_1, \dots, \alpha_q)$ with $\alpha_i \geq 0$ and a non-singular $q \times q$ matrix S such that the limit matrix:*

$$\bar{G}(y) = \lim_{\lambda \rightarrow \infty} \text{diag}(\lambda^{\alpha_i}) SG(y/\lambda) \quad (15)$$

is a finite non-zero matrix. Then for $\bar{\alpha}$ for which (13) holds we get $\sum_{i=1}^q \alpha_i \leq \bar{\alpha}$. $\bar{G}(y)$ is non-singular if and only if

$$\sum_{i=1}^q \alpha_i = \bar{\alpha}.$$

When $\bar{G}(y)$ exists for some a and some matrix S , the matrix S can always be chosen such that $0 \leq \alpha_1 \leq \dots \leq \alpha_q \leq \bar{\alpha}$.

Definition 4.1. *A $q \times p$ matrix of polynomials $G(y)$ satisfies the "continuity of lower degree ranks property" (CLDR) if for some non-singular $q \times q$ matrix S and for some $\alpha = (\alpha_1, \dots, \alpha_q)$ such that $\sum_{i=1}^q \alpha_i = \bar{\alpha}$, $0 \leq \alpha_1 \leq \dots \leq \alpha_q \leq \bar{\alpha}$, (15) provides a rank q matrix of polynomials $\bar{G}(y)$.*

Essentially, the CLDR property holds if for some S the transformed $SG(y)$ is such that the stabilizing rate \bar{a} for the determinant is shared between the rows of the matrix $SG(y)$ according to (15), and the limit matrix is non-singular.

The matrix $\bar{G}(y)$ depends upon the choice of the matrix S . Indeed in Example 4.1 $\bar{\alpha} = 4$ but it is clear that for $S = I$ Lemma 4.1 does not hold. This is a consequence of the fact that there is a linear dependence between the degree one polynomial terms in the rows of the matrix. However setting

$$S = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & -1 \end{bmatrix}$$

yields $SG(y/\lambda)$ as

$$\begin{bmatrix} 2y_1/\lambda & 0 & 3y_3^2/\lambda^2 & 0 \\ 0 & 2y_2/\lambda & 0 & 3y_4^2/\lambda^2 \\ 0 & 0 & 3y_3^2/\lambda^2 & 3y_4^2/\lambda^2 \end{bmatrix}$$

and CLDR holds with this S and $\alpha = (1, 1, 2)$.

The next example demonstrates that the CLDR property may not hold for some $G(y)$ even with $q = p$.

Example 4.2. *Consider*

$$G(y) = \begin{bmatrix} y_1 & 0 \\ (c + y_2)^2 & y_1(c + y_2) \end{bmatrix}$$

with $c \neq 0$. Then $\bar{\alpha} = 2$. Consider an arbitrary 2×2 matrix $S = (s_{ij})_{1 \leq i, j \leq 2}$.

Three possibilities could arise for $\bar{\alpha} = 2$ if CLDR were to hold so that $\bar{a} = \alpha_1 + \alpha_2$.

First, $\alpha_1 = \alpha_2 = 1$, then

$$\lim_{\lambda \rightarrow \infty} \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} SG(y/\lambda)$$

does not exist, except if $s_{12} = s_{22} = 0$, which is precluded for non-singular matrix S .

Second, $\alpha_1 = 2, \alpha_2 = 0$, then

$$\lim_{\lambda \rightarrow \infty} \begin{bmatrix} \lambda^2 & 0 \\ 0 & 1 \end{bmatrix} SG(y/\lambda)$$

does not exist for any non-zero matrix S .

Third, $\alpha_1 = 0, \alpha_2 = 2$, then

$$\lim_{\lambda \rightarrow \infty} \begin{bmatrix} 1 & 0 \\ 0 & \lambda^2 \end{bmatrix} SG(y/\lambda)$$

does not exist, except if $s_{21} = s_{22} = 0$, which is precluded for non-singular matrix S .

Now that we see that some matrices of polynomials satisfy the CLDR property and some do not, we further characterize the difference between the two possibilities.

Lemma 4.2. *Given a matrix $G(y)$ with the corresponding \bar{a} , for any non-singular matrix S and $a' = (\alpha'_1, \dots, \alpha'_q)$ with $0 \leq \alpha'_1 \leq \dots \leq \alpha'_q \leq \bar{a}$ and $\sum_{i=1}^q \alpha'_i = \bar{a}$ either (i) CLDR property holds with this S and a' , or (ii) no finite limit exists for*

$$\left[\text{diag}(\lambda^{\alpha'_i}) SG(y/\lambda) \right],$$

or (iii) if a finite limit does exist

$$\text{rank} \lim_{\lambda \rightarrow \infty} \left[\text{diag}(\lambda^{\alpha'_i}) SG(y/\lambda) \right] < q. \quad (16)$$

Thus if S and a are such that a finite limit (15) exists then either the CLDR property holds for such S, a or the limit matrix $\bar{G}(y)$ has a deficient rank. If the limit matrix $\bar{G}(y)$ has a deficient rank for some S, a , it has a deficient rank for any other S', a' . We can thus say that $G(y)$ is either CLDR or deficient rank. To determine whether there exist some S and a for which CLDR property holds we provide a recursive construction of S and a that either gives the CLDR property or results in a deficient rank.

Lemma 4.3. *Given a $q \times p$ matrix $G(y)$ of polynomials, there is a recursive construction that provides the pair S and a , such that either CLDR property is satisfied for this pair or the deficient rank property holds.*

The construction in the proof implies that we can write:

$$SG(y) = \bar{G}(y) + \bar{R}(y) \tag{17}$$

where for $i = 1, \dots, q$, the row i of $\bar{R}(y)$ contains no homogeneous polynomial of order smaller or equal to α_i .

4.2 Vectors of polynomial functions, Jacobian matrices and the Wald statistic

Consider the $q \times 1$ vector of polynomial functions, $g(y)$ with $g(0) = 0$ and the Jacobian matrix of polynomials, $G(y) = \frac{\partial g}{\partial y'}(y)$.

Consider a non-singular S that satisfies (17) for $G(y)$ (and $\sum_{i=1}^q \alpha_i \leq \bar{\alpha}$).

Then

$$Sg(y) = \bar{g}(y) + \bar{r}(y) \quad (18)$$

where for every $i = 1, \dots, q$

$$\begin{aligned} \bar{g}_i(y) &= \int_0^y \bar{G}(x)_i dx; \\ \bar{r}_i(y) &= \int_0^y \bar{R}(x)_i dx, \end{aligned}$$

where the integration of the gradient along any continuous curve from 0 to y provides each component of g, r .

Each $\bar{g}_i(y)$ of $\bar{g}(y)$ is a homogeneous polynomial of order $(\alpha_i + 1)$ and, by Euler formula:

$$\bar{g}(y) = \Lambda \bar{G}(y) y \quad (19)$$

with:

$$\Lambda = \text{diag} \left(\frac{1}{\alpha_i + 1} \right).$$

Each element $\bar{r}_i(y)$ of $\bar{r}(y)$ contains no homogeneous polynomial of order smaller or equal to $(\alpha_i + 1)$.

In particular, when λ goes to infinity:

$$\begin{aligned}
diag(\lambda^{\alpha_i})SG(y/\lambda) &= \bar{G}(y) + O(1/\lambda); \\
diag(\lambda^{\alpha_i})S\lambda g(y/\lambda) &= \bar{g}(y) + O(1/\lambda).
\end{aligned} \tag{20}$$

Define now for some positive definite matrix Ω a quadratic form

$$W(y, g, \lambda, \Omega) = \lambda^2 g'(y/\lambda) [G(y/\lambda) \Omega G'(y/\lambda)]^{-1} g(y/\lambda). \tag{21}$$

Note that $W(y, g, \lambda, \Omega) = W(y, Mg, \lambda, \Omega)$ for any non-singular matrix M ; we can choose $M = S(\lambda) = diag(\lambda^{\alpha_i})S$. This provides

$$W(y, g, \lambda, \Omega) = g'(y/\lambda) \lambda S'(\lambda) [S(\lambda) G(y/\lambda) \Omega G'(y/\lambda) S'(\lambda)]^{-1} S(\lambda) \lambda g(y/\lambda). \tag{22}$$

Suppose that $\Omega = \Omega(\lambda)$ with the property that as $\lambda \rightarrow \infty$ the matrix $\Omega = \Omega^0 + o(1)$, with Ω^0 a non-singular matrix. Then we can write as $\lambda \rightarrow \infty$

$$\begin{aligned}
&W(y, g, \lambda, \Omega) \\
&= [\bar{g}(y) + O(1/\lambda)]' \left\{ [\bar{G}(y) + O(1/\lambda)] [\Omega^0 + o(1)] [\bar{G}(y) + O(1/\lambda)]' \right\}^{-1} [\bar{g}(y) + O(1/\lambda)].
\end{aligned}$$

If CLDR property holds for G , $\bar{G}(y)$ is full rank and then

$$\begin{aligned}
\lim_{\lambda \rightarrow \infty} W(y, g, \lambda, \Omega) &= [\bar{g}(y)]' \left\{ [\bar{G}(y)] \Omega^0 [\bar{G}(y)]' \right\}^{-1} [\bar{g}(y)] \quad (23) \\
&= W_\infty(y, g, \Omega^0).
\end{aligned}$$

Next, we demonstrate tht if CLDR property does not hold $W(y, g, \lambda, \Omega)$ diverges to infinity as $\lambda \rightarrow \infty$.

Suppose that CLDR property does not hold, then find a for which (14) provides a finite matrix, by lack of CLDR in that case $\Sigma \alpha_i < \bar{a}$.

Then recall that $[G(y/\lambda)\Omega G'(y/\lambda)]^{-1}$ can be represented as the ratio of the adjoint matrix, denoted $[G(y/\lambda)\Omega G'(y/\lambda)]^*$, to the determinant, $\det[G(y/\lambda)\Omega G'(y/\lambda)]$. Write (22) as

$$\frac{g'(y/\lambda)\lambda S'(\lambda)[S(\lambda)G(y/\lambda)\Omega G'(y/\lambda)S'(\lambda)]^* S(\lambda)\lambda g(y/\lambda)}{\det[S(\lambda)G(y/\lambda)\Omega G'(y/\lambda)S'(\lambda)]};$$

this is

$$\frac{[\bar{g}(y) + O(1/\lambda)]' \left\{ [\bar{G}(y) + O(1/\lambda)] [\Omega^0 + o(1)] [\bar{G}(y) + O(1/\lambda)]' \right\}^* [\bar{g}(y) + O(1/\lambda)]}{\det[S(\lambda)G(y/\lambda)\Omega G'(y/\lambda)S'(\lambda)]}.$$

The numerator has a finite limit.

In the denominator we have

$$\begin{aligned}\det[S(\lambda)G(y/\lambda)\Omega G'(y/\lambda)S'(\lambda)] &= \lambda^{2\Sigma\alpha_i} \det[SG(y/\lambda)\Omega G'(y/\lambda)S'] \\ &= \lambda^{2[\Sigma\alpha_i - \bar{\alpha}]} \lambda^{2\bar{\alpha}} \det[SG(y/\lambda)\Omega G'(y/\lambda)S'].\end{aligned}$$

Thus as $\lambda \rightarrow \infty$, when the CLDR property is not fulfilled, while $\lambda^{2\bar{\alpha}} \det[SG(y/\lambda)\Omega G'(y/\lambda)S']$ has a finite limit for every Ω , $\lambda^{2[\Sigma\alpha_i - \bar{\alpha}]}$ converges to zero and

$$W(y, g, \lambda, \Omega) \xrightarrow[\lambda \rightarrow \infty]{} \infty. \quad (24)$$

Thus CLDR property plays a very important role in the existence of a limit for the Wald statistic.

4.3 The limit distribution of the Wald statistic

Define $y = A(\theta - \bar{\theta})$ for some non-degenerate matrix A ; with this substitution under the assumption $g(\bar{\theta}) = 0$ the polynomial function $g(\theta)$ becomes $g(A^{-1}y + \bar{\theta}) = g_{\bar{\theta}}(y)$, a polynomial function with $g_{\bar{\theta}}(0) = 0$. The Jacobian polynomial matrix G gets multiplied by the nonsingular matrix A^{-1} to provide the new Jacobian $G_{\bar{\theta}}(y)$ with respect to y and $G(\theta) = G_{\bar{\theta}}(y)A$. Note the role that the nonsingular matrix A plays: it does not change the order of polynomial function g ; if CLDR property holds for $G_{\bar{\theta}}$ defined with some nonsingular A , it holds for any other nonsingular A . In the notation for the function and the Jacobian we do not emphasize then the role of A .

The Wald test statistic in (6) for $\theta = \hat{\theta}_T, \lambda = \lambda_T, \Omega = \hat{V}_T$ and with $y_T = A(\hat{\theta}_T - \bar{\theta})$ can be written as

$$\lambda_T^2 g'_\theta(y_T) [G_\theta(y_T) A \hat{V}_T A' G'_\theta(y_T)]^{-1} g_\theta(y_T).$$

Consider $g_{\bar{\theta},0}(y)$ for $A_0 = V^{-\frac{1}{2}}$, and $g_{\bar{\theta},1}(y)$ for some nonsingular A ; then the distribution of $g_{\bar{\theta},1}(V^{\frac{1}{2}}AZ)$ is the same as $g_{\bar{\theta},0}(Z)$. The distribution of $G_{\bar{\theta},1}(V^{\frac{1}{2}}AZ)$ is the same as $G_{\bar{\theta},0}(Z)V^{-\frac{1}{2}}A^{-1}$, since we consider non-singular reparametrizations of continuous functions. Thus the limit distribution does not depend on A . Below we write $g_{\bar{\theta}}, G_{\bar{\theta}}$ for $g_{\bar{\theta},0}$ and $G_{\bar{\theta},0}$.

Theorem 4.1. *Under the Assumptions 2.1a, 2.2, 2.3 and 2.4 if (a) at $\bar{\theta}$ the CLDR property holds for $G_\theta(y)$ (for any non-singular A) then the limit distribution of W_T as $T \rightarrow \infty$ is given by the distribution of*

$$[\bar{g}_\theta(Z)]' \{ [\bar{G}_\theta(Z)] [\bar{G}_\theta(Z)]' \}^{-1} [\bar{g}_\theta(Z)]; \quad (25)$$

if (b) at $\bar{\theta}$ the deficient rank property holds, then W_T diverges to infinity as $T \rightarrow \infty$.

Corollary 4.1. *When the CLDR property holds the limit distribution can be represented as the distribution of*

$$Z' [\bar{G}_\theta(Z)]' \Lambda_{\bar{\theta}} \{ [\bar{G}_\theta(Z)] [\bar{G}_\theta(Z)]' \}^{-1} \Lambda_{\bar{\theta}} [\bar{G}_\theta(Z)] Z. \quad (26)$$

This follows from the Euler formula (19).

The Example below illustrates the applicability of parts (a) and (b) of Theorem 4.1.

Example 4.3. Recall Example 3.4 with $g(\theta) = \begin{bmatrix} \theta_1^2 \\ \theta_1 \theta_2^2 \end{bmatrix}$. Then, the set of possible values of $\bar{\theta}$ under the null is the line $(0, \bar{\theta}_2)'$, $\bar{\theta}_2 \in \mathbb{R}$, and for $A = I$

$$G(y) = \begin{bmatrix} 2y_1 & 0 \\ (\bar{\theta}_2 + y_2)^2 & 2y_1(\bar{\theta}_2 + y_2) \end{bmatrix}.$$

When $\bar{\theta}_2 \neq 0$ $\bar{G} = \begin{bmatrix} 2y_1 & 0 \\ \bar{\theta}_2^2 & 0 \end{bmatrix}$; the CLDR property does not hold, (b) of the Theorem applies. By contrast, if $\bar{\theta}_2 = 0$, we have $\bar{\alpha} = 3$, and the sharing rule $\alpha = (1, 2)$ for CLDR immediately follows and (a) applies.

When there is only one restriction there is only one $\alpha_1 = \alpha_{\bar{\theta}}$. Here CLDR always holds and thus under the assumptions of Theorem 4.1 the convergence of W_T to

$$\frac{1}{(1 + \alpha_{\bar{\theta}})^2} \frac{(Z' \bar{G}_{\bar{\theta}}(Z))'^2}{\bar{G}_{\bar{\theta}}(Z) \bar{G}_{\bar{\theta}}(Z)'} \equiv \frac{\|\bar{g}(Z)\|^2}{\|\bar{G}(Z)\|^2}$$

always obtains.

In the case of multiple restrictions violation of the CLDR property is possible; in such a case the statistic may diverge under the null. One could consider replacing the restrictions by a set of equivalent restrictions that preclude violation of CLDR property. This is always possible.

Indeed, for any vector g of q restrictions $g(\theta) = 0$, the restrictions are

equivalent to a single restriction

$$\|g(\theta)\|^2 = \sum_{i=1}^q g_i^2(\theta) = 0.$$

Since the CLDR property is not an issue with one constraint a possible strategy is to replace the q restriction by the single restriction and consider the corresponding test statistic.

Of course, this simplification may have an important cost in terms of power since it does not take into account the fact that for an estimator $\hat{\theta}_T$ the components may be highly correlated. Then, the naive norm $\|g(\hat{\theta}_T)\|$ of the vector $g(\hat{\theta}_T)$ may not be the efficient way to assess its distance from zero; some weighting may be advantageous.

5 Bounds on the statistic and bounds on critical values

Sometimes the asymptotic distribution of the Wald statistic under the null, even when non-standard, can be uniquely determined; this is the case in Example 3.3. But typically under conditions of Theorem 4.1 with possible singularity the asymptotic distribution under the null is not uniquely determined. Because the asymptotic distribution of the Wald statistic may be discontinuous in the true values it is useful to establish uniform bounds on the asymptotic distribution of the statistic, or on the critical values for the

test.

Denote by α the smallest α_i (usually α_1) in Definition 4.1. Below we show that $\frac{1}{(1+\alpha)^2} \|Z\|^2$ (distributed $\frac{1}{(1+\alpha)^2} \chi_p^2$ under normality) provides a uniform upper bound on the asymptotic null-distribution always under conditions of the Theorem 4.1(a), i.e. when CLDR property holds. If $\alpha = 0$, then there may be no singularity in which case with normality the usual asymptotic χ_q^2 distribution holds; in general the overall uniform bound with or without singularity under Theorem 4.1.(a) is χ_p^2 .

It is possible to improve on the χ_p^2 bound when $\alpha \geq 1$; sometimes the form of the restrictions may provide $\alpha \geq 1$. When this is not the case, it may be possible to establish that $\alpha \geq 1$ by the adaptive strategy proposed in the next section that would eliminate the possibility that $\alpha = 0$.

However, for testing it may be sufficient to bound the distribution in the tail rather than everywhere, and so uniform bounds on critical values are also of interest. Gouriéroux and Jasiak (2013) discuss a bound on critical values for a test of a determinant.

Here in Theorem 5.1 we first establish general bounds on an asymptotic distribution derived for a particular vector of true parameter values, when there may be a singularity at that value. We separately examine the case of one restriction. We also examine a relation between critical values at different α . In the case of one restriction it is possible to provide the number of variables for which generally the standard critical values deliver a conservative test.

5.1 A general uniform upper bound

Start with the representation of the asymptotic distribution from (26):

$$W(Z) = Z' \bar{G}(Z)' \Lambda [\bar{G}(Z) \bar{G}(Z)']^{-1} \Lambda \bar{G}(Z) Z.$$

This distribution depends on the singularity properties that are exhibited at the true value $\bar{\theta}, V$.

As the theorem below states a bound that depends only on α is possible in all cases when CLDR holds.

Theorem 5.1. *Under the conditions of Theorem 4.1(a), the asymptotic distribution of the Wald statistic under the null (that depends on the singularity properties at $\bar{\theta}$) is bounded from above by the distribution of $\frac{1}{(1+\alpha)^2} \|Z\|^2$; under the normality Assumption 2.3 this bound is $\frac{1}{(1+\alpha)^2} \chi_p^2$.*

Thus under conditions of the Theorem 5.1 there is always a general upper bound on the distribution of the Wald statistic under the null given by χ_p^2 .

Remark 5.1. *When $\Lambda = I$ implying that $\bar{G}(Z)$ does not depend on Z and is a $q \times p$ rank q matrix of constants the projection is onto a q -dimensional subspace, the limit distribution is standard and is given under normality by χ_q^2 .*

In the case of one restriction under normality the upper bound is either the usual χ_1^2 , if $\alpha = 0$, or else for some $\alpha > 0$ the bound is $\frac{1}{(1+\alpha)^2} \chi_p^2$. If all that is known is that $\alpha > 0$, then the bound $\frac{1}{4} \chi_p^2$ applies for any such α .

In Example 3.3 the limiting distribution is $\frac{1}{4}\chi_p^2$; thus this bound can be attained.

In the special case $p = q$ and under CLDR \bar{G} is invertible a.e. and $Z'\bar{G}(Z)'\Lambda [\bar{G}(Z)\bar{G}(Z)']^{-1}\Lambda\bar{G}(Z)Z = Z'\bar{G}'\Lambda\bar{G}'^{-1}\bar{G}^{-1}\Lambda\bar{G}Z$. Then

$$W(Z) \leq \|\Lambda\|^2 \|Z\|^2,$$

since the norm of a similar matrix is the same as for Λ . Under normality the bound is $\frac{1}{(1+\alpha)^2}\chi_q^2$. In this case the asymptotic distribution is bounded from above by the usual distribution and under normality the distribution χ_q^2 provides a conservative test.

5.2 Bounds on critical values for purely singular cases

$\alpha \geq 1$ under normality

A conservative test for a given level may be given by the standard critical values, even in the non-standard cases considered here since then dominance by the standard distribution is required only in the tail and not everywhere. The following Lemma demonstrates that when the distribution is purely singular ($\alpha \geq 1$) there is always a level, γ_0 , such that using the standard critical values for any $\gamma \leq \gamma_0$ provides a conservative asymptotic test. Indeed, there exists γ_0 such that $\Pr(\frac{1}{(1+\alpha)^2}\chi_p^2 > \chi_q^2(\gamma_0)) < \gamma_0$, where $\chi_q^2(\gamma_0)$ denotes the critical value.

The following lemma establishes this tail dominance.

Lemma 5.1. *Consider two random variables $T \sim \chi_{p_1}^2/\alpha_1$ and $S \sim \chi_{p_2}^2/\alpha_2$, where $p_2 > p_1, \alpha_2 > \alpha_1 > 0$. Then there exists y_0 such that for $y > y_0$ we have*

$$p.d.f._S(y) < p.d.f._T(y).$$

This makes it possible to rely only on p and q in indicating when standard critical values provide a conservative test.

When there may be a singularity with $\alpha \geq 1$ the critical value coming from the standard test will at some level result in a conservative Wald test; the question is whether this holds for conventional test levels. Abstracting from the specific form of restrictions the answer depends on α , p and q ; the higher the α and the closer together p and q , the easier to obtain conservative asymptotic tests at conventional levels. Since p and q are given by the restrictions, all that is required is to establish α .

Comparing the values of $p.d.f._{\chi_q^2}(y_{.05})$ with $p.d.f._{\chi_p^2/(1+\alpha)^2}(y_{.05})$ where $y_{.05}$ is the critical value for χ_q^2 at .05 level we determine for which $\max p$ we get a smaller value for the second $p.d.f.$; because of monotonicity in the tail this indicates smaller probability and a conservative test.

For one restriction the standard test based on χ_1^2 critical value is conservative at .05 level for $p \leq 6$ but may not be for $p = 7$. At .01 level this test is conservative for $p \leq 10$, but may not be for $p = 11$. To show this only a computation of the critical values for χ_1^2 and for overall bound $\frac{1}{4}\chi_p^2$ is

required.

When CLDR holds for $q = 2$, if $\alpha = 1$ at .05 level we get $\max p = 11$, for $q = 3$ and $\alpha = 1$ we get $\max p = 17$.

These computations show that in many situations the standard test is conservative.

6 An adaptive strategy for determining the asymptotic distribution and the bounds

From (26) it follows that the asymptotic distribution requires the knowledge of $\bar{G}_{\bar{\theta}}$ and $\Lambda_{\bar{\theta}}$. For bounds determining the lowest value on the diagonal of $\Lambda_{\bar{\theta}}$ is sufficient. The construction in proof of Lemma 4.3 makes it clear that the main issue for finding the elements of $\Lambda_{\bar{\theta}}$ is deciding on the lowest order of the homogeneous polynomial that enters non-trivially into a polynomial (that represents a matrix entry or a determinant of a polynomial matrix), to find $\bar{G}_{\bar{\theta}}$ homogeneous polynomials (their coefficients) of the corresponding lowest orders would have to be consistently estimated.

6.1 Adaptive estimation of polynomial functions and orders

For $\bar{\theta}$ consider $P_{\bar{\theta}}(\theta)$ that is a polynomial function of order m_P in components of a $p \times 1$ vector $\bar{\theta}$ with the representation in terms of components of $\theta - \bar{\theta}$

given by

$$\begin{aligned}
P_{\bar{\theta}}(\theta) &= \bar{P}_0(\bar{\theta}) + \sum_{k=1}^{m_P} \bar{P}_k(\theta - \bar{\theta}) \\
&= \bar{P}(0, \dots, 0, \bar{\theta}) + \sum_{k=1}^{m_P} \left[\sum_{i_1 + \dots + i_p = k} \bar{P}(i_1, \dots, i_p, \bar{\theta}) (\theta_1 - \bar{\theta}_1)^{i_1} \dots (\theta_p - \bar{\theta}_p)^{i_p} \right],
\end{aligned} \tag{27}$$

where the corresponding coefficients $\bar{P}(i_1, \dots, i_p, \bar{\theta})$ are values of a polynomial in components of $\bar{\theta}$ and the constant term $\bar{P}_0(\bar{\theta})$ can be represented as a coefficient, $\bar{P}(0, \dots, 0, \bar{\theta})$; $P_{\bar{\theta}}(\bar{\theta}) = \bar{P}_0(\bar{\theta})$.

Consider a linear substitution with a nonsingular matrix $A : y = A(\theta - \bar{\theta})$, with it

$$(\theta_1 - \bar{\theta}_1)^{i_1} \dots (\theta_p - \bar{\theta}_p)^{i_p} = \sum_{\substack{i'_1 + \dots + i'_p = k \\ 1 \leq i'_v \leq i_v}} \bar{A}(i'_1, \dots, i'_p, A) y_1^{i'_1} \dots y_p^{i'_p},$$

with some coefficients $\bar{A}(i'_1, \dots, i'_p, A)$ that are polynomials in the matrix ele-

ments of the matrix A^{-1} . Then the polynomial $P_{\bar{\theta}}(\theta)$ becomes

$$P_{\bar{\theta}}(y) = \bar{P}(0, \dots, 0, \bar{\theta}) + \sum_{k=1}^{m_P} \left[\sum_{i_1+\dots+i_p=k} \bar{P}(i_1, \dots, i_p, \bar{\theta}) \sum_{\substack{i'_1+\dots+i'_p=k \\ 1 \leq i'_v \leq i_v}} \bar{A}_{i_1, \dots, i_p}(i'_1, \dots, i'_p, A) y_1^{i'_1} \dots y_p^{i'_p} \right] \quad (28)$$

$$= \bar{P}(0, \dots, 0, \bar{\theta}) + \sum_{k=1}^{m_P} P_k(y) \quad (29)$$

$$= \bar{P}(0, \dots, 0, \bar{\theta}) + \sum_{k=1}^{m_P} \left[\sum_{i'_1+\dots+i'_p=k} \left(\sum_{\substack{i_1+\dots+i_p=k \\ i'_v \leq i_v}} \bar{A}_{i_1, \dots, i_p}(i'_1, \dots, i'_p, A) \bar{P}(i_1, \dots, i_p, \bar{\theta}) \right) y_1^{i'_1} \dots y_p^{i'_p} \right]$$

$$= \bar{P}(0, \dots, 0, \bar{\theta}) + \sum_{k=1}^{m_P} \left[\sum_{i_1+\dots+i_p=k} \bar{P}(i_1, \dots, i_p, \bar{\theta}, A) y_1^{i_1} \dots y_p^{i_p} \right],$$

where the coefficients $\bar{P}(i_1, \dots, i_p, \bar{\theta}, A) = \sum_{\substack{i'_1+\dots+i'_p=k \\ i_v \leq i'_v}} \bar{A}_{i_1, \dots, i_p}(i'_1, \dots, i'_p, A) \bar{P}(i'_1, \dots, i'_p, \bar{\theta})$.

For estimator $\hat{\theta}_T$ of $\bar{\theta}$ define estimators of the coefficients $\bar{P}(i_1, \dots, i_k, \bar{\theta})$

by

$$\hat{P}(i_1, \dots, i_p, \bar{\theta}) = \begin{cases} \bar{P}(i_1, \dots, i_p, \hat{\theta}_T) & \text{if } \left| \bar{P}(i_1, \dots, i_p, \hat{\theta}_T) \right| \geq \frac{c}{\lambda_T^\delta}; \\ 0 & \text{if } \left| \bar{P}(i_1, \dots, i_p, \hat{\theta}_T) \right| < \frac{c}{\lambda_T^\delta} \end{cases} \quad (30)$$

for $0 < \delta < 1$ and some $c > 0$.

If $A = I$, no further estimation is required.

For $A = V^{-\frac{1}{2}}$ and an estimator \hat{V}_T of V estimate $\bar{A}_{i'_1, \dots, i'_p}(i_1, \dots, i_p, V^{-\frac{1}{2}})$ by $\hat{A}_{i'_1, \dots, i'_p}(i_1, \dots, i_p, V^{-\frac{1}{2}}) = \bar{A}_{i'_1, \dots, i'_p}(i_1, \dots, i_p, \hat{V}_T^{-\frac{1}{2}})$.

Combining according to (29) we can obtain the estimator $\hat{P}(i_1, \dots, i_p, \bar{\theta}, A)$ of $\bar{P}(i_1, \dots, i_p, \bar{\theta}, A)$.

Define (as in (10))

$$k_P = \min_{0 \leq k \leq m_P} \{k : \bar{P}(i_1, \dots, i_p, \bar{\theta}, A) \neq 0 \text{ for } i_1 + \dots + i_p = k\}, \quad (31)$$

and correspondingly \hat{k}_P for the polynomial P with estimated coefficients $\hat{P}(i_1, \dots, i_p, \bar{\theta}, A)$. Note that k_p does not depend on A .

Lemma 6.1. *For $\hat{\theta}_T$ and \hat{V}_T that satisfy Assumptions 2.3 and 2.4*

$$\hat{P}(i_1, \dots, i_p, \bar{\theta}, V) - \bar{P}(i_1, \dots, i_p, \bar{\theta}, V) \rightarrow_p 0,$$

moreover if $\bar{P}(i_1, \dots, i_p, \bar{\theta}) = 0$,

$$\Pr(\hat{P}(i_1, \dots, i_p, \bar{\theta}) = 0) \rightarrow 1$$

and

$$\Pr(\hat{k}_P = k_P) \rightarrow 1.$$

The result implies that for any polynomial \bar{P} with probability approaching one the lowest order of homogeneous polynomials entering into \bar{P} can be determined and also for each coefficient it can be decided whether it is zero or not with probability approaching 1; each non-zero coefficient can be consistently estimated.

6.2 Adaptively estimated asymptotic Wald statistic

The case of one restriction is given by the following Lemma.

With $q = 1$ represent each component $\{G(\theta)\}_i$ of $G(\theta)$ as a polynomial of form (29) and consider the corresponding k_{G_i} defined in (31) and the corresponding estimator, \hat{k}_{G_i} . Then define $\hat{k} = \min \{\hat{k}_{G_i}\}$ and the corresponding vector $\bar{G}_{\hat{k}}(y)$ with components given by the homogeneous polynomials of order \hat{k} (some could be zero).

Lemma 6.2. *Under the Assumptions of Theorem 4.1 if (a) $\hat{k} = 0$, then with probability approaching 1 as $T \rightarrow \infty$ there is no singularity, the distribution of the asymptotic statistic is standard and under the normality assumption is χ_1^2 ; if (b) $\hat{k} > 0$ the estimated asymptotic statistic is*

$$\hat{W}_T = \frac{1}{(\hat{k} + 1)^2} \frac{(Z' \bar{G}_{\hat{k}}(Z))^2}{\bar{G}_{\hat{k}}(Z) \bar{G}_{\hat{k}}(Z)'},$$

and its distribution converges to the non-standard asymptotic distribution for the Wald statistic at $\bar{\theta}$ as $T \rightarrow \infty$.

The next theorem considers the general case of the Wald test for several restrictions. Denote by \hat{k}_{\det} the estimator of k_P applied to $\bar{P}_{\bar{\theta}}(\theta)$ in (27) that represents the polynomial $\det[G(\theta)G(\theta)']$; as $T \rightarrow \infty$ $\Pr(\hat{k}_{\det} = k_{\det}) \rightarrow 1$. Set $A = I$ then for every $q \times q$ submatrix $\hat{G}_l(y)$ of $\hat{G}(y)$ the estimator of $k_{\det,l}$ for the corresponding determinant polynomial as $T \rightarrow \infty$ equals the true $k_{\det,l}$ with probability approaching 1, and so does then the estimated value of \bar{a} , as well as the estimated α_i defined in proof of Lemma 4.3. It

follows that thus one can determine with probability approaching 1 whether the CLDR property holds and if it does estimate the matrix $\Lambda_{\bar{\theta}}$ with probability approaching 1. Then the corresponding consistent estimator, $\hat{G}_{\bar{\theta}}(y)$ of $\bar{G}_{\bar{\theta}}(y)$ is obtained for $A = I$. A consistent estimator of the corresponding polynomial, $\bar{G}_{\bar{\theta}}(Ay)$ for $A = V^{-\frac{1}{2}}$ can be obtained by substituting $\tilde{y} = \hat{V}^{-\frac{1}{2}}y$ into the estimator $\hat{G}_{\bar{\theta}}(y)$ to obtain $\tilde{G}_{\bar{\theta}}(\tilde{y})$.

Theorem 6.1. *Under the Assumptions of Theorem 4.1 if (a) the corresponding estimated $\hat{k}_{\text{det}} = 0$, then with probability approaching 1 as $T \rightarrow \infty$ the distribution of the asymptotic statistic is standard and under normality is χ_q^2 ; if (b) for $A = I$, $\hat{k}_{\text{det}} \neq 0$ but the estimated $\hat{G}_{\bar{\theta}}(y)$ has deficient rank as $T \rightarrow \infty$, then the statistic diverges to infinity; if (c) with $A = I$, $\hat{k}_{\text{det}} \geq 1$ and the estimated $\bar{G}_{\bar{\theta}}(y)$ satisfies the CLDR property with the estimated matrix $\hat{\Lambda}_{\bar{\theta}}$, then the limit distribution is consistently estimated by*

$$Z'[\tilde{G}_{\bar{\theta}}(Z)]'\hat{\Lambda}_{\bar{\theta}} \left\{ \left[\tilde{G}_{\bar{\theta}}(Z) \right] [\tilde{G}_{\bar{\theta}}(Z)]' \right\}^{-1} \hat{\Lambda}_{\bar{\theta}}[\tilde{G}_{\bar{\theta}}(Z)]Z.$$

6.3 Conservative tests with adaptively estimated bounds

From Theorem 5.1 it follows that if the CLDR property holds the bound is provided by

$$\frac{1}{(1 + \alpha)^2} \|Z'Z\|, \text{ with } \|Z'Z\| \text{ distributed as } \chi_p^2 \text{ under normality.}$$

For one restriction CLDR always holds and $\hat{a} = \hat{k}$ as defined in Lemma 6.2 provides α with probability approaching 1.

For several restrictions estimate $\hat{k}_{\text{det}}, \hat{G}_{\bar{\theta}}(y)$ as defined in Theorem 6.1. and then if CLDR property holds it is sufficient to define the estimate of α as the smallest diagonal element of $\hat{\Lambda}_{\bar{\theta}}$ as in Theorem 6.1. This estimator will equal the true α with probability approaching 1.

Use the bound $\frac{1}{(\hat{a}+1)^2} \chi_p^2$.

7 Appendix: Proofs

Proof of lemma 4.1. We first note that for the submatrix $\tilde{G}_l(y)$ defined by (13),

$$\bar{G}_l(y) = \lim_{\lambda \rightarrow +\infty} \text{diag}(\lambda^{\alpha_i}) S \tilde{G}_l(y/\lambda)$$

is a submatrix of $\bar{G}(y)$. Then

$$\lim_{\lambda \rightarrow +\infty} \lambda^{\sum_{i=1}^q \alpha_i} \det(S) \det(\tilde{G}_l(y/\lambda))$$

exists and since S is nonsingular

$$\lim_{\lambda \rightarrow +\infty} \lambda^{\sum_{i=1}^q \alpha_i} \det(\tilde{G}_l(y/\lambda)) < \infty.$$

Then this is

$$\lim_{\lambda \rightarrow +\infty} \left(\lambda^{\sum_{i=1}^q \alpha_i - \bar{a}} \right) \lambda^{\bar{a}} \det(\tilde{G}_l(y/\lambda))$$

and it follows that $\sum_{i=1}^q \alpha_i - \bar{a} \leq 0$. If $\sum_{i=1}^q \alpha_i - \bar{a} = 0$ then $\bar{G}_l(y)$ is full rank and so is $\bar{G}(y)$. If $\bar{G}(y)$ is full rank then there is a square submatrix $\bar{G}_l(y)$ of full rank, for the corresponding submatrix in $SG(y)$

$$\lim_{\lambda \rightarrow +\infty} \lambda^{\sum_{i=1}^q \alpha_i} \det(\tilde{G}_l(y/\lambda)) > 0$$

and $\sum_{i=1}^q \alpha_i - \bar{a} \geq 0$. The equality follows. ■

Proof of Lemma 4.2. By the property (11) some a for which (13) holds exists and by the Lemma 4.1 either CLDR holds or a is such that $\sum_{i=1}^q \alpha_i \neq \bar{a}$. Then by the condition on a' if CLDR does not hold then for some i we have $\alpha'_i > \alpha_i$, or $\alpha'_i < \alpha_i$. Then for any $\{i, j\}$ matrix entry in the matrix $SG(y/\lambda)$, given by $S'_i G_{\cdot j}(y/\lambda)$ (where for a matrix A , A_i denotes the i th row and $A_{\cdot j}$ - the j th column)

$$\lambda^{\alpha'_i} S'_i G_{\cdot j}(y/\lambda) = \lambda^{\alpha'_i - \alpha_i} \lambda^{\alpha_i} S'_i G_{\cdot j}(y/\lambda)$$

In the first case $\alpha'_i > \alpha_i$, this matrix entry diverges to infinity. In the second case $\alpha'_i < \alpha_i$ and as $\lambda \rightarrow \infty$ this matrix entry converges to zero for every j , thus the limit matrix $\lim_{\lambda \rightarrow +\infty} [diag(\lambda^{\alpha'_i}) SG(y/\lambda)]$ has deficient rank. ■

Proof of Lemma 4.3. Start with a $q_v \times p$ matrix of polynomials $G^v(y)$.

For each matrix element, $\{G^v(y)\}_{ij}$, which is a polynomial, define the lowest order of homogeneous polynomial, $\bar{k}_{\{G^v(y)\}_{ij}}$. Then select $\bar{k}_v = \min_{i,j} \{\bar{k}_{\{G^v(y)\}_{ij}}\}$.

Consider a polynomial matrix, $\tilde{G}_v(y)$ such that

$$\left\{ \tilde{G}_v(y) \right\} = \begin{cases} \{G^v(y)\}_{ij, \bar{k}_v} & \text{when this polynomial is non-zero} \\ 0 & \text{otherwise.} \end{cases}$$

In other words, the ij element of $\tilde{G}_v(y)$ is either a non-zero polynomial of order \bar{k}_v , that entered into $\{G^v(y)\}_{ij}$, or zero. Next, consider all square submatrices $r_v \times r_v$, $r_v \leq q_v$ of $\tilde{G}_v(y)$, for at least one of those determinant is non-zero; select the largest \bar{r}_v with the property that some submatrix of this dimension has a non-zero determinant, and (i) either $\bar{r}_v = q_v$, or (ii) determinant of any submatrix with $q_v \geq r_v > \bar{r}_v$ is zero.

In case (i) define $S^v = I_{q_v}$. In case (ii) construct a non-singular matrix S^v , such that for some $\bar{r}_v \times p$ full row rank matrix of polynomials, $\bar{G}^v(y)$,

$$S^v \tilde{G}_v(y) = \begin{bmatrix} \bar{G}^v(y) \\ 0 \end{bmatrix}.$$

Such a matrix always exists. Then $S^v G^v(y)$ has the representation

$$\begin{bmatrix} \bar{G}^v(y) + N^v(y) \\ G^{v+1}(y) \end{bmatrix}$$

where if $N^v(y)$ is non-zero the polynomial entries in the matrix $N^v(y)$ have homogeneous polynomial terms of order no less than \bar{k}_v ; and the non-zero $(q_v - \bar{r}_v) \times p$ matrix $G^{v+1}(y)$ has polynomial terms only of order $\geq \bar{k}_v + 1$.

Consider now the original matrix $G(y)$, denote it $G^1(y)$ with $q_1 = q$ and employ the construction recursively until for some m it ends: $\sum_{v=1}^m \bar{r}_v = q$.

Denote by \bar{S}^v the matrix $\begin{bmatrix} I_{\bar{r}_1 + \dots + \bar{r}_{v-1}} \\ S^v \end{bmatrix}$ and define $S = \bar{S}^m \dots \bar{S}^1$. Set $a = (\alpha_1, \dots, \alpha_q) = (\bar{k}_1, \dots, \bar{k}_1, \dots, \bar{k}_m, \dots, \bar{k}_m)$, where each \bar{k}_v enters \bar{r}_v times. Then for this a and S

$$\lim_{\lambda \rightarrow \infty} [\text{diag}(\lambda^{\alpha_i}) S G(y/\lambda)]$$

is a finite matrix $\bar{G}(y) = \begin{bmatrix} \bar{G}^1(y) \\ \vdots \\ \bar{G}^m(y) \end{bmatrix}$; if $\sum_{i=1}^q \alpha_i = \bar{\alpha}$, then CLDR property

holds, if $\sum_{i=1}^q \alpha_i < \bar{\alpha}$ then the limit matrix has deficient rank. ■

Proof of Theorem 4.1. Consider $y_T^* = \lambda_T y_T$ and the quadratic form similar to (21)

$$W(y_T^*, g_{\bar{\theta}}, \lambda_T, A \hat{V}_T A') = \lambda_T^2 g_{\bar{\theta}}'(y_T^*/\lambda_T) [G_{\bar{\theta}}(y_T^*/\lambda_T) A \hat{V}_T A' G_{\bar{\theta}}'(y_T^*/\lambda_T)]^{-1} g_{\bar{\theta}}(y_T^*/\lambda_T).$$

From Assumption 2.3 if $\lambda = \lambda_T$ and $\theta = \hat{\theta}_T$ then the probability limit of corresponding $V^{-\frac{1}{2}} A^{-1} y_T^*$ is Z with distribution $Q(\bar{\theta})$; from Assumption 2.4 $\hat{V}_T = V + o_p(1)$. From (20) and convergence it follows that

$$\text{diag}(\lambda_T^{\alpha_i}) S G_{\bar{\theta}}(y_T^*/\lambda_T) = \bar{G}_{\bar{\theta}}(y_T^*) + O_p(1/\lambda_T); \quad (32)$$

$$\text{diag}(\lambda_T^{\alpha_i}) S \lambda_T g_{\bar{\theta}}(y_T^*/\lambda_T) = \bar{g}_{\bar{\theta}}(y_T^*) + O_p(1/\lambda_T).$$

Then $W(y_T^*, g_{\bar{\theta}}, \lambda_T, A\hat{V}_T A') =$

$$\begin{aligned} & [\bar{g}_{\bar{\theta}}(y_T^*) + O_p(1/\lambda_T)]' \left\{ [\bar{G}_{\bar{\theta}}(y_T^*) + O_p(1/\lambda_T)] A[V + o_p(1)] A' [\bar{G}_{\bar{\theta}}(y_T^*) + O_p(1/\lambda_T)]' \right\}^{-1} \\ & \times [\bar{g}_{\bar{\theta}}(y_T^*) + O_p(1/\lambda_T)]. \end{aligned}$$

(a) If CLDR holds then W_T by continuity of the determinants of polynomials and polynomial matrices converges to

$$[\bar{g}_{\bar{\theta}}(Z)]' \left\{ [\bar{G}_{\bar{\theta}}(Z)] A V A' [\bar{G}_{\bar{\theta}}(Z)]' \right\}^{-1} [\bar{g}_{\bar{\theta}}(Z)];$$

substituting the reparametrized functions for $A = V^{-\frac{1}{2}}$ we get the result.

(b) Follows by continuity of the determinants of polynomial matrices and (24). ■

Proof of Theorem 5.1. Consider the asymptotically equivalent statistic:

$$\begin{aligned} & Z' \bar{G}(Z)' \Lambda [\bar{G}(Z) \bar{G}(Z)']^{-1} \Lambda \bar{G}(Z) Z \\ = & Z' \bar{G}(Z)' (\bar{G}(Z) \bar{G}(Z)')^{-\frac{1}{2}} \\ & \times \left[(\bar{G}(Z) \bar{G}(Z)')^{\frac{1}{2}} \Lambda (\bar{G}(Z) \bar{G}(Z)')^{-\frac{1}{2}} (\bar{G}(Z) \bar{G}(Z)')^{-\frac{1}{2}} \Lambda (\bar{G}(Z) \bar{G}(Z)')^{\frac{1}{2}} \right] \\ & \times (\bar{G}(Z) \bar{G}(Z)')^{-\frac{1}{2}} \bar{G}(Z) Z \\ \leq & \left\| (\bar{G}(Z) \bar{G}(Z)')^{-\frac{1}{2}} \bar{G}(Z) Z \right\|^2 \left\| (\bar{G}(Z) \bar{G}(Z)')^{\frac{1}{2}} \Lambda (\bar{G}(Z) \bar{G}(Z)')^{-\frac{1}{2}} \right\| \\ & \times \left\| (\bar{G}(Z) \bar{G}(Z)')^{-\frac{1}{2}} \Lambda (\bar{G}(Z) \bar{G}(Z)')^{\frac{1}{2}} \right\| \\ \leq & \|\Lambda\|^2 \|Z\|^2 \sim \frac{1}{(1 + i_0)^2} \chi_p^2, \end{aligned}$$

since for similar matrices the eigenvalues are the same, so eigenvalues of $(\bar{G}(Z)\bar{G}(Z)')^{\frac{1}{2}} \Lambda (\bar{G}(Z)\bar{G}(Z)')^{-\frac{1}{2}}$ are the same as for Λ regardless of Z and the norm is given by the largest eigenvalue, and finally,

$$\left\| (\bar{G}(Z)\bar{G}(Z)')^{-\frac{1}{2}} \bar{G}(Z)Z \right\|^2 = \left(Z' \bar{G}(Z)' (\bar{G}(Z)\bar{G}(Z)')^{-1} \bar{G}(Z)Z \right)$$

where for every value of Z the corresponding constant matrix $\bar{G}'(Z) (\bar{G}(Z)\bar{G}(Z)')^{-1} \bar{G}(Z)$ is a projection and thus its norm is always bounded by 1. ■

Proof of Lemma 5.1. Express the p.d.f. of $\chi_{p_1}^2/\alpha_1$:

$$p.d.f. \chi_{p_1}^2/\alpha_1(y) = \frac{\alpha_1}{2^{p_1/2} \Gamma\left(\frac{p_1}{2}\right)} \exp(-\alpha_1 y/2) (\alpha_1 y)^{\frac{p_1}{2}-1},$$

and similarly for $\chi_{p_2}^2/\alpha_2$. The ratio $\frac{p.d.f. \chi_{p_1}^2/\alpha_1(y)}{p.d.f. \chi_{p_2}^2/\alpha_2(y)}$ is

$$2^{\frac{p_2-p_1}{2}} \left(\Gamma\left(\frac{p_2}{2}\right) / \Gamma\left(\frac{p_1}{2}\right) \right) \frac{\alpha_2^{\frac{p_2}{2}-1}}{\alpha_1^{\frac{p_1}{2}-1}} y^{\frac{p_1-p_2}{2}} \exp\left(\frac{y}{2}(\alpha_2 - \alpha_1)\right).$$

Since $\alpha_2 > \alpha_1$ for large enough y this expression is larger than 1. ■

Proof of Lemma 6.1. First consider $\hat{P}(i_1, \dots, i_p, \bar{\theta})$ as defined in (30). By polynomial structure and the convergence rate in Assumption 2.3. $\hat{P}(i_1, \dots, i_p, \bar{\theta}) = P(i_1, \dots, i_p, \bar{\theta}) + O_p(\lambda^{-1})$. Two consequence are (a) from Assumption 2.4 then $\hat{P}(i_1, \dots, i_p, \bar{\theta}, V) - \bar{P}(i_1, \dots, i_p, \bar{\theta}, V) \rightarrow_p 0$; (b) when $P(i_1, \dots, i_p, \bar{\theta}) = 0$, $\Pr(\hat{P}(i_1, \dots, i_p, \bar{\theta}) = 0) \rightarrow 1$ by construction (30). Since $\hat{k}_P - 1$ can be defined as the highest order of polynomial with $\hat{P}(i_1, \dots, i_p, \bar{\theta}) = 0$ it follows

$\Pr(\hat{k}_P = k_P) \rightarrow 1$; note that \hat{k}_P for a polynomial constructed for $A = I$ is the same as for any non-singular A . ■

Proof of Lemma 6.2. Apply Lemma 6.1 to each of the estimated polynomials to determine with probability approaching 1 the lowest order k_P of the non-zero homogeneous polynomial and to obtain the consistent estimators of the polynomial vector functions, $\bar{G}(\cdot)$. Substituting the limit Z provides the consistent estimator of the asymptotic distribution. ■

Proof of Theorem 6.1. The proof follows by application of Lemma 1 to each polynomial that is estimated. ■

References

- [1] Andrews, D. W. K. (1987), ‘Asymptotic results for generalized Wald tests’, *Econometric Theory* 3, 348–358.
- [2] Boudjellaba, H., Dufour, J.-M. and Roy, R. (1992), ‘Testing causality between two vectors in multivariate ARMA models’, *Journal of the American Statistical Association* 87(420), 1082–1090.
- [3] Boudjellaba, H., Dufour, J.-M. and Roy, R. (1994), ‘Simplified conditions for non-causality between two vectors in multivariate ARMA models’, *Journal of Econometrics* 63, 271–287.

- [4] Breusch, T. S. and Schmidt, P. (1988), ‘Alternative forms of the Wald test: How long is a piece of string?’, *Communications in Statistics, Theory and Methods* 17, 2789–2795.
- [5] Dagenais, M. G. and Dufour, J.-M. (1991), ‘Invariance, nonlinear models and asymptotic tests’, *Econometrica* 59, 1601–1615.
- [6] Dufour, J.-M. (1997), ‘Some impossibility theorems in econometrics, with applications to structural and dynamic models’, *Econometrica* 65, 1365–1389.
- [7] Dufour, J.-M. (2003), ‘Identification, weak instruments and statistical inference in econometrics’, *Canadian Journal of Economics* 36(4), 767–808.
- [8] Dufour, J.-M. (2005), ‘Monte Carlo tests with nuisance parameters: A general approach to finite sample inference and nonstandard asymptotics in econometrics’, *Journal of Econometrics* forthcoming.
- [9] Dufour, J.-M., Pelletier, D. and Renault, É. (2006), ‘Short run and long run causality in time series: Inference’, *Journal of Econometrics*, 132, 337–362.
- [10] Dufour, J.-M. and Renault, E. (1998), ‘Short-run and long-run causality in time series: Theory’, *Econometrica* 66, 1099–1125.

- [11] Galbraith, J.W. and Zinde-Walsh, V. (1997), ‘On some simple, autoregression-based estimation and identification techniques for ARMA models’, *Biometrika* 84, 685–696.
- [12] Glonek, G. F. V. (1993). On the behaviour of Wald statistics for the disjunction of two regular hypotheses. *J. Roy. Statist. Soc. Ser. B* 55 749–755.
- [13] Gouriéroux, C. and J. Jasiak (2013) ”Size Distortion in the Analysis of Volatility and Covolatility Effects” in *Advances in Intelligent Systems and Computing*, 200, ”Uncertainty Analysis in Econometrics with Applications”, Huynh, V.N.; Kreinovich, V., Sriboonchita, S., and Suriya, K. (ed), Springer, p91-118.
- [14] Gouriéroux, C., Monfort, A. and Renault, É. (1988), *Contraintes bilinéaires: estimation et tests*, in P. Champsaur, M. Deleau, J.-M. Grandmont, C. Henry, J.-J. Laffont, G. Laroque, J. Mairesse, A. Monfort and Y. Yonès, eds, ‘*Mélanges économiques. Essais en l’honneur de Edmond Malinvaud*’, Economica, Paris.
- [15] Gouriéroux, C., Monfort, A. and Renault, E. (1989), ‘Testing for common roots’, *Econometrica* 57, 171–185.
- [16] Gouriéroux, C., Monfort, A. and Renault, E. (1993), ‘Tests sur le noyau, l’image et le rang de la matrice des coefficients d’un modèle linéaire multivarié’, *Annales d’Économie et de Statistique* 11, 81–111.

- [17] Gregory, A. and Veall, M. (1985), ‘Formulating Wald tests of nonlinear restrictions’, *Econometrica* 53, 1465–1468.
- [18] Lütkepohl, H. and Burda, M.M. (1997), ‘Modified Wald tests under non-regular conditions’, *Journal of Econometrics* 78, 315–332.
- [19] Phillips, P. C. B. and Park, J. Y. (1988), ‘On the formulation of Wald tests of nonlinear restrictions’, *Econometrica* 56, 1065–1083.
- [20] Sargan, J. D. (1980), ‘Some tests of dynamic specification for a single equation’, *Econometrica* 48(4), 879–897.