Wald tests when restrictions are locally singular *

Jean-Marie Dufour † McGill University

Eric Renault ‡ Brown University

Victoria Zinde-Walsh § McGill University

First version: September 2016
Revised: February 2017, November 2017
This version: November 2017
Compiled: December 21, 2017, 23:13

*The authors thank Tianyu He, Sid Kankanala, Vinh Nguyen, Masaya Takano, Purevdorj Tuvaandorj, and numerous seminar participants for their comments. This work was supported by the William Dow Chair in Political Economy (McGill University), the Bank of Canada (Research Fellowship), the Toulouse School of Economics (Pierre-de-Fermat Chair of excellence), the Universitad Carlos III de Madrid (Banco Santander de Madrid Chair of excellence), a Guggenheim Fellowship, a Konrad-Adenauer Fellowship (Alexander-von-Humboldt Foundation, Germany), the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], 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).

† William Dow Professor of Economics, McGill University, Centre interuniversitaire de recherche en analyse des organisations (CIRANO), and Centre interuniversitaire de recherche en économie quantitative (CIREQ). Mailing address: Department of Economics, McGill University, Leacock Building, Room 414, 855 Sherbrooke Street West, Montréal, Québec H3A 2T7, Canada. TEL: (1) 514 398 6071; FAX: (1) 514 398 4800; e-mail: `jean-marie.dufour@mcgill.ca`. Web page: `http://www.jeanmariedufour.com`

‡C.V. Starr Professor of Economics, Brown University. Mailing address: Department of Economics, Brown University, Box B, 64 Waterman Street, Providence, RI 02912, USA. TEL: (1) 401 863-3519; e-mail: `eric_renault@brown.edu`. Web page: `https://sites.google.com/a/alumni.brown.edu/eric-renault-personal-website`

§ Department of Economics, McGill University, and Centre interuniversitaire de recherche en économie quantitative (CIREQ). Mailing address: Department of Economics, McGill University, Leacock Building, Room 414, 855 Sherbrooke Street West, Montréal, Québec H3A 2T7, Canada. TEL: (1) 514-398-4400 Ext 00782; FAX: (1) 514 398 4800; e-mail: `victoria.zinde-walsh@mcgill.ca`. Web page: `https://www.mcgill.ca/economics/victoria-zinde-walsh`

ABSTRACT

This paper provides an exhaustive characterization of the asymptotic null distribution of Wald-type statistics for testing restrictions given by polynomial functions – which may involve asymptotic singularities – when the limiting distribution of the parameter estimators is absolutely continuous (*e.g.*, Gaussian). In addition to the well-known finite-sample non-invariance, there is also an asymptotic non-invariance (non-pivotality): with standard critical values, the test may either under-reject or over-reject, and may even diverge under the null hypothesis. The asymptotic distributions of the test statistic can vary under the null hypothesis and depends on the true unknown parameter value. All these situations are possible in testing restrictions which arise in the statistical and econometric literatures, *e.g.* for examining the specification of ARMA models, causality at different horizons, indirect effects, zero determinant hypotheses on matrices of coefficients, and many other situations when singularity in the restrictions cannot be excluded. We provide the limit distribution and general bounds for the single restriction case. For multiple restrictions, we give a necessary and sufficient condition for the existence of a limiting distribution and the form of the limit distribution whenever it exists.

Key words: nonlinear restriction; deficient rank; singular covariance matrix; Wald test**;** nonstandard asymptotic theory; bound.
Journal of Economic Literature classification: C3.
AMS 2000 Subject classification: 60E05; 62H10.

Contents

List of Tables

List of Definitions, Assumptions, Propositions and Theorems

1. Introduction

We consider the problem of testing q nonlinear restrictions on a parameter vector $\theta = (\theta_1, \ldots, \theta_p)' \in \Theta \subseteq \mathbb{R}^p$:

$$\mathscr{H}_0 : g(\theta) = 0 \tag{1.1}$$

where $g(\theta) = [g_1(\theta), \ldots, g_q(\theta)]'$ is a $q \times 1$ vector of polynomial functions and $q \leq p$. We denote by $\bar{\theta}$ the "true" parameter vector, so $g(\bar{\theta}) = 0$ under $\mathscr{H}_0$. Each polynomial $g_l(\theta)$ has order m_l in the components of θ:

$$g_l(\theta) = \sum_{i=0}^{m_l} \Big\{ \sum_{j_1+\cdots+j_p=i} c_l(j_1, \ldots, j_p) \prod_{k=1}^{p} \theta_k^{j_k} \Big\}, \quad l = 1, \ldots, q, \tag{1.2}$$

where $\sum_{j_1+\cdots+j_p=i}$ represents the sum over all the distinct sets $\{j_1, \ldots, j_p\}$ such that $j_1, \ldots, j_p$ are nonnegative integers and $j_1 + \cdots + j_p = i$ (with the convention $\theta^0 = 1$). $\mathscr{H}_0$ defines an algebraic variety $\Theta_0 \subseteq \mathbb{R}^p$. Further, we suppose that a consistent (typically asymptotically normal) estimator $\hat{\theta}_T$ of $\bar{\theta}$ is available (as $T \to \infty$), so it is natural to test $\mathscr{H}_0$ by using a Wald-type test statistic. On the other hand, a completely specified model (like a likelihood function) may not be available, so other types of tests – such as likelihood ratio (LR) or score-type tests – may not be applicable.

Many statistical problems lead one to consider tests of polynomial restrictions:

1. "collapsibility" and unions of conditional independence hypotheses in contingency tables [Simpson (1951), Whittemore (1978), Ducharme and Lepage (1986), Davis (1989), Glonek (1993)];

2. tests of dynamic specification in time series [Sargan (1980), Gouriéroux, Monfort and Renault (1989), Galbraith and Zinde-Walsh (1997)];

3. tests for indirect effects and "mediation" analysis [Sobel (1982, 1986), Baron and Kenny (1986), MacKinnon, Lockwood, Hoffman, West and Sheets (2002), MacKinnon, Lockwood and Williams (2006), MacKinnon (2008), Emsley, Dunn and White (2010), Hayes (2013), Steyer, Mayer and Fiege (2014), VanderWeele (2015), Koo, Leite and Algina (2016), Tofighi and MacKinnon (2016)];

4. tests on matrices of coefficients, *e.g.* for the rank (including singularity), the kernel or the image of such matrices [Gouriéroux, Monfort and Renault (1990, 1993), Robin and Smith (2000), Al-Sadoon (2017)];

5. collapsibility and unconfoundness hypotheses in regression models [Clogg, Petkova and Shihadeh (1992), Clogg, Petkova and Haritou (1995), Alison (1995), Steyer, Gabler and Rucai (1996), Gaffke, Steyer and von Davier (1999), Gaffke, Heiligers and Offinger (2002)];

6. tests of Granger noncausality restrictions in VARMA models [Boudjellaba, Dufour and Roy (1992, 1994)];

7. noncausality at multiple horizons in multivariate time series [Lütkepohl and Burda (1997), Dufour and Renault (1998), Dufour, Pelletier and Renault (2006)];

8. tests on tetrads and other nonlinear hypotheses in graphical causal modelling and factor analysis [Bollen and Ting (2000), Spirtes, Glymour and Scheines (2000), Hipp and Bollen (2003), Silva, Scheines, Glymour and Spirtes (2006), Drton, Sturmfels and Sullivant (2007, 2009), Johnson and Bodner (2007), Drton, Massam and Olkin (2008), Bollen, Lennox and Dahly (2009), Sullivant, Talaska and Draisma (2010), Zwiernik and Smith (2012)];

9. tests on volatility and covolatility in financial time series [Gouriéroux and Jasiak (2013)].

Under standard regularity conditions, the asymptotic distributions of the classical test statistics, such as the likelihood ratio and the Wald-type statistic are χ^2_q, and there is asymptotic (local) equivalence between these tests. In many cases, Wald-type tests are relatively convenient because they allow one to test a wide spectrum of null hypotheses using a single asymptotically normal estimator $\hat{\theta}_T$. This feature may be important when the likelihood function is not available [or is difficult to maximize under the relevant restrictions]. The same remark holds in models estimated by pseudo-likelihood, estimating functions, or generalized method-of-moments (GMM) methods.

Even though Wald-type tests are not generally invariant to equivalent reformulations of the null hypothesis and reparameterizations [Gregory and Veall (1985), Breusch and Schmidt (1988), Phillips and Park (1988), Dagenais and Dufour (1991, 1994), Dufour and Dagenais (1992), Critchley, Marriott and Salmon (1996), Dufour, Trognon and Tuvaandorj (2017)] and tend to be strongly affected by identification problems [Dufour (1997, 2003)], their convenience makes them difficult to avoid in many circumstances. Since Wald-type tests depend crucially on the parameterization considered (which may reflect parameters of interest from a subject-specific viewpoint, such as economic theory), their power also depends on the parameterization, which allows one to achieve relatively high power in the "directions" associated with parameters of interest. In the case of single restrictions $(q = 1)$, t-type statistics [obtained through the division of a parameter estimate by a "standard error"] may be interpreted as "signed" Wald-type statistics and easily yield one-sided tests: such tests explicitly aim at increasing power in a specific direction.

The standard regularity conditions fail when there are singularities in the algebraic structure of the restrictions. These are characterized by rank deficiency of the Jacobian matrix of the restrictions. For singularities of the type we consider here in Wald-type statistic, Drton (2009) examined likelihood ratio tests using the tools of algebraic statistics.[1] In particular, real algebraic varieties and their tangent cones play a crucial role in describing the asymptotic distribution of the LR statistic. The set Θ_0 in $\mathbb{R}^p$ where several real polynomial functions are zero is called a real algebraic variety, the tangent cone in this case is fully determined by the Jacobian; the rank of the Jacobian matrix at a point determines the dimension of the tangent cone. If it is of full rank at a point $\bar{\theta} \in \Theta_0$, the dimension of the tangent cone is $\dim(\Theta_0)$ and there is no singularity at $\bar{\theta}$. If however there is a rank deficiency at $\bar{\theta}$, the tangent cone at that point has dimension lower than $\dim(\Theta_0)$ and this is a singular point of the real algebraic variety. The limit distributions associated with restrictions (and

[1] For other work on LR statistics in such nonregular contexts, see Chernoff (1954), Ritz and Skovgaard (2005), Azaïs, Gassiat and Mercadier (2006), Kato and Kuriki (2013).

algebraic varieties) involving such singularities are no longer pivotal and take different forms depending on whether the parameters at which they are evaluated define a regular or a singular point. Moreover, the asymptotic equivalence between the test statistics no longer holds.

For Wald tests, the fact that the asymptotic distribution can be non-standard was pointed out by Sargan (1980), Andrews (1987) and Glonek (1993) for the problem of testing an hypothesis of the form $\mathscr{H}_0 : \theta_1\theta_2 = 0$, when $\theta_1 = \theta_2 = 0$. Glonek (1993) also showed that usual critical values based on the χ_1^2 distribution are conservative. More recently, Drton and Xiao (2016) and Pillai and Meng (2016) studied the hypothesis $\mathscr{H}_0 : \theta_1^{\nu_1}\theta_2^{\nu_2}\cdots\theta_p^{\nu_p} = 0$, and the distribution of the corresponding Wald-type statistic in the special case where $\theta_1 = \theta_2 = \cdots = \theta_p = 0$ and $\nu_1, \ldots, \nu_p$ are positive integers. Note the null hypothesis $\mathscr{H}_0$ holds whenever $\bar{\theta}_{i_1} = \cdots = \bar{\theta}_{i_k} = 0$ for some subset $\{i_1, \ldots, i_k\}$ of $\{1, \ldots, p\}$, each of which may entail a different limiting distribution under $\mathscr{H}_0$.

This paper establishes a full characterization of the limit distribution of the Wald-type statistic for restrictions defined by polynomials in several variables, when $T^{\frac{1}{2}}(\hat{\theta}_T - \bar{\theta})$ converges to some asymptotic distribution. For monomial restrictions such as $\mathscr{H}_0 : \theta_1^{\nu_1}\theta_2^{\nu_2}\cdots\theta_p^{\nu_p} = 0$, our results allow for cases where only a subset of the elements of $(\theta_1, \ldots, \theta_p)'$ are zero. We also derive stochastic dominance results and bounds on critical values. After emphasizing that "anything can happen" (underrejection or overrejection when regular chi-square critical values are used, or even divergence) – even when $T^{\frac{1}{2}}(\hat{\theta}_T - \bar{\theta})$ is asymptotically Gaussian – we study in turn the case of a single polynomial restriction, and then several restrictions of this type. The fact that these two types of situations lead to qualitatively different results is also underscored.

For the case of a single restriction involving only one term or a quadratic form, Drton and Xiao (2016) provided the form of the limit distribution at a singular point and a generic bound on the distribution. We consider here general polynomials – involving several terms and variables – and provide a detailed form of the limit distribution which explicitly reflects the degree of singularity at any given singular point, together with a more precise bound at such a point. The issue of the non-existence of a unique asymptotic distribution is underscored, while the existence of the uniform bound allows to control test level asymptotically (leading to a possibly conservative test). We show that standard χ_p^2 critical values are uniformly conservative, provided the number of parameters is not too large. Our results also entail that divergence does not occur when only one restriction is tested. Special tighter bounds applicable when the function can be expressed as a product polynomials are also derived. We also observe that some of these bounds remain valid even when the asymptotic covariance matrix of model parameters is not fully known: for so-called "diagonal Wald-type statistics" where some covariances between parameter estimates are (arbitrarily) set to zero, we show that the null distribution of statistics for products of polynomials can be bounded in a surprisingly tight way.

The case of several polynomial restrictions is also fully characterized. The dependence of the limiting distribution of the Wald-type statistic on the unknown true parameter value is even more crucial here. We show that, even under $\mathscr{H}_0$, the statistic can diverge to $+\infty$ at a singular point, so no uniform generic bound exists for the Wald-type statistic. Theoretically, the restrictions could be examined to verify whether at some singularity point divergence could actually occur. In any but the most trivial cases of nonlinearity, this is a difficult and cumbersome undertaking. These considerations lead us to conclude that application of the Wald test to any multiple nonlinear restrictions is quite problematic for level control. We thus propose that in such cases one replace the test of multiple restrictions by testing a single restriction that provides an identical algebraic variety. Our results extend those of Gaffke et al. (1999) and Gaffke et al. (2002), which are based on rank assumptions on the Jacobian or Hessian matrices of the restrictions (in the context of testing unconfoundedness) and exclude the possibility that higher-order terms could play a role.

The paper is organized as follows. Section 2 describes the framework considered to study the asymptotic distribution of Wald-type statistics with local singularities. Simple examples illustrating the types of problems which can arise in this context are also presented. Section 3 provides limit results for the case of a single restriction. Section 4 derives bounds for the case of a single restriction. Sections 5 and 6 are devoted to several restrictions: in Section 5, we characterize the limit distribution when it exists, while in Section 6 we give a necessary and sufficient condition for the existence of the limit distribution. Section 7 concludes. The proofs are in the Appendix.

2. Framework

We consider a general probability model where the parameter space is an open subset of $\mathbb{R}^p$. Further, we have consistent parameter estimators whose distribution converges to an absolutely continuous (possibly Gaussian) distribution. Throughout the paper, $\theta = (\theta_1, \ldots, \theta_p)'$ and $\bar{\theta} = (\bar{\theta}_1, \ldots, \bar{\theta}_p)'$ represent $p \times 1$ vectors of fixed real coefficients, $\hat{\theta}_T = (\hat{\theta}_{1T}, \ldots, \hat{\theta}_{pT})'$ is a $p \times 1$ real random vector, and T is an integer such that $T \geq T_0 \geq 1$, and $\xrightarrow[T\to\infty]{\mathrm{p}}$ represents convergence in probability as $T \to \infty$. Absolute continuity is defined with respect to the Lebesgue measure.

Assumption 2.1 ESTIMATOR ASYMPTOTIC DISTRIBUTION. *The sequence $\{\hat{\theta}_T : T \geq T_0\}$ satisfies*

$$T^{\frac{1}{2}}(\hat{\theta}_T - \bar{\theta}) \xrightarrow[T\to\infty]{\mathrm{p}} JZ \tag{2.1}$$

where J is a full-rank $p \times p$ fixed matrix and Z is a $p \times 1$ real random vector. The distributions of Z and $\hat{\theta}_T$ for $T \geq T_0$ are absolutely continuous.

Assumption 2.2 CONVERGENCE OF PARAMETER COVARIANCE ESTIMATOR. *$\{\hat{V}_T\}$ is a sequence of $p \times p$ full-rank random matrices such that*

$$\underset{T\to\infty}{\mathrm{plim}}\, \hat{V}_T = V \text{ where } V = JJ'. \tag{2.2}$$

Assumption 2.1 means that $T^{\frac{1}{2}}(\hat{\theta}_T - \bar{\theta})$ has an asymptotic distribution characterized by a (typically unknown) "scaling matrix" J and the distribution of the random vector Z. In the important special case where $Z \sim \mathrm{N}[0, I_p]$, the asymptotic distribution of $T^{\frac{1}{2}}(\hat{\theta}_T - \bar{\theta})$ is $\mathrm{N}[0, JJ']$. The scaling matrix J then determines the asymptotic covariance matrix of $T^{\frac{1}{2}}(\hat{\theta}_T - \bar{\theta})$. J may represent the square root $V^{1/2}$, a lower triangular matrix (the Cholesky factor of V), or any other appropriate matrix. The form of J may not be identifiable when JZ is Gaussian, but with non-Gaussian distributions it can correspond to non-trivial distributional assumptions. The multiplicative representation

JZ allows one to consider cases where the distribution of Z depends on additional nuisance parameters, including random covariance and location parameters. The mean of Z need not be zero. Various non-Gaussian distributions, such as mixtures of normal distributions and spherically symmetric distributions are allowed.

Assumption 2.2 postulates the existence of a consistent estimate $\hat{V}_T$ of $V = JJ'$. When JJ' is the asymptotic covariance matrix of $T^{\frac{1}{2}}(\hat{\theta}_T - \bar{\theta})$, $\hat{V}_T$ is a consistent parameter covariance estimator. More generally, unless stated otherwise, V represents a general positive definite matrix, which may differ from the asymptotic covariance matrix of $T^{\frac{1}{2}}(\hat{\theta}_T - \bar{\theta})$.

We define the usual Wald-type test statistic:

$$W_T(\hat{\theta}_T; g, \hat{V}_T) = T\, g(\hat{\theta}_T)' \left[G(\hat{\theta}_T)\hat{V}_T G(\hat{\theta}_T)'\right]^{-1} g(\hat{\theta}_T) \tag{2.3}$$

where $G(\theta) := \frac{\partial g}{\partial \theta'}(\theta)$. We study here situations where the matrix $G(\hat{\theta}_T)\hat{V}_T G(\hat{\theta}_T)'$ is nonsingular in finite samples (with probability one), so the Wald-type statistic is well defined with probability one. For the case of a single restriction, we also consider the corresponding Student-type statistic based on dividing $g(\hat{\theta}_T)$ by the usual asymptotic standard error:

$$t_T(\hat{\theta}_T; g, \hat{V}_T) = \frac{T^{1/2} g(\hat{\theta}_T)}{\left[G(\hat{\theta}_T)\hat{V}_T G(\hat{\theta}_T)'\right]^{1/2}}\,. \tag{2.4}$$

$t_T(\hat{\theta}_T; g, \hat{V}_T)$ allows one to perform one-sided tests, while $W_T(\hat{\theta}_T; g, \hat{V}_T)$ yields two-sided tests. The notations $t_T(\hat{\theta}_T; g, \hat{V}_T)$ and $W_T(\hat{\theta}_T; g, \hat{V}_T)$ underscore the fact that these test statistics depend crucially on three arguments: the function $g(\cdot)$, and the "estimates" $\hat{\theta}_T$ and $\hat{V}_T$. When there is no ambiguity, we may write t_T and W_T instead of $t_T(\hat{\theta}_T; g, \hat{V}_T)$ and $W_T(\hat{\theta}_T; g, \hat{V}_T)$. Note that the two statistics $W_T(\hat{\theta}_T; g, \hat{V}_T)$ and $t_T(\hat{\theta}_T; g, \hat{V}_T)$ are invariant to multiplication of $g(\theta)$ by a nonsingular fixed matrix A: the test statistics remain the same if we consider the hypothesis $\mathscr{H}_0 : Ag(\theta) = 0$.

Assumptions 2.1 - 2.2 allow one to consider cases where $\hat{V}_T$ does not converge to the asymptotic covariance matrix of $T^{\frac{1}{2}}(\hat{\theta}_T - \bar{\theta})$. Indeed, $T^{\frac{1}{2}}(\hat{\theta}_T - \bar{\theta})$ may not even possess an asymptotic variance. Since Z is not restricted to follow the $\mathrm{N}(0, I_p)$ distribution, Z can be redefined in a way that allows V to differ from the asymptotic covariance matrix. For example, if $T^{\frac{1}{2}}(\hat{\theta}_T - \bar{\theta}) \xrightarrow[T\to\infty]{\mathrm{p}} U$ where $U \sim \mathrm{N}[0, \Sigma_0]$ and Σ_0 is nonsingular, we can can define $Z = J^{-1}U$ where $V = JJ'$. This feature can be useful to allow for alternative variants of the Wald-type statistic whose distribution may be more easily established or bounded. In particular, in Section 4, we will observe that bounds obtain in important cases where the asymptotic covariance matrix of model parameters is not fully known: for *diagonal Wald-type statistics* where some covariances between parameter estimates are (wrongly) set to zero, the null distribution of statistics for testing products of polynomials can be bounded in a remarkably tight way.

Consider now a $q \times p$ matrix $F(x)$ whose elements are polynomial functions of $x \in \mathbb{R}^p$ [a *matrix of polynomials* or a *polynomial matrix* in x]. We say that a square polynomial matrix $F(x)$ is nonsingular if its determinant is non-zero *a.e.* (in $\mathbb{R}^p$). More generally, we say that $F(x)$ has *full row rank* [or *full rank*] if $F(x)F(x)'$ is nonsingular *a.e.* We define the rank of the $q \times p$ matrix $F(x)$ as the largest dimension of a square nonsingular submatrix.

The following properties will play an important role in the rest of this paper:

1. a polynomial function is either identically zero or different from zero *a.e.* [see Caron and Traynor (2005), Mityagin (2015)];

2. since $\Delta(x) := \det[F(x)F(x)']$ is a polynomial function, it is non-zero *a.e.* as long as it is distinct from zero at one point x; so, if $\Delta(x) \neq 0$ for some x, $F(x)F(x)'$ is nonsingular *a.e.*;

3. the determinants of square submatrices of $F(x)$ are either zero everywhere or non-zero *a.e.*; thus, if the largest nonsingular submatrix of $F(x)$ has dimension $r_F \times r_F$, its determinant is non-zero *a.e.*, and any larger square submatrix have zero determinant everywhere: the rank of $F(x)$ is constant and equal to r_F for almost all $x \in \mathbb{R}^p$.

If a polynomial matrix function has rank q, the matrices formed by the numerical values of the polynomial functions have rank q *a.e.* on $\mathbb{R}^p$. But $F(x)$ may not have full rank at points in a set of Lebesgue measure zero. If $F(x)$ has full row rank *a.e.*, its rows must be *linearly independent functions* (of x): for $\lambda \in \mathbb{R}^p$,

$$\lambda' F(x) = 0 \text{ for all } x \in \mathbb{R}^p \Rightarrow \lambda = 0\,. \tag{2.5}$$

In this case, we also say that the rows of $F(x)$ are *linearly independent vectors of polynomials* (or *polynomial vectors*). The converse does not however hold: if the rows of $F(x)$ are linearly independent vectors of polynomials, the rank of $F(x)$ may be less than q for all x.

Assumption 2.3 FULL-RANK JACOBIAN MATRIX. *The $q \times p$ polynomial matrix $\frac{\partial g}{\partial \theta'}(\theta)$ has full rank q* a.e.

Assumptions 2.2 and 2.3 ensure the existence of the Wald-type statistic with probability one. However, it does not preclude the presence of singularities, which are defined as follows.

Definition 2.1 SINGULARITY. *If Assumption 2.3 holds and $g(\bar{\theta}) = 0$, but $\frac{\partial g}{\partial \theta'}(\bar{\theta})$ does not have full rank, we say that $\bar{\theta}$ is a* singularity (*or a* singular point) *of the null hypothesis $\mathscr{H}_0 : g(\theta) = 0$.*

When $Z \sim \mathrm{N}(0, I_p)$, the standard asymptotic χ^2_q distribution holds as long as the matrix $\frac{\partial g}{\partial \theta'}(\bar{\theta})$ has rank q. But this distributional result may not hold when $\bar{\theta}$ is a singular point. We will now discuss a number of simple examples which show that "anything can happen".

Singular points do not occur with linear restrictions, *i.e.* when $g(\theta) = A\theta - a$. If $\frac{\partial g}{\partial \theta} = A$ has full rank q, there is **no rank deficiency** at any θ, hence no singularity. If $rank(A) < q$, then $\frac{\partial g}{\partial \theta}$ has the same reduced rank everywhere, and we can use a generalized inverse [Andrews (1987)], so there is no singularity (as defined above) in this case.

In the examples below, we consider nonlinear restrictions with $T^{\frac{1}{2}}(\hat{\theta}_T - \bar{\theta}) \xrightarrow[T \to \infty]{\mathrm{p}} Z \sim \mathrm{N}[0, V]$. Unless stated otherwise, we take $J = V = \hat{V}_T = I_p$ in these examples. Details on the derivations are available in the Appendix A.

Example 2.1 Asymptotic distribution depends on the form of the restriction. Consider the two (equivalent) null hypotheses: (i) $\theta_1 = 0$, and (ii) $\theta_1^2 = 0$. In case (i), we have $g(\theta) = \theta_1$ and the asymptotic distribution of the Wald-type statistic $W_T(\hat{\theta}_T; g, \hat{V}_T)$ is χ_1^2 under $\mathscr{H}_0$. In case (ii), the asymptotic null distribution of $W_T(\hat{\theta}_T; g, \hat{V}_T)$ is $\frac{1}{4}\chi_1^2$. Thus, using the usual χ_1^2 asymptotic distribution in case (ii) would lead to underrejections under $\mathscr{H}_0$ in large samples (asymptotically conservative tests).

Example 2.2 Non-pivotal but conservative Wald-type statistic. Let $g(\theta) = \theta_1\theta_2$. If either θ_1 or θ_2 is non-zero the limiting distribution is χ_1^2. When $\theta_1 = \theta_2 = 0$,

$$W_T(\hat{\theta}_T; g, \hat{V}_T) \xrightarrow[T\to\infty]{\text{p}} \frac{Z_1^2 Z_2^2}{Z_1^2 + Z_2^2} \,. \tag{2.6}$$

In the latter case, the limit distribution here is not chi-square [Andrews (1987)], but is given by the $\frac{1}{4}\chi_1^2$ distribution [Glonek (1993), Drton and Xiao (2016), Pillai and Meng (2016)]. Thus the limit distribution is *not pivotal* for the null hypothesis $\theta_1\theta_2 = 0$, but the χ_1^2 distribution provides a *uniformly valid bound* asymptotically.

Example 2.3 Five asymptotic distributions. Let $g(\theta) = \theta_1\theta_2\theta_3$ and $\hat{V}_T = V$, where $V = JJ' = [\sigma_{ij}]_{i,j=1.2.3}$ is a general positive definite matrix, $X = JZ$, and $Z \sim \mathrm{N}[0, I_3]$. Under $\mathscr{H}_0$, we have $\bar{\theta} = \left(\bar{\theta}_1, \bar{\theta}_2, \bar{\theta}_3\right)'$ where at least one $\bar{\theta}_i$ is equal to zero. Then, the asymptotic distribution (under $\mathscr{H}_0$) depends on the number of zero coefficients:

$$\begin{array}{lll} W_T(\hat{\theta}_T; g, \hat{V}_T) & \to X_i^2/\sigma_{ii} \sim \chi_1^2 & \text{if } \theta_i = 0 \text{ and } \theta_j \neq 0 \text{ for } j \neq i \\ & \to W_i & \text{if } \theta_i \neq 0 \text{ and } \theta_j = 0 \text{ for } j \neq i\,,\ 1 \leq i \leq 3 \\ & \to W_0 & \text{if } \theta_1 = \theta_2 = \theta_3 \neq 0 \end{array} \tag{2.7}$$

where

$$W_1 = \frac{X_2^2 X_3^2}{\Delta_1}\,, \quad W_2 = \frac{X_1^2 X_3^2}{\Delta_2}\,, \quad W_3 = \frac{X_1^2 X_2^2}{\Delta_3}\,, \quad W_0 = \frac{X_1^2 X_2^2 X_3^2}{\Delta_0}\,, \tag{2.8}$$

$$\Delta_1 = \sigma_{22}X_3^2 + \sigma_{33}X_2^2 + 2\sigma_{23}X_2X_3\,,\ \Delta_2 = \sigma_{11}X_3^2 + \sigma_{33}X_1^2 + 2\sigma_{13}X_1X_3\,, \tag{2.9}$$

$$\Delta_3 = \sigma_{11}X_2^2 + \sigma_{22}X_1^2 + 2\sigma_{12}X_1X_2\,, \quad \Delta_0 = G_0VG_0', \quad G_0 = [X_2X_3,\, X_1X_3,\, X_1X_2]\,. \tag{2.10}$$

In this case, five different asymptotic distributions are possible. This example shows that the asymptotic distribution depends on V. Further, if $V = I_3$, we still have three different asymptotic distributions.

Example 2.4 Oversized test. Consider $g(\theta) = \theta_1^2 + \cdots + \theta_p^2$. Then the limit distribution is $\frac{1}{4}\chi_p^2$. If p is large enough, χ_1^2 critical values lead to overrejections.

Example 2.5 Asymptotic non-equivalence of Wald and LR tests. Consider $g(\theta) = \theta_1^2 + \theta_1^3 - \theta_2^2$. If $\bar{\theta} = 0$, then $g(\bar{\theta}) = 0$ and the limit distribution of the $W_T(\hat{\theta}_T; g, \hat{V}_T)$ statistic is $\frac{1}{4}\chi_1^2$. On the other hand, the asymptotic distribution of the LR statistic is given by the distribution of the minimum of two independent χ_1^2; see Drton (2009, Example 1.1). Thus the two limit distributions are different.

Example 2.6 Divergence under $\mathscr{H}_0$. Suppose that $q = 2$ and $g(\theta) = \left[\theta_1^2,\, \theta_1\theta_2^2\right]'$. Then,

$$W_T(\hat{\theta}_T;\, g,\, \hat{V}_T) = T\,\frac{4\hat{\theta}_1^2 + \hat{\theta}_2^2}{16}. \tag{2.11}$$

(i) If $\bar{\theta}_1 = \bar{\theta}_2 = 0$, the asymptotic distribution is $\frac{1}{4}Z_1^2 + \frac{1}{16}Z_2^2$. This is a linear combination of two independent χ_1^2, bounded by $\frac{1}{4}\chi_2^2$. (ii) However, if $\bar{\theta}_1 = 0$ and $\bar{\theta}_2 \neq 0$, the null hypothesis also holds, but the Wald-type statistic diverges to $+\infty$ as $T \to \infty$.[2] In Section 3, we show that **divergence cannot happen in the case of one restriction.**

3. The case of a single restriction

In this section, we focus on the case where only one restriction is tested ($q = 1$). Then, $\bar{\theta}$ is a singular point if and only if

$$G(\bar{\theta}) = \frac{\partial g}{\partial \theta'}(\bar{\theta}) = 0 \tag{3.1}$$

where $G(\bar{\theta})$ is a $1 \times p$ row vector. It will be useful to reexpress the polynomial $g(\cdot)$ in terms of the difference $\theta - \bar{\theta}$:

$$g(\theta) = \sum_{i=0}^{m} g[\theta - \bar{\theta};\, i,\, \bar{\theta}] \tag{3.2}$$

where

$$g[x;\, i,\, \bar{\theta}] = \sum_{j_1+\cdots+j_p=i} c(j_1, \ldots, j_p;\, \bar{\theta}) \prod_{k=1}^{p} x_k^{j_k}, \quad x = (x_1, \ldots, x_p)' \in \mathbb{R}^p. \tag{3.3}$$

By convention, we set $0^0 = 1$.

When $\bar{\theta}$ satisfies the null hypothesis ($\mathscr{H}_0$), we must have:

$$g[x;\, 0,\, \bar{\theta}] = c_0(0, \ldots, 0;\, \bar{\theta}) = 0\,. \tag{3.4}$$

Of course, other coefficients could be zero, and in fact it is possible that all the coefficients $c(j_1, \ldots, j_p; \bar{\theta})$ with $j_1 + \cdots + j_p \leq i$ be zero for some i. If this happens for $i > 1$, then $\frac{\partial g}{\partial \theta'}(\bar{\theta}) = 0$ and $\bar{\theta}$ is a singular point of $g(\theta)$. Let us denote by $s(\bar{\theta})$ the integer that satisfies

$$s(\bar{\theta}) = \min\left\{i : c(j_1, \ldots, j_p;\, \bar{\theta}) \neq 0 \text{ for some } (j_1, \ldots, j_p) \text{ with } \sum_{k=1}^{p} j_k = i\right\}. \tag{3.5}$$

In other words, under $\mathscr{H}_0$, the lowest degree term in the centered form of $g(\theta)$ [in (3.2)] has degree

[2] The polynomials in the restrictions of this example form a Grobner basis of the algebraic variety. This demonstrates that rewriting restrictions in the Grobner basis form does not solve the problem of divergence.

$s(\bar{\theta})$. Note also that

$$s(\bar{\theta}) = \min\{i : g[\theta - \bar{\theta}; i, \bar{\theta}] \neq 0 \text{ for some } \theta\} \geq 1 \tag{3.6}$$

where the inequality is entailed by the full-rank Assumption 2.3. If $s(\bar{\theta}) = 1$, $\bar{\theta}$ is a *regular point*; if $s(\bar{\theta}) > 1$, $\bar{\theta}$ is a *singular point*. Consequently, we call

$$\gamma(\bar{\theta}) := s(\bar{\theta}) - 1 \tag{3.7}$$

the *singularity order* at $\bar{\theta}$. We focus here on cases where the singularity order is larger than zero $[\gamma(\bar{\theta}) > 0]$: $\bar{\theta}$ is a singular point if and only if $\gamma(\bar{\theta}) > 0$.

Let us gather all the polynomials corresponding to this lowest degree, and define

$$\bar{g}(\theta - \bar{\theta}) := g[\theta - \bar{\theta}; s(\bar{\theta}), \bar{\theta}] = \sum_{j_1 + \cdots + j_p = s(\bar{\theta})} c(j_1, \ldots, j_p; \bar{\theta}) \prod_{k=1}^{p} \left(\theta_k - \bar{\theta}_k\right)^{j_k} . \tag{3.8}$$

$s(\bar{\theta})$ and $\bar{g}(\theta)$ depend on the value of $\bar{\theta}$ and may not be the same for different values of $\bar{\theta}$, even when the latter satisfies $\mathscr{H}_0$. The function $g[\theta - \bar{\theta}; s(\bar{\theta}), \bar{\theta}]$ is homogeneous of degree $s(\bar{\theta})$ in $\theta - \bar{\theta}$.

When $g(\bar{\theta}) = 0$, we can write:

$$g(\theta) = \bar{g}(\theta - \bar{\theta}) + \bar{r}(\theta - \bar{\theta}), \tag{3.9}$$

$$\bar{r}(\theta - \bar{\theta}) = \sum_{i=s(\bar{\theta})+1}^{m} g[\theta - \bar{\theta}; i, \bar{\theta}]. \tag{3.10}$$

Set

$$\bar{G}(x) = \frac{\partial g[x; s(\bar{\theta}), \bar{\theta}]}{\partial x'}, \; x \in \mathbb{R}^p. \tag{3.11}$$

Using the Euler formula for homogeneous polynomials of degree $s(\bar{\theta}) \geq 1$, we get the following identity:

$$\bar{g}(x) = \frac{1}{s(\bar{\theta})} \bar{G}(x) x \tag{3.12}$$

where $\bar{G}(x)$ is a $1 \times p$ row vector. Each element of $\bar{G}(x)$ is a homogenous polynomial of degree $s(\bar{\theta}) - 1$ [including possibly zeros]. Note the zero constant function is interpreted as a polynomial of degree zero (like any other constant function) and it is homogeneous of any degree. The main result of this section is the following theorem, where $\|\cdot\|^2$ represents the Euclidean norm (so $\|\cdot\|^2 = xx'$ when x is a row vector).

Theorem 3.1 ASYMPTOTIC DISTRIBUTION OF WALD STATISTICS: ONE RESTRICTION. *Suppose the Assumptions 2.1, 2.2 and 2.3 hold. If $g(\theta)$ is a polynomial function of θ as given in* (1.1) - (1.2) *with $q = 1$, and if the true unknown value $\bar{\theta}$ satisfies $g(\bar{\theta}) = 0$, then the Student-type* *statistic* $t_T(\hat{\theta}_T; g, \hat{V}_T)$ *defined in* (2.4) *converges in probability to*

$$t(\bar{\theta}; g, J) = \frac{1}{1+\gamma(\bar{\theta})} \frac{\bar{G}^*(Z)Z}{\left[\bar{G}^*(Z)\bar{G}^*(Z)'\right]^{1/2}} = \frac{1}{1+\gamma(\bar{\theta})} \frac{\bar{G}^*(Z)Z}{\left\|\bar{G}^*(Z)\right\|} \tag{3.13}$$

and the Wald-type statistic $W_T(\hat{\theta}_T; g, \hat{V}_T)$ *in* (2.3) *converges in probability to*

$$W(\bar{\theta}; g, J) = t(\bar{\theta}; g, J)^2 = \frac{1}{[1+\gamma(\bar{\theta})]^2} \frac{[\bar{G}^*(Z)Z]^2}{\left\|\bar{G}^*(Z)\right\|^2} \tag{3.14}$$

where $\bar{G}^*(Z) = \bar{G}(JZ)J$, *with* $\gamma(\bar{\theta})$ *and* $\bar{G}(\cdot)$ *defined in* (3.7) *and* (3.11).

From Theorem 3.1, it is clear that the asymptotic distributions of $t_T(\hat{\theta}_T; g, \hat{V}_T)$ and $W_T(\hat{\theta}_T; g, \hat{V}_T)$ depend on nuisance parameters: the scaling matrix J, the unknown value of $\bar{\theta}$ [through $s(\bar{\theta})$ and the coefficients $c(j_1, \ldots, j_p; \bar{\theta})$] and the distribution of Z (if it is not specified). These limit distributions are identical to those of the "pseudo test statistics"

$$\bar{t}_T(\hat{\theta}_T; \bar{g}, V) = \frac{T^{1/2}\,\bar{g}(\hat{\theta}_T - \bar{\theta})}{\left[\bar{G}(\hat{\theta}_T - \bar{\theta})V\bar{G}(\hat{\theta}_T - \bar{\theta})'\right]^{1/2}}, \tag{3.15}$$

$$\begin{aligned} \bar{W}_T(\hat{\theta}_T; \bar{g}, V) &= T\,\bar{g}(\hat{\theta}_T - \bar{\theta})' \left[\bar{G}(\hat{\theta}_T - \bar{\theta})V\bar{G}(\hat{\theta}_T - \bar{\theta})'\right]^{-1} g(\hat{\theta}_T - \bar{\theta}) \\ &= T \frac{\bar{g}(\hat{\theta}_T - \bar{\theta})^2}{\bar{G}(\hat{\theta}_T - \bar{\theta})V\bar{G}(\hat{\theta}_T - \bar{\theta})'}, \end{aligned} \tag{3.16}$$

where $V = JJ'$, for testing the "pseudo hypothesis"

$$\bar{\mathscr{H}}_0 : \bar{g}(\theta - \bar{\theta}) = 0 \tag{3.17}$$

instead of $\mathscr{H}_0 : g(\theta) = 0$. Of course, $\bar{t}_T$ and $\bar{W}_T$ cannot be computed in practice, because $\bar{\theta}$ is typically unknown. However, the latter interpretation underscores the dependence of the null distributions on the unknown true parameter value $\bar{\theta}$. Note also that the components of $\bar{G}(x)$ and $\bar{G}^*(x)$ are homogeneous of degree $\gamma(\bar{\theta})$. In regular cases, we have $\gamma(\bar{\theta}) = 0$, so $\bar{G}^*(Z)$ is a non-zero constant vector. Further, if $Z \sim \mathrm{N}(0, I_p)$, we have: $t(\bar{\theta}; g, V) \sim \mathrm{N}(0, 1)$ and $W(\bar{\theta}; g, V) \sim \chi_1^2$.

If the distribution of Z is symmetric with respect to zero [$Z \sim -Z$] and $s(\bar{\theta})$ is an odd integer [$s(\bar{\theta}) = 1, 3, 5, \ldots$], the distribution of $t(\bar{\theta}; g, V)$ is symmetric around zero. Further, when Z is Gaussian, or more generally, if Z has a spherically symmetric distribution, it is possible to represent $t(\bar{\theta}; g, V)$ and $W(\bar{\theta}; g, V)$ as products of independent variables.

Theorem 3.2 FACTORIZATION OF WALD-TYPE STATISTICS: ONE RESTRICTION. *Under the* *conditions of Theorem 3.1,*

$$t(\bar{\theta};g,J)=\frac{1}{1+\gamma(\bar{\theta})}\frac{\bar{G}^{*}(U)U}{\left\|\bar{G}^{*}(U)\right\|}\left\|Z\right\|, \tag{3.18}$$

$$W(\bar{\theta};g,J)=\frac{1}{[1+\gamma(\bar{\theta})]^{2}}\frac{[\bar{G}^{*}(U)U]^{2}}{\left\|\bar{G}^{*}(U)\right\|^{2}}\left\|Z\right\|^{2}, \tag{3.19}$$

where $U=Z/\left\|Z\right\|$. If Z follows a spherically symmetric distribution, then U and $\left\|Z\right\|$ are independent, and U follows a uniform distribution on the unit sphere in $\mathbb{R}^{p}$.

The factorization result in Theorem 3.2 shows that the limit $t-$type and Wald-type statistics can be represented as a product of three factors: the first one is a function of the singularity order at $\bar{\theta}$, the second one $\bar{G}^{*}(U)U/\left\|\bar{G}^{*}(U)\right\|$ represents the orientation of the random vector Z, and the third one given by the norm $\left\|Z\right\|$ (or its square). If the distribution of Z is spherically symmetric, the factors are independent and U is uniformly distributed on the unit sphere in $\mathbb{R}^{p}$. If $Z\sim N(0,I_{p})$ then $\left\|Z\right\|^{2}\sim\chi_{p}^{2}$, but for non-Gaussian Z the distribution of $\left\|Z\right\|^{2}$ could be fat-tailed (or thin-tailed). This representation may be convenient for simulating the distribution of the limit statistic. In regular cases, $\gamma(\bar{\theta})=0$ and $\bar{G}^{*}(Z)=c'$ is a non-zero row vector of constants, so $\bar{G}^{*}(U)=c'$ and

$$\frac{\bar{G}^{*}(U)U}{\left\|\bar{G}^{*}(U)\right\|}\left\|Z\right\|=\frac{c'Z/\left\|Z\right\|}{\left\|c\right\|}\left\|Z\right\|=\frac{c'Z}{\left\|c\right\|}; \tag{3.20}$$

if $Z\sim N(0,I_{p})$, we have $t(\bar{\theta};g,V)\sim\mathrm{N}(0,\,1)$ and $W(\bar{\theta};g,V)\sim\chi_{1}^{2}$, as expected.

It is important to note that θ may not represent all the parameters of the model, only those involved in the restrictions of interest or (regular) transformations of these. For example, suppose the original parameter vector of the model is a $p_{0}\times1$ vector β, with parameter estimate $\hat{\beta}_{T}$ such that

$$T^{\frac{1}{2}}(\hat{\beta}_{T}-\bar{\beta})\xrightarrow[T\to\infty]{\mathrm{p}}Z_{0}\sim\mathrm{N}[0,\,\Sigma_{0}],\quad\det(\Sigma_{0})\neq0, \tag{3.21}$$

and $\theta=C\beta$ where C is a full-rank $p\times p_{0}$ fixed matrix with $1\leq p<p_{0}$; $\xrightarrow[T\to\infty]{\mathrm{d}}$ means convergence in distribution as $T\to\infty$. We can then take $\hat{\theta}_{T}=C\hat{\beta}_{T}$, $\bar{\theta}=C\bar{\beta}$, and

$$T^{\frac{1}{2}}(\hat{\theta}_{T}-\bar{\theta})\xrightarrow[T\to\infty]{\mathrm{p}}CZ_{0}\sim\mathrm{N}[0,\,V]\text{ where }V=C\Sigma_{0}C'. \tag{3.22}$$

Clearly, p can be much smaller than p_{0}, and p is the relevant degree-of-freedom number to be used in the bound $\chi_{p}^{2}/[1+\gamma(\bar{\theta})]^{2}$.

Similarly, suppose $g(\theta)$ has the form

$$g(\theta)=g_{1}(C_{1}\theta) \tag{3.23}$$

where C_{1} is a $p_{1}\times p$ full-rank matrix ($1\leq p_{1}\leq p$), *e.g.* the subvector $\theta_{1}=C_{1}\theta$ of $\theta=(\theta_{1}',\theta_{2}')'$. In other words, the restrictions can be expressed in terms of the linear parameter transformation

$\theta_1^* = C_1\theta$:

$$\mathscr{H}_{10}^* : g_1(\theta_1^*) = 0\,. \tag{3.24}$$

On setting $\hat{\theta}_{1T}^* = C_1\hat{\theta}_T$ and $\bar{\theta}_1^* = C_1\bar{\theta}$, the t and Wald-type statistics for testing $\mathscr{H}_{10}^*$ are then $t_T(\hat{\theta}_{1T}^*; g_1, C_1\hat{V}_T C_1)$ and $W_T(\hat{\theta}_{1T}^*; g_1, C_1\hat{V}_T C_1)$. Theorem 3.1 entails that this Wald-type statistic converges in probability (under $\mathscr{H}_{10}^*$) to

$$\begin{aligned} t(\bar{\theta}_1^*; g_1, C_1 J) &= \frac{1}{[1+\gamma_1(\bar{\theta}_1^*)]}\frac{\bar{G}_1^*(Z)Z}{\left\|\bar{G}_1^*(Z)\right\|}\,, & (3.25)\\ W(\bar{\theta}_1^*; g_1, C_1 J) &= t(\bar{\theta}_1^*; g_1, C_1 J)^2, & (3.26)\end{aligned}$$

where $\gamma_1(\bar{\theta}_1^*)$ is the singularity order of g_1 at $\bar{\theta}_1^*$ [which only depends on $\bar{\theta}_1^*$], $\bar{G}_1^*(Z) = \bar{G}_1(C_1 JZ)C_1 J$, and

$$\bar{G}_1(x_1) = \frac{\partial g_1[x_1; s(\bar{\theta}_1^*), \bar{\theta}_1^*]}{\partial x_1'}\,,\ x_1 \in \mathbb{R}^{p_1}\,. \tag{3.27}$$

Here the distribution of the test statistic only depends on g_1, the $p_1\times 1$ parameter $\bar{\theta}_1^*$, and the $p_1\times 1$ random vector $C_1 JZ$, rather than the higher-dimensional $p\times 1$ vectors $\bar{\theta}$ and JZ.

Another important invariance case is the one where $g(\theta)$ can be represented as the product of polynomials:

$$g(\theta) = h_1(\theta)\, h_2(\theta) \tag{3.28}$$

where $h_1(\theta)$ and $h_2(\theta)$ are polynomials. Suppose $h_2(\bar{\theta}) = c \neq 0$, *i.e.* the centered polynomial $h_2(\theta)$ expressed as a function of $\theta - \bar{\theta}$ [as in (3.2)] has a non-zero constant term [$s(\bar{\theta}) = 0$]. Consequently, the coefficients of $h_2(\theta)$ only contribute to higher-order terms of the polynomial $\bar{g}(\theta-\bar{\theta})$ [in (3.8)], so the asymptotic null distributions of the $t_T(\hat{\theta}_T; g, \hat{V}_T)$ and $W_T(\hat{\theta}_T; g, \hat{V}_T)$ statistics only depend on h_1:

$$t(\bar{\theta}; g, J) = t(\bar{\theta}; h_1, J)\,, \quad W(\bar{\theta}; g, J) = W(\bar{\theta}; h_1, J)\,. \tag{3.29}$$

If furthermore $h_1(\theta)$ only depends on $\theta_1^* = C_1\theta$ as in (3.23), *i.e.*

$$h_1(\theta) = h_1^*(C_1\theta)\,, \tag{3.30}$$

(3.26) entails that

$$t(\bar{\theta}; g, J) = t(\bar{\theta}_1^*; h_1^*, C_1 J)\,, \quad W(\bar{\theta}; g, J) = W(\bar{\theta}_1^*; h_1^*, C_1 J)\,. \tag{3.31}$$

4. Bounds for Wald tests of a single restriction

Despite the fact that the asymptotic null distributions of $t_T(\hat{\theta}_T; g, \hat{V}_T)$ and $W_T(\hat{\theta}_T; g, \hat{V}_T)$ generally depend on several nuisance parameters, it is of interest to note that these distributions are bounded by nuisance-parameter-free distributions. In this section, we give "universal bounds" which hold for general polynomials irrespective of the singularity order. We also examine cases where the restriction function involves a product of differentiable functions or polynomials. In such cases, the universal bound can be tightened.

4.1. Universal bounds

We first give general bounds applicable to general polynomials of any order.

Theorem 4.1 BOUNDS FOR SINGLE-RESTRICTION WALD STATISTICS. *Under the conditions of Theorem 3.1, the following properties hold.*

(i) *If* $\gamma(\bar{\theta}) = 0$ [i.e., $\bar{\theta}$ *is not a singularity point of* $g(\theta)$],

$$\left|t(\bar{\theta};\, g,\, V)\right| \le \left|\tau(\bar{\theta})Z\right| \le \|Z\|\,, \quad W(\bar{\theta};\, g,\, V) = [\tau(\bar{\theta})Z]^2 \le \|Z\|^2\,, \tag{4.1}$$

where $\tau(\bar{\theta}) = \bar{G}^*(Z)\,/\,\left\|\bar{G}^*(Z)\right\|$ *is a* $p\times 1$ *unit-norm fixed vector; if furthermore* $Z \sim \mathrm{N}(0,\, I_p)$, *we have* $\tau(\bar{\theta})Z \sim \mathrm{N}(0,\, 1)$ *and* $W(\bar{\theta};\, g,\, V) \sim \chi^2_1$.

(ii) *If* $\gamma(\bar{\theta}) \ge 1$,

$$\left|t(\bar{\theta};\, g,\, V)\right| \le \frac{1}{1+\gamma(\bar{\theta})}\|Z\| \le \frac{1}{2}\|Z\| \le \|Z\|\,, \tag{4.2}$$

$$W\left(\bar{\theta}; g, V\right) \le \frac{1}{[1+\gamma(\bar{\theta})]^2}\|Z\|^2 \le \frac{1}{4}\|Z\|^2 \le \|Z\|^2\,. \tag{4.3}$$

(iii) *If* $Z \sim \mathrm{N}(0,\, I_p)$, *we have: for all* $y \ge 0$,

$$\mathbb{P}[\left|t(\bar{\theta};\, g,\, V)\right| > y] \le \max\{\mathbb{P}[|Z_0| > y],\, \mathbb{P}[\chi_p/2 > y]\}\,, \tag{4.4}$$

$$\mathbb{P}[W\left(\bar{\theta}; g, V\right) > y] \le \max\{\mathbb{P}[\chi^2_1 > y],\, \mathbb{P}[\chi^2_p/4 > y]\}\,, \tag{4.5}$$

where $Z_0 \sim \mathrm{N}(0,\, 1)$ *and* $\chi_p \sim (\chi^2_p)^{1/2}$.

In all cases where $T^{\frac{1}{2}}(\hat{\theta}_T - \bar{\theta})$ is asymptotically Gaussian, the asymptotic distribution of the Wald-type statistic $W_T(\hat{\theta}_T;\, g,\, \hat{V}_T)$ is dominated by $\chi^2_p/[1+\gamma(\bar{\theta})]^2$ distribution. In Example 2.4 with $\gamma(\bar{\theta}) = 1$, this bound is sharp. In singular cases $[\gamma(\bar{\theta}) \ge 1]$, the bound $\chi^2_p/4$ is thus applicable irrespective of $\bar{\theta}$ and $\gamma(\bar{\theta})$. When $\bar{\theta}$ may or may not be regular, valid (possibly conservative) asymptotic p-values can be obtained by computing $\mathrm{p}_{\max}[W_T(\hat{\theta}_T;\, g,\, \hat{V}_T)]$ with

$$\mathrm{p}_{\max}[y] = \max\{\mathbb{P}[\chi^2_1 > y],\, \mathbb{P}[\chi^2_p/4 > y]\}\,,\ y \in \mathbb{R}. \tag{4.6}$$

The critical region $\mathrm{p}_{\max}[W_T(\hat{\theta}_T;\, g,\, \hat{V}_T)] \le \alpha$ has level α (or lower) asymptotically, irrespective of the singularity order of $\bar{\theta}$ $(0 < \alpha < 1)$.

Theorem 4.1 shows that, for a given level, standard (regular) critical values can be conservative in non-standard cases, for dominance by the standard distribution is required only in the tail and not everywhere. The following proposition shows that there exists $\bar{\alpha}$ such that $\alpha \le \bar{\alpha}$ and $\mathrm{p}_{\max}[y] \le \alpha$ entails $\mathbb{P}[\chi^2_1 > y] \le \alpha$, for all y. In other words, the χ^2_1 critical value leads to a conservative test

Table 1
Maximum number of parameters $p_{\max}$ in $g(\theta)$
for which χ_1^2 critical values are conservative for scalar Wald statistics

Test level	0.1	0.05	0.025	0.01	0.001
χ_1^2 critical value	2.706	3.841	5.024	6.635	10.828
$p_{\max}$ for conservative test	6	7	9	12	18

at level α. The proposition is somewhat more general in that it derives tail dominance for scaled χ^2 distributions for various degrees of freedom: in addition to examining critical values for Wald tests of one restriction, it can also be applied to Wald statistics for testing several restrictions. We can compare the critical values for χ_q^2/ζ_q and for χ_p^2/ζ_p with $q \leq p$ and $\zeta_p \geq \zeta_q$. Without loss of generality, we take $\zeta_q = 1$ and $\zeta_p = \zeta > 1$.

Proposition 4.2 TAIL CROSSING OF DIFFERENT CHI-SQUARE DISTRIBUTIONS. *Let p and q be two positive integers,* $\zeta > 1$, *and* $0 < \bar{\alpha} < 1$. *If* $\bar{\alpha}$ *is small enough, there always exists* $\bar{y} > 0$ *such that* $\mathbb{P}[\chi_p^2/\zeta > \bar{y}] = \bar{\alpha}$ *and*

$$\mathbb{P}[\chi_q^2 > y] \geq \mathbb{P}[\chi_p^2/\zeta > y] \quad \text{for } y \geq \bar{y}\,. \tag{4.7}$$

The latter proposition entails that, whenever $\zeta > 1$, the χ_p^2/ζ distribution is dominated by the χ_q^2 distribution in the upper tail, irrespective of the values of p and q (even if $q < p$). Consequently, for $y \geq \bar{y}$, $\mathbb{P}[\chi_q^2 > y] \leq \bar{\alpha}$ implies $\mathbb{P}[\chi_p^2/\zeta > y] \leq \bar{\alpha}$. When the test statistic follows a χ_p^2/ζ distribution, critical values based on a non-scaled χ_q^2 distribution can be conservative.

Table 1 shows when χ_1^2 critical values are conservative at different levels. For example, at level 0.05, the usual χ_1^2 critical value is conservative as long as the number of parameters does not exceed 7. In Table 2, we provide critical values for $1 \leq p \leq 20$ at standard significance levels ($\alpha = 0.1$, 0.05, 0.025, 0.01, 0.001). We see from the latter table that the bound grows slowly with the number of parameters.

As observed in (3.21) - (3.26), θ may represent all the parameters in a model, only those involved in the restriction tested. If the latter can be formulated in terms of a relatively small number of parameter transforms, tighter bounds can be achieved. For similar observations on testing nonlinear hypotheses in regression models, see Dufour (1989).

We can also achieve better bounds by taking into account the form of the restriction $g(\theta) = 0$ or by considering special matrices $\hat{V}_T$ or V. For example, this happens if $g(\theta)$ is the product of two polynomials, involving different coefficients. We will now discuss such cases.

4.2. Diagonal Wald statistics and product restrictions

We will now examine cases where $g(\theta)$ is a product of differentiable functions,

$$g(\theta) = h_1(\theta_{1\cdot})\, h_2(\theta_{2\cdot}) \cdots h_n(\theta_{n\cdot}) \tag{4.8}$$

Table 2
Universal bound critical values for testing a single restriction $g(\theta)$
Based on the uniform bound $\max\{\chi_1^2, \chi_p^2/4\}$

Test level	.1	.05	.025	.01	.001
$p \le 6$	2.706	3.841	5.024	6.635	10.828
$p = 7$	**3.004**	3.841	5.024	6.635	10.828
$p = 8$	**3.341**	**3.877**	5.024	6.635	10.828
$p = 9$	**3.671**	**4.230**	5.024	6.635	10.828
$p = 10$	**3.997**	**4.577**	**5.121**	6.635	10.828
$p = 11$	**4.319**	**4.919**	**5.480**	6.635	10.828
$p = 12$	**4.637**	**5.256**	**5.834**	**6.554**	10.828
$p = 13$	**4.953**	**5.591**	**6.184**	**6.992**	10.828
$p = 14$	**5.266**	**5.921**	**6.530**	**7.285**	10.828
$p = 15$	**5.577**	**6.249**	**6.872**	**7.645**	10.828
$p = 16$	**5.886**	**6.574**	**7.211**	**8.000**	10.828
$p = 17$	**6.192**	**6.897**	**7.548**	**8.352**	10.828
$p = 18$	**6.497**	**7.217**	**7.882**	**8.701**	10.828
$p = 19$	**6.801**	**7.536**	**8.213**	**9.048**	**10.955**
$p = 20$	**7.103**	**7.852**	**8.543**	**9.392**	**11.329**

with $\theta = \left(\theta_{1\cdot}', \theta_{2\cdot}', \ldots, \theta_{n\cdot}'\right)'$ and $\bar{\theta} = (\bar{\theta}_{1\cdot}', \bar{\theta}_{2\cdot}', \ldots, \bar{\theta}_{n\cdot}')'$, where $\theta_{i\cdot}$ and $\bar{\theta}_{i\cdot}$ are $p_i \times 1$ vectors, and $p_1 + p_2 + \cdots + p_n = p$. Clearly, $g(\theta) = 0$ if and only if at least one of the functions satisfies $h_i(\theta_{i\cdot}) = 0$. We have the estimator $\hat{\theta}_T = (\hat{\theta}_{1\cdot T}', \hat{\theta}_{2\cdot T}', \ldots, \hat{\theta}_{n\cdot T}')'$, where $\hat{\theta}_{i\cdot T}$ is a $p_i \times 1$ random vector, and a "covariance matrix estimator" $\tilde{\Sigma}_T$ for $\hat{\theta}_T$. Further, we focus on Wald-type statistics where $\tilde{\Sigma}_T$ is restricted to be block diagonal. This fits naturally the case where $\hat{\theta}_{1\cdot T}, \ldots, \hat{\theta}_{n\cdot T}$ are asymptotically uncorrelated. However, we will observe below that the asymptotic distribution of the Wald-type statistic can be bounded even if $\hat{\theta}_{1\cdot T}, \ldots, \hat{\theta}_{n\cdot T}$ are asymptotically correlated, possibly with unknown covariances across $\hat{\theta}_{1\cdot T}, \ldots, \hat{\theta}_{n\cdot T}$.

We consider Wald-type statistics of the form

$$W_T(\hat{\theta}_T; g, \tilde{\Sigma}_T) = T\, g(\hat{\theta}_T)' \left[G(\hat{\theta}_T)\tilde{\Sigma}_T G(\hat{\theta}_T)'\right]^{-1} g(\hat{\theta}_T) = T\, W(\hat{\theta}_T; g, \tilde{\Sigma}_T) \tag{4.9}$$

where $\hat{\theta}_T = (\hat{\theta}_{1\cdot T}', \hat{\theta}_{2\cdot T}', \ldots, \hat{\theta}_{n\cdot}')'$, $\hat{\theta}_{i\cdot T}$ is a $p_i \times 1$ random vector,

$$G(\theta) = \frac{\partial g}{\partial \theta'} = \left[G_1(\theta), G_2(\theta), \ldots, G_n(\theta)\right], \; G_i(\theta) = \frac{\partial g}{\partial \theta_{i\cdot}'}\,, \; i = 1, \ldots, n, \tag{4.10}$$

$$W(x; g, \tilde{\Sigma}_T) \equiv \frac{g(x)^2}{G(x)\, \tilde{\Sigma}_T\, G(x)'}\,, \; x \in \mathbb{R}^p, \tag{4.11}$$

$$\tilde{\Sigma}_T = \mathrm{diag}[\tilde{\Sigma}_{1T}, \tilde{\Sigma}_{2T}, \ldots, \tilde{\Sigma}_{nT}] \tag{4.12}$$

is a block-diagonal matrix, and each $\tilde{\Sigma}_{iT}$ is a (possibly random) $p_i \times p_i$ positive semidefinite matrix. For example, $\tilde{\Sigma}_{iT}$ may be a consistent estimator of the asymptotic covariance matrix of $T^{\frac{1}{2}}(\hat{\theta}_{i\cdot T} - \bar{\theta}_i)$. We call Wald statistics based on such block diagonal covariance estimates "diagonal Wald statistics".

Set

$$H_i(x_i) = \frac{\partial h_i}{\partial x_i'}(x_i)\,, \quad \Delta_i(x_i) = H_i(x_i)\,\tilde{\Sigma}_{iT}\,H_i(x_i)'\,,\; x_i \in \mathbb{R}^{p_i},\; i = 1, 2, \ldots, n. \tag{4.13}$$

We can then show the following stochastic dominance property, which holds both in finite samples and asymptotically.

Proposition 4.3 DOMINANCE PROPERTY FOR PRODUCT RESTRICTIONS. *Suppose $g(\theta)$ satisfies (4.8) along with Assumption 2.3, $Y := (Y_1', Y_2', \ldots, Y_n')'$ a random vector where each Y_i has dimension $p_i \times 1$, with $G_i(\cdot)$, $\tilde{\Sigma}_{iT}$, $W(x;\, g,\, \tilde{\Sigma}_T)$ and $\Delta_i := \Delta_i(Y_i)$ defined as in (4.9) - (4.13), $i = 1, \ldots, n$. Then:*

(i) *conditional on $\Delta_i > 0$,*

$$W(Y;\, g,\, \tilde{\Sigma}_T) \leq W(Y_i;\, h_i,\, \tilde{\Sigma}_{iT}) \quad \textit{for any } i \in \{1, \ldots, n\}\,; \tag{4.14}$$

(ii) *conditional on $g(Y) \neq 0$ and $\Delta_i > 0$ for all $i \in \{1, \ldots, n\}$,*

$$W(Y;\, g,\, \tilde{\Sigma}_T) \leq \min_{1 \leq i \leq n} W(Y_i;\, h_i,\, \tilde{\Sigma}_{iT})\,, \tag{4.15}$$

$$\begin{aligned} \frac{1}{n} \min_{1 \leq i \leq n} W(Y_i;\, h_i,\, \tilde{\Sigma}_{iT}) \quad &\leq \quad W(Y;\, g,\, \tilde{\Sigma}_T) \leq \frac{1}{n}\Big[\prod_{i=1}^{n} W(Y_i;\, h_i,\, \tilde{\Sigma}_{iT})\Big]^{1/n} \\ &\leq \quad \frac{1}{n}\sum_{i=1}^{n} W(Y_i;\, h_i,\, \tilde{\Sigma}_{iT}) \leq \frac{1}{n} \max_{1 \leq i \leq n} W(Y_i;\, h_i,\, \tilde{\Sigma}_{iT}) \end{aligned} \tag{4.16}$$

where

$$W(x_i;\, h_i,\, \tilde{\Sigma}_{iT}) \equiv \frac{h_i(x_i)^2}{H_i(x_i)\,\tilde{\Sigma}_{iT}\,H_i(x_i)'}\,,\; x_i \in \mathbb{R}^{p_i}. \tag{4.17}$$

From (4.14) - (4.17), it is clear that:

$$\mathbb{P}[W(Y;\, g,\, \tilde{\Sigma}_T) \leq W(Y_i;\, h_i,\, \tilde{\Sigma}_{iT})\,|\,\Delta_i > 0] = 1\,, \quad i = 1, \ldots, n\,. \tag{4.18}$$

When

$$\mathbb{P}[\Delta_i(Y_i) > 0] = 1\,, \tag{4.19}$$

we can replace the conditional probabilities in (4.18) by the corresponding unconditional probabilities:

$$\mathbb{P}[W(Y;\, g,\, \tilde{\Sigma}_T) \leq W(Y_i;\, h_i,\, \tilde{\Sigma}_{iT})] = 1, \quad i = 1, \ldots, n\,; \tag{4.20}$$

this condition allows $\mathbb{P}[\Delta_j(Y_j) > 0] < 1$ for some $j \neq i$. If $\mathbb{P}[\Delta_i(Y_i) > 0] = 1$ for all $i = 1, \ldots, n$, we also have

$$\mathbb{P}[W(Y;\, g,\, \tilde{\Sigma}_T) \leq \min_{1\leq i\leq n} W(Y_i;\, h_i,\, \tilde{\Sigma}_{iT})] = 1\,, \tag{4.21}$$

which yields a tighter bound. (4.21) also underscores the fact that better bounds can be achieved by factoring $h(\theta)$ into the largest possible number of factors, *i.e.* by letting n be large. Similarly, if

$$\mathbb{P}[g(Y) \neq 0,\, \Delta_1 > 0, \ldots, \Delta_n > 0] = 1\,, \tag{4.22}$$

(4.16) holds with probability one.

When (4.19) holds, both $W_g(Y)$ and $W_{h_i}(Y_i)$ are well defined (finite) with probability one. The test statistic

$$W_T(\hat{\theta}_{i\cdot T};\, h_i,\, \tilde{\Sigma}_{iT}) = T\, W(\hat{\theta}_{i\cdot T};\, h_i,\, \tilde{\Sigma}_{iT}) = T\, \frac{h_i(\hat{\theta}_{i\cdot T})^2}{H_i(\hat{\theta}_{i\cdot T})\, \tilde{\Sigma}_{iT}\, H_i(\hat{\theta}_{i\cdot T})'} \tag{4.23}$$

can be interpreted as a Wald-type statistic for testing $h_i(\theta_{i\cdot}) = 0$ using the parameter estimates $\hat{\theta}_{i\cdot T}$ and $\tilde{\Sigma}_{iT}$. So the distribution of $W_T(\hat{\theta}_T;\, g,\, \tilde{\Sigma}_T)$ can be bounded by the distribution of $W_T(\hat{\theta}_{i\cdot T};\, h_i,\, \tilde{\Sigma}_{iT})$, under both the null and alternative hypotheses.

It is also remarkable that this dominance property holds without any assumption on the distribution of Y. In particular, the distribution of Y_j for $j \neq i$ is irrelevant. Under $\mathscr{H}_0$, we may have $h_i(\theta_{i\cdot}) = 0$ along with $h_j(\theta_{j\cdot}) \neq 0$ for $j \neq i$, without the bound being affected. Note also that none of the functions $h_i(\theta_{i\cdot})$ need be a polynomial.

Let us consider the case where at least one of the functions in $g(\theta)$ is a polynomial, along with conditions similar to those in Assumptions *2.1 - 2.3*.

Assumption 4.1 POLYNOMIAL FACTOR REGULARITY. *Let $1 \leq i \leq n$. The function $h_i(\theta_{i\cdot})$ is a polynomial, and the estimator $\hat{\theta}_{i\cdot T}$ in* (4.8) - (4.9) *satisfies Assumptions 2.1 - 2.3 upon replacing p by p_i, J by J_i, Z by Z_i, V by $V_i = J_i J_i'$, and $\hat{V}_T$ by $\hat{V}_{iT}$. Further, $\tilde{\Sigma}_{iT} = \hat{V}_{iT}$ and $\gamma_i(\bar{\theta}_{i\cdot})$ is the singularity order of $h_i(\theta_{i\cdot})$ at $\bar{\theta}_{i\cdot}$ as defined in* (3.6) - (3.7).

We can then get the following asymptotic dominance result on the asymptotic distribution of $W_T(\hat{\theta}_T;\, g,\, \tilde{\Sigma}_T)$ as characterized in Theorem 3.1 (with $\hat{V}_T = \tilde{\Sigma}_T$).

Proposition 4.4 POLYNOMIAL FACTOR BOUNDS FOR WALD STATISTICS OF SINGLE RESTRICTIONS. *Under the conditions of Proposition 4.3, let $V = \mathrm{diag}[V_1, V_2, \ldots, V_n]$ where each V_i is a $p_i \times p_i$ positive semidefinite matrix, and $W(\bar{\theta};\, g,\, V)$ is defined as in* (3.14).

(1) *If $h_{i_0}(\bar{\theta}_{i_0\cdot}) = 0$ and Assumption 4.1 holds for $i = i_0$ where $i_o \in \{1, 2, \ldots, n\}$, then conditionally on $\Delta_{i_0}(Y_{i_0}) > 0$,*

$$W(\bar{\theta};\, g,\, V) \leq \|Z_{i_0}\|^2 / [\gamma_{i_0}(\bar{\theta}_{i_0\cdot}) + 1]^2 \leq \|Z_{i_0}\|^2 \tag{4.24}$$

where $\gamma_{i_0}(\bar{\theta}_{i_0\cdot})$ is the singularity order of $h_{i_0}(\theta_{i_0\cdot})$ at $\bar{\theta}_{i_0\cdot}$, and when $\bar{\theta}_{i_0\cdot}$ is a singular point of $h_{i_0}(\theta_{i_0\cdot})$,

$$W(\bar{\theta};\, g,\, V) \leq \frac{1}{4}\, \|Z_{i_0}\|^2\,. \tag{4.25}$$

(2) *If* $g(\bar{\theta}) = 0$ *and Assumption 4.1 holds for all* $i \in \{1, \ldots, n\}$, *then conditionally on* $\Delta_i(Y_i) > 0$ *for all* $i \in \{1, \ldots, n\}$,

$$W(\bar{\theta}; g, V) \leq \|Z_i\|^2 / [\gamma_i(\bar{\theta}_{i\cdot}) + 1]^2 \leq \|Z_i\|^2 \quad \textit{for some } i \in \{1, \ldots, n\} \tag{4.26}$$

and, when $\bar{\theta}$ *is a singular point of* $g(\theta)$,

$$W(\bar{\theta}; g, V) \leq \frac{1}{4}\, \|Z_i\|^2 \quad \textit{for some } i \in \{1, \ldots, n\}. \tag{4.27}$$

If $Z_{i_0} \sim \mathrm{N}[0, I_{p_0}]$ and $h_{i_0}(\theta_{i_0\cdot}) = 0$, (4.24) entails that

$$W(\bar{\theta}; g, V) \leq \frac{1}{[\gamma_{i_0}(\bar{\theta}_{i_0\cdot}) + 1]^2} \chi^2_{p_0} \tag{4.28}$$

so the number of degrees of freedom in the bound can be substantially reduced with respect to the χ^2_p distribution. Similarly, when $Z \sim \mathrm{N}[0, I_p]$ and $h_1(\theta_{1\cdot}), \ldots, h_n(\theta_{n\cdot})$ are all polynomials, (4.26) yields:

$$W(\bar{\theta}; g, V) \leq \chi^2_{p_i} / [\gamma_i(\bar{\theta}_{i\cdot}) + 1]^2 \leq \chi^2_{p_i} \text{ for some } i \in \{1, 2, \ldots, n\}.$$

If furthermore $\bar{\theta}$ is a singular point of $g(\theta)$, we can write:

$$W(\bar{\theta}; g) \leq \frac{1}{4} \chi^2_{p_i} \text{ for some } i \in \{1, 2, \ldots, n\}. \tag{4.29}$$

Thus, when $Z \sim \mathrm{N}[0, I_p]$ and $g(\theta)$ is a product of polynomials of the form (4.8), valid asymptotic p-values can be obtained by computing $\mathrm{p}_{\max}[W_T(\hat{\theta}_T; g, \hat{V}_T)]$ using

$$\mathrm{p}_{\max}[y, n] \equiv \max\{\mathbb{P}[\chi^2_1 > y], \mathrm{pv}[y; p_1, \gamma_1(\bar{\theta}_{1\cdot})], \ldots, \mathrm{pv}[y; p_n, \gamma_n(\bar{\theta}_{n\cdot})]\}, \; y \in \mathbb{R}, \tag{4.30}$$

where

$$\mathrm{pv}[y; p_i, \gamma_i] = \max\{\mathbb{P}[\chi^2_{p_i} / [\gamma_i + 1]^2 > y]\}, \; i = 1, \ldots, n. \tag{4.31}$$

If the singularity orders $\gamma_i(\bar{\theta}_{i\cdot})$ are unknown, we can replace $\mathrm{pv}[y; p_i, \gamma_i]$ by the upper bound $\mathrm{pv}[y; p_i, 1] = \max\{\mathbb{P}[\chi^2_{p_i}/4 > y]\}$, which holds whenever $h_i(\theta_{i\cdot})$ is singular at $\bar{\theta}_{i\cdot}$: this yields the p-value function upper bound $p^U_{\max}[W_T(\hat{\theta}_T; g)]$ where

$$\mathrm{p}^U_{\max}[y, n] \equiv \max\{\mathbb{P}[\chi^2_1 > y], \; \mathrm{pv}[y; p_1, 1], \ldots, \; \mathrm{pv}[y; p_n, 1]\} \geq \mathrm{p}_{\max}[y, n]. \tag{4.32}$$

4.3. Monomials

An important special case of the above problem is the one where $g(\theta)$ is a monomial:

$$g(\theta) = c\, \theta_1^{\nu_1} \cdots \theta_p^{\nu_p} \tag{4.33}$$

where $\nu_1, \ldots, \nu_p$ are positive integers ($\nu_i > 0,\ i = 1, \ldots, n$) and c is non-zero constant. In this case, $\bar{\theta}_1^{\nu_1} \cdots \bar{\theta}_p^{\nu_p} = 0$ under $\mathscr{H}_0$. Clearly, $\mathscr{H}_0$ holds if and only if at least one of the p parameters $\bar{\theta}_1, \ldots, \bar{\theta}_p$ is equal to zero. Let $r = r(\bar{\theta})$ be the number of such zero parameters. Under $\mathscr{H}_0$, $1 \leq r \leq p$. Without loss of generality, we can assume that $\bar{\theta}_i = 0$ for $1 \leq i \leq r$, and $\bar{\theta}_i \neq 0$ if $i > r$. This allows us to write:

$$g(\theta) = \left[\bar{\theta} + (\theta - \bar{\theta})\right]^{\nu_1} \cdots \left[\bar{\theta} + (\theta - \bar{\theta})\right]^{\nu_p} = \prod_{i=1}^{r} (\theta_i - \bar{\theta}_i)^{\nu_i} \prod_{i=r+1}^{p} \left[\bar{\theta}_i + (\theta_i - \bar{\theta}_i)\right]^{\nu_i} \tag{4.34}$$

from which it is easy to see that

$$s(\bar{\theta}) = 1 + \gamma(\bar{\theta}) = \nu_1 + \cdots + \nu_r. \tag{4.35}$$

In other words, $s(\bar{\theta})$ is the sum of the exponents associated with the zero components of $\bar{\theta}$. The larger the number of zero coefficients, the tighter the bounds given by Theorem 4.1. If $Z \sim \mathrm{N}(0, I_p)$, $W(\bar{\theta}; g, V)$ is bounded from above by the $\chi^2_p/(\nu_1 + \cdots + \nu_r)^2$ distribution. For monomials, this bound can however be improved.

Consider first the case of diagonal Wald statistics. Monomial functions of the form (4.33) correspond to the case where $n = p$ and $h_i = \theta_i^{\nu_i}$, $i = 1, \ldots, p$, in (4.8). We then get the following stochastic dominance property.

Proposition 4.5 BOUND ON DIAGONAL WALD STATISTICS FOR MONOMIALS. *Let $Y = (Y_1, \ldots, Y_p)$ be a $p \times 1$ real random vector. Then, for any monomial function of the form (4.33), we have:*

$$\mathbb{P}[W(Y; g, \tilde{\Sigma}_T) \geq z] \leq \mathbb{P}\left[\frac{1}{\nu_i^2}\left(\frac{Y_i}{\tilde{\sigma}_i}\right)^2 \geq z\right], \text{ for } i = 1, \ldots, p, \text{ and } z \in \mathbb{R}, \tag{4.36}$$

where $W(Y; g, \tilde{\Sigma}_T)$ is defined by (4.11) and $\tilde{\Sigma}_T = \mathrm{diag}[\tilde{\sigma}_1^2, \tilde{\sigma}_2^2, \ldots, \tilde{\sigma}_p^2]$.

The above result holds without any assumption on the distribution of Y. For example, $Y_1, \ldots, Y_p$ need not be Gaussian or independent. However, if one of the components of Y, say Y_1, follows a $\mathrm{N}(0, \sigma_1^2)$ distribution, the distribution of $W_g(Y)$ is bounded by the χ_1^2/ν_1^2 distribution:

$$\mathbb{P}[W(Y; g, \tilde{\Sigma}_T) \geq z] \leq \mathbb{P}\left[\frac{1}{\nu_1^2}\chi_1^2 \geq z\right], \text{ for all } z \in \mathbb{R}. \tag{4.37}$$

If $Y_j \sim \mathrm{N}(0, \sigma_j^2)$ distribution and $\tilde{\sigma}_j^2 = \sigma_j^2$ for at least one of the variables $Y_1, \ldots, Y_p$, but we do not know which one, we can write:

$$\mathbb{P}[W(Y; g, \tilde{\Sigma}_T) \geq z] \leq \max_{1 \leq i \leq p} \mathbb{P}\left[\frac{1}{\nu_i^2}\chi_1^2 \geq z\right] = \mathbb{P}\left[\frac{1}{\min\limits_{1 \leq i \leq p} \nu_i^2}\chi_1^2 \geq z\right], \text{ for all } z \in \mathbb{R}. \tag{4.38}$$

The result given by Proposition 4.5 is in fact a finite-sample one. It shows that the distribution of a Wald-type statistic can be bounded (under appropriate conditions) even when the distribution of some component of Y is unknown. If $Y_j \sim \mathrm{N}(0, \sigma_j^2)$ distribution and $\tilde{\sigma}_j^2 = \sigma_j^2$ for all $j = 1, \ldots, p$, we have:

$$\mathbb{P}[W(Y; g, \tilde{\Sigma}_T) \geq z] \leq \min_{1 \leq i \leq p} \mathbb{P}\left[\frac{\chi_1^2}{\nu_i^2} \geq z\right] = \mathbb{P}\left[\frac{\chi_1^2}{\max\limits_{1 \leq i \leq p} \nu_i^2} \geq z\right], \text{ for all } z \in \mathbb{R}. \tag{4.39}$$

This result complements earlier ones obtained under more restrictive conditions. The bounds in (4.38) - (4.39) apply to a relatively specific type of restriction (monomials), but hold under weaker distributional assumptions than the universal bound given in Section 4.1. Here, all cases where the null hypothesis $g(\theta) = 0$ holds are covered, not only the one where $\theta = 0$. Since $\max\limits_{1 \leq i \leq p} \nu_i^2 \leq \big(\sum\limits_{j=1}^{p} \nu_j\big)^2$, we have:

$$\mathbb{P}\left[\chi_1^2 / \big(\sum_{j=1}^{p} \nu_j\big)^2 \geq z\right] \leq \mathbb{P}\left[\chi_1^2 / \max_{1 \leq i \leq p} \nu_i^2 \geq z\right] \leq \mathbb{P}\left[\chi_1^2 / \min_{1 \leq i \leq p} \nu_i^2 \geq z\right], \text{ for all } z \in \mathbb{R}. \tag{4.40}$$

Proposition 4.5 has the following analogue for the asymptotic distribution of $W_T(\hat{\theta}_T; g, \tilde{\Sigma}_T)$ when g is a monomial function.

Corollary 4.6 BOUNDS FOR MONOMIAL WALD STATISTICS. *Under the conditions of Proposition 4.3, suppose Assumption 4.1 also holds. Then, for any monomial function as defined in* (4.33), *we have:*

$$\mathbb{P}[W(\bar{\theta}; g, V) \geq z] \leq \mathbb{P}\left[\frac{1}{\nu_i^2}\left(\frac{Z_i}{\sigma_i}\right)^2 \geq z\right], \textit{ for } i = 1, \ldots, p, \textit{ and } z \in \mathbb{R}, \tag{4.41}$$

where $W(\bar{\theta}; g, V)$ *is defined by* (3.14) *and* $V = \mathrm{diag}[\sigma_1^2, \sigma_2^2, \ldots, \sigma_p^2]$.

Finally, we will give a result applicable to non-diagonal Wald-type statistics.

Proposition 4.7 BOUNDS FOR NON-DIAGONAL MONOMIAL WALD STATISTICS. *Under the assumptions of Theorem 3.1, suppose* $g(\theta)$ *is a monomial function of* θ *as defined in* (4.33), *with* $\theta = \big[\theta_{[1,r]}; \theta_{[r+1,p]}\big]$ *and* $\theta_{[1,r]} = (\theta_1, \ldots, \theta_r)'$. *If* $\bar{\theta}_i = 0$ *for* $1 \leq i \leq r$, *and* $\bar{\theta}_i \neq 0$ *if* $i > r$, *then the Student-type statistic* $t_T(\hat{\theta}_T; g, \hat{V}_T)$ *defined in* (2.4) *converges in probability to* $t(\bar{\theta}; h_1, J)$ *and the Wald-type statistic* $W_T(\hat{\theta}_T; g, \hat{V}_T)$ *in* (2.3) *converges in probability to* $W(\bar{\theta}; h_1, J)$, *with* $h_1(\theta) = \theta_1^{\nu_1} \cdots \theta_r^{\nu_r}$, *where* $t(\bar{\theta}; h_1, J)$ *and* $W(\bar{\theta}; h_1, J)$ *are defined as in* (3.13) - (3.14). *If furthermore* $Z \sim \mathrm{N}(0, I_p)$,

$$W(\bar{\theta}; h_1, J) \sim \frac{1}{\big(\sum\limits_{j=1}^{r} \nu_j\big)^2} \chi_1^2 \tag{4.42}$$

and the distribution of $(\chi_1^2 / \min\limits_{1 \leq i \leq p} \nu_i^2)$ *provides a uniform upper bound regardless of r.*

The distributional result (4.42) can be viewed as an asymptotic extension of the finite-sample results given by Pillai and Meng (2016), applied to the subvector $\theta_{[1,r]}$, and function $h_1\left(\theta_{[1,r]}\right)$. Related finite-sample results on monomial restrictions are also given by Drton and Xiao (2016), and (for $p=2$) Shepp (1962, 1964), Cohen (1981), Reid (1987), Glonek (1993), and Quine (1994); see also DasGupta and Shepp (2004). These however hold provided Y has mean zero, while the results presented in this section allow for cases where only a subset of the elements of $\bar{\theta}$ may be zero. When we do not know how many and which elements of $\bar{\theta}$ are zero, several asymptotic distributions (generally up to 2^p-1) are possible. Thus, when $Z \sim N(0,I_p)$, the bound $\chi_1^2/\big(\min_{1\le i\le p} v_i^2\big)$ provides a uniform upper bound on the asymptotic distribution of the Wald-type statistic.

5. Several restrictions

In this section, we consider the general problem of testing q restrictions when $q \geq 1$:

$$g_l(\theta)=0, \quad l=1,\ldots,q, \tag{5.1}$$

where each $g_l(\theta)$ is a polynomial function of degree m_l. (3.2) - (3.4) then hold with $g(\theta)$, $g(x,\,i,\,\bar{\theta})$ and $c(j_1,\ldots,\,j_p;\,\bar{\theta})$ subscripted by l. We maintain the assumption that the true unknown value $\bar{\theta}$ fulfills the null hypothesis $[g_l(\bar{\theta})=0,\,l=1,\ldots,q]$. For each l, we denote by $s_l(\bar{\theta})$ the order

$$s_l(\bar{\theta})=\min\left\{i : c_l(j_1,\ldots,\,j_p;\,\bar{\theta})\neq 0 \text{ for some } (j_1,\ldots,\,j_p) \text{ with } \Sigma_{i=1}^{p} j_k = i\right\} \tag{5.2}$$

and by $\bar{g}_l(\theta)$ the polynomial that gathers the monomials of lowest degree in g_l:

$$\bar{g}_l(\theta-\bar{\theta}) := g_l[\theta-\bar{\theta};\,s_l(\bar{\theta}),\,\bar{\theta}] = \sum_{\Sigma_{i=1}^{p} j_k = s_l(\bar{\theta})} c_l(j_1,\ldots,\,j_p;\,\bar{\theta}) \prod_{k=1}^{p}\left(\theta_k-\bar{\theta}_k\right)^{j_k}, \quad 1\leq l\leq q. \tag{5.3}$$

We call $\bar{g}(\theta)=[\bar{g}_1(\theta),\ldots,\bar{g}_q(\theta)]'$ the *vector of lowest degree polynomials* associated with $g(\theta)$. Note that the situation where $g_l(x)=0$ for all x [and thus $\bar{g}(x)=0$] is precluded by the assumption that $G(x)$ has full rank *a.e.*.

As in the one restriction case, each polynomial depends on $\bar{\theta}$ through its order and its coefficients. The Jacobian matrix of $\bar{g}(\theta)$ is then

$$\bar{G}(x)=\frac{\partial \bar{g}}{\partial \theta'}(x)=\begin{bmatrix} \bar{G}_1(x) \\ \bar{G}_2(x) \\ \vdots \\ \bar{G}_q(x) \end{bmatrix} \tag{5.4}$$

where each row $\bar{G}_l(x)$ satisfies (3.11) - (3.12) with $\bar{g}$ replaced by $\bar{g}_l$. $\bar{G}(x)$ will be called the *matrix of lowest degree polynomials* for the Jacobian matrix of $g(\theta)$. The elements of the row vector $\bar{G}_l(x)$ are homogeneous polynomials of degree $s_l(\bar{\theta})-1$, for $1\leq l\leq q$. Further, from the Euler homogeneous-function theorem, we have:

$$\bar{g}(\theta-\bar{\theta}) = \Lambda(\bar{\theta})\bar{G}(\theta-\bar{\theta})(\theta-\bar{\theta}), \quad \Lambda(\bar{\theta}) := \mathrm{diag}\left[\frac{1}{s_1(\bar{\theta})}, \ldots, \frac{1}{s_q(\bar{\theta})}\right], \tag{5.5}$$

where $\Lambda(\bar{\theta})$ is the $q\times q$ diagonal matrix. We study again the Wald-type statistic $W_T(\hat{\theta}_T; g, \hat{V}_T)$ and the "pseudo-statistic"

$$\bar{W}_T(\hat{\theta}_T; \bar{g}, V) = T\,\bar{g}(\hat{\theta}_T-\bar{\theta})'\left[\bar{G}(\hat{\theta}_T-\bar{\theta})V\,\bar{G}(\hat{\theta}_T-\bar{\theta})'\right]^{-1}\bar{g}(\hat{\theta}_T-\bar{\theta}) \tag{5.6}$$

for testing $\bar{\mathscr{H}}_0 : \bar{g}(\theta-\bar{\theta}) = 0$.

We will now see that there is a fundamental difference between considering the q restrictions in (5.1) individually and jointly. The key issue is the rank of the Jacobian matrix $\bar{G}(x)$. While we know from Assumption 2.3 that the matrix of polynomials $G(\theta)$ has full row rank q, this does not necessarily hold for the matrix $\bar{G}(\theta-\bar{\theta})$ which only contains the lowest-degree terms. This feature has a strong impact on the distribution of the Wald-type test statistic in large samples, and may jeopardize the existence of an asymptotic distribution.

5.1. Full rank reached at lowest degrees

We will now show that the asymptotic distribution of the Wald-type statistic for q polynomial restrictions is fully determined by the lowest-degree polynomials $\bar{g}_l(\theta)$, $l = 1, \ldots, q$, when $\bar{G}(\theta-\bar{\theta})$ has full row rank q (*a.e.* in $\mathbb{R}^p$), as expressed by the following assumption.

Assumption 5.1 FULL RANK AT LOWEST DEGREE (FRALD). *The $q\times p$ matrix $\bar{G}(x)$ of lowest degree polynomials for the Jacobian matrix of $g(\theta)$ has full rank q* (a.e.).

When the above condition holds, we say that $g(\cdot)$ satisfies has *full rank at lower degree* (FRALD) property. We can then formulate the following theorem.

Theorem 5.1 ASYMPTOTIC DISTRIBUTION OF WALD STATISTIC: FRALD RESTRICTIONS. *Suppose the Assumptions 2.1, 2.2, 2.3 and 5.1 hold. If $g(\theta)$ is a polynomial function of θ as given in* (1.1) - (1.2)*, and if the true unknown value $\bar{\theta}$ satisfies $g(\bar{\theta}) = 0$, then the Wald-type statistics $W_T(\hat{\theta}_T; g, \hat{V}_T)$ and $\bar{W}_T(\hat{\theta}_T; \bar{g}, V)$* [*in* (2.3) *and* (5.6)] *both converge in probability to*

$$W(\bar{\theta}; g, J) = Z'G^*(Z)'\Lambda(\bar{\theta})\left[G^*(Z)G^*(Z)'\right]^{-1}\Lambda(\bar{\theta})G^*(Z)Z \tag{5.7}$$

where $G^*(Z) = \bar{G}(JZ)J$.

Since the FRALD condition holds trivially with only one restriction ($q = 1$), Theorem 5.1 is a generalization of Theorem 3.1. Indeed, when all the singularity orders $s_l(\bar{\theta})$ are the same, *i.e.*

$$s_l(\bar{\theta}) = s(\bar{\theta}),\ l = 1, \ldots, q, \tag{5.8}$$

we have $\Lambda(\bar{\theta}) = s(\bar{\theta})^{-1} I_q$, and Theorem 5.1 provides a limit distribution quite similar to the one of Theorem 3.1, except for the fact that several restrictions are now allowed ($q \geq 1$). However, when (5.8) does not hold, the weighting matrix $\Lambda(\bar{\theta})$ can substantially modify the result.

Theorem 5.1 warrants the same kind of comments as Theorem 3.1. The limit distribution of the Wald-type statistic under $\mathscr{H}_0$ is also the limit distribution of $\bar{W}_T(\hat{\theta}_T; \bar{g}, V)$. In other words, this asymptotic distribution again coincides with the asymptotic distribution of the Wald-type statistic for testing the null hypothesis $\mathscr{H}_0^* : \bar{g}(\theta - \bar{\theta}) = 0$. As in the case of a single restriction, this pseudo-null hypothesis and the corresponding asymptotic distribution depend on the true (unknown) parameter value $\bar{\theta}$.

To allow for more than one restriction, the additional FRALD condition on the rank of $\bar{G}(\theta)$ is required. This condition may be violated. To make things more complicated, the FRALD condition may indeed hold at some value of $\bar{\theta}$ that satisfies the null, but not at another one. We demonstrate this by reconsidering Example 2.6.

Example 5.1 FRALD failure for some $\bar{\theta}$ [Example 2.6 continued]. Let $g(\theta) = \left[\theta_1^2, \theta_1\theta_2^2\right]'$ and $V = I$. Then

$$G(\theta) = \begin{bmatrix} 2\theta_1 & 0 \\ \theta_2^2 & 2\theta_1\theta_2 \end{bmatrix}. \tag{5.9}$$

As already noted, the null hypothesis clearly holds when $\bar{\theta}_1 = \bar{\theta}_2 = 0$. In this case, $\bar{G}(\theta - \bar{\theta})$ is the same as $G(\theta)$ and has full rank: the FRALD condition is satisfied, and Theorem 5.1 is applicable. The null hypothesis also holds for $\bar{\theta}$ with $\bar{\theta}_1 = 0$ and $\bar{\theta}_2 = a \neq 0$. In this case,

$$\bar{G}(\theta - \bar{\theta}) = \begin{bmatrix} 2\theta_1 & 0 \\ a^2 & 0 \end{bmatrix} \tag{5.10}$$

does not have full rank, even though the rows of $\bar{G}(x)$ are linearly independent polynomial vectors: a constant and a linear function. The FRALD condition used by Theorem 5.1 does not hold.

In the above example, the Wald-type statistic diverges at points where the FRALD condition fails. We shall now address the question whether this is always the case.

5.2. Full rank reached at lowest degrees after linear transformation (FRALD-T)

It is well known that the Wald-type test statistic is numerically invariant with respect to non-degenerate linear transformations of the restriction vector. However, it is important to realize that Theorem 5.1 is not invariant to linear transformations of the matrix $G(\theta)$: the rank assumption on the matrix $\bar{G}(\theta - \bar{\theta})$ and its conclusion may change on applying a linear transformation to the vector function $g(\theta)$. The key reason for this feature is the following: if S is a nonsingular matrix of size q, we may have

$$S[\bar{G}(\theta - \bar{\theta})] \neq \overline{SG}(\theta - \bar{\theta}). \tag{5.11}$$

In particular, even though $\bar{G}(\theta - \bar{\theta})$ and $S[\bar{G}(\theta - \bar{\theta})]$ do not have full rank q, the matrix $\overline{SG}(\theta - \bar{\theta})$ of lowest degree polynomials inside $S[G(\theta - \bar{\theta})]$ can have rank q. Example 5.2 below illustrates such a case.

Example 5.2 FRALD failure for some $\bar{\theta}$: FRALD holds for a linear transformation of $g(\theta)$. Consider $g(\theta) = \left[\theta_1 + \theta_2^3,\ \theta_1 + \theta_2^2\right]'$. Clearly, $\bar{\theta} = 0$ and $\bar{\theta} = (-1,\,1)'$ both satisfy $g(\bar{\theta}) = 0$. Then

$$G(\theta) = \begin{pmatrix} 1 & 3\theta_2^2 \\ 1 & 2\theta_2 \end{pmatrix}. \tag{5.12}$$

For $\bar{\theta} = 0$, we have $\bar{G}(\theta - \bar{\theta}) = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$, and the FRALD condition is not satisfied for $g(\theta)$; however, for $S = \begin{pmatrix} 1 & 0 \\ 1 & -1 \end{pmatrix}$, we have:

$$\begin{aligned} SG(\theta - \bar{\theta}) &= SG(\theta) = \begin{pmatrix} 1 & 3\theta_2^2 \\ 0 & 3\theta_2^2 - 2\theta_2 \end{pmatrix}, && (5.13) \\ \overline{SG}(\theta) &= \overline{SG}(\theta - \bar{\theta}) = \begin{pmatrix} 1 & 0 \\ 0 & -2\theta_2 \end{pmatrix}, && (5.14) \end{aligned}$$

so that the FRALD condition holds for $Sg(\theta)$ at $\bar{\theta} = 0$. For $\bar{\theta} = (-1,\,1)'$, $\det[G(\bar{\theta})] = -1 \neq 0$, so there is no singularity at this point.

We thus consider the following extension of the FRALD condition.

Assumption 5.2 FRALD-T. *There exists a nonsingular $q \times q$ matrix S such that the $q \times p$ matrix $\overline{SG}(x)$ of lowest degree polynomials for the Jacobian matrix of $Sg(\theta)$ satisfies*

$$\operatorname{rank}\{\overline{SG}(x)\} = q \quad \textit{for almost all } x \in \mathbb{R}^p. \tag{5.15}$$

When the above condition is satisfied, we say that the *FRALD-T property* holds for $g(\theta)$. This leads to a straightforward but important extension of Theorem 5.1.

Theorem 5.2 ASYMPTOTIC DISTRIBUTION OF WALD STATISTIC: TRANSFORMED FRALD RESTRICTIONS. *Suppose the Assumptions 2.1, 2.2 and 2.3 hold. If $g(\theta)$ is a polynomial function of θ as given in* (1.1) - (1.2)*, if the true unknown value $\bar{\theta}$ satisfies $g(\bar{\theta}) = 0$, and if Assumption 5.2 is satisfied, then the Wald-type statistic $W_T(\hat{\theta}_T;\, g,\, \hat{V}_T)$ in* (2.3) *converges in probability to*

$$W(\bar{\theta};\, g, J, S) = Z'\bar{G}_S^*(Z)'\Lambda_S(\bar{\theta})\left[\bar{G}_S^*(Z)\bar{G}_S^*(Z)'\right]^{-1}\Lambda_S(\bar{\theta})\bar{G}_S^*(Z)Z \tag{5.16}$$

where S is any matrix satisfying (5.15) *and $\bar{G}_S^*(Z) = \overline{SG}(JZ)J$.*

Under the null hypothesis, the limit distribution for the Wald-type statistic for $g(\bar{\theta}) = 0$ [in (5.16)] is also the asymptotic distribution of the Wald-type statistic to test $\mathscr{H}_0(S) : h(\theta) = 0$, with $h(\theta) = Sg(\theta)$. It is the same for any choice of S compatible with the FRALD-T condition. In Section 6, Lemma 6.2 provides a way to derive such a matrix S. The asymptotic distribution given by Theorem 5.2 is also the one applicable to the Wald-type statistic for $\mathscr{H}_0^*(S) : \bar{h}(\theta) = 0$. The latter, however, may not be identical to the asymptotic distribution of the Wald-type statistic for testing $\mathscr{H}_0^* : \bar{g}(\theta) = 0$.

6. Divergence

As observed in Example 2.6, the asymptotic distribution of $W_T(\hat{\theta}_T; g, \hat{V}_T)$ may not exist under $\mathscr{H}_0$. In Theorem 5.2, we showed that the FRALD-T condition is sufficient to ensure the existence of a non-degenerate asymptotic distribution for the Wald-type statistic under $\mathscr{H}_0$. We will now show the condition is also necessary, so the FRALD-T condition is both necessary and sufficient to have a non-degenerate asymptotic distribution of $W_T(\hat{\theta}_T; g, \hat{V}_T)$. When this condition fails, the Wald-type statistic diverges towards infinity at the true unknown value $\bar{\theta}$, even though $\mathscr{H}_0$ holds.

6.1. Characterization of the FRALD-T condition

We first establish auxiliary results on polynomial matrices which will help verify whether the FRALD-T condition holds. We denote by $\mathscr{F}_q$ the space of all nonsingular $q \times q$ matrices, and by $\mathscr{P}_q$ the subspace of $q \times q$ permutation matrices. The elements of $\mathscr{P}_q$ are obtained by permuting the rows of the identity matrix I_q, and these constitute orthogonal matrices; see Harville (1997, section 8.4c). For a general $q \times p$ polynomial matrix $F(x) = [F_{kl}(x)]$ with

$$F_{kl}(x) = \sum_{i=0}^{m_{kl}} \Big\{ \sum_{j_1+\cdots+j_p=i} c_{kl}(j_1,\ldots,j_p) \prod_{n=1}^{p} x_n^{j_n} \Big\}, \tag{6.1}$$

denote $F_{k\cdot}(x) = [F_{k1}(x), \ldots, F_{kp}(x)]$ the k-th row of $F(x)$, $\bar{s}_{kl}$ the lowest degree of the non-zero terms in $F_{kl}(x)$ [setting $\bar{s}_{kl} = 0$ if $F_{kl}(x)$ is a non-zero constant, and $\bar{s}_{kl} = +\infty$ if $F_{kl}(x) = 0$ for all x], and

$$\begin{aligned} \bar{s}_k &= \min\{\bar{s}_{kl} : F_{kl}(x) \neq 0 \text{ and } 1 \leq l \leq p\} && \text{if } F_{k\cdot}(x) \neq 0 \\ &= +\infty && \text{if } F_{k\cdot}(x) = 0 \end{aligned} \tag{6.2}$$

the lowest degree among the polynomials of $F_{k\cdot}(x)$. Overbar on a polynomial matrix $F(x)$ means that each row of $\bar{F}(x)$ only contains the terms with the lowest order $\bar{s}_k$ on the k-th row of $F(x)$, *i.e.*, $\bar{F}(x) = [\bar{F}_{kl}(x)]$ where

$$\begin{aligned} \bar{F}_{kl}(x) &= \sum_{i=0}^{\bar{s}_k} \Big\{ \sum_{j_1+\cdots+j_p=i} c_{kl}(j_1,\ldots,j_p) \prod_{n=1}^{p} x_n^{j_n} \Big\} \\ &= \sum_{j_1+\cdots+j_p=\bar{s}_k} c_{kl}(j_1,\ldots,j_p) \prod_{n=1}^{p} x_n^{j_n}, \end{aligned} \tag{6.3}$$

$k = 1, \ldots, q$. Clearly, $\bar{F}_{kl}(x) = 0$ if the orders of the terms of $F_{kl}(x)$ are all greater than $\bar{s}_k$ [or if $F_{kl}(x) = 0$]. The zero constant function is interpreted as polynomial of degree zero (like any other constant function) and it is homogeneous of any degree. If $P \in \mathscr{P}_q$, it is useful to observe that

$$\overline{PF}(x) = P\bar{F}(x) \tag{6.4}$$

since $PF(x)$ simply involves a permutation of the rows of $F(x)$; see Harville (1997, section 8.4c).

The lemma below gives a linear independence property for transformations of polynomial matrices, while the following lemma provides a construction of an "echelon-type" form for such matrices.

Lemma 6.1 SEPARATION OF LOWEST ORDER POLYNOMIAL ROWS. *Let $F(x)$ be a $q \times p$ non-zero matrix of polynomial functions where $x \in \mathbb{R}^p$, and let s_1 be the minimum degree over all the non-zero polynomials of $\bar{F}(x)$. Then, there is a matrix $S \in \mathscr{F}_q$ such that*

$$\overline{SF}(x) = \begin{bmatrix} [\overline{SF}(x)]_1 \\ [\overline{SF}(x)]_2 \end{bmatrix} \tag{6.5}$$

where $[\overline{SF}(x)]_1$ is a matrix whose rows are linearly independent vectors of homogeneous polynomials all with degree s_1, $[\overline{SF}(x)]_2$ is a matrix whose non-zero elements are polynomials with lowest degree larger than s_1, $\mathrm{row}\{[\overline{SF}(x)]_1\} \geq 1$, $\mathrm{row}\{[\overline{SF}(x)]_2\} \geq 0$, and $\mathrm{row}\{\cdot\}$ denotes the number of rows in a matrix.

Lemma 6.2 ECHELON POLYNOMIAL-DEGREE FORM. *Let $F(x)$ be a $q \times p$ non-zero matrix of polynomial functions where $x \in \mathbb{R}^p$. Then there is a matrix $S \in \mathscr{F}_q$ such that*

$$\overline{SF}(x) = \begin{bmatrix} [\overline{SF}(x)]_1 \\ \vdots \\ [\overline{SF}(x)]_\nu \end{bmatrix} \tag{6.6}$$

where each submatrix $[\overline{SF}(x)]_i$ only contains homogeneous polynomials of degree s_i, for $i = 1, \ldots, \nu$, with $0 \leq s_1 < \cdots < s_i < \cdots < s_\nu$ (where $s_\nu = +\infty$ if the corresponding rows are zero), and the rows of the matrix $\left[[\overline{SF}(x)]_1', \ldots, [\overline{SF}(x)]_{\nu-1}'\right]'$ are linearly independent functions. Further, if $F(x)$ has full rank a.e.*, all the rows of $\overline{SF}(x)$ are linearly independent functions.*

Using the above lemmas, we can now establish that full rank of the "echelon-type" form is a necessary and sufficient condition for the FRALD-T property to hold for a polynomial matrix.

Proposition 6.3 FRALD-T PROPERTY CHARACTERIZATION. *Let $g(\theta)$ be a polynomial function of θ as given in* (1.1) - (1.2)*, suppose Assumption 2.3 holds, and let $S \in \mathscr{F}_q$ be any matrix such that $\overline{SG}(x)$ has the form* (6.6) *with $F = G$. Then, the FRALD-T property is satisfied if and only*

$$\mathrm{rank}\{\overline{SG}(x)\} = q \quad \textit{for almost all } x \in \mathbb{R}^p. \tag{6.7}$$

6.2. Characterization of convergence for Wald statistics

We now establish that failure of the FRALD-T condition entails that Wald-type statistics diverge under $\mathscr{H}_0$.

Theorem 6.4 DIVERGENCE CONDITION. *Suppose the Assumptions 2.1, 2.2 and 2.3 hold, and $g(\bar{\theta}) = 0$ for the true unknown value $\bar{\theta}$. If $g(\theta)$ is a polynomial function as given in* (1.1) - (1.2) *but*

Assumption 5.2 [FRALD-T property] is not satisfied for $g(\theta)$, then $W_T(\hat{\theta}_T; g, \hat{V}_T)$ in (2.3) diverges in probability to $+\infty$.

On combining Theorems 5.2 and 6.4, we finally see that the FRALD-T property is both *necessary and sufficient* for a limit distribution of $W_T(\hat{\theta}_T; g, \hat{V}_T)$ to exist under $\mathscr{H}_0$.

Corollary 6.5 NECESSARY AND SUFFICIENT CONDITION FOR WALD STATISTIC CONVERGENCE. *Suppose the Assumptions 2.1, 2.2 and 2.3 hold. If $g(\theta)$ is a polynomial function as given in (1.1) - (1.2) and if the true unknown value $\bar{\theta}$ satisfies $g(\bar{\theta}) = 0$, then $W_T(\hat{\theta}_T; g, \hat{V}_T)$ in (2.3) converges if and only if Assumption 5.2 is satisfied.*

7. Conclusion

This paper provides a complete characterization of the limit properties of the Wald statistic for testing polynomial restrictions under assumptions that include the standard asymptotic Gaussianity. General distributional results in the case of one restriction demonstrate that uniform bounds on the distribution and on critical values can provide tests with correct asymptotic level. Derivations for some specific cases (such as product restrictions) show that the bounds can be significantly tightened in many cases. When there is more than one polynomial restriction divergence of the Wald statistic under the null is possible, even with asymptotically Gaussian parameter estimates. A necessary and sufficient condition for convergence of the distribution of the statistic under the null hypothesis is provided, and a construction that would verify whether the condition holds is outlined in the proofs of Lemmas 6.1 and 6.2.

Thus only a full investigation of the restrictions at every point in the algebraic variety induced by the null can indicate: (1) whether singular points exist, (2) whether (with several restrictions) there are any points at which divergence occurs, and (3) what forms the limit distributions can take on the algebraic variety studied. A practical implementation of such an investigation would require establishing minimal degrees of homogeneity of polynomial functions based on estimated polynomials. This could be implemented by using "superconsistent estimators", but implementing such methods goes beyond the scope of this paper.

Barring such a complete investigation, we note that, due to the possibility of divergence, employing the Wald statistic to test more than one restriction is risky as the test may have size arbitrarily close to one. There are various methods for combining several restrictions into one such as combining results from tests of individual restrictions or combining the multiple restrictions into one that defines the same algebraic variety.

Finally, for one restriction the results on the bounds on the distribution of the test statistic and on the critical values make it possible to implement conservative tests based on the Wald statistic.

Appendix

A. Underrejections and overrejections: examples

The derivations in the examples below assume $Z \sim N(0,I)$ and (except for Example 2.3) $J = V = \hat{V}_T = I$.

Example 2.1. (i) $\theta_1 = 0$ is a regular linear hypothesis, so the asymptotic distribution under $\mathscr{H}_0$ is χ_1^2. (ii) For $g(\theta) = \theta_1^2$, we have:

$$W_T(\hat{\theta}_T; g, \hat{V}_T) = T\frac{\hat{\theta}_1^2}{4\hat{\theta}_1^2} = \frac{1}{4}T\,\hat{\theta}_1^2 = \frac{1}{4}[T^{\frac{1}{2}}(\hat{\theta}_1 - \bar{\theta})]^2 \xrightarrow[T\to\infty]{\text{p}} \frac{1}{4}Z^2 \sim \frac{1}{4}\chi_1^2. \tag{A.1}$$

The usual critical value is conservative in this case.

Example 2.2. For $g(\theta) = \theta_1\theta_2$, we have $G(\theta) = (\theta_2, \theta_1)$ and

$$\begin{aligned} W_T(\hat{\theta}_T; g, \hat{V}_T) &= T g'(\hat{\theta}_T)\left[G(\hat{\theta}_T)\hat{V}_T G(\hat{\theta}_T)'\right]^{-1} g(\hat{\theta}_T) \\ &= T\frac{\hat{\theta}_1^2\hat{\theta}_2^2}{(\hat{\theta}_2,\, \hat{\theta}_1)(\hat{\theta}_2,\, \hat{\theta}_1)'} = T\,\frac{\hat{\theta}_1^2\hat{\theta}_2^2}{\hat{\theta}_1^2 + \hat{\theta}_2^2}\,. \end{aligned} \tag{A.2}$$

(i) If $\theta_1 = 0$ and $\theta_2 \neq 0$, then

$$W_T(\hat{\theta}_T; g, \hat{V}_T) = T\,\frac{\hat{\theta}_1^2\hat{\theta}_2^2}{\hat{\theta}_1^2 + \hat{\theta}_2^2} \xrightarrow[T\to\infty]{\text{p}} Z_2^2 \sim \chi_1^2 \tag{A.3}$$

and similarly if $\theta_1 \neq 0$ and $\theta_2 = 0$. These represent regular cases. (ii) If $\theta_1 = \theta_2 = 0$,

$$W_T(\hat{\theta}_T; g, \hat{V}_T) \xrightarrow[T\to\infty]{\text{p}} \frac{Z_1^2 Z_2^2}{Z_1^2 + Z_2^2}\,. \tag{A.4}$$

The above limit random variable does not follow a chi-square distribution. However, it can be bounded a chi-square distribution. The spherically distributed (standard Gaussian) vector $(Z_1, Z_2)'$ can be viewed in polar coordinates: $r(\sin\phi, \cos\phi)'$, where $r = \left(Z_1^2 + Z_2^2\right)^{1/2}$ and $\sin\phi = Z_1/r, \cos\phi = Z_2/r$. ϕ is distributed uniformly over $[0, 2\pi]$ and independently of r. Then

$$\frac{Z_1^2 Z_2^2}{Z_1^2 + Z_2^2} = \frac{r^4}{r^2}\sin^2\phi\cos^2\phi = r^2\frac{1}{4}\sin^2(2\phi) = \frac{1}{4}r^2\sin^2 2\phi \tag{A.5}$$

where the distribution of $\sin 2\phi$ is the same as that of $\sin\phi$ and thus the limit distribution is the same as for

$$\frac{1}{4}r^2\sin^2\phi = \frac{1}{4}Z_1^2 \sim \frac{1}{4}\chi_1^2. \tag{A.6}$$

Example 2.3. The limit forms of the statistic in the non-standard cases when more than one $\bar{\theta}_i$ is

zero are obtained by substituting the limit variables X for $\hat{\theta}_T$ and the limit elements of the $\hat{V}_T$ matrix into the expression for the Wald statistic. Convergence follows from Slutsky's theorem and the fact that $\Delta_i \neq 0$ with probability one for $i = 0, 1, 2, 3$.
Example 2.4. For $g(\theta) = \theta_1^2 + \cdots + \theta_p^2$, we have:

$$\begin{aligned} W_T(\hat{\theta}_T; g, \hat{V}_T) &= T g'(\hat{\theta}_T)\left[G(\hat{\theta}_T)G(\hat{\theta}_T)'\right]^{-1} g(\hat{\theta}_T) \\ &= T\frac{\left(\hat{\theta}_1^2 + \cdots + \hat{\theta}_p^2\right)^2}{4\left(\hat{\theta}_1^2 + \cdots + \hat{\theta}_p^2\right)} = \frac{1}{4}T\left(\hat{\theta}_1^2 + \cdots + \hat{\theta}_p^2\right) \\ &\xrightarrow[T\to\infty]{\mathrm{p}} \frac{1}{4}\left(Z_1^2 + \cdots + Z_p^2\right) \sim \frac{1}{4}\chi_p^2. \end{aligned} \tag{A.7}$$

Example 2.5. For $g(\theta) = \theta_1^2 + \theta_1^3 - \theta_2^2$, we have $G = \begin{pmatrix} 3\theta_1^2 + 2\theta_1 & -2\theta_2 \end{pmatrix}$. Then

$$W_T = T\frac{\left(\hat{\theta}_1^3 + \hat{\theta}_1^2 - \hat{\theta}_2^2\right)^2}{9\hat{\theta}_1^4 + 12\hat{\theta}_1^3 + 4\hat{\theta}_1^2 + 4\hat{\theta}_2^2}. \tag{A.8}$$

At $\bar{\theta}_1 = \bar{\theta}_2 = 0$ we get that $W_T \xrightarrow[T\to\infty]{\mathrm{p}} \frac{\left(Z_1^2 - Z_2^2\right)^2}{4(Z_1^2 + Z_2^2)}$. As in Example 2.2, on substituting the same spherical coordinates $r(\sin\phi, \cos\phi)$ for (Z_1, Z_2), we get

$$\frac{\left(Z_1^2 - Z_2^2\right)^2}{4(Z_1^2 + Z_2^2)} = \frac{1}{4}r^2\cos^2 2\phi \sim \frac{1}{4}Z_2^2 \sim \frac{1}{4}\chi_1^2. \tag{A.9}$$

Example 2.6. For $q = 2$ and $g(\theta) = \left[\theta_1^2,\, \theta_1\theta_2^2\right]'$, we have:

$$G(\theta) = \begin{bmatrix} 2\theta_1 & 0 \\ \theta_2^2 & 2\theta_1\theta_2 \end{bmatrix}, \tag{A.10}$$

$$\begin{gathered} W_T(\hat{\theta}_T; g, \hat{V}_T) = T g'(\hat{\theta}_T)\left[G(\hat{\theta}_T)G(\hat{\theta}_T)'\right]^{-1} g(\hat{\theta}_T) \\ = T\begin{pmatrix} \hat{\theta}_1^2 \\ \hat{\theta}_1\hat{\theta}_2^2 \end{pmatrix}'\left[\begin{pmatrix} 2\hat{\theta}_1 & 0 \\ \hat{\theta}_2^2 & 2\hat{\theta}_1\hat{\theta}_2 \end{pmatrix}\begin{pmatrix} 2\hat{\theta}_1 & 0 \\ \hat{\theta}_2^2 & 2\hat{\theta}_1\hat{\theta}_2 \end{pmatrix}'\right]^{-1}\begin{pmatrix} \hat{\theta}_1^2 \\ \hat{\theta}_1\hat{\theta}_2^2 \end{pmatrix} = T\frac{4\hat{\theta}_1^2 + \hat{\theta}_2^2}{16}. \end{gathered} \tag{A.11}$$

(i) If $\bar{\theta}_1 = \bar{\theta}_2 = 0$,

$$W_T(\hat{\theta}_T; g, \hat{V}_T) = T\frac{4\hat{\theta}_1^2 + \hat{\theta}_2^2}{16} \xrightarrow[T\to\infty]{\mathrm{p}} \frac{1}{4}Z_1^2 + \frac{1}{16}Z_2^2 \sim \frac{1}{4}\chi_1^2 + \frac{1}{16}\chi_1^2. \tag{A.12}$$

The asymptotic distribution is $\frac{1}{4}Z_1^2 + \frac{1}{16}Z_2^2$, is a linear combination of two independent χ_1^2 . Since $\frac{1}{4}Z_1^2 + \frac{1}{16}Z_2^2 \leq \frac{1}{4}(Z_1^2 + Z_2^2)$, its distribution is bounded by $\frac{1}{4}\chi_2^2$ distribution. (ii) If $\bar{\theta}_1 = 0$ and $\bar{\theta}_2 \neq 0$, the null hypothesis holds, and

$$W_T(\hat{\theta}_T; g, \hat{V}_T) = T\frac{4\hat{\theta}_1^2 + \hat{\theta}_2^2}{16} = \frac{4T\theta_1^2 + T\bar{\theta}_2^2}{16} \to \frac{4Z_1^2 + T\bar{\theta}_2^2}{16}. \tag{A.13}$$

As $T \to \infty$ $W_T(\hat{\theta}_T; g, \hat{V}_T)$ diverges to $+\infty$.

B. Proofs

PROOF OF THEOREM 3.1 Using the representation given by (3.2) - (3.8), we can write:

$$g(\hat{\theta}_T) = \bar{g}(\hat{\theta}_T - \bar{\theta}) + \bar{r}(\hat{\theta}_T - \bar{\theta}) \tag{B.1}$$

where

$$\bar{g}(\hat{\theta}_T - \bar{\theta}) = g[\hat{\theta}_T - \bar{\theta}; \bar{s}, \bar{\theta}] = \sum_{j_1+\cdots+j_p=\bar{s}} c(j_1, \ldots, j_p; \bar{\theta}) \prod_{k=1}^{p} (\hat{\theta}_{kT} - \bar{\theta}_k)^{j_k}, \tag{B.2}$$

$$\bar{r}(\hat{\theta}_T - \bar{\theta}) = \sum_{i=\bar{s}+1}^{m} g[\hat{\theta}_T - \bar{\theta}; i, \bar{\theta}], \tag{B.3}$$

with $\bar{s} = s(\bar{\theta})$. On multiplying both sides of (B.1) by $T^{\bar{s}/2}$, we get:

$$T^{\bar{s}/2} g(\hat{\theta}_T) = T^{\bar{s}/2}\bar{g}(\hat{\theta}_T - \bar{\theta}) + T^{\bar{s}/2}\bar{r}(\hat{\theta}_T - \bar{\theta}), \tag{B.4}$$

$$T^{\bar{s}/2}\bar{g}(\hat{\theta}_T - \bar{\theta}) = \sum_{j_1+\cdots+j_p=\bar{s}} c(j_1, \ldots, j_p; \bar{\theta}) \prod_{k=1}^{p} \left[T^{1/2}(\hat{\theta}_{kT} - \bar{\theta}_k)\right]^{j_k}, \tag{B.5}$$

$$T^{\bar{s}/2}\bar{r}(\hat{\theta}_T - \bar{\theta}) = \sum_{i=\bar{s}+1}^{m} T^{(\bar{s}-i)/2} \sum_{j_1+\cdots+j_p=i} c(j_1, \ldots, j_p; \bar{\theta}) \prod_{k=1}^{p} \left[T^{1/2}(\hat{\theta}_{kT} - \bar{\theta}_k)\right]^{j_k}. \tag{B.6}$$

By Assumption 2.1, $\sqrt{T}(\hat{\theta}_T - \bar{\theta}) \xrightarrow[T\to\infty]{\mathrm{p}} Y$ where $Y = JZ = (Y_1, \ldots, Y_p)'$, hence

$$\prod_{k=1}^{p} \left[T^{1/2}(\hat{\theta}_{kT} - \bar{\theta}_k)\right]^{j_k} \xrightarrow[T\to\infty]{\mathrm{p}} \prod_{k=1}^{p} Y_k^{j_k}, \tag{B.7}$$

$$T^{\bar{s}/2}\bar{g}(\hat{\theta}_T - \bar{\theta}) \xrightarrow[T\to\infty]{\mathrm{p}} \sum_{j_1+\cdots+j_p=\bar{s}} c(j_1, \ldots, j_p; \bar{\theta}) \prod_{k=1}^{p} Y_k^{j_k} = g[Y; \bar{s}; \bar{\theta}] = \bar{g}(Y), \tag{B.8}$$

$$T^{\bar{s}/2}\bar{r}(\hat{\theta}_T - \bar{\theta}) = o_{\mathrm{p}}(1). \tag{B.9}$$

The last equation follows from (B.7) and the fact that $T^{(\bar{s}-i)/2} \to 0$ for $i > \bar{s}$. Thus,

$$T^{\bar{s}/2} g(\hat{\theta}_T) \xrightarrow[T\to\infty]{\mathrm{p}} \bar{g}(Y) = \bar{g}(JZ), \tag{B.10}$$

$$T^{\bar{s}/2}\left[g(\hat{\theta}_T) - \bar{g}(\hat{\theta}_T - \bar{\theta})\right] = T^{\bar{s}/2}\bar{r}(\hat{\theta}_T - \bar{\theta}) = o_{\mathrm{p}}(1). \tag{B.11}$$

By differentiation of the functions in (B.1), we can write:

$$G(\theta) = \bar{G}(\theta - \bar{\theta}) + \bar{R}(\theta - \bar{\theta})\,, \tag{B.12}$$

with

$$\bar{G}(\theta - \bar{\theta}) = \frac{\partial \bar{g}(\theta - \bar{\theta})}{\partial \theta'} = \frac{\partial \bar{g}(\theta)}{\partial \big(\theta - \bar{\theta}\big)'} = \sum_{j_1 + \cdots + j_p = \bar{s}} c(j_1, \dots, j_p; \bar{\theta}) M_k[\theta - \bar{\theta}; j_1, \dots, j_p]'\,, \tag{B.13}$$

$$\bar{R}(\theta - \bar{\theta}) = \sum_{i=\bar{s}+1}^{m} \sum_{j_1 + \cdots + j_p = i} c(j_1, \dots, j_p; \bar{\theta}) M_k[\theta - \bar{\theta}; j_1, \dots, j_p]'\,, \tag{B.14}$$

where $M[x; j_1, \dots, j_p] = (M_1[x; j_1, \dots, j_p], \dots, M_p[x; j_1, \dots, j_p])'$ and

$$M_k[x; j_1, \dots, j_p] = \frac{\partial}{\partial x_k}\left(\prod_{h=1}^{p} x_h^{j_k}\right) = j_k x_k^{j_k - 1} \prod_{h=1, h\neq k}^{p} x_h^{j_h}\,, \quad x \in \mathbb{R}^p,\ k = 1, \dots, p. \tag{B.15}$$

Since the polynomials in $\bar{G}(\theta - \bar{\theta})$ have degree $\bar{\gamma} = \bar{s} - 1$, we have:

$$T^{\bar{\gamma}/2} G(\hat{\theta}_T) = T^{\bar{\gamma}/2} \bar{G}(\hat{\theta}_T - \bar{\theta}) + T^{\bar{\gamma}/2} \bar{R}(\hat{\theta}_T - \bar{\theta})\,, \tag{B.16}$$

$$\begin{aligned} T^{\bar{\gamma}/2} \bar{G}(\hat{\theta}_T - \bar{\theta}) &= \sum_{j_1 + \cdots + j_p = \bar{s}} c(j_1, \dots, j_p; \bar{\theta}) T^{\bar{\gamma}/2} M_k(\hat{\theta}_T - \bar{\theta}; j_1, \dots, j_p)' \\ &\xrightarrow[T\to\infty]{\mathrm{p}} \sum_{j_1 + \cdots + j_p = \bar{s}} c(j_1, \dots, j_p; \bar{\theta}) M_k(Y; j_1, \dots, j_p)' \equiv \bar{G}(Y) = \bar{G}(JZ)\,, \end{aligned} \tag{B.17}$$

$$\begin{aligned} T^{\bar{\gamma}/2} \bar{R}(\hat{\theta}_T - \bar{\theta}) &= \sum_{i=\bar{s}+1}^{m} T^{(\bar{\gamma}-i)/2} \sum_{j_1 + \cdots + j_p = \bar{s}} c(j_1, \dots, j_p; \bar{\theta}) M_k\big(T^{1/2}(\hat{\theta}_T - \bar{\theta}); j_1, \dots, j_p\big)' \\ &= o_{\mathrm{p}}(1) \end{aligned} \tag{B.18}$$

where again $T^{(\bar{\gamma}-i)/2} \to 0$ for $i > \bar{s}$. Thus

$$T^{\bar{\gamma}/2} G(\hat{\theta}_T) \xrightarrow[T\to\infty]{\mathrm{p}} \bar{G}(Y) = \bar{G}(JZ)\,. \tag{B.19}$$

Now, the Student-type statistic can be rewritten as

$$t_T(\hat{\theta}_T; g, \hat{V}_T) = \frac{T^{1/2} g(\hat{\theta}_T)}{\big[G(\hat{\theta}_T) \hat{V}_T G(\hat{\theta}_T)'\big]^{1/2}} \tag{B.20}$$

can be rewritten as

$$t_T(\hat{\theta}_T; g, \hat{V}_T) = \frac{T^{(\bar{\gamma}+1)/2} g(\hat{\theta}_T)}{\left[\left(T^{\bar{\gamma}/2} G(\hat{\theta}_T)\right)\hat{V}_T\left(T^{\bar{\gamma}/2} G(\hat{\theta}_T)\right)'\right]^{1/2}}. \tag{B.21}$$

Using Assumptions 2.1 - 2.2 and the limits obtained above, we get

$$t_T(\hat{\theta}_T; g, \hat{V}_T) \xrightarrow[T\to\infty]{\text{p}} t(\bar{\theta}; g, V), \tag{B.22}$$

$$t(\bar{\theta}; g, V) := \frac{\bar{g}(Y)}{\left[\bar{G}(Y) V \bar{G}(Y)'\right]^{1/2}}. \tag{B.23}$$

Further, since $\bar{g}(x)$ is homogeneous of degree $\bar{s}$, Euler's theorem yields the identity:

$$\bar{g}(Y) = \left(\frac{1}{\bar{s}}\right) \bar{G}(Y)\, Y, \tag{B.24}$$

hence

$$t(\bar{\theta}; g) = \frac{1}{\bar{s}} \frac{\bar{G}(Y)Y}{\left[\bar{G}(Y)V\bar{G}(Y)'\right]^{1/2}} = \left(\frac{1}{\bar{s}}\right) \frac{\bar{G}(JZ)JZ}{\left[\bar{G}(JZ)JJ'\bar{G}(JZ)'\right]^{1/2}} = \left(\frac{1}{\bar{s}}\right) \frac{\bar{G}^*(Z)Z}{\left[\bar{G}^*(Z)\,\bar{G}^*(Z)'\right]^{1/2}}, \tag{B.25}$$

where $\bar{G}^*(Z) = \bar{G}(JZ)J$. The result for the Wald-type statistic $W_T(\hat{\theta}_T; g, \hat{V}_T)$ follows on observing that

$$\begin{aligned} W_T(\hat{\theta}_T; g, \hat{V}_T) &= T g(\hat{\theta}_T)'[G(\hat{\theta}_T)\hat{V}_T G(\hat{\theta}_T)']^{-1} g(\hat{\theta}_T) \\ &= T \frac{g(\hat{\theta}_T)^2}{G(\hat{\theta}_T)\hat{V}_T G(\hat{\theta}_T)'} = t_T(\hat{\theta}_T; g, \hat{V}_T)^2 \end{aligned} \tag{B.26}$$

hence

$$W_T(\hat{\theta}_T; g, \hat{V}_T) \xrightarrow[T\to\infty]{\text{p}} t(\bar{\theta}; g)^2 = \left(\frac{1}{\bar{s}}\right)^2 \frac{1}{\bar{s}^2} \frac{[\bar{G}^*(Z)Z]^2}{\bar{G}^*(Z)\,\bar{G}^*(Z)'} = \left(\frac{1}{\bar{s}}\right)^2 \frac{[\bar{G}^*(Z)Z]^2}{\left\|\bar{G}^*(Z)\right\|^2}. \tag{B.27}$$

This concludes the proof. □

PROOF OF THEOREM 3.2 Since $\bar{g}(\theta) := g[\theta - \bar{\theta}; s(\bar{\theta}), \bar{\theta}]$ in (3.8) is homogeneous of degree $1+\gamma(\bar{\theta})$ with respect to $\theta - \bar{\theta}$, the functions $\bar{G}(z)$ defined in (3.11) and $\bar{G}^*(z) = \bar{G}(Jz)J$ are homogeneous of degree $\gamma(\bar{\theta})$. Thus

$$\bar{G}^*(U) = \bar{G}^*(Z/\|Z\|) = \|Z\|^{-\gamma(\bar{\theta})}\, \bar{G}^*(Z), \tag{B.28}$$

$$\frac{\bar{G}^*(U)U}{\left\|\bar{G}^*(U)\right\|} = \frac{\|Z\|^{-\gamma(\bar{\theta})}\,\bar{G}^*(Z)Z}{\|Z\|^{-\gamma(\bar{\theta})}\left\|\bar{G}^*(Z)\right\|}\frac{1}{\|Z\|} = \frac{\bar{G}^*(Z)Z}{\left\|\bar{G}^*(Z)\right\|}\frac{1}{\|Z\|}\,, \tag{B.29}$$

and the identities (3.18)- (3.19) follow. When Z is spherically symmetric, the fact that U and $\|Z\|$ are independent is a consequence of the fact the density of Z is constant on spheres. Finally, when $Z \sim \mathrm{N}(0, I_p)$, $\|Z\|^2$ is a sum of p independent χ_1^2 variables, so $\|Z\|^2 \sim \chi_p^2$. □

PROOF OF THEOREM 4.1 By (3.13), we get on using the Cauchy-Schwarz inequality:

$$\left|t(\bar{\theta}; g, V)\right| = \frac{1}{1+\gamma(\bar{\theta})}\frac{\left|\bar{G}^*(Z)Z\right|}{\left\|\bar{G}^*(Z)\right\|} \le \frac{1}{1+\gamma(\bar{\theta})}\frac{\left\|\bar{G}^*(Z)\right\|\,\|Z\|}{\left\|\bar{G}^*(Z)\right\|} = \frac{1}{1+\gamma(\bar{\theta})}\,\|Z\|\,. \tag{B.30}$$

If $\gamma(\bar{\theta}) = 0$, $\bar{G}^*(Z)$ is a non-zero fixed vector by the definition of $\gamma(\bar{\theta})$ [see (3.6) - (3.7)], and the inequalities (4.1) follow on observing that $W(\bar{\theta}; g, V) = t(\bar{\theta}; g, V)^2$. When $Z \sim \mathrm{N}(0, I_p)$, we get using (3.20) that $\tau(\bar{\theta})Z \sim \mathrm{N}(0, 1)$ and $W(\bar{\theta}; g, V) \sim \chi_1^2$. If $\gamma(\bar{\theta}) \ge 1$, (4.2) and (4.3) also follow from (B.30). Since $\|Z\|^2 \sim \chi_p^2$ when $Z \sim \mathrm{N}(0, I_p)$, the global inequalities (4.4) -(4.5) follow by combining (4.1), (4.2) and (4.3). □

PROOF OF PROPOSITION 4.2 If $q \ge p$, the result is trivial because the χ_q^2 distribution dominates the χ_p^2 distribution. So we consider $q < p$. If y is large, the density functions of χ_q^2 and χ_p^2/ζ random variables are monotonically decreasing. Let us denote by $f_X(x)$ the density function of a random variable X. We will now establish that we can find $\bar{y}_0 > 0$ for which

$$f_{\chi_q^2}(y) \ge f_{\chi_p^2/\zeta}(y) \quad \text{when } y \ge \bar{y}_0\,. \tag{B.31}$$

The probability density functions of the χ_q^2 and χ_p^2/ζ distributions are:

$$f_{\chi_q^2}(y) = \frac{1}{2^{q/2}\Gamma(q/2)}y^{(q/2)-1}\exp(-y/2)\,, \tag{B.32}$$

$$f_{\chi_p^2/\zeta}(y) = \frac{\zeta}{2^{p/2}\Gamma(p/2)}(\zeta y)^{(p/2)-1}\exp(-\zeta y/2)\,, \tag{B.33}$$

hence the ratio

$$\begin{aligned}\frac{f_{\chi_q^2}(y)}{f_{\chi_p^2/\zeta}(y)} &= 2^{\frac{p-q}{2}}\frac{\Gamma(p/2)}{\Gamma(q/2)}\frac{1}{\zeta^{(p/2)}}y^{(q-p)/2}\exp[(\zeta-1)y/2] \\ &= 2^{\frac{p-q}{2}}\frac{\Gamma(p/2)}{\Gamma(q/2)}\frac{1}{\zeta^{(p/2)}}\exp\{[(\zeta-1)y-(p-q)\ln(y)]/2\}\end{aligned} \tag{B.34}$$

Since $(\zeta-1) > 0$ and $(p-q) > 0$, this ratio goes to zero as $y \to +\infty$, so (B.31) holds for y large enough, say for $y \ge \bar{y}_0$, hence

$$\mathbb{P}[\chi_q^2 > y] \ge \mathbb{P}[\chi_p^2/\zeta > y] \quad \text{for } y \ge \bar{y}_0\,. \tag{B.35}$$

If $\bar{\alpha} = \mathbb{P}\left[\chi_p^2/\zeta > \bar{y}\right]$ for some $\bar{y} \geq \bar{y}_0$, we have:

$$\mathbb{P}[\chi_q^2 > y] \geq \mathbb{P}[\chi_p^2/\zeta > y] \quad \text{for } y \geq \bar{y}\,. \tag{B.36}$$

□

PROOF OF PROPOSITION 4.3 The proof of this dominance result is based on observing that

$$G_i(\theta) = \frac{\partial g}{\partial \theta_{i\cdot}'} = H_i(\theta_{i\cdot})\, h_{(i)}\,,\; h_{(i)} = \prod_{\substack{j=1 \\ j\neq i}}^{n} h_j\,(\theta_{j\cdot})\,,\; H_i(\theta_{i\cdot}) = \frac{\partial h_i}{\partial \theta_{i\cdot}'}(\theta_{i\cdot})\,, \tag{B.37}$$

for $i = 1, \ldots, n$, hence

$$W(y;\, g,\, \tilde{\Sigma}_T) = \frac{g(y)^2}{G(y)\, \tilde{\Sigma}_T\, G(y)'} = \frac{h_1\,(y_1)^2 \cdots h_n\,(y_n)^2}{H_1(y_1)\tilde{\Sigma}_{1T} H_1(y_1)' h_{(1)}^2 + \cdots + H_n(y_n)\tilde{\Sigma}_{nT} H_n(y_n)' h_{(n)}^2}\,. \tag{B.38}$$

Let $1 \leq i \leq n$. If $H_i(y_i)\tilde{\Sigma}_{iT} H_i(y_i)' > 0$, we can consider two distinct cases: if $h_{(i)} = 0$, we have

$$W(y;\, g,\, \tilde{\Sigma}_T) = 0 \leq \frac{h_i(y_i)^2}{H_i(y_i)\, \tilde{\Sigma}_{iT}\, H_i(y_i)'} = W(Y_i;\, h_i,\, \tilde{\Sigma}_{iT})\,; \tag{B.39}$$

if $h_{(i)} \neq 0$, we have

$$\begin{aligned} W(y;\, g,\, \tilde{\Sigma}_T) &= \frac{h_i(y_i)^2}{H_i(y_i)\, \tilde{\Sigma}_{iT}\, H_i(y_i)' + \sum\limits_{j\neq i} [H_j(y_j)\tilde{\Sigma}_j H_j(y_j)' h_{(j)}^2 / h_{(i)}^2]} \\ &\leq \frac{h_i\,(y_i)^2}{H_i(y_i)\tilde{\Sigma}_{iT} H_i(y_i)'} = W(y_i;\, h_i,\, \tilde{\Sigma}_{iT})\,. \end{aligned} \tag{B.40}$$

Thus $W(Y;\, g,\, \tilde{\Sigma}_T) \leq W(Y_i;\, h_i,\, \tilde{\Sigma}_{iT})$ when $\Delta_i > 0$, and (4.14) is established. When $g(Y) \neq 0$ and $\Delta_1 > 0, \ldots, \Delta_n > 0$, (4.14) holds for all $1 \leq i \leq n$ and (4.15) follows. Further, in this case, we can observe that

$$W(Y;\, g,\, \tilde{\Sigma}_T) = \left\{ \sum_{i=1}^{n} [H_i(Y_i)\tilde{\Sigma}_{iT} H_i(Y_i)' / h_i\,(Y_i)^2] \right\}^{-1} = \left\{ \sum_{i=1}^{n} \frac{1}{W(Y_i;\, h_i,\, \tilde{\Sigma}_{iT})} \right\}^{-1} \tag{B.41}$$

so $n W(Y;\, g,\, \tilde{\Sigma}_T)$ is the harmonic mean of $W(Y_1;\, h_1,\, \tilde{\Sigma}_{1T}), \ldots, W(Y_n;\, h_n,\, \tilde{\Sigma}_{nT})$. (4.16) follows on applying classical inequalities between harmonic, geometric and arithmetic means; see Mitrinović and Vasić (1970, Section 2.1) or Cloud and Drachman (1998, Section 3.4). □

PROOF OF PROPOSITION 4.4 This result follows directly on applying Proposition 4.3 to $W(Y;\, g,\, \tilde{\Sigma}_T)$, and Proposition 4.1 to the statistics $W(Y_i;\, h_i,\, \tilde{\Sigma}_{iT})$ associated with different factors. □

PROOF OF PROPOSITION 4.5 If $g(y) = y_1^{\nu_1} \cdots y_p^{\nu_p}$, we see easily that $G(y) = \left[\frac{\partial g}{\partial y_1}, \ldots, \frac{\partial g}{\partial y_p}\right]$ with

$$\frac{\partial g}{\partial y_j} = n_j y_j^{\nu_j - 1} \prod_{h=1, h \neq j}^{p} y_h^{\nu_h} = \frac{\nu_j g(y)}{y_j}, \; j = 1, \ldots, p, \tag{B.42}$$

so that

$$G(y)G(y)' = g(y)^2 \sum_{j=1}^{p} \left(\frac{\nu_j}{y_j}\right)^2, \tag{B.43}$$

$$W(y; g) = \frac{g(y)^2}{G(y)G(y)'} = \left[\sum_{j=1}^{p} \left(\frac{\nu_j}{y_j}\right)^2\right]^{-1} \leq \left(\frac{\nu_j}{y_j}\right)^{-2} = \left(\frac{y_j}{\nu_j}\right)^2, \; j = 1, \ldots, p. \tag{B.44}$$

This entails (4.36). □

PROOF OF PROPOSITION 4.7 This result is a direct consequence of (3.31) along with Theorem 2.1 of Pillai and Meng (2016). Since $\min_{1 \leq i \leq p} \nu_i^2 \leq \big(\sum_{j=1}^{r} \nu_j\big)^2$ for $1 \leq r \leq p$, the bound $(\chi_1^2 / \min_{1 \leq i \leq p} \nu_i^2)$ dominates the expression in (4.1) for any $1 \leq r \leq p$. □

PROOF OF THEOREM 5.1 We want to show that the Wald-type statistic $W_T(\hat{\theta}_T; g, \hat{V}_T)$ has the same asymptotic distribution as

$$\bar{W}_T(\hat{\theta}_T; g, V) = T\, \bar{g}(\hat{\theta}_T)' \left[\bar{G}(\hat{\theta}_T - \bar{\theta}) V\, \bar{G}(\hat{\theta}_T - \bar{\theta})'\right]^{-1} \bar{g}(\hat{\theta}_T). \tag{B.45}$$

By considering the q components of the function $g(\theta)$, we can reproduce component by component the proof of Theorem 3.1: for $l = 1, \ldots, q$, we get

$$T^{s_l/2} g_l(\hat{\theta}_T) \xrightarrow[T \to \infty]{\mathrm{p}} \bar{g}_l(Y), \quad T^{s_l/2} \left[g_l(\hat{\theta}_T) - \bar{g}_l(\hat{\theta}_T - \bar{\theta})\right] = o_{\mathrm{p}}(1), \tag{B.46}$$

where $\bar{s}_l := s_l(\bar{\theta}) = 1 + \bar{\gamma}_l$ and

$$\bar{g}_l(Y) = \sum_{j_1 + \cdots + j_p = \bar{s}_l} c_l(j_1, \ldots, j_p; \bar{\theta}) \prod_{k=1}^{p} [Y_k]^{j_k}, \tag{B.47}$$

$$T^{\bar{\gamma}_l/2} G_l(\hat{\theta}_T) \xrightarrow[T \to \infty]{\mathrm{p}} \bar{G}_l(Y), \quad T^{\bar{\gamma}_l/2} \left[G_l(\hat{\theta}_T) - \bar{G}_l(\hat{\theta}_T - \bar{\theta})\right] = o_{\mathrm{p}}(1), \tag{B.48}$$

$$\bar{G}_l(\theta - \bar{\theta}) = \frac{\partial \bar{g}_l(\theta)}{\partial \theta'} = \frac{\partial \bar{g}_l(\theta)}{\partial \big(\theta - \bar{\theta}\big)'} = \sum_{j_1 + \cdots + j_p = \bar{s}_l} c_l(j_1, \ldots, j_p; \bar{\theta}) M_k[\theta - \bar{\theta}; j_1, \ldots, j_p]', \tag{B.49}$$

$$\bar{G}_l(Y) = \sum_{j_1 + \cdots + j_p = \bar{s}_l} c_l(j_1, \ldots, j_p; \bar{\theta}) M_k(Y; j_1, \ldots, j_p)'. \tag{B.50}$$

where $M_k(\cdot)$ is defined in (B.15). Setting $\Delta_T(\bar{\theta}) := \operatorname{diag}[T^{\bar{\gamma}_1/2}, \ldots, T^{\bar{\gamma}_q/2}]$, we then have:

$$T^{1/2}\Delta_T(\bar{\theta})\, g(\hat{\theta}_T) \xrightarrow[T\to\infty]{\text{p}} \bar{g}(Y)\,, \quad \Delta_T(\bar{\theta})\, G(\hat{\theta}_T) \xrightarrow[T\to\infty]{\text{p}} \bar{G}(Y)\,, \tag{B.51}$$

$$T^{1/2}\Delta_T(\bar{\theta})\, \bar{g}(\hat{\theta}_T - \bar{\theta}) \xrightarrow[T\to\infty]{\text{p}} \bar{g}(Y)\,, \quad \Delta_T(\bar{\theta})\, G(\hat{\theta}_T - \bar{\theta}) \xrightarrow[T\to\infty]{\text{p}} \bar{G}(Y)\,. \tag{B.52}$$

By Assumption 2.1 and 2.2, $\hat{V}_T \xrightarrow[T\to\infty]{\text{p}} V$ where $V = JJ'$ is invertible, so that

$$[\Delta_T(\bar{\theta})G(\hat{\theta}_T)]\hat{V}_T\,[\Delta_T(\bar{\theta})G(\hat{\theta}_T)]' \xrightarrow[T\to\infty]{\text{p}} \bar{G}(Y)\,V\,\bar{G}(Y)'\,, \tag{B.53}$$

$$[\Delta_T(\bar{\theta})\bar{G}(\hat{\theta}_T - \bar{\theta})]\hat{V}_T\,[\Delta_T(\bar{\theta})\bar{G}(\hat{\theta}_T - \bar{\theta})]' \xrightarrow[T\to\infty]{\text{p}} \bar{G}(Y)V\bar{G}(Y)'. \tag{B.54}$$

Since the matrix of polynomials $\bar{G}(\theta)$ has full rank q (as a matrix of polynomials) and the vector Y is absolutely continuous, the random matrix $\bar{G}_l(Y)\,V\,\bar{G}_l(Y)'$ is almost surely nonsingular. By Slutsky's theorem, we thus have

$$\{[\Delta_T(\bar{\theta})G(\hat{\theta}_T)]\hat{V}_T\,\big(\Delta_T(\bar{\theta})G(\hat{\theta}_T)\big)'\}^{-1} \xrightarrow[T\to\infty]{\text{p}} \{\bar{G}(Y)\,V\,\bar{G}(Y)'\}^{-1}, \tag{B.55}$$

$$\{[\Delta_T(\bar{\theta})\bar{G}(\hat{\theta}_T - \bar{\theta})]\hat{V}_T\,[\Delta_T(\bar{\theta})\bar{G}(\hat{\theta}_T - \bar{\theta})]'\}^{-1} \xrightarrow[T\to\infty]{\text{p}} \{\bar{G}(Y)\,V\,\bar{G}(Y)'\}^{-1}, \tag{B.56}$$

hence

$$\begin{aligned} W_T(\hat{\theta}_T; g, \hat{V}_T) &= T\,g(\hat{\theta}_T)'\left[G(\hat{\theta}_T)\,\hat{V}_T\,G(\hat{\theta}_T)'\right]^{-1} g(\hat{\theta}_T) \\ &= [\Delta_T(\bar{\theta})T^{1/2}g(\hat{\theta}_T)]'\{[\Delta_T(\bar{\theta})G(\hat{\theta}_T)]\,\hat{V}_T\,[\Delta_T(\bar{\theta})G(\hat{\theta}_T)]'\}^{-1}[\Delta_T(\bar{\theta})T^{1/2}g(\hat{\theta}_T)] \\ &\xrightarrow[T\to\infty]{\text{p}} W(\bar{\theta}; g, J)\,, \end{aligned} \tag{B.57}$$

$$\begin{aligned} \bar{W}_T(\hat{\theta}_T; \bar{g}, V) &= T\,\bar{g}(\hat{\theta}_T - \bar{\theta})'\left[\bar{G}(\hat{\theta}_T - \bar{\theta})V\,\bar{G}(\hat{\theta}_T - \bar{\theta})'\right]^{-1} \bar{g}(\hat{\theta}_T - \bar{\theta}) \\ &= [\Delta_T(\bar{\theta})T^{1/2}\bar{g}(\hat{\theta}_T - \bar{\theta})]'[\tilde{G}_T\,V\,\tilde{G}_T']^{-1}[\Delta_T(\bar{\theta})T^{1/2}\bar{g}(\hat{\theta}_T - \bar{\theta})] \\ &\xrightarrow[T\to\infty]{\text{p}} W(\bar{\theta}; g, J)\,, \end{aligned} \tag{B.58}$$

where $\tilde{G}_T := \Delta_T(\bar{\theta})\bar{G}(\hat{\theta}_T - \bar{\theta})$ and, using (5.5),

$$\begin{aligned} W(\bar{\theta}; g, J) &= \bar{g}(Y)'\left[\bar{G}(Y)\,V\,\bar{G}(Y)'\right]^{-1}\bar{g}(Y) \\ &= Y'\bar{G}(Y)'\Lambda(\bar{\theta})\left[\bar{G}(Y)\,V\,\bar{G}(Y)'\right]^{-1}\Lambda(\bar{\theta})\bar{G}(Y)Y\,. \end{aligned} \tag{B.59}$$

Finally, by the definition of Y and $\bar{G}^*$, we have $\bar{G}^*(Z) = \bar{G}(JZ)J$ and

$$W(\bar{\theta}; g, J) = Z'\bar{G}^*(Z)'\Lambda(\bar{\theta})\left[\bar{G}^*(Z)\,\bar{G}^*(Z)'\right]^{-1}\Lambda(\bar{\theta})\bar{G}^*(Z)\,Z\,. \tag{B.60}$$

□

PROOF OF LEMMA 6.1 Since $F(x)$ is non-zero, it has at least one row containing a non-zero homogeneous polynomial of degree s_1, and similarly for $\bar{F}(x)$. Further, we can write

$$F(x) = \bar{F}(x) + R(x) \tag{B.61}$$

where $R(x) := [R_{kl}(x)]$ and each row $R_{k\cdot}(x) = [R_{k1}(x), \ldots, R_{kp}(x)]$ of $R(x)$ only contains polynomials with lowest order larger than $\bar{s}_k$ (or zeros). The proof is split into three main steps.
(1) Let $n_1 \geq 1$ be the number of non-zero rows of $F(x)$ for which the minimum lowest degree of any polynomial is s_1, and $I(s_1) = \{i_1, \ldots, i_{n_1}\}$ the indices of the corresponding rows of $F(x)$. Then there is a $q \times q$ permutation matrix P_1 such that

$$P_1 F(x) = \begin{bmatrix} H_1(x) \\ H_2(x) \end{bmatrix} \tag{B.62}$$

where $H_1(x) := [P_1 F(x)]_1$ is an $n_1 \times p$ matrix containing all the rows with numbers in $I(s_1)$, and $H_2(x) := [P_1 F(x)]_2$ contains the other rows of $F(x)$: each row of $H_1(x)$ contains at least one non-zero homogenous polynomial of degree s_1 (plus possibly higher-order homogeneous polynomials and zeros) [*i.e.*, the rows with numbers in $I(s_1)$], while the non-zero elements of $H_2(x)$ are polynomials with lowest degree larger than s_1. Then

$$P_1 F(x) = P_1 \bar{F}(x) + P_1 R(x)\,, \quad \overline{P_1 F}(x) = P_1 \bar{F}(x) = \begin{bmatrix} \bar{H}_1(x) \\ \bar{H}_2(x) \end{bmatrix}, \tag{B.63}$$

where $\bar{H}_1(x)$ only contains homogeneous polynomials of degree s_1, $\bar{H}_2(x)$ only contains zeros and non-zero homogeneous polynomials of order greater than s_1, and $P_1 R(x)$ only contains polynomials with lowest degree larger than s_1 (or zeros).
(2) Let r_1 be the number of linearly independent polynomial (row) vectors in $\bar{H}_1(x)$. Clearly, $1 \leq r_1 \leq n_1$. If $r_1 = n_1$, the rows of $\bar{H}_1(x)$ are linearly independent functions, and (6.5) holds on taking $S = P_1$, so that $\bar{H}_1(x) = [\overline{SF}(x)]_1$ and $\bar{H}_2(x) = [\overline{SF}(x)]_2$.
(3) If $r_1 < n_1$, $\bar{H}_1(x)$ contains r_1 linearly independent polynomial rows, while the remaining $k_1 = n_1 - r_1$ rows are linear combinations of the rows in the first group. Then we can find a $q \times q$ permutation matrix P_2 such that

$$P_2 \begin{bmatrix} \bar{H}_1(x) \\ \bar{H}_2(x) \end{bmatrix} = \begin{bmatrix} \bar{H}_{11}(x) \\ \bar{H}_{21}(x) \\ \bar{H}_2(x) \end{bmatrix} \tag{B.64}$$

where $\bar{H}_{11}(x)$ is an $r_1 \times p$ matrix obtained by taking r_1 linearly independent polynomial rows of $\bar{H}_1(x)$, and $\bar{H}_{21}(x)$ is a $k_1 \times p$ matrix containing the other rows of $\bar{H}_1(x)$. By the definition of linear independence between polynomial vectors [see (2.5)], we can find a $k(s_1) \times q$ constant matrix C_{21}

such that $\bar{H}_{21}(x) = C_{21}\bar{H}_{11}(x)$. Set

$$Q = \begin{bmatrix} I_{r_1} & 0 & 0 \\ -C_{21} & I_{k_1} & 0 \\ 0 & 0 & I_{q-n_1} \end{bmatrix}, \; S = QP_2P_1\,, \; SR(x) = \begin{bmatrix} R_{11}(x) \\ R_{21}(x) \\ R_{22}(x) \end{bmatrix}, \; R_2(x) = \begin{bmatrix} R_{21}(x) \\ R_{22}(x) \end{bmatrix}, \tag{B.65}$$

where the matrices $R_{11}(x)$, $R_{21}(x)$ and $R_{22}(x)$ have dimensions $r_1 \times p$, $k_1 \times p$ and $[q-n_1] \times p$ respectively, and only contain polynomial terms with order larger than s_1. Clearly Q and S are nonsingular, and

$$S\bar{F}(x) = Q \begin{bmatrix} \bar{H}_{11}(x) \\ \bar{H}_{21}(x) \\ \bar{H}_2(x) \end{bmatrix} = \begin{bmatrix} \bar{H}_{11}(x) \\ \bar{H}_{21}(x) - C_{21}\bar{H}_{11}(x) \\ \bar{H}_2(x) \end{bmatrix} = \begin{bmatrix} \bar{H}_{11}(x) \\ 0 \\ \bar{H}_2(x) \end{bmatrix}, \tag{B.66}$$

$$SF(x) = S\bar{F}(x) + SR(x) = \begin{bmatrix} \bar{H}_{11}(x) + R_{11}(x) \\ R_{21}(x) \\ R_{22}(x) \end{bmatrix} = \begin{bmatrix} \bar{H}_{11}(x) + R_{11}(x) \\ R_2(x) \end{bmatrix}. \tag{B.67}$$

It is then easy to see that

$$\overline{SF}(x) = \begin{bmatrix} \bar{H}_{11}(x) \\ \bar{R}_2(x) \end{bmatrix} = \begin{bmatrix} [\overline{SF}(x)]_1 \\ [\overline{SF}(x)]_2 \end{bmatrix}, \tag{B.68}$$

where $\bar{H}_{11}(x)$ is a matrix whose rows are linearly independent vectors of homogeneous polynomials all with degree s_1, and $\bar{R}_2(x)$ is a matrix whose non-zero elements are polynomials with lowest degree larger than s_1. Thus (6.5) holds on setting $[\overline{SF}(x)]_1 := \bar{H}_{11}(x)$ and $[\overline{SF}(x)]_2 := \bar{R}_2(x)$. □

PROOF OF LEMMA 6.2 *Let* $s_1 \geq 0$ be the lowest degree of any non-zero polynomial in $F(x)$. By Lemma 6.1, we can find a matrix $S_1 \in \mathscr{F}_q$ such that

$$\overline{S_1F}(x) = \begin{bmatrix} [\overline{S_1F}(x)]_1 \\ [\overline{S_1F}(x)]_2 \end{bmatrix} \tag{B.69}$$

where $[\overline{S_1F}(x)]_1$ is an $r_1 \times q$ matrix whose rows are linearly independent vectors of homogeneous polynomials all with degree s_1, $1 \leq r_1 \leq q$, and $[\overline{SF}(x)]_2$ is a matrix whose non-zero elements are polynomials with lowest degree larger than s_1. If $r_1 = q$ or $[\overline{S_1F}(x)]_2 = 0$, (6.6) is satisfied with $S = S_1$. Otherwise, let $F_2(x) = [S_1F(x)]_2$, a $(q-r_1) \times p$ matrix, and s_2 the lowest degree of any non-zero polynomial in $F_2(x)$. Clearly $s_2 > s_1$, and we can apply Lemma 6.1 to $F_2(x)$: we can find a matrix $Q_2 \in \mathscr{F}_{q-r_1}$ such that

$$\overline{Q_2F_2}(x) = \begin{bmatrix} [\overline{Q_2F_2}(x)]_1 \\ [\overline{Q_2F_2}(x)]_2 \end{bmatrix} \tag{B.70}$$

where $[\overline{Q_2F_2}(x)]_1$ is an $r_2 \times q$ matrix whose rows are linearly independent vectors of homogeneous polynomials all with degree s_2, $1 \leq r_2 \leq q - r_1$, and $[\overline{Q_2F_2}(x)]_2$ is a matrix whose non-zero elements are polynomials with lowest degree larger than s_2. Then, for

$$S_2 = \begin{bmatrix} I_{r_1} & 0 \\ 0 & Q_2 \end{bmatrix} S_1\,, \tag{B.71}$$

we have:

$$S_2F(x) = \begin{bmatrix} I_{r_1} & 0 \\ 0 & Q_2 \end{bmatrix} \begin{bmatrix} [S_1F(x)]_1 \\ [S_1F(x)]_2 \end{bmatrix} = \begin{bmatrix} [S_1F(x)]_1 \\ Q_2[S_1F(x)]_2 \end{bmatrix} = \begin{bmatrix} [S_1F(x)]_1 \\ Q_2F_2(x) \end{bmatrix}, \tag{B.72}$$

hence on setting $[\overline{S_2F}(x)]_1 = [\overline{S_1F}(x)]_1$, $[\overline{S_2F}(x)]_2 = [\overline{Q_2F_2}(x)]_1$ and $[\overline{S_2F}(x)]_3 = [\overline{Q_2F_2}(x)]_2$,

$$\overline{S_2F}(x) = \begin{bmatrix} [\overline{S_1F}(x)]_1 \\ \overline{Q_2F_2}(x) \end{bmatrix} = \begin{bmatrix} [\overline{S_1F}(x)]_1 \\ [\overline{Q_2F_2}(x)]_1 \\ [\overline{Q_2F_2}(x)]_2 \end{bmatrix} = \begin{matrix} [\overline{S_2F}(x)]_1 \\ [\overline{S_2F}(x)]_2 \\ [\overline{S_2F}(x)]_3 \end{matrix} \tag{B.73}$$

where $[\overline{S_1F}(x)]_1$ and $[\overline{S_2F}(x)]_2$ only contain linearly independent vectors of homogeneous polynomials all with degree s_1 and s_2, and $[\overline{S_2F}(x)]_3$ is a matrix whose non-zero elements are polynomials with lowest degree larger than s_2.
The same process is repeated as long as the last block $\left[\overline{S_{\nu-1}F}(x)\right]$ either contains non-zero linearly independent polynomial rows of degree s_ν with $0 \leq s_1 < s_2 < \ldots < s_{\nu-1} < s_\nu < \infty$, or contains only zeros in which case the corresponding rows in $S_{\nu-1}F(x)$ are zeros. (6.6) then follows on taking $S = S_{\nu-1}$.
Finally, if $F(x)$ has *full rank q* a.e., $SF(x)$ also has full rank q *a.e.*, and the last block $[S_{\nu-1}F(x)]_\nu$ has full row rank a.e.. *so all the rows of* $[S_{\nu-1}F(x)]_\nu$ must be non-zero, and the lowest degree of any non-zero polynomial is $s_\nu < +\infty$. Further, all the rows of $[\overline{S_{\nu-1}F}(x)]_\nu$ must be linearly independent functions, for otherwise $[S_{\nu-1}F(x)]_\nu$ would not be the last block. Since the different blocks must be linearly independent functions, it follows that all the rows of $\overline{SF}(x)$ all linearly independent functions. □

PROOF OF PROPOSITION 6.3 When the FRALD-T condition does not hold, the rank of $\overline{SG}(y)$ constructed in Lemma 6.2 has to be less than q. If the FRALD-T condition holds for $G(y)$ with some matrix $\bar{S}$, then for any $S \in \mathscr{F}_q$ the FRALD-T property also holds for $SG(y)$ on selecting the matrix $S_{SG} := \bar{S}S^{-1}$. Consider the matrix S defined in the proof of Lemma 6.2. Then $SG(y) = \overline{SG}(y) + R(y)$, where by construction $\overline{SG}(y)$ has q linearly independent rows and, in each row of $R(y)$, the degree of the polynomials is higher than for the corresponding row of $\overline{SG}(y)$. Multiplying $SG(y)$ by S_{SG}, we then see that $\overline{S_{SG}SG}(y) = S_{SG}\overline{SG}(y)$. By the FRALD-T property, the rank of $\overline{S_{SG}SG}(y)$ is q. Multiplication by S_{SG}^{-1} does not change the rank, so the rank of $\overline{SG}(y)$ is q as well. □

PROOF OF THEOREM 6.4 By applying Lemma 6.2 to the Jacobian matrix $G\left(\theta - \bar{\theta}\right)$, we can find a nonsingular $q \times q$ matrix S such that

$$\overline{SG}(x) = \begin{bmatrix} [\overline{SG}(x)]_1 \\ \vdots \\ [\overline{SG}(x)]_\nu \end{bmatrix} \tag{B.74}$$

where each submatrix $[SG(x)]_k$ is an $n_k \times q$ matrix which only contains homogeneous polynomials of degree s_i, with $0 \leq s_1 < \cdots < s_k < \cdots < s_\nu$, and all the rows of $\overline{SG}(x)$ are linearly independent functions of x. Since the Wald test statistic is invariant with respect to linear transformations, we can assume without loss of generality that S is the identity matrix: $S = I_q$. This just means that we have already applied a nonsingular linear transformation to the restrictions under test [since $g(\theta) = 0 \Leftrightarrow Sg(\theta) = 0$].

If the FRALD-T property does not hold, we have $\mathrm{rank}\{\overline{G}(x)\} = r$ for $r < q$ *a.e.*. Write the Wald-type test statistic as follows:

$$W_T(\hat{\theta}_T; g, \hat{V}_T) = \frac{Tg(\hat{\theta}_T)' \left[G(\hat{\theta}_T - \bar{\theta})\hat{V}_T G(\hat{\theta}_T - \bar{\theta})'\right]^{\#} g(\hat{\theta}_T)}{\det\left[G(\hat{\theta}_T - \bar{\theta})\hat{V}_T G(\hat{\theta}_T - \bar{\theta})'\right]} \tag{B.75}$$

where $A^{\#}$ stands for the adjoint matrix (transpose of the cofactor matrix) of a matrix A. Note that the adjoint matrix and the cofactor matrix actually coincide (by symmetry) in the case of (B.75).

Denote by Δ_T the diagonal matrix of size q with i-th diagonal term $T^{s_k/2}$ when $\Sigma_{j=1}^{k-1} n_j < i \leq \Sigma_{j=1}^{k} n_j$, $k = 2, \ldots, \nu$. We can rewrite the Wald test statistic as follows:

$$W_T(\hat{\theta}_T; g, \hat{V}_T) = \frac{(\Delta_T \sqrt{T} g(\hat{\theta}_T))' \left[\left(\Delta_T G(\hat{\theta}_T - \bar{\theta})\right)\hat{V}_T \left(G(\hat{\theta}_T - \bar{\theta})'\Delta_T\right)\right]^{\#} (\Delta_T \sqrt{T} g(\hat{\theta}_T))}{T^{\alpha} \det\left[G(\hat{\theta}_T - \bar{\theta})\hat{V}_T G(\hat{\theta}_T - \bar{\theta})'\right]} \tag{B.76}$$

where $\alpha = \sum_{i=1}^{\nu} n_i s_i$. By a straightforward application of the derivation in the proof of Theorem 5.1 using (B.51) - (B.52), $\hat{V}_T \xrightarrow[T\to\infty]{\text{p}} V = JJ'$, and the continuity of the polynomial matrix function $[A(x)]^{\#}$, we have

$$\begin{aligned}(\Delta_T \sqrt{T} g(\hat{\theta}_T))' \left[\left(\Delta_T G(\hat{\theta}_T - \bar{\theta})\right)\hat{V}_T \left(G(\hat{\theta}_T - \bar{\theta})'\Delta_T\right)\right]^{\#} (\Delta_T \sqrt{T} g(\hat{\theta}_T)) \\ \xrightarrow[T\to\infty]{\text{p}} \bar{g}(Y)' \left[\bar{G}(Y) V \bar{G}(Y)'\right]^{\#} \bar{g}(Y)\end{aligned} \tag{B.77}$$

and, by the definition of Y and $\bar{G}^*$, we get the limit for the numerator as

$$Z'\bar{G}^*(Z)'\Lambda(\bar{\theta}) \left[\bar{G}^*(Z)\bar{G}^*(Z)'\right]^{\#} \Lambda(\bar{\theta})\bar{G}^*(Z)Z. \tag{B.78}$$

Note that this quadratic form is non-zero almost surely. Indeed by Assumption 2.1, Z is absolutely continuous. Z is thus non-zero almost surely; by construction the rows of $\bar{G}^*(Z)$ and thus of $\Lambda(\bar{\theta})\bar{G}^*(Z)$ are linearly independent as row vectors of polynomials and thus linearly independent almost surely and the polynomial matrix $\left[\bar{G}^*(Z)\bar{G}^*(Z)'\right]^{\#}$ is non-zero almost surely. Thus the limit of the numerator is almost surely non-zero.

By an argument similar to the one leading to (B.77), we get for the denominator:

$$\begin{aligned}T^{\alpha} \det\left[G(\hat{\theta}_T - \bar{\theta})\hat{V}_T G(\hat{\theta}_T - \bar{\theta})'\right] &= \det\left[\left(\Delta_T G(\hat{\theta}_T - \bar{\theta})\right)\hat{V}_T \left(\Delta_T G(\hat{\theta}_T - \bar{\theta})\right)'\right] \\ &\xrightarrow[T\to\infty]{\text{p}} \det\left[\bar{G}^*(Z)\bar{G}^*(Z)'\right].\end{aligned} \tag{B.79}$$

Since $\bar{G}^*(Z) = \bar{G}(JZ)J$, the rank of $\bar{G}^*(Z)$ equals the rank of $\bar{G}(JZ)$ and is the same as rank$\{\bar{G}(\theta - \bar{\theta})\} = r^* < q$ when FRALD-T does not hold.
Thus the limit matrix $\left[\bar{G}^*(Z)\bar{G}^*(Z)'\right]$ does not have full rank q, and the limit determinant has to be zero almost everywhere. Then the denominator converges to zero and the statistic diverges to $+\infty$. $\square$

References

Al-Sadoon, M. M. (2017), 'A unifying theory of tests of rank', *Journal of Econometrics* **199**, 49–62.

Alison, P. D. (1995), 'The impact of random predictors on comparisons of coefficients between models: Comment on Clogg, Petkova, and Haritou', *American Journal of Sociology* **100**(5), 1294–1305.

Andrews, D. W. K. (1987), 'Asymptotic results for generalized Wald tests', *Econometric Theory* **3**, 348–358.

Azaïs, J.-M., Gassiat, E. and Mercadier, C. (2006), 'Asymptotic distribution and local power of the log-likelihood ratio test for mixtures: Bounded and unbounded cases', *Bernoulli* **12**(2), 775–799.

Baron, R. M. and Kenny, D. A. (1986), 'The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations', *Journal of Personality and Social Psychology* **51**(6), 1173–1182.

Bollen, K. A., Lennox, R. D. and Dahly, D. L. (2009), 'Practical application of the vanishing tetrad test for causal indicator measurement models: An example from health-related quality of life', *Statistics in Medicine* **28**, 1524–1536.

Bollen, K. A. and Ting, K.-F. (2000), 'A tetrad test for causal indicators', *Psychological Methods* **5**, 3–22.

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.

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.

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.

Caron, R. and Traynor, T. (2005), The zero set of a polynomial, Technical report, Department of Mathematics and Statistics, University of Windsor, Winsdor, Ontario.

Chernoff, H. (1954), 'On the distribution of the likelihood ratio', *Annals of Mathematical Statistics* **25**, 573–578.

Clogg, C. C., Petkova, E. and Haritou, A. (1995), 'Statistical methods for comparing regression coefficients between models', *American Journal of Sociology* **100**(5), 1261–1293.

Clogg, C. C., Petkova, E. and Shihadeh, E. S. (1992), 'Statistical methods for analyzing collapsibility in regression models', *Journal of Educational Statistics* **17**(1), 51–74.

Cloud, M. J. and Drachman, B. C. (1998), *Inequalities: With Applications to Engineering*, Springer, New York.

Cohen, Jr., E. A. (1981), 'A note on normal functions of normal random variables', *Computers & Mathematics with Applications* **7**(5), 395–400.

Critchley, F., Marriott, P. and Salmon, M. (1996), 'On the differential geometry of the Wald test with nonlinear restrictions', *Econometrica* **64**, 1213–1222.

Dagenais, M. G. and Dufour, J.-M. (1991), 'Invariance, nonlinear models and asymptotic tests', *Econometrica* **59**, 1601–1615.

Dagenais, M. G. and Dufour, J.-M. (1994), 'Pitfalls of rescaling regression models with Box-Cox transformations', *Review of Economics and Statistics* **76**, 571–575.

DasGupta, A. and Shepp, L. (2004), Chebyshev polynomials and G-distributed functions of F-Distributed variables, *in* A. DasGupta, ed., 'A Festschrift for Herman Rubin', number 45 *in* 'Lecture Notes - Monograph Series', Institute of Mathematical Statistics, Beachwood, Ohio, pp. 153–163.

Davis, L. J. (1989), 'Intersection union tests for strict collapsibility in three-dimensional contingency tables', *The Annals of Statistics* **17**(4), 1693–1708.

Drton, M. (2009), 'Likelihood ratio tests and singularities', *The Annals of Statistics* **37**(2), 975–1012.

Drton, M., Massam, H. and Olkin, I. (2008), 'Moments of minors of Wishart matrices', *The Annals of Statistics* **36**, 2261–2283.

Drton, M., Sturmfels, B. and Sullivant, S. (2007), 'Algebraic factor analysis: Tetrads, pentads, and beyond', *Probability Theory and Related Fields* **138**(3), 463–493.

Drton, M., Sturmfels, B. and Sullivant, S. (2009), *Lectures on Algebraic Statistics*, Vol. 39 of *Oberwolfach Seminars*, Birkaüser, Basel.

Drton, M. and Xiao, H. (2016), 'Wald tests of singular hypotheses', *Bernoulli* **22**, 38–59.

Ducharme, G. R. and Lepage, Y. (1986), 'Testing collapsibility in contingency tables', *Journal of The Royal Statistical Society Series B* **17**, 197–205.

Dufour, J.-M. (1989), 'Nonlinear hypotheses, inequality restrictions, and non-nested hypotheses: Exact simultaneous tests in linear regressions', *Econometrica* **57**, 335–355.

Dufour, J.-M. (1997), 'Some impossibility theorems in econometrics, with applications to structural and dynamic models', *Econometrica* **65**, 1365–1389.

Dufour, J.-M. (2003), 'Identification, weak instruments and statistical inference in econometrics', *Canadian Journal of Economics* **36**(4), 767–808.

Dufour, J.-M. and Dagenais, M. G. (1992), 'Nonlinear models, rescaling and test invariance', *Journal of Statistical Planning and Inference* **32**, 111–135.

Dufour, J.-M., Pelletier, D. and Renault, É. (2006), 'Short run and long run causality in time series: Inference', *Journal of Econometrics* **132**(2), 337–362.

Dufour, J.-M. and Renault, E. (1998), 'Short-run and long-run causality in time series: Theory', *Econometrica* **66**, 1099–1125.

Dufour, J.-M., Trognon, A. and Tuvaandorj, P. (2017), 'Invariant tests based on M-estimators, estimating functions, and the generalized method of moments', *Econometric Reviews* **36**(1-3), 182–204.

Emsley, R., Dunn, G. and White, I. R. (2010), 'Mediation and moderation of treatment effects in randomised controlled trials of complex interventions', *Statistical Methods in Medical Research* **19**, 237–270.

Gaffke, N., Heiligers, B. and Offinger, R. (2002), 'On the asymptotic null-distribution of the Wald statistic at singular parameter points', *Statistics & Decisions* **20**, 379–398.

Gaffke, N., Steyer, R. and von Davier, A. A. (1999), 'On the asymptotic null-distribution of the Wald statistic at singular parameter points', *Statistics & Decisions* **17**(4), 339–358.

Galbraith, J. W. and Zinde-Walsh, V. (1997), 'On some simple, autoregression-based estimation and identification techniques for ARMA models', *Biometrika* **84**, 685–696.

Glonek, G. F. V. (1993), 'On the behaviour of Wald statistics for the disjunction of two regular hypotheses', *Journal of The Royal Statistical Society Series B* **55**(3), 749–755.

Gouriéroux, C. and Jasiak, J. (2013), Size distortion in the analysis of volatility and covolatility effects, *in* V. N. Huynh, V. Kreinovich, S. Sriboonchita and K. Suriya, eds, 'Uncertainty Analysis in Econometrics with Applications', Springer, Berlin and New York, pp. 91–118.

Gouriéroux, C., Monfort, A. and Renault, É. (1989), 'Testing for common roots', *Econometrica* **57**, 171–185.

Gouriéroux, C., Monfort, A. and Renault, É. (1990), Bilinear constraints: Estimation and test, *in* P. Champsaur, M. Deleau, J.-M. Grandmont, R. Guesnerie, C. Henry, J.-J. Laffont, G. Laroque, J. Mairesse, A. Monfort and Y. Younès, eds, 'Essays in Honor of Edmond Malinvaud, Volume 3: Empirical Economics', The MIT Press, Cambridge, Massachusetts, pp. 166–191.

Gouriéroux, C., Monfort, A. and Renault, É. (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.

Gregory, A. and Veall, M. (1985), 'Formulating Wald tests of nonlinear restrictions', *Econometrica* **53**, 1465–1468.

Harville, D. A. (1997), *Matrix Algebra from a Statistician's Perspective*, Springer-Verlag, New York.

Hayes, A. F. (2013), *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*, The Guilford Press, New York and London.

Hipp, J. R. and Bollen, K. A. (2003), 'Model fit in structural equation models with censored, ordinal, and dichotomous variables: Testing vanishing tetrads', *Sociological Methodology* **3**, 267–305.

Johnson, T. R. and Bodner, T. E. (2007), 'A note on the use of bootstrap tetrad tests for covariance structures', *Structural Equations Modelling* **14**, 113–124.

Kato, N. and Kuriki, S. (2013), 'Likelihood ratio tests for positivity in polynomial regressions', *Journal of Multivariate Analysis* **115**, 334–346.

Koo, N., Leite, W. L. and Algina, J. (2016), 'Mediated effects with the parallel process latent growth model: An evaluation of methods for testing mediation in the presence of nonnormal data', *Structural Equation Modeling: A Multidisciplinary Journal* **23**(1), 32–44.

Lütkepohl, H. and Burda, M. M. (1997), 'Modified Wald tests under nonregular conditions', *Journal of Econometrics* **78**, 315–332.

MacKinnon, D. P. (2008), *Introduction to Statistical Mediation Analysis*, Lawrence Erlbaum Associates, New York and London.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G. and Sheets, V. (2002), 'A comparison of methods to test mediation and other intervening variable effects', *Journal of Personality and Social Psychology* **7**(1), 83–104.

MacKinnon, D. P., Lockwood, C. M. and Williams, J. (2006), 'Confidence limits for the indirect effect: Distribution of the product and resampling methods', *Multivariate Behavioral Research* **39**(1), 99–128.

Mitrinović, D. S. and Vasić (1970), *Analytic Inequalities*, Springer-Verlag, Berlin and New York.

Mityagin, B. S. (2015), The zero set of a real analytic function, Technical report, Department of Mathematics, The Ohio State University, Columbus, OH. arXiv:1:1512.07276v1.

Phillips, P. C. B. and Park, J. Y. (1988), 'On the formulation of Wald tests of nonlinear restrictions', *Econometrica* **56**, 1065–1083.

Pillai, N. H. and Meng, X.-L. (2016), 'An unexpected encounter with Cauchy and Lévy', *The Annals of Statistics* **44**(5), 2089–2097.

Quine, M. P. (1994), 'A result of Shepp', *Applied Mathematics Letters* **7**(6), 33–34.

Reid, J. G. (1987), 'Normal functions of normal random variables', *Computers & Mathematics with Applications* **14**(3), 157–160.

Ritz, C. and Skovgaard, I. M. (2005), 'Likelihood ratio tests in curved exponential families with nuisance parameters present only under the alternative', *Biometrika* **92**, 507–517.

Robin, J.-M. and Smith, R. J. (2000), 'Tests of rank', *Econometric Theory* **16**, 151–175.

Sargan, J. D. (1980), 'Some tests of dynamic specification for a single equation', *Econometrica* **48**(4), 879–897.

Shepp, L. (1962), 'Problem 62-9. Normal functions of normal random variables', *SIAM Review* **6**(3), 255–256.

Shepp, L. (1964), 'Problem 62-9. Normal functions of normal random variables', *SIAM Review* **6**(4), 459–460.

Silva, R., Scheines, R., Glymour, C. and Spirtes, P. (2006), 'Learning the structure of linear latent variable models', *Journal of Machine Learning Research* **7**, 191–246.

Simpson, E. H. (1951), 'The interpretation of interaction in contingency tables', *Journal of The Royal Statistical Society Series B* **13**, 238–241.

Sobel, M. E. (1982), 'Asymptotic confidence intervals for indirect effects in structural equation models', *Sociological Methodology* **13**, 290–312.

Sobel, M. E. (1986), 'Some new results on indirect effects and their standard errors in covariance structure models', *Sociological Methodology* **16**, 159–186.

Spirtes, P., Glymour, C. and Scheines, R. (2000), *Causation, Prediction, and Search. Adaptive Computation and Machine Learning*, MIT Press, Cambridge, MA.

Steyer, R., Gabler, S. and Rucai, A. A. (1996), Individual causal effects, average causal effects, and unconfoundedness in regression models, *in* F. Faulbaum and W. Bandilla, eds, 'Softstat'95: Advances in Statistical Software 5', Lucius & Lucius, Heidelberg, pp. 203–210.

Steyer, R., Mayer, A. and Fiege, C. (2014), Causal inference on total, direct, and indirect effects, *in* A. C. Michalos, ed., 'Encyclopedia of Quality of Life Research', Springer, Dordrecht, pp. 606–630.

Sullivant, S., Talaska, K. and Draisma, J. (2010), 'Trek separation for Gaussian graphical models', *The Annals of Statistics* **38**, 1665–1685.

Tofighi, D. and MacKinnon, D. P. (2016), 'Monte Carlo confidence intervals for complex functions of indirect effects', *Structural Equation Modeling: A Multidisciplinary Journal* **23**(2), 194–205.

VanderWeele, T. (2015), *Explanation in Causal Inference Methods for Mediation and Interaction*, Oxford University Press, Oxford and New York.

Whittemore, A. S. (1978), 'Collapsibility of multidimensional contingency tables', *Journal of The Royal Statistical Society Series B* **40**, 328–340.

Zwiernik, P. and Smith, J. Q. (2012), 'Tree cumulants and the geometry of binary tree models', *Bernoulli* **18**, 290–321.