Testing joint hypotheses when one of the alternatives is one-sided

Karim Abadir and Walter Distaso*

University of York

May 31, 2002

ABSTRACT: We propose a class of statistics where the direction of one of the alternatives is incorporated. It is founded on modifying a class of multivariate tests with elliptical confidence regions, not necessarily arising from Normal-based distribution theory. The resulting statistics are easy to compute, they do not require the re-estimation of models subject to inequality restrictions, and their distributions do not require bounds-based inference. We derive exact explicit distributions, then prove some desirable properties of our class of modified tests. We then illustrate the relevance of the method to practical problems by applying it to devising an improved test of random walks in autoregressive models with deterministic components.

KEYWORDS: mixed one-sided and two-sided alternatives, convolution, test powers, size-unadjusted and size-adjusted powers, unit-root tests.

Corresponding author:

Karim Abadir
Department of Mathematics and Department of Economics
University of York
Heslington
YORK YO10 5DD
UK

(E-mail: kma4@york.ac.uk)

* We would like to thank Les Godfrey, Patrick Marsh, Michael Perlman, and seminar participants at the LSE and Université Libre de Bruxelles for helpful comments. This paper is the basis for one of the chapters of the second author's Doctoral dissertation at the University of York.

1 Introduction

Suppose the parameters of a model are summarized by a vector $\boldsymbol{\theta}$ of fixed and finite dimensions $m \times 1$. A question that arises in some applications takes the form of testing restrictions on $\boldsymbol{\theta}$. The restrictions may be of the type $\mathrm{H}_0 : \boldsymbol{h}(\boldsymbol{\theta}) = \mathbf{0}$; where $\boldsymbol{h}(.)$ is an $(r+1) \times 1$ function which is once-differentiable with respect to $\boldsymbol{\theta}$, and $r + 1 \leq m$. We shall write $\boldsymbol{h}(\boldsymbol{\theta}) \equiv [h_1(\boldsymbol{\theta}) : \boldsymbol{h}_2(\boldsymbol{\theta})']'$ to separate the first scalar component of $\boldsymbol{h}(.)$.

If the alternatives to these restrictions are of the form $\mathrm{H}_1 : \boldsymbol{h}(\boldsymbol{\theta}) \neq \mathbf{0}$, then well-known classical tests exist and their asymptotic optimality has been established in some circumstances. However, suppose that the alternatives were to take the different form $\mathrm{H}_1 : \boldsymbol{h}(\boldsymbol{\theta}) \leq \mathbf{0}$ and $\boldsymbol{\iota}'\boldsymbol{h}(\boldsymbol{\theta}) < 0$, where $\boldsymbol{\iota} \equiv [1 \dots 1]'$. Then, it would improve the powers of classical tests, for a given size, if the one-sided nature of the alternatives is exploited to modify the decision rules; for example, see Gouriéroux, Holly and Monfort (1982), Rogers (1986), Dufour (1989). See also Berger (1989) or Liu and Berger (1995) for the equivalent formulation of such alternative hypotheses as sign tests, ordering tests, etc. when $\boldsymbol{h}(\boldsymbol{\theta})$ is linear in $\boldsymbol{\theta}$.

We now wish to investigate the different setting of mixed alternatives of the form

$$(1) \qquad \mathrm{H}_1 : \ h_1(\boldsymbol{\theta}) < \underset{1\times 1}{0} \ \text{or} \ \boldsymbol{h}_2(\boldsymbol{\theta}) \neq \underset{r\times 1}{\mathbf{0}},$$

which include the special subset $\mathrm{H}_1^{\cap} : h_1(\boldsymbol{\theta}) < 0$ and $\boldsymbol{h}_2(\boldsymbol{\theta}) \neq \mathbf{0}$, of relevance to the application in our paper. We wish to do so in generality, without assuming the linearity of $\boldsymbol{h}(\boldsymbol{\theta})$ or the Normality of estimators of $\boldsymbol{\theta}$. Few authors have investigated (1) or variants thereof; for example, see Andrews (1998) in the case of linear $\boldsymbol{h}(.)$, and Wolak (1989), Kodde and Palm (1986), Silvapulle (1991). Some treatments deal with the related inequality $h_1(\boldsymbol{\theta}) \leq 0$ for H_1, which would not extend to $h_1(\boldsymbol{\theta}) < 0$ because of a problem relating to optimizing over a parameter space which is not closed. Additionally, see the criticisms, about critical regions and least favourable distributions, in Perlman and Wu (1999). These difficulties do not apply to the different setup in (1); see Perlman and Wu (1999, Section 10).

In this paper, we propose a class of statistics where the direction of the alternative on $h_1(\boldsymbol{\theta})$ is incorporated. It is based on modifying, for the one-sided component $h_1(\boldsymbol{\theta})$, a *class* of multivariate tests having confidence regions which are elliptical in the test's univariate components. That the class of statistics leads to decision rules based on ellipses for its components does *not* mean that it is based on estimators which are elliptically distributed, the latter being a much stronger requirement. The class of statistics leading to elliptical confidence regions includes well-known statistics like the conventional Wald (W) Lagrange Multiplier (LM), and normally Likelihood Ratio (LR), even when their distributions are nonstandard and possibly distinct from each other; e.g. see Abadir (1993b). Our approach to solving the inference problem will not presuppose the asymptotic Normality of the estimator(s) of $\boldsymbol{\theta}$. The resulting class of test statistics is easy to compute, as they do not require the re-estimation of models subject to inequality restrictions, and their distributions are unique (i.e. do not necessitate bounds-based inference). For the Normal special case, we also give exact explicit formulae for the new tests' null distributions (hence quantiles) and power functions. Having done so, we can prove the unbiasedness and consistency of our decision rules, and their invariance to some groups of transformations. We can then also prove analytically, without resort to simulations, that the unmodified decision rules are inadmissible. Apart from proving consistency, earlier papers dealing with general nonlinear $\boldsymbol{h}(.)$ restrictions have not proved analytically the uniform superiority (if at all) of their one-sided inference procedures over the unadjusted counterparts.

One motivation for alternative hypotheses of the forms $\mathrm{H}_1^{\cap}$ or H_1 of (1) may be given by a simple Auto-Regressive (AR) model with deterministic components. It is often of interest, e.g. in macroeconomics and finance, to test economic efficiency. The hypothesis of strong efficiency usually take the form of a series following a random walk with no deterministic components. Alternative hypotheses could include trend-stability of a series (no unit root), weak efficiency (unit root with deterministic components), as well as others. The main feature of such inference is that the component of the test which is one-sided is the largest AR root, ruling out explosive roots in a frequentist setup, while the remaining AR roots and deterministic variables (e.g. drift and/or time trend) have a two-sided component if it is not the objective of the test to determine whether the series is trending up or down. AR models are also known to give rise to problems of low powers and lack of similarity and of pivots,[1] and are therefore a useful tool for illustration. By using our procedure, we give a novel application of mixed one-sided and two-sided inference in this time series problem.[2]

In some cases, approaches alternative to ours may be based on separate testing of the components $h_1(.)$ and $\boldsymbol{h}_2(.)$. First, one may use the standard methods of testing multiple hypotheses; for example, see Savin (1984) for an introduction. A second possibility would be a sequential (conditional) procedure for separate testing of $h_1(.)$ and $\boldsymbol{h}_2(.)$. When similar tests do not exist for either of the individual components of H_0, the approach based on separate testing of $h_1(.)$ and $\boldsymbol{h}_2(.)$ raises a number of difficulties, and conservative bounds-based decision rules would be needed. These would result in a loss of power, which was supposed to have been avoided by the new procedure. We therefore opted for joint inference to tackle our problem, and the reasons for our choice will be illustrated further

[1]For the definitions of pivots and similarity, see, for example, Cox and Hinkley (1974) or Lehmann (1986).

[2]Kim and Newbold (2001) also give an application of one-sided inference and inequality-restricted estimation to a related time series problem.

by the application of our method in the latter part of our paper.

In Section 2, we present the general setup and give the main theorems of the paper. Section 3 contains an application of our approach to devising improved inference procedures for tests of random walks in AR models with deterministic components. Section 4 concludes. The proofs of our theorems are collected in an Appendix, where we use the concept of a size-unadjusted power function in order to prove an inequality of size-adjusted power functions. These concepts are often used in simulations, but here we use them unconventionally for analytical derivations. Finally, we use the following notation frequently throughout the paper. The indicator function $1_{\mathcal{K}}$ gives 1 if condition $\mathcal{K}$ is satisfied and zero otherwise. The standard Normal density and distribution functions are defined by $\phi(.)$ and $\Phi(.)$, respectively, and a noncentral χ^2 variate with ν degrees of freedom and noncentrality parameter δ is denoted by $\chi^2_\nu(\delta)$.

2 A class of modified statistics

Suppose n observations are available for the model whose parameters are summarized by $\boldsymbol{\theta}$. Let $\hat{\boldsymbol{\theta}}$ be some consistent estimator of the parameter vector $\boldsymbol{\theta}$, and $\boldsymbol{\mathcal{I}}^{-1}$ be its asymptotic variance matrix, which we assume is finite. When standard regularity conditions hold, one may think of $\boldsymbol{\mathcal{I}}$ as Fisher's information matrix, but this need not be the case for our paper. However, we need to assume that the parameters are locally identified, implying that $\boldsymbol{\mathcal{I}}$ is nonsingular; see Rothenberg (1971), Catchpole and Morgan (1997). We will also need to assume that $\partial \boldsymbol{h}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}'$ has rank $r+1$.

Let $\boldsymbol{h}(\hat{\boldsymbol{\theta}})$ be the function $\boldsymbol{h}(.)$, with argument $\hat{\boldsymbol{\theta}}$ instead of $\boldsymbol{\theta}$. By consistency of $\hat{\boldsymbol{\theta}}$, the delta method gives the asymptotic variance of $\boldsymbol{h}(\hat{\boldsymbol{\theta}})$ as

$$\boldsymbol{V} \equiv \begin{bmatrix} v_{11} & \boldsymbol{v}_{21}' \\ \boldsymbol{v}_{21} & \boldsymbol{V}_{22} \end{bmatrix} \equiv \frac{\partial \boldsymbol{h}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \boldsymbol{\mathcal{I}}^{-1} \frac{\partial \boldsymbol{h}(\boldsymbol{\theta})'}{\partial \boldsymbol{\theta}},$$

implying that

$$\boldsymbol{h}_{2_\perp 1}\left(\hat{\boldsymbol{\theta}}\right) \equiv \boldsymbol{h}_2\left(\hat{\boldsymbol{\theta}}\right) - \frac{h_1(\hat{\boldsymbol{\theta}})}{v_{11}} \boldsymbol{v}_{21} \quad (2)$$

is uncorrelated with $h_1(\hat{\boldsymbol{\theta}})$. The function $\boldsymbol{h}_{2_\perp 1}(\hat{\boldsymbol{\theta}})$ contains v_{11} and $\boldsymbol{v}_{21}$ which will, in general, depend on the unknown parameter $\boldsymbol{\theta}$ (but not on the random $\hat{\boldsymbol{\theta}}$). The function $\boldsymbol{h}_{2_\perp 1}(\boldsymbol{\theta})$ is defined as in (2), but with $\hat{\boldsymbol{\theta}}$ replaced by $\boldsymbol{\theta}$. The asymptotic variance of $\boldsymbol{h}_{2_\perp 1}(\hat{\boldsymbol{\theta}})$ is

$$\boldsymbol{V}_{2_\perp 1} \equiv \boldsymbol{V}_{22} - \frac{1}{v_{11}} \boldsymbol{v}_{21} \boldsymbol{v}_{21}'.$$

In some models, analytical derivation of $\boldsymbol{\mathcal{I}}$ may be intractable, and a resort to numerical methods may be required in devising $\boldsymbol{h}_{2_\perp 1}(\hat{\boldsymbol{\theta}})$.

In general, orthogonality and independence will coincide only to second order, and an application involving non-Normal densities will follow in Section 3. This caveat aside, the canonical form of the class of statistics we suggest is

$$\tau \equiv \tau_2^2 + 1_{\tau_1<0}\tau_1^2, \tag{3}$$

with critical region of size α defined by $\tau > c_\alpha^2$ and where:

1. the component τ_2^2 tests the joint hypotheses $\boldsymbol{h}_{2_\perp 1}(\boldsymbol{\theta}) = \mathbf{0}$ [by applying $\boldsymbol{h}(\boldsymbol{\theta}) = \mathbf{0}$ to (2)] versus $\boldsymbol{h}_{2_\perp 1}(\boldsymbol{\theta}) \neq \mathbf{0}$; and
2. the component τ_1 is a one-sided statistic for testing $h_1(\boldsymbol{\theta}) = 0$ versus $h_1(\boldsymbol{\theta}) < 0$.

We now explain this choice of canonical form, starting with a general justification in the next subsection then a formal assessment in the cases of standard and nonstandard asymptotics for $\hat{\boldsymbol{\theta}}$.

2.1 Design of the class of statistics

The class of statistics in (3) is a modification of the usual critical regions of the form

$$\tilde{\tau} \equiv \tau_2^2 + \tau_1^2 > \tilde{c}_\alpha^2, \tag{4}$$

where $c_\alpha^2 < \tilde{c}_\alpha^2$ for $\alpha \in (0,1)$. In the two-dimensional decision space of the two components for testing the scalar $h_1(.)$ and the vector $\boldsymbol{h}_{2_\perp 1}(.)$, the critical regions defined by (4) are bounded by circles. These would have been ellipses in the non-orthogonalized space which would have been based on $h_1(.)$ and $\boldsymbol{h}_2(.)$. Rotating the axes in that latter space, namely orthogonalizing the components into $h_1(\hat{\boldsymbol{\theta}})$ and $\boldsymbol{h}_{2_\perp 1}(\hat{\boldsymbol{\theta}})$, implies that the critical regions can now be fully characterized by the sum of a function of τ_1 and another function of τ_2, but no cross-products like $\tau_1\tau_2$. This has simplified the modification which we propose in (3).

The critical regions implied by (3) and (4) can be represented graphically in $\tau_1|\tau_2$ space, as in Figure 1. The curve C_1 represents the boundary for the usual critical region implied by (4). This region is unduly conservative, in the light of the partially one-sided nature of H_1 in (1). It tends to reject the null hypothesis when

$$h_1(\boldsymbol{\theta}) > 0 \quad \text{and} \quad \boldsymbol{h}_{2_\perp 1}(\boldsymbol{\theta}) = \mathbf{0}, \tag{5}$$

which is clearly not intersecting $\mathrm{H}_1^{\cap}$ at all. Furthermore, as we will now show, the design of optimal test procedures for inference on H_1 will give rise to regions which will again not intersect with the one in (5), thus causing an unnecessarily large Type I error when using C_1. If one were to use a region of the form of C_1

with a fixed size, a lower Type I error would come at the expense of an increased Type II error (i.e. reduced power), which is not desirable.

The intersection of the set in (5) with H_1 is the (possibly empty) manifold[3] defined by

$$\frac{h_1(\boldsymbol{\theta})}{v_{11}}\boldsymbol{v}_{21} = \boldsymbol{h}_2(\boldsymbol{\theta}) \neq \mathbf{0}. \tag{6}$$

The complement of the set given by H_1 requires $h_1(\boldsymbol{\theta}) \geq 0$, and the least favourable distribution [e.g. Lehmann (1986, Section 3.8)] is reached at $h_1(\boldsymbol{\theta}) = 0$. Optimal inference should therefore fix $h_1(\boldsymbol{\theta}) = 0$ and exclude $h_1(\boldsymbol{\theta}) > 0$ from the critical region. This amounts to excluding the rectangular region separating $\boldsymbol{h}_{2_\perp 1}(\boldsymbol{\theta}) = \mathbf{0}$ from $\boldsymbol{h}_{2_\perp 1}(\boldsymbol{\theta}) \neq \mathbf{0}$, whenever $h_1(\boldsymbol{\theta}) > 0$. Not only would (5) not overlap with $\mathrm{H}_1^{\cap}$, it would also lead now to the manifold in (6) being empty (because of the inequality there) hence not intersecting at all with H_1 either. The critical region bounded by C_1 is therefore not optimally-sized, as it contains some events which are incompatible with both $\mathrm{H}_1^{\cap}$ and H_1.

Removing the region (5) from the old critical region modifies the boundary C_1 into C_2 which, in Figure 1, is composed of the dotted lines for $\tau_1 > 0$ and the old semicircle for $\tau_1 \leq 0$. However, excluding the set in (5) from the critical region bounded by C_1 has reduced its size (Type I error). So, for a fixed size α, the dotted curve C_3 gives the new boundary for the critical region of comparable size to the original C_1. The parameterization of the boundary of our new critical region is given by (3), and the decision rule implied by it is to reject H_0 if either:

1. $\tau_2^2 > c_\alpha^2$; or

2. $\tau_2^2 + \tau_1^2 > c_\alpha^2$ if $\tau_1 < 0$.

The decision rule is in terms of one rectangular coordinate in the right half of the graph, while in terms of the radius (one polar coordinate) in the left half of the graph. In the latter half, the decision rule is the same as the one in (4), albeit with a smaller quantile c_α^2 instead of $\tilde{c}_\alpha^2$, which should help in achieving a power gain for τ over $\tilde{\tau}$. An analysis of power gains will follow in the next subsection.

The statistic τ leaves some open choices, just as $\tilde{\tau}$ would. For example, if $\boldsymbol{\mathcal{I}}$ were Fisher's Information, one may replace it by minus the Hessian (second derivative of the model's log-likelihood, also called Observed Information), and/or different estimates of $\boldsymbol{V}$ (e.g. H_0-restricted versus unrestricted) may be used. Under classical conditions, these choices are asymptotically equivalent under H_0. In general, however, this is not always the case; for an example in time series, see Abadir (1993b). These aspects will be discussed further in our application in Section 3.

[3] We investigate this intersection because set (5) is in terms of $\boldsymbol{h}_{2_\perp 1}(\boldsymbol{\theta})$, while H_1 is in terms of $\boldsymbol{h}_2(\boldsymbol{\theta})$.

2.2 Assessment of performance for the class of statistics, Normal case

The question arises as to what type of distribution τ takes under the standard classical assumptions that lead to asymptotic Normality of $\hat{\boldsymbol{\theta}}$; see Le Cam (1986) and also Ploberger (1999). Having orthogonalized the two components and fixed the form of τ, we consider asymptotically-optimal tests of each component. Optimal directed tests of $h_1(\boldsymbol{\theta}) = 0$ are asymptotically standard Normal under H_0, and are distributed independently from optimal tests of $\boldsymbol{h}_{2_\perp 1}(\boldsymbol{\theta}) = \mathbf{0}$ which are asymptotically χ^2_r; whereas at least one of them is not properly centred under H_1. For either hypothesis, the density of τ of (3) is the convolution of two independent variates, $\chi^2_r(\delta)$ and the square of a positive-censored $\mathrm{N}(\lambda, 1)$ where $\lambda \leq 0$. The following theorem derives the distribution function of τ under both hypotheses

$$\begin{aligned} \mathrm{H}_0 &: \lambda = \delta = 0 \\ \mathrm{H}_1 &: \lambda < 0 \text{ or } \delta \neq 0 \end{aligned} \tag{7}$$

Theorem 1 *For $\tau_1 \sim \mathrm{N}(\lambda, 1)$ independently from $\tau_2^2 \sim \chi^2_r(\delta)$, the distribution function of $\tau \equiv \tau_2^2 + 1_{\tau_1 < 0}\tau_1^2$ is*

$$G_{\lambda,\delta}(\tau) = \frac{\left(\frac{\tau}{2}\right)^{\frac{\tau}{2}}}{\sqrt{2\pi}} e^{-\frac{\delta}{2}-\frac{\tau}{2}} \sum_{j=0}^{\infty} \frac{\tau^{\frac{j}{2}} D^{-}_{j-1}(-\lambda)}{j!} \sum_{k=0}^{\infty} \frac{\Gamma\left(\frac{j}{2}+1\right)\left(\frac{\delta\tau}{4}\right)^k}{\Gamma\left(k+\frac{j+r}{2}+1\right)k!} \, {}_1F_1\left(\frac{j}{2}+1; k+\frac{j+r}{2}+1; \frac{\tau}{2}\right)$$

where $\Gamma(.)$ is the Gamma (generalized factorial) function and ${}_1F_1$ is Kummer's function. [For an introduction, see Erdélyi (1953) or Abadir (1999).] The function $D^{-}_{j-1}(-\lambda) \equiv e^{-\lambda^2/4} D_{j-1}(-\lambda)$ is the modified parabolic cylinder function, whose series expansions are derived in Abadir (1993c).

The power function of τ follows from this formula as $1 - G_{\lambda,\delta}(c^2_\alpha)$, where c^2_α is the quantile leading to a test of size α. The (limiting) distributions are the same as existing ones for the one-sided LR of, for example, Gouriéroux et al. (1982, p.68) and Conaway et al. (1990) when $r = 0$, but not for $r > 0$. Our theorem gives their explicit formulae for the first time. Strictly speaking, the LR and W tests are not applicable here to testing (7). First, the LR test would require the maximum likelihood estimation procedure to restrict the maximum likelihood estimator of λ, say $\hat{\lambda}$, to be strictly negative under the alternative H_1. Such estimates need not exist, because the parameter space is open as $\hat{\lambda} \to 0^-$. The results of Kudô (1963, pp.417-418), Wolak (1989, pp.26-27) and Silvapulle (1991, pp.388-390) concerning LR tests are therefore not usable here. Second, the same difficulty occurs with a Wald test, if one interprets it in the sense of being based on estimation under H_1. The derivations of Kodde and Palm (1986) for W would then be relating to the alternative comprising $h_1(\boldsymbol{\theta}) \leq 0$, rather than their stated $h_1(\boldsymbol{\theta}) < 0$, or we could interpret W as a pseudo-Wald test.

It is now possible to analyse the distributions $G_{\lambda,\delta}(\tau)$ and their implied densities. For example, for $j \in \mathbb{Z}$,

$$D_{j-1}^{-}(-\lambda) \equiv \sqrt{2\pi} \int \ldots \int \Phi(\lambda) (\mathrm{d}\lambda)^{-j}, \tag{8}$$

which indicates the relation of our distribution to functionals of the Normal. Note that (8) is a multiple of the Hermite polynomials when $j \in \mathbb{N}$ and negative orders of integration are interpreted as derivatives. One may also wish to establish the rates of decay of the density as τ tends to extreme values. More specifically, the leading (dominant) term for the tails of the distribution are, as $\tau \to 0$,

$$G_{\lambda,\delta}(\tau) \sim \frac{\Phi(\lambda)}{\Gamma\left(\frac{r}{2}+1\right)} \left(\frac{\tau}{2}\right)^{\frac{r}{2}} e^{-\frac{\delta}{2}-\frac{\tau}{2}}$$

which is a scaling of the distribution of χ_r^2; and, as $\tau \to \infty$,

$$G_{\lambda,\delta}(\tau) \sim \Phi\left(\lambda + \sqrt{\tau}\right) \sim 1 - \frac{\phi\left(\lambda + \sqrt{\tau}\right)}{\lambda + \sqrt{\tau}}$$

by (18) of the Appendix and by the large-argument expansion of Kummer functions. The effect of large and/or small $-\lambda, \delta$ and/or r may be analysed in a similar way.

Once a class of tests is proposed, it is natural to investigate whether it satisfies the main desirable properties, such as unbiasedness, consistency, invariance and optimality. The following theorems do so, in the Normal setup for $\hat{\boldsymbol{\theta}}$.

Theorem 2 *The power function of the test of (7) based on τ of (3) is a monotone nondecreasing function of $-\lambda$ and δ, and the test is therefore unbiased.*

Theorem 3 *The test of (7) based on τ of (3) is consistent as $\lambda \to -\infty$ or $\delta \to \infty$.*

Theorem 4 *The test of (7) based on τ of (3) is invariant to the subgroup of the Affine Group of transformations given by a Jacobian matrix which takes the block diagonal form $\operatorname{diag}(a, \boldsymbol{A})$, where $a \in \mathbb{R}_+$ and $\boldsymbol{A}$ is $r \times r$ positive definite.*

Theorem 5 *The power of the test of (7) based on τ is no less than the power of the corresponding unmodified one based on $\tilde{\tau}$, whenever the sizes (Type I errors) of the two tests are equal. This also implies that the unmodified pseudo-W, LR and LM are therefore inadmissible.*

The invariance group of Theorem 4 is obviously a restricted class of the one in Kariya and Cohen (1992). Theorems 2 and 5 stand in particular contrast with earlier results in the literature, where such performances were not necessarily achievable; e.g. see the warning remark in Perlman (1969, p.558). In the standard setting of this subsection, our procedure based on modified confidence regions achieves unbiasedness and provides uniform power gains over the unmodified counterparts. See also Eaton (1970) for the case when $\boldsymbol{h}(\boldsymbol{\theta})$ is linear, which relies on an exact global (not local) conic representation of the parameter space. Of course, the power gains are not limited to $\alpha/2$, the potential savings in Type I error. To illustrate the potential for power gains we take the following simple canonical example.

Suppose a random sample of $i = 1, 2, \ldots, n$ observations is available for $\boldsymbol{y}_i \sim \mathrm{IN}(\boldsymbol{\mu}, \boldsymbol{\Omega})$, where $\boldsymbol{\mu} \equiv [\mu_1 : \mu_2]'$ and $\boldsymbol{\Omega}$ is a known 2×2 matrix. For the equivalence of testing hypotheses in this canonical model and in standard linear regressions, see Gouriéroux et al. (1982, pp. 66-67). Let $\boldsymbol{\Omega}$ be the identity matrix of order 2, without loss of generality because of our orthogonalization inherent in the component τ_2 of (3) and (4). For both statistics, set $\alpha = 0.10$, $n = 50$, $\mu_1 = -0.3$ and $\mu_2 = 0$. Then, the power of tests based on $\tilde{\tau}$ is 0.59, while the power of tests based on τ is 0.67. These numbers have been obtained by the evaluation of the exact sizes and power functions in Theorem 1 and the noncentral χ^2 distribution; e.g. see Abadir (1999, p.300). The total time taken for both calculations was a fraction of a second on a PC running on a Pentium III processor, which illustrates how efficient these formulae are. The series in j and k of Theorem 1 converge exponentially fast; see Abadir (1999) for more details. Users of Scientific Workplace need not even programme these formulae: this typesetting package interprets the formulae in Maple and computes them numerically when required.

2.3 Assessment of performance for the class of statistics, general case

We now consider the behaviour of our class of tests when not restricting ourselves to the (asymptotically) Normal setting of the previous theorems. Less can be said analytically about the general properties in nonstandard setups (such as in the application of the following section), and we cannot resort to $G_{\lambda,\delta}(\tau)$ of Theorem 1 for investigating all the desirable properties in Theorems 2-5. We can, however, show the following.

Theorem 6 *For τ_1 a consistent tests of $h_1(\boldsymbol{\theta}) = 0$ versus $h_1(\boldsymbol{\theta}) < 0$, and τ_2^2 a consistent tests of $\boldsymbol{h}_{2_\perp 1}(\boldsymbol{\theta}) = \boldsymbol{0}$ versus $\boldsymbol{h}_{2_\perp 1}(\boldsymbol{\theta}) \neq \boldsymbol{0}$, tests based on τ of (3) are also consistent for testing H_0 versus H_1 of (1) or any subsets thereof (such as $\mathrm{H}_1^{\cap}$).*

Theorem 7 *For τ_1 and τ_2 invariant to a class $\mathcal{T}$ of data-transformations, the test based on τ of (3) is also invariant to $\mathcal{T}$. This class $\mathcal{T}$ includes the subgroup of the Orthogonal Group whose induced transformation leaves the direction of $h_1(\boldsymbol{\theta})$ unchanged.*

Our class of statistics is geared towards exploiting the one-sidedness of $h_1(\boldsymbol{\theta})$ in the hypotheses outlined in and after (1). It is really best at discriminating between H_0 and $\mathrm{H}_1^{\cap}$; because $\mathrm{H}_1^{\cap}$ is true only if $h_1(\boldsymbol{\theta}) = 0$ is violated, whereas H_1 could be true in spite of $h_1(\boldsymbol{\theta}) = 0$. This makes it particularly suited to the application of the following section, where H_0 and $\mathrm{H}_1^{\cap}$ are the most relevant parts of the parameter space in the economic applications of that model.

3 An application: tests of random walks in AR models with deterministic components

The linear model with autocorrelated errors has a long history in econometrics; e.g. see the discussion of and references on common factor models in Hendry, Pagan and Sargan (1984). We now consider a special case of it, namely Bhargava's (1986)

$$\begin{aligned} y_t &= \beta_1 + \beta_2 \left(t - 1 - \frac{n}{2}\right) + u_t, \\ u_t &= \rho u_{t-1} + \varepsilon_t, \quad t = 1, 2, \ldots, n \end{aligned} \tag{9}$$

with $\{\varepsilon_t\} \sim \mathrm{IN}(0, \sigma^2)$ and we set $u_0 = 0$. Here we choose to centre the time variable around zero to enhance the ease of parameter interpretations in terms of mean and trend of the series of interest $\{y_t\}$, but also because this will simplify the calculation of the elements of the Hessian matrix later on.

Model (9) can be more conveniently rewritten as

$$\begin{aligned} y_1 &= \beta_1 - \beta_2 \frac{n}{2} + \varepsilon_1, \\ \nabla y_t &= \gamma y_{t-1} - \beta_1 \gamma + \beta_2 (\gamma + 1) - \beta_2 \gamma \left(t - 1 - \frac{n}{2}\right) + \varepsilon_t, \quad t = 2, \ldots, n \end{aligned} \tag{10}$$

where $\gamma + 1 \equiv \rho$ and $\nabla y_t \equiv y_t - y_{t-1}$, and the parameters of the model are

$$\boldsymbol{\theta} \equiv \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{bmatrix} = \begin{bmatrix} \gamma \\ \beta_1 \\ \beta_2 \\ \sigma^2 \end{bmatrix}. \tag{11}$$

Then, we can formulate the following null hypothesis

$$\mathrm{H}_0 : \boldsymbol{h}(\boldsymbol{\theta}) \equiv \begin{bmatrix} h_1(\boldsymbol{\theta}) \\ h_2(\boldsymbol{\theta}) \end{bmatrix} = \begin{bmatrix} \gamma \\ -\beta_1 \gamma + \beta_2 (\gamma + 1) \end{bmatrix} = \mathbf{0} \tag{12}$$

versus the alternative hypothesis H_1 given by (1). The null hypothesis is that of a random walk without deterministic components, while the alternative allows for either a dynamically-stable root or deterministic components. The subset

$H_1^{\cap}$, where both components of the null are violated, is one of trend stationarity. We now need to settle a couple of choices regarding the statistic which we shall modify into our τ statistic analysed earlier.

The first choice is about using $-\boldsymbol{\mathcal{H}}$, Fisher's $\boldsymbol{\mathcal{I}}$ or some other variants. There is a growing body of evidence suggesting that procedures using the Hessian are preferable to ones using the Information matrix; e.g. see Wang and McDermott (1998) in the context of one-sided multivariate tests, and Lindsay and Li (1997). Under our null hypotheses, limiting distributions in this model are the same functionals of Brownian motions as in the model which is linear in its parameters. Therefore, orthogonalization will achieve approximate asymptotic independence of components for Hessian-based statistics but not for Information-based ones; see Abadir (1993a, pp.1068-1069; 1995a, pp.787-788), Larsson (1995) and Gonzalo and Pitarakis (1998) for the distribution functions and their relation to shifted and rescaled Normals. We shall therefore focus on modifying $\boldsymbol{\mathcal{H}}$-based statistics in our application.

The second choice is about the testing principle. LM (both $\boldsymbol{\mathcal{I}}$ and $\boldsymbol{\mathcal{H}}$-based) has lower power in this unit-root setting, because of the discontinuous and non-standard nature of limit theory for the two hypotheses; see Abadir (1993b). We therefore do not consider it further in this section. As for W, it can have some unpleasant features which affect size, such as lack of invariance; see the differential-geometric interpretations in Critchley, Marriott and Salmon (1996) for W based on $\boldsymbol{\mathcal{I}}$ (but not on $\boldsymbol{\mathcal{H}}$), and the impossibility theorems in Dufour (1997). These problems will not arise in the case studied in this section, where we use $\boldsymbol{\mathcal{H}}$-based W and the hypotheses formulate a fixed coordinate system for the time series $\{y_t\}$. Notice that the second component of the hypothesis is not regarding β_1 *per se*, so the important problem highlighted by Dufour (1997, p.1380) does not arise in our setup. In fact, as we shall see in Theorem 8 below, inference for (12) is invariant to β_1. Also notice that the parameters are always identified. This is true for β_1 which is identified through the existence of a separate treatment of the initial condition y_1, though it cannot be consistently estimated when H_0 is true.

We now need to set up the likelihood before analysing estimators and test statistics. The log-likelihood function is

$$
\begin{aligned}
(13)\quad \ell(\boldsymbol{\theta}) \;=\; & -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\left(\sigma^2\right) - \frac{\left(y_1 - \beta_1 + \beta_2 \frac{n}{2}\right)^2}{2\sigma^2} \\
& - \sum_{t=2}^{n} \frac{\left[\triangledown y_t - \gamma y_{t-1} + \beta_1\gamma - \beta_2(\gamma+1) + \beta_2\gamma\left(t - 1 - \frac{n}{2}\right)\right]^2}{2\sigma^2}.
\end{aligned}
$$

The Hessian with respect to the parameters (11) of the model is given by

$$\boldsymbol{\mathcal{H}} \equiv \begin{bmatrix} \mathcal{H}_{11} & \mathcal{H}_{12} & \mathcal{H}_{13} & \mathcal{H}_{14} \\ \mathcal{H}_{12} & \mathcal{H}_{22} & \mathcal{H}_{23} & \mathcal{H}_{24} \\ \mathcal{H}_{13} & \mathcal{H}_{23} & \mathcal{H}_{33} & \mathcal{H}_{34} \\ \mathcal{H}_{14} & \mathcal{H}_{24} & \mathcal{H}_{34} & \mathcal{H}_{44} \end{bmatrix}, \tag{14}$$

where, by using $\sum_{t=2}^{n}\left(t-1-\frac{n}{2}\right)=0$, we have

$$\begin{aligned} \mathcal{H}_{11} &= \frac{1}{\sigma^2}\Bigg[-\left(\sum_{t=2}^{n} y_{t-1}^2\right)+2\left(\beta_1-\beta_2\right)\left(\sum_{t=2}^{n} y_{t-1}\right)+2\beta_2\left(\sum_{t=2}^{n}(t-1-\frac{n}{2})y_{t-1}\right) \\ &\quad -(n-1)\left(\beta_1-\beta_2\right)^2-\beta_2^2\sum_{t=2}^{n}(t-1-\frac{n}{2})^2\Bigg], \end{aligned}$$

$$\mathcal{H}_{12} = \frac{1}{\sigma^2}\left[2\gamma\left(\sum_{t=2}^{n} y_{t-1}\right)-\left(\sum_{t=2}^{n}\nabla y_t\right)-2\beta_1\gamma\left(n-1\right)+\beta_2\left(n-1\right)\left(2\gamma+1\right)\right]$$

$$\begin{aligned} \mathcal{H}_{13} &= \frac{1}{\sigma^2}\Bigg[-\left(2\gamma+1\right)\left(\sum_{t=2}^{n} y_{t-1}\right)+\left(\sum_{t=2}^{n}\nabla y_t\right)+2\gamma\left(\sum_{t=2}^{n}(t-1-\frac{n}{2})y_{t-1}\right) \\ &\quad -\left(\sum_{t=2}^{n}(t-1-\frac{n}{2})\nabla y_t\right)+(n-1)\beta_1(2\gamma+1) \\ &\quad -2(n-1)\beta_2(\gamma+1)-2\beta_2\gamma\sum_{t=2}^{n}(t-1-\frac{n}{2})^2\Bigg], \end{aligned}$$

$$\mathcal{H}_{22} = \frac{1}{\sigma^2}\left[-\left(n-1\right)\gamma^2-1\right],$$

$$\mathcal{H}_{23} = \frac{1}{\sigma^2}\left[(n-1)\gamma(\gamma+1)+\frac{n}{2}\right],$$

$$\mathcal{H}_{33} = \frac{1}{\sigma^2}\left[-(n-1)(\gamma+1)^2-\frac{n^2}{4}-\gamma^2\sum_{t=2}^{n}(t-1-\frac{n}{2})^2\right].$$

By Normality of $\{\varepsilon_t\}$ and the corresponding least-squares orthogonality decomposition, $\hat{\boldsymbol{\mathcal{H}}}$ is block-diagonal with respect to the last parameter σ^2. Moreover, $\partial h_{\bullet}\left(\boldsymbol{\theta}\right)/\partial\sigma^2=0$. So, one need not work out $\mathcal{H}_{\bullet 4}$ explicitly for the purpose of the subsequent analysis of our H_0.

We maximize (13) numerically, obtaining the estimates of the parameters (11). We can write accordingly the Hessian-based Wald statistic, $\mathrm{W}_{\mathcal{H}}$, as

$$\mathrm{W}_{\mathcal{H}} = \boldsymbol{h}\left(\hat{\boldsymbol{\theta}}\right)'\hat{\boldsymbol{V}}^{-1}\boldsymbol{h}\left(\hat{\boldsymbol{\theta}}\right),$$

where

$$\hat{\boldsymbol{V}} = \left[\begin{array}{cc} \hat{v}_{11} & \hat{v}_{21} \\ \hat{v}_{21} & \hat{v}_{22} \end{array} \right] = \left.\frac{\partial \boldsymbol{h}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \left(-\hat{\boldsymbol{\mathcal{H}}}\right)^{-1} \left.\frac{\partial \boldsymbol{h}(\boldsymbol{\theta})'}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}.$$

Now, we modify $\mathrm{W}_{\mathcal{H}}$ into the τ statistic defined in Section 2. After having calculated

$$\widehat{h_{2_\perp 1}\left(\hat{\boldsymbol{\theta}}\right)} \equiv h_2\left(\hat{\boldsymbol{\theta}}\right) - \frac{\hat{v}_{21} h_1\left(\hat{\boldsymbol{\theta}}\right)}{\hat{v}_{11}},$$

the test statistic τ is given by

$$\begin{aligned}
\tau &= \left[\left.\frac{\partial h_{2_\perp 1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \left(-\hat{\boldsymbol{\mathcal{H}}}\right)^{-1} \left.\frac{\partial h_{2_\perp 1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \right]^{-1} \left(\widehat{h_{2_\perp 1}\left(\hat{\boldsymbol{\theta}}\right)}\right)^2 \\
&\quad + 1_{h_1(\hat{\boldsymbol{\theta}})<0} \left[\left.\frac{\partial h_1(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \left(-\hat{\boldsymbol{\mathcal{H}}}\right)^{-1} \left.\frac{\partial h_1(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \right]^{-1} \left(h_1\left(\hat{\boldsymbol{\theta}}\right)\right)^2 \\
&= \left[\left.\frac{\partial h_{2_\perp 1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \left(-\hat{\boldsymbol{\mathcal{H}}}\right)^{-1} \left.\frac{\partial h_{2_\perp 1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \right]^{-1} \left(\widehat{h_{2_\perp 1}\left(\hat{\boldsymbol{\theta}}\right)}\right)^2 + 1_{\hat{\gamma}<0} \left(-\hat{\mathcal{H}}^{11}\right)^{-1} \hat{\gamma}^2,
\end{aligned}$$

where $\mathcal{H}^{11}$ refers to the first diagonal element of the inverse of (14). We finally need a theorem deriving the explicit expression for the function

$$h_{2_\perp 1}(\boldsymbol{\theta}) \equiv h_2(\boldsymbol{\theta}) - \frac{v_{21} h_1(\boldsymbol{\theta})}{v_{11}},$$

where $\boldsymbol{V}$ is now based on the negative of the Hessian, instead of $\boldsymbol{\mathcal{I}}$ used at the beginning of Section 2. Recall from the discussion earlier in this section that the choice between $\boldsymbol{\mathcal{I}}$ and $\boldsymbol{\mathcal{H}}$ is only needed for the construction of the statistic, and does not alter the definition of the hypotheses to be tested. This distinction should be borne in mind because the components of the Hessian-based $\boldsymbol{V}$ are only 'fixed' after conditioning on $\boldsymbol{\mathcal{H}}$.

Theorem 8 *The function $h_{2_\perp 1}(\boldsymbol{\theta})$ based on the Hessian matrix is*

$$\begin{aligned}
h_{2_\perp 1}(\boldsymbol{\theta}) &= \left[n(n-1)(n-2) S_1 \gamma^5 + n(n-1)(n-2)(\beta_2(n-1) - S_2)\gamma^4 \right. \\
&\quad +3\,(n-2)((n-2)S_1 + 2S_3)\gamma^3 + \left(\beta_2\left(4n^3 - 18n^2 + 26n - 12\right)\right. \\
&\quad \left. -12(n-2)S_1 - 3(n-2)^2 S_2 - 12S_3 - 6(n-2)S_4\right)\gamma^2 \\
&\quad \left. +6\,(2S_1 - 2\beta_2(n-1)(n-2) + (n-2)S_2 + 2S_4)\gamma + 12\beta_2\,(n-1)\right] \\
&\quad \div \left[(n-1)\left(12 + n(n-1)(n-2)\gamma^4 + 2\,(2n-3)(n-2)\gamma^2 - 12(n-2)\gamma\right)\right],
\end{aligned}$$

where

$$S_1 \equiv \sum_{t=2}^{n} \sum_{j=0}^{t-3} (\gamma+1)^j \varepsilon_{t-1-j},$$

$$S_2 \equiv \sum_{t=2}^{n} \varepsilon_t,$$

$$S_3 \equiv \sum_{t=2}^{n} \sum_{j=0}^{t-3} (\gamma+1)^j \left(t-1-\frac{n}{2}\right) \varepsilon_{t-1-j},$$

$$S_4 \equiv \sum_{t=2}^{n} \left(t-1-\frac{n}{2}\right) \varepsilon_t.$$

One of the implications of this theorem is to complete the proof of invariance, with respect to β_1, of both components $h_1(.)$ and $h_{2\perp 1}(.)$ of our test τ. The former invariance follows from the definition $h_1(\boldsymbol{\theta}) = \gamma$ from (12), while the latter follows from this theorem.

A Monte-Carlo experiment based on 100,000 replications has been carried out to evaluate the power of the proposed testing procedures for

$$\begin{aligned} \rho &= 0.6, 0.7, 0.8, 0.9, 0.95, 0.99, 1 \\ \beta_2 &= 0, 0.25, 0.5 \\ n &= 25, 50, 100 \\ \alpha &= 5\%, 10\% \end{aligned}$$

where α is the size of the test. By the invariance to β_1 implied by Theorem 8, we set $\beta_1 = 0$ in our simulations. Table I summarizes the resulting quantiles of the $\mathrm{W}_{\mathcal{H}}$ and τ statistics. Furthermore, Table II expands the simulation to a larger number of sample sizes, in order to fit response surfaces for the 5% and 10% quantiles. These are approximated parsimoniously by the following functional forms:

$$c_{0.05}^2 \simeq 4.7945 \exp\left(\frac{9.1869}{n} - \frac{3.5652}{n^2}\right) \tag{15}$$

for the 5% critical value, and

$$c_{0.10}^2 \simeq 3.7670 \exp\left(\frac{8.0603}{n} + \frac{26.0097}{n^2}\right) \tag{16}$$

for the 10% critical value. As Table II shows the approximation is very accurate, and can be used by practitioners to generate quantiles for any sample size. These exponential functional forms have been used in the related context of approximating unit-root estimators' moments in Abadir (1995b), based on derivations in earlier work by the same author.

In Table III, we report the values of the power of the $\mathrm{W}_{\mathcal{H}}$ and τ tests. They are qualitatively analogous to the powers based on a size of 10%, which are not reported here for space considerations. The one-sided modification of $\mathrm{W}_{\mathcal{H}}$ into τ achieves uniform power gains, some of them substantial (e.g. 0.57 to 0.63, when $\rho = 0.6$ and $\beta_2 = 0$). When the time trend vanishes ($\beta_2 \to 0$) and the AR root is close to unit boundary ($\rho \to 1^-$), the usual biases of unit root tests reappear, thought they are less severe in the case of our modified τ test.

Finally, the model we have considered can be extended in many different directions, with the same qualitative conclusions holding. For example, one may wish to consider higher-order AR models or correct for higher order dynamics nonparametrically; see the parametric reformulation of high-order ARs into an AR(1) in Dickey and Fuller (1981), or the nonparametric Phillips and Perron (1988) type of correction.

4 Concluding comments

In illustrating our new class of tests, we have chosen an application involving a model which is nonlinear in its parameters. There are many other possibly simpler applications to which our method may be applied, but they are too many to tackle in this single paper. The applications are not restricted to time series either. For example, Silvapulle's (1991) paper was motivated by a model of limited dependent variables, and our methods can be applied to microeconometric problems too. Finally, other potential applications to nonstandard problems (i.e. ones with non-Normal underlying distribution theory) include:

1. Conditional heteroskedasticity models; e.g. see Bollerslev, Engle and Nelson (1994). For example, one-sided inference on the sum of the GARCH roots may be required jointly with inference on other components of the model, or vice versa. More specifically, in testing the market efficiency hypothesis in financial data, many authors have employed AR models with GARCH disturbances. The new procedure can provide tests for the null hypothesis of a unit root and no ARCH effect, versus the alternative that the model is (asymptotically) stationary or that there is some Auto-Regressive Conditional Heteroskedasticity in the data.

2. One more application that can be studied is inference in the type of random-coefficients model introduced by McCabe and Tremayne (1995), extended by Smith and Taylor (2001), and used in the empirical study by Leybourne, McCabe and Tremayne (1996) where the model is applied successfully to important economic datasets. Theirs is a model where the autoregressive parameter is stochastic, and distributed around a fixed mean of 1. Their H_0 is that of an exact (deterministic) unit root as opposed to H_1 of a random root with mean 1. Fixing this latter mean to be 1 is restrictive, and the alternative can be generalized to allow for a random root with mean less than 1.

3. Yet another possible application is to test for structural breaks [e.g. see Perron (1989) and the large subsequent literature] when a specific direction is suspected for one of the breaks. This is often the case in practice, when one identifies an event which has caused a break with the past of the process

to occur. For example, regarding the oil shock which happened in the early 1970's, the suspected direction of its effect was not positive. Ceteris paribus, the shock could not have been a stimulus to consumption in the short to medium run. Similarly for the possible break induced by the October 1997 stock market crash.

References

Abadir, K. M. (1993a): "The limiting distribution of the autocorrelation coefficient under a unit root," *Annals of Statistics*, 21, 1058-1070.

——— (1993b): "On the asymptotic power of unit root tests," *Econometric Theory*, 9, 189-221.

——— (1993c): "Expansions for some confluent hypergeometric functions," *Journal of Physics, Series A*, 26, 4059-4066 [Corrigendum for printing error (1993) 7663].

——— (1995a): "The limiting distribution of the t ratio under a unit root," *Econometric Theory*, 11, 775-793.

——— (1995b): "Unbiased estimation as a solution to testing for random walks," *Economics Letters*, 47, 263-268.

——— (1999): "An introduction to hypergeometric functions for economists," *Econometric Reviews*, 18, 287-330.

Andrews, D. W. K. (1998): "Hypothesis testing with a restricted parameter space," *Journal of Econometrics*, 84, 155-199.

Berger, R. L. (1989): "Uniformly more powerful tests for hypotheses concerning linear inequalities and normal means," *Journal of the American Statistical Association*, 84, 192-199.

Bhargava, A. (1986): "On the theory of testing for unit roots in observed time series," *Review of Economic Studies*, 53, 369-384.

Bollerslev, T., R. F. Engle and D. B. Nelson (1994): "ARCH models," in *Handbook of Econometrics, vol.4*, ed. by R. F. Engle and D. L. McFadden. Amsterdam: North-Holland, Ch.49.

Catchpole, E. A. and B. J. T. Morgan (1997): "Detecting parameter redundancy," *Biometrika*, 84, 187-196.

Conaway, M., C. Pillers, T. Robertson and J. Sconing (1990): "The power of the circular cone test: a noncentral chi-bar-squared distribution," *The Canadian Journal of Statistics*, 18, 63-70.

Cox, D. R., and D. V. Hinkley (1974): *Theoretical statistics.* London: Chapman & Hall.

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

Dickey, D. A., and W. A. Fuller (1981): "Likelihood ratio statistics for autoregressive time series with a unit root," *Econometrica*, 49, 1057-1072.

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

——— (1997): "Some impossibility theorems in econometrics with applications to structural and dynamic models," *Econometrica*, 65, 1365-1387.

Eaton, M. L. (1970): "A complete class theorem for multidimensional one-sided alternatives," *Annals of Mathematical Statistics*, 41, 1884-1888.

Erdélyi, A. (Ed.) (1953): *Higher transcendental functions, volumes 1-2.* New York: Mc.Graw-Hill.

Gonzalo, J. and J.-Y. Pitarakis (1998): "On the exact moments of asymptotic distributions in an unstable AR(1) with dependent errors," *International Economic Review*, 39, 71-88.

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

Hendry, D. F., A. R. Pagan and J. D. Sargan (1984): "Dynamic specification," in *Handbook of Econometrics, vol.2*, ed. by Z. Griliches and M. D. Intriligator. Amsterdam: North-Holland, Ch.18.

Kariya, T. and A. Cohen (1992): "On the invariance structure of the one-sided testing problem for a multivariate normal mean," *Statistica Sinica*, 2, 221-236.

Kim, T. H. and P. Newbold (2001): "Unit root tests based on inequality-restricted estimators," School of Economics, University of Nottingham, mimeo.

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

Kudô, A. (1963): "A multivariate analogue of the one-sided test," *Biometrika*, 50, 403-418.

Larsson, R. (1995): "The asymptotic distribution of some test statistics in near-integrated AR processes," *Econometric Theory*, 11, 306-330.

Le Cam, L. (1986): *Asymptotic methods in statistical decision theory.* Berlin: Springer-Verlag.

Lehmann, E. L. (1986): *Testing statistical hypotheses.* New York: John Wiley and sons.

Leybourne, S. J., B. P. M. McCabe and A. R. Tremayne (1996): "Can economic time series be differenced to stationarity?" *Journal of Business and Economic Statistics*, 14, 435-446.

Lindsay, B. G. and B. Li (1997): "On second-order optimality of the observed Fisher information," *Annals of Statistics*, 25, 2172-2199.

Liu, H. and R. L. Berger (1995): "Uniformly more powerful, one-sided tests for hypotheses about linear inequalities," *Annals of Statistics*, 23, 55-72.

McCabe, B. P. M. and A. R. Tremayne (1995): "Testing a time series for difference stationarity," *Annals of Statistics*, 23, 1015-1028.

Mood, A. M., F. A. Graybill, and D. Boes (1974): *Introduction to the Theory of Statistics, 3rd edn.* Singapore: McGraw-Hill.

Perlman, M. D. (1969): "One-sided testing problems in multivariate analysis," *Annals of Mathematical Statistics*, 40, 549-567 [Correction (1971), 1777].

Perlman, M. D. and L. Wu (1999): "The emperor's new tests," *Statistical Science*, 14, 355-381 [with discussion].

Perron, P. (1989): "The great crash, the oil price shock, and the unit root hypothesis," *Econometrica*, 57, 1361-1401.

Phillips, P. C. B. and P. Perron (1988): "Testing for a unit root in time series regression," *Biometrika*, 75, 335-346.

Ploberger, W. (1999): "A complete class of tests when the likelihood is locally asymptotically quadratic," Department of Economics, University of Rochester, mimeo.

Rogers, A. J. (1986): "Modified Lagrange multiplier tests for problems with one-sided alternatives," *Journal of Econometrics*, 31, 341-361.

Rothenberg, T. J. (1971): "Identification in parametric models," *Econometrica*, 39, 577-591.

Savin, N. E. (1984): "Multiple hypothesis testing," in *Handbook of Econometrics, vol.2*, ed. by Z. Griliches and M. D. Intriligator. Amsterdam: North-Holland, Ch.14.

Silvapulle, M. J. (1991): "On limited dependent variable models: maximum likelihood estimation and test of one-sided hypothesis," *Econometric Theory*, 7, 385-395.

Smith, R. J. and A. M. R. Taylor (2001): "Tests of the seasonal unit root hypothesis against heteroscedastic seasonal integration," *Journal of Business and Economic Statistics*, forthcoming.

Wang, Y. and M. P. McDermott (1998): "A conditional test for a non-negative mean vector based on a Hotelling's T^2-type statistic," *Journal of Multivariate Analysis, 66, 64-70.*

Wolak, F. A. (1989): "Local and global testing of linear and nonlinear inequality constraints in nonlinear econometric models," *Econometric Theory*, 5,1-35.

APPENDIX

Proof of Theorem 1: We will use freely standard statistical theorems that can be found in, for example, Mood, Graybill and Boes (1974), and the transcendental functions and their properties in Erdélyi (1953) or Abadir (1999). We denote the generalized hypergeometric function by ${}_pF_q$.

First, we need the distribution of the square of a positive-censored $\mathrm{N}(\lambda, 1)$, say $\xi \equiv 1_{\tau_1 \leq 0}\tau_1^2 \in [0, \infty)$, which is $\Phi\left(\lambda + \sqrt{\xi}\right)$. Knowing that the density of a

$\chi_r^2(\delta)$ variate, say ζ, is

$$\frac{1}{2\Gamma\left(\frac{r}{2}\right)}\left(\frac{\zeta}{2}\right)^{\frac{r}{2}-1}e^{-\frac{\delta}{2}-\frac{\zeta}{2}}\,{}_0F_1\left(\frac{r}{2};\frac{\delta}{4}\zeta\right),$$

we can then apply the convolution theorem for independent variates to $\tau\equiv\zeta+\xi$ as

$$(17)\qquad G_{\lambda,\delta}(\tau)=\frac{1}{2\Gamma\left(\frac{r}{2}\right)}e^{-\frac{\delta}{2}}\int_0^\tau\Phi\left(\lambda+\sqrt{\tau-\zeta}\right)\left(\frac{\zeta}{2}\right)^{\frac{r}{2}-1}e^{-\frac{\zeta}{2}}\,{}_0F_1\left(\frac{r}{2};\frac{\delta}{4}\zeta\right)\mathrm{d}\zeta$$
$$=\frac{1}{2\Gamma\left(\frac{r}{2}\right)}e^{-\frac{\delta}{2}-\frac{\tau}{2}}\int_0^\tau\Phi\left(\lambda+\sqrt{\zeta}\right)\left(\frac{\tau-\zeta}{2}\right)^{\frac{r}{2}-1}e^{\frac{\zeta}{2}}\,{}_0F_1\left(\frac{r}{2};\frac{\delta}{4}(\tau-\zeta)\right)\mathrm{d}\zeta$$

by a change of variable. By the addition theorem

$$(18)\qquad \Phi\left(\lambda+\sqrt{\zeta}\right)\equiv\frac{1}{\sqrt{2\pi}}D_{-1}^{-}\left(-\lambda-\sqrt{\zeta}\right)=\frac{1}{\sqrt{2\pi}}\sum_{j=0}^{\infty}\frac{D_{j-1}^{-}(-\lambda)}{j!}\zeta^{\frac{j}{2}}$$

and the expansion of ${}_0F_1$, we can write

$$G_{\lambda,\delta}(\tau)=\frac{1}{\sqrt{8\pi}}e^{-\frac{\delta}{2}-\frac{\tau}{2}}\sum_{j=0}^{\infty}\frac{D_{j-1}^{-}(-\lambda)}{j!}\sum_{k=0}^{\infty}\frac{\left(\frac{\delta}{2}\right)^k}{\Gamma\left(k+\frac{r}{2}\right)k!}\int_0^\tau\zeta^{\frac{j}{2}}\left(\frac{\tau-\zeta}{2}\right)^{k+\frac{r}{2}-1}e^{\frac{\zeta}{2}}\mathrm{d}\zeta$$
$$=\frac{\left(\frac{\tau}{2}\right)^{\frac{r}{2}}}{\sqrt{2\pi}}e^{-\frac{\delta}{2}-\frac{\tau}{2}}\sum_{j=0}^{\infty}\frac{\tau^{\frac{j}{2}}D_{j-1}^{-}(-\lambda)}{j!}\sum_{k=0}^{\infty}\frac{\Gamma\left(\frac{j}{2}+1\right)\left(\frac{\delta\tau}{4}\right)^k}{\Gamma\left(k+\frac{j+r}{2}+1\right)k!}\,{}_1F_1\left(\frac{j}{2}+1;k+\frac{j+r}{2}+1;\frac{\tau}{2}\right)$$

by the integral representation of ${}_1F_1$. By analysing the orders of magnitude in terms of j and k, the double series are absolutely convergent for all values of the parameters and variable. Q.E.D.

PROOF OF THEOREM 2: The power function is given by $1-G_{\lambda,\delta}\left(c_\alpha^2\right)$ when the size α of the test is such that $c_\alpha^2>0$. From (17), its derivative with respect to $-\lambda$ is

$$\frac{1}{2\Gamma\left(\frac{r}{2}\right)}e^{-\frac{\delta}{2}-\frac{\tau}{2}}\int_0^\tau\phi\left(\lambda+\sqrt{\zeta}\right)\left(\frac{\tau-\zeta}{2}\right)^{\frac{r}{2}-1}e^{\frac{\zeta}{2}}\,{}_0F_1\left(\frac{r}{2};\frac{\delta}{4}(\tau-\zeta)\right)\mathrm{d}\zeta.$$

This integral is nonnegative everywhere, which establishes the required monotonicity with respect to $-\lambda$. The monotonicity with respect to δ follows by rewriting the convolution with the roles of ξ and ζ swapped, and noting that the cumulative distribution function of a $\chi_r^2(\delta)$ is a decreasing function of δ. Unbiasedness follows by definition. Q.E.D.

Proof of Theorem 3: Consider (17). As $\lambda \to -\infty$, we have $\Phi\left(\lambda + \sqrt{\zeta}\right) \to 0$ so that $1 - G_{\lambda,\delta}\left(c_\alpha^2\right) \to 1$. The monotonicity with respect to δ follows by rewriting the convolution with the roles of ξ and ζ swapped, and noting that the cumulative distribution function of a $\chi_r^2(\delta)$ tends to zero as $\delta \to \infty$. Q.E.D.

Proof of Theorem 4: The subgroup mentioned in the theorem can be represented by the transformation

$$\hat{\boldsymbol{\theta}} \mapsto \boldsymbol{b} + \operatorname{diag}(a, \boldsymbol{A})\,\hat{\boldsymbol{\theta}} \sim \mathrm{N}\left(\boldsymbol{b} + \operatorname{diag}(a, \boldsymbol{A})\,\boldsymbol{\theta}, \operatorname{diag}(a, \boldsymbol{A})\,\boldsymbol{\mathcal{I}}^{-1}\operatorname{diag}(a, \boldsymbol{A}')\right)$$

where $\boldsymbol{b}$ is $(r+1) \times 1$. The direction of the first component of the mean vector is preserved, and the orthogonalization inherent in τ_2 preserves the independence of the components of the test τ. Q.E.D.

Proof of Theorem 5: The unmodified statistics are deterministic monotone transforms of $\tilde{\tau} \equiv \tau_1^2 + \tau_2^2$. The distribution of $\tilde{\tau}$ is the convolution of a $\chi_1^2\left(\lambda^2\right)$ with an independent $\chi_r^2(\delta)$, which is a $\chi_{r+1}^2\left(\lambda^2 + \delta\right)$. The distribution function of $\tilde{\xi} \equiv \tau_1^2$ is

$$\Phi\left(\lambda + \sqrt{\tilde{\xi}}\right) - \Phi\left(\lambda - \sqrt{\tilde{\xi}}\right);$$

e.g. see Abadir (1999, p.300). For the comparisons to follow, derivations similar to the ones in Theorem 1 allow us to write $\chi_{r+1}^2\left(\lambda^2 + \delta\right)$ as the convolution

$$(19)\quad \tilde{G}_{\lambda,\delta}(\tilde{\tau}) = \frac{1}{2\Gamma\left(\frac{r}{2}\right)} e^{-\frac{\delta}{2}-\frac{\tilde{\tau}}{2}} \int_0^{\tilde{\tau}} \left[\Phi\left(\lambda + \sqrt{\zeta}\right) - \Phi\left(\lambda - \sqrt{\zeta}\right)\right] \left(\frac{\tilde{\tau} - \zeta}{2}\right)^{\frac{r}{2}-1} e^{\frac{\zeta}{2}}\, {}_0F_1\left(\frac{r}{2}; \frac{\delta}{4}(\tilde{\tau} - \zeta)\right) \mathrm{d}\zeta.$$

The variate $\tilde{\tau}$ stochastically-dominates τ, because $\tilde{\tau} = \tau + 1_{\tau_1>0}\tau_1^2 > \tau$ where the inequality holds almost surely (i.e. with probability 1). Recalling $G_{\lambda,\delta}(\tau)$ of (17), the excess of the power function of τ over that of $\tilde{\tau}$ is given by

$$\Xi(\lambda, \delta) \equiv \left[1 - G_{\lambda,\delta}\left(c_\alpha^2\right)\right] - \left[1 - \tilde{G}_{\lambda,\delta}\left(\tilde{c}_\alpha^2\right)\right] = \tilde{G}_{\lambda,\delta}\left(\tilde{c}_\alpha^2\right) - G_{\lambda,\delta}\left(c_\alpha^2\right)$$

where $\tilde{c}_\alpha^2 > c_\alpha^2$. The equality of the sizes of the two tests gives $G_{0,0}\left(c_\alpha^2\right) = \tilde{G}_{0,0}\left(\tilde{c}_\alpha^2\right) = 1 - \alpha$, hence $\Xi(0,0) = 0$. We will now show that, for $-\lambda$ and δ finite and nonnegative, the difference $\Xi(\lambda, \delta)$ is minimized at $\Xi(0,0)$, so that $\Xi(\lambda, \delta) \geq 0$. A direct proof of this inequality seems hard to establish, so we adopt an indirect route.

Consider the size-unadjusted difference of power functions

$$\tilde{\Xi}(\lambda, \delta, \tilde{\tau}) \equiv \tilde{G}_{\lambda,\delta}(\tilde{\tau}) - G_{\lambda,\delta}\left(c_\alpha^2\right),$$

where the components are given by (19) and (17), respectively. Recall that, by the definition of a c.d.f., $\tilde{G}(.)$ and $G(.)$ are monotone functions of their arguments, e.g. $\tilde{\tau}$, as they vary. We will now show that the magnitude of the adjustment of $\tilde{\tau}$ in order to go from $\tilde{\Xi}(\lambda, \delta, c_\alpha^2) < 0$ to $\tilde{\Xi}(\lambda, \delta, \tilde{\tau}) = 0$ is maximized when $\lambda = \delta = 0$, i.e. under the null hypothesis where $\Xi(0,0) = 0$. Therefore, making the sizes of the two tests equal will lead to $\Xi(\lambda, \delta) > 0$ for all positive and finite $-\lambda$ and δ. One may wish to visualize this algebraic manipulation as a size adjustment of the power function of the $\tilde{\tau}$-based test, with the 'horizontal' axes being $-\lambda \in \mathbb{R}_+$ and $\delta \in \mathbb{R}_+$. This adjustment shifts the function horizontally until it touches the power function of the τ-based test at any chosen coordinates of λ and δ. It will be shown that the maximal such displacement of this function is the one that leads to the sizes of the two tests being equal, so that the size-adjusted power function of the $\tilde{\tau}$-based test lies below that of τ everywhere in the space where λ and δ are finite.

We need to solve the optimization problem

$$\max_{\lambda,\delta} \tilde{\tau} \quad \text{subject to } \tilde{G}_{\lambda,\delta}(\tilde{\tau}) = G_{\lambda,\delta}\left(c_\alpha^2\right) \tag{20}$$

for $-\lambda$ and δ finite and nonnegative. We know that equality is achieved for the quantile $\tilde{\tau} = \tilde{c}_\alpha^2$ when $\lambda = \delta = 0$, which is now our starting point. Recalling (17) and (19), a more negative λ reduces the extent of the inequality

$$e^{-\lambda\sqrt{\zeta}} > e^{-\lambda\sqrt{\zeta}} - e^{\lambda\sqrt{\zeta}} \tag{21}$$

which implies that the limit of integration $\tilde{\tau}$ needs to be reduced in (19) to keep the equality $\tilde{G}_{\lambda,\delta}(\tilde{\tau}) = G_{\lambda,\delta}(c_\alpha^2)$ as λ decreases. This proves that the maximum $\tilde{\tau}$ in (20) is achieved for $\lambda = 0$. As for $\delta > 0$, rewrite the convolutions (17) and (19) with the roles of τ_1 and τ_2 swapped. As δ increases, the distribution function $\chi_r^2(\delta)$ of τ_2^2 declines. This dampens the inequality of powers caused by (21), thus requiring a smaller $\tilde{\tau}$ to equate them. This proves that the maximum $\tilde{\tau}$ in (20) is achieved for $\delta = 0$. Q.E.D.

Proof of Theorem 6: Consider the situation of H_0 not true. When $\boldsymbol{v}_{21} = \mathbf{0}$, either (or both) $1_{\tau_1<0}\tau_1^2$ or τ_2^2 diverge to ∞. When $\boldsymbol{v}_{21} \neq \mathbf{0}$, the component τ_2^2 diverges except on the manifold $\boldsymbol{h}_{2\perp 1}(\boldsymbol{\theta}) = \mathbf{0}$. But in this case, $1_{\tau_1<0}\tau_1^2 \xrightarrow{p} \infty$, so that $\tau \xrightarrow{p} \infty$ too. Q.E.D.

Proof of Theorem 7: The proof follows directly from the definitions of τ_1 and τ_2, by the orthogonalization to $h_1(.)$ inherent in τ_2. Q.E.D.

Proof of Theorem 8: We can solve the model for y_t in terms of y_1 and the sequence $\{\varepsilon_t\}$ by recursive substitution. We obtain

$$
\begin{aligned}
(22) \qquad y_t &= \rho^{t-1} y_1 + [\beta_1 (1-\rho) + \beta_2 \rho] \left[\sum_{j=0}^{t-2} \rho^j \right] \\
&\quad + \beta_2 (1-\rho) \left[\sum_{j=0}^{t-2} \left(t - 1 - \frac{n}{2} - j \right) \rho^j \right] + \sum_{j=0}^{t-2} \rho^j \varepsilon_{t-j}.
\end{aligned}
$$

For $\rho \neq 1$, the deterministic summations can be simplified by means of geometric series, giving

$$
\begin{aligned}
y_t &= \rho^{t-1} y_1 + [\beta_1 (1-\rho) + \beta_2 \rho] \frac{\rho^{t-1} - 1}{\rho - 1} \\
&\quad + \beta_2 (1-\rho) \left[\frac{t - 1 - n/2}{1-\rho} - \frac{\rho (1 - \rho^{t-2})}{(1-\rho)^2} - \frac{(1 - n/2) \rho^{t-1}}{1-\rho} \right] + \sum_{j=0}^{t-2} \rho^j \varepsilon_{t-j},
\end{aligned}
$$

and

$$
\begin{aligned}
\nabla y_t &= y_1 \rho^{t-2} (\rho - 1) + [\beta_1 (1-\rho) + \beta_2 \rho] \rho^{t-2} \\
&\quad + \beta_2 (1-\rho) \left[\frac{1}{1-\rho} - \frac{\rho^{t-2}}{(1-\rho)} + \left(1 - \frac{n}{2} \right) \rho^{t-2} \right] + \varepsilon_t + (\rho - 1) \sum_{j=0}^{t-3} \rho^j \varepsilon_{t-1-j}.
\end{aligned}
$$

The result follows by substituting into the components of $\boldsymbol{\mathcal{H}}$ and simplifying. For the simpler case of $\rho = 1$, summing the deterministic components in (22) and proceeding in the same way gives the same formula for $h_{2_{\perp}1}(\boldsymbol{\theta})$, but with $\gamma = \rho - 1 = 0$ substituted in. Q.E.D.

TABLE I.
Quantiles of τ and of $W_{\mathcal{H}}$ [in brackets].

n	1%	2.5%	5%	10%	50%	90%	95%	97.5%	99%
25	0.18	0.26	0.35	0.49	1.89	5.41	6.88	8.42	10.54
	[0.22]	[0.29]	[0.38]	[0.53]	[1.97]	[5.60]	[7.16]	[8.76]	[11.00]
50	0.20	0.27	0.36	0.51	1.63	4.50	5.77	7.05	8.81
	[0.22]	[0.29]	[0.38]	[0.53]	[1.67]	[4.67]	[6.05]	[7.53]	[9.75]
100	0.20	0.28	0.37	0.52	1.57	4.08	5.25	6.52	8.71
	[0.22]	[0.29]	[0.38]	[0.53]	[1.59]	[4.18]	[5.46]	[7.00]	[10.29]

TABLE II.
Simulated and approximate critical values of τ.

n	5% Critical values		10% Critical values	
	Simulated	Approx.	Simulated	Approx.
25	6.88	6.88	5.41	5.42
50	5.77	5.75	4.50	4.47
100	5.25	5.25	4.08	4.09
150	5.09	5.10	3.97	3.98
200	5.01	5.02	3.90	3.92
300	4.92	4.94	3.86	3.87
400	4.89	4.91	3.82	3.84
500	4.91	4.88	3.81	3.83
600	4.87	4.87	3.83	3.82
700	4.90	4.86	3.84	3.81
800	4.85	4.85	3.83	3.81
1000	4.81	4.84	3.81	3.80

TABLE III.
Powers of τ and of W_B [in brackets], both for a size of 5%.

ρ	$\beta_2 = 0$			$\beta_2 = 0.25$			$\beta_2 = 0.5$		
	$n = 25$	$n = 50$	$n = 100$	$n = 25$	$n = 50$	$n = 100$	$n = 25$	$n = 50$	$n = 100$
0.6	0.15 [0.13]	0.63 [0.57]	1.00 [0.99]	0.24 [0.21]	0.85 [0.81]	1.00 [1.00]	0.60 [0.56]	1.00 [1.00]	1.00 [1.00]
0.7	0.09 [0.07]	0.35 [0.30]	0.94 [0.92]	0.15 [0.14]	0.63 [0.58]	1.00 [0.99]	0.49 [0.45]	1.00 [1.00]	1.00 [1.00]
0.8	0.06 [0.05]	0.15 [0.12]	0.58 [0.53]	0.11 [0.09]	0.37 [0.32]	0.96 [0.95]	0.41 [0.37]	0.97 [0.94]	1.00 [1.00]
0.9	0.04 [0.03]	0.05 [0.04]	0.13 [0.11]	0.08 [0.07]	0.19 [0.16]	0.73 [0.69]	0.35 [0.32]	0.89 [0.86]	1.00 [1.00]
0.95	0.04 [0.03]	0.04 [0.03]	0.05 [0.03]	0.09 [0.08]	0.15 [0.12]	0.52 [0.47]	0.35 [0.32]	0.83 [0.80]	1.00 [1.00]
0.99	0.04 [0.04]	0.04 [0.04]	0.03 [0.03]	0.11 [0.10]	0.18 [0.16]	0.41 [0.37]	0.37 [0.35]	0.77 [0.73]	1.00 [1.00]
1	0.05 [0.05]	0.05 [0.05]	0.05 [0.05]	0.13 [0.12]	0.21 [0.20]	0.44 [0.41]	0.40 [0.37]	0.75 [0.72]	0.99 [0.98]

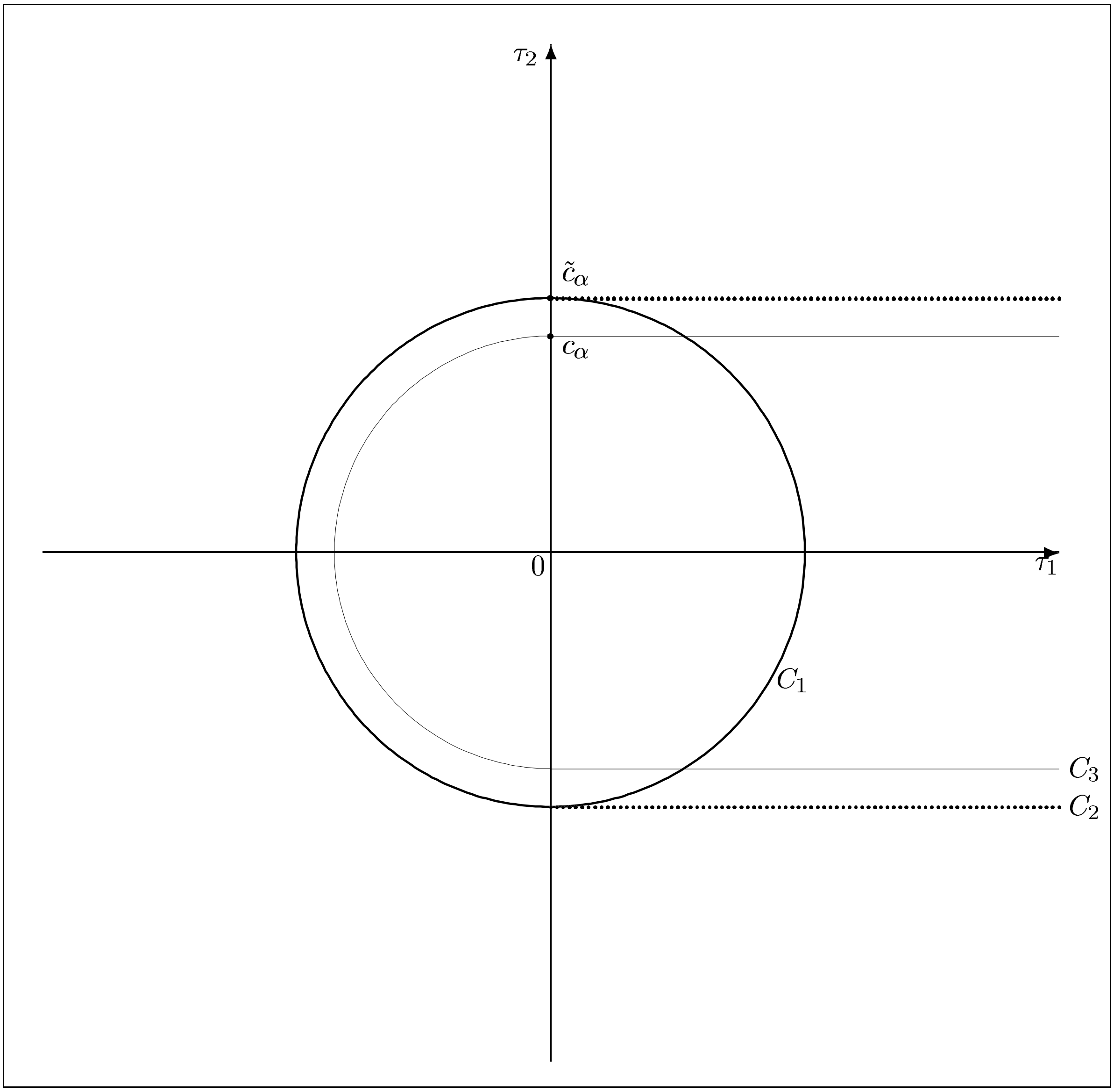

Figure 1: Boundaries of the critical regions, C_1 (unmodified), C_2 (modified but size-unadjusted), C_3 (modified and size-adjusted).