EUI WORKING PAPERS

EUI WORKING PAPER No. 89/424

On the Differential Geometry of the Wald Test with Nonlinear Restrictions

FRANK CRITCHLEY
PAUL MARRIOTT
MARK SALMON

European University Institute, Florence

EUROPEAN UNIVERSITY INSTITUTE, FLORENCE

DEPARTMENT OF ECONOMICS

EUI WORKING PAPER No. 89/424

On the Differential Geometry of the Wald Test with Nonlinear Restrictions

FRANK CRITCHLEY
PAUL MARRIOTT
MARK SALMON*

Departments of Statistics, Mathematics and Economics, University of Warwick.
*Department of Economics, European University Institute, Florence.

This research has been supported, in part, by an ESRC grant on "The application of Differential Geometry to Econometrics". We are grateful for a number of comments received on earlier versions of this paper when presented at the Universities of Cambridge, Oxford, Manchester and the ESRC Econometric study group conference at Warwick.

BADIA FIESOLANA, SAN DOMENICO (FI)

Printed in Italy in January 1990
European University Institute
Badia Fiesolana
– 50016 San Domenico (FI) –
Italy

On the Differential Geometry of the Wald Test with Nonlinear Restrictions

by

Frank Critchley, Paul Marriott and Mark Salmon†

Departments of Statistics, Mathematics and Economics,
University of Warwick.

This paper provides a geometric analysis of the Wald statistic. Testing nonlinear hypotheses using the Wald test suffers from the severe practical problem that, although equivalent asymptotically, the finite sample behaviour of different Wald statistics W(g) for a given null hypothesis depends critically on the non-cannonical choice of the algebraic expression, g, used to formulate the null ; H_0: $g(\theta)=0$. We show that the problem arises because W(g) is not , in general, a geometric quantity in the natural statistical manifold. Hence Wald statistics are , in general, not invariant to changes in the coordinate system in which the null hypothesis is expressed. We identify a "Geodesic Statistic" that naturally corresponds to the Wald statistic and yet is a true geometric quantity. We describe the conditions under which this Geodesic statistic coincides with the Wald statistic. We also provide a methodology to calculate error bounds on the use of the Wald test that indicate the degree to which it diverges from the Geodesic test. The formal mathematical analysis suggests readily applicable graphical techniques for determining whether the problems with the use of the Wald statistic are likely to be severe. Finally we establish, under certain regularity conditions , an inequality between the Wald and Geodesic statistic which ensures unambiguous inference from the Wald test even with nonlinear restrictions.

Version : September 1989

This research has been supported ,in part, by an ESRC grant on "The application of Differential Geometry to Econometrics". We are grateful for a number of comments received on earlier versions of this paper when presented at the Universities of Cambridge, Oxford, Manchester and the ESRC Econometric study group conference at Warwick.

† Department of Economics, European University Institute, Badia Fiesolana, C.P.N.2330, I-50100 Firenze Ferrovia, Italy.

1: Introduction

Several papers (in particular Gregory and Veall (1985)(1986) and LaFontaine and White (1986)) have recently noted the existence of a serious difficulty in the application of the Wald test to nonlinear restrictions in finite samples. Essentially the problem lies in that a given hypothesis may be written in any number of ways which are algebraically equivalent under the null but differ nontrivially under any particular alternative. Since the Wald statistic is based on a first order Taylor series expansion of the function defining the null hypothesis the test statistic is not invariant to the particular algebraic form chosen to represent the null . Thus , in general, different algebraic expressions of precisely the same null hypothesis lead to different test statistics which have different rejection regions at the same asymptotic significance level. The corresponding tests therefore have different exact significance levels and different powers at the same alternative. Gregory and Veall (1986) conclude their Monte Carlo study of a particular nonlinear example by emphasizing "the need for an analytical resolution to the problem of Wald test sensitivity".

In this paper we attempt to provide such an analysis through the use of differential geometry which has recently found considerable application within mathematical statistics (see for instance the references in Barndorff Nielsen, Cox and Reid(1986) and Amari et al. (1987)). We have two distinct goals in this paper ; one is to provide a clear theoretical understanding of why the Wald statistic behaves as it does in the non-linear case and the second is to produce practical solutions to the problem. Both of these objectives are most easily achieved using the techniques and insight provided by differential geometry.

We start our analysis with a critical look at the assumptions and justification of the Wald test particularly in the non-linear context. We show how the Wald statistic corresponds to a hybrid geometric quantity, in that it considers a vector in a statistical manifold and yet measures its length using a metric which is only appropriate to a tangent space. We then follow this theoretical analysis by defining a truely geometric test in the correct space which is a direct generalistion of the Wald test to our non-linear case. This new geometric test we call *The Geodesic test*. We can therefore reach a theoretical resolution of the problems of the Wald statistic by showing how the dependence of the statistic on the form of the restriction function is viewed geometrically as a failure of the statistic to transform correctly under a change of coordinates of the underlying manifold. The geodesic test is shown to transform correctly and we can further show that the two tests coincide under the classical assumptions of the General Linear Model with linear restrictions and in this case the Wald test is reliable. The two tests are in any case asymptotically equivalent as we show below. We continue the geometric approach by using the tools of curvature and the related notion of the Christoffel symbols of a metric to compare the two statistics in general .

As a first practical result we show how these Christoffel symbols can be used to compare different forms of the restriction function and hence how a best selection can be made. As an example of these methods we consider the discrimination between the two choices of restriction function used by Gregory and Veall (1986).

We continue by studying how the choice of restriction function effects the behaviour of the Wald statistic using graphical techniques suggested by the previous analysis. Finally we establish, under certain regularity conditions an inequality between the Geodesic and Wald statistics that indicates when reliable inference is possible using the Wald test. This last practical application of our geometric approach completes the paper.

Three papers, in particular, have appeared recently that are related to our analysis. Moolgavkar and Venson (1987), have employed a geometrical analysis of Wald confidence intervals for a simple hypothesis in nonlinear regression and more generally curved exponential families. Their object is to reparametrise the model so that it looks as much like " uncurved" Euclidean space as possible. One way to achieve this is to use geodesic normal coordinates but as we shall see below, given the difficulty of calculating geodesics in practise they are forced to use approximations that while they improve on the Wald confidence regions they do not correspond with the geodesic regions that follow from the geodesic statistic we introduce below.

Veath (1985) also considers the use of reparametrisation in the exponential model , however the restriction of his analysis to the one dimensional case avoids much of the difficulty of the multidimensional problem that we consider below.The results of both these papers are encompassed by in our more general geometric proceedures below.

Phillips and Park (1988) have also considered the issue by means of calculating Edgeworth expansions to investigate alternative forms of the Wald statistic with nonlinear restrictions. These expansions are able to explain , to a degree ,the observed behaviour of the test as the higher order terms account for the deviations from the asymptotic distribution and also to provide corrections to the test that indicate transformations of the restrictions which accelerate convergence to the asymptotic distribution. However the analysis is limited to the $O(T^{-1})$ terms in the expansions and hence their correction factors are similarly limited unlike the geometric analysis and Geodesic test introduced below.

2: The Wald Test

The algebraic development of the Wald statistic may be found in any standard text such as Silvey (1975) or Cox and Hinkley(1973) and assumes a model summarised in a log likelihood function $l(.,\theta)$ together with an estimator $\hat{\theta}$ for the unknown parameter $\theta \in \mathbf{R}^p$, which is distributed at least asymptotically as multivariate normal $N_p(\theta, I_\theta^{-1})$. This happens of course in all regular likelihood problems where we can identify I_θ with Fisher's information matrix. The null hypothesis is specified as the zero level set of a vector valued function g;

$$H_0=g^{-1}(0)=\{\theta\in\Theta \mid g(\theta)=0\} \tag{1}$$

where Θ denotes the parameter space and $g=(g_1,.....,g_r)$ is a vector of real valued functions , one for each restriction. The Wald statistic , W(g) , is then defined as

$$W(g)= g(\hat{\theta})^T\{\text{var } g(\hat{\theta})\}^{-1}g(\hat{\theta}) \tag{2}$$

in which the estimated variance covariance matrix of $g(\hat{\theta})$ is given by

$$\text{var } g(\hat{\theta})=\{Dg(\hat{\theta})\}^T I_{\hat{\theta}}^{-1}\{Dg(\hat{\theta})\} \tag{3}$$

where $Dg(\hat{\theta})$ is the evaluation at $\theta = \hat{\theta}$ of the pxr matrix

$$Dg(\theta) = \left(\frac{\partial g_i(\theta)}{\partial \theta_j}\right) \tag{4}$$

and $I_{\hat{\theta}}$ is the evaluation of I_θ at $\theta = \hat{\theta}$.

Since W(g) depends solely upon quantities evaluated at $\hat{\theta}$, it is particularly well suited for use in situations where the unrestricted estimate $\hat{\theta}$ is easy to compute but the restricted maximum likelihood estimate , $\tilde{\theta}$, under H_0 is not. This is likely to be the case when the restriction function $g(\theta)$ is nonlinear and yet it is precisely in this case that the difficulties with the use of the Wald test appear.

The distribution and properties of the Wald Statistic ,which is based on an expansion of the restriction function $g(\theta)$, rest on three fundamental approximations:

(i). Ignore any non-normality in the finite sample distribution of $\hat{\theta}$, in other words work effectively only with the asymptotic distribution ,

$$\sqrt{n}(\hat{\theta}-\theta) \sim N_p(0, B_{\theta_0}^{-1}),$$

where B_{θ_0} is the information matrix for a single observation, and θ_0 is the assumed true value of θ.

(ii). Ignore all terms beyond the linear one in the Taylor expansion of $g(\theta)$ about θ_0 evaluated at $\hat{\theta}$.

$$\sqrt{n}\{g(\hat{\theta})-g(\theta_0)\} = \sqrt{n}[Dg(\theta_0)]^T(\hat{\theta}-\theta_0) + O(\sqrt{n}\,|(\hat{\theta}-\theta_0)|^2)$$

in other words work effectively with

$$\sqrt{n}\{g(\hat{\theta})-g(\theta)\} \sim N_r(0,[Dg(\theta_0)]^T B_{\theta_0}^{-1}[Dg(\theta_0)])$$

(iii). Finally to gain an operational statistic, ignore the dependence in the covariance matrix of $(g(\hat{\theta})-g(\theta_0))$ on the unknown θ_0 and replace θ_0 by $\hat{\theta}$. In other words use

$$\sqrt{n}\{g(\hat{\theta})-g(\theta)\} \sim N_r(0,[Dg(\hat{\theta})]^T \hat{B}_{\theta}^{-1}[Dg(\hat{\theta})])$$

Under these conditions W(g) is asymptotically a χ_r^2 random variable under H_0.

The applicability of the Wald statistic is critically determined by the validity of these approximations. In particular (i) covers any regular maximum likelihood problem and those where a central limit theorem may be applied. Approximation (ii) is exact only if $g(\theta)$ is affine ,i.e. $g(\theta)=A\theta +b$, and (iii) is exact only if $g(\theta)$ is affine and I_θ is independent of θ. This latter condition we refer to as the constant metric case below. Critically for our present concern it is the linearization in (ii) that leads to the lack of invariance with respect to reparameterizations.

The previous argument leads to the standard Wald statistic as implemented empirically. However for our geometric analysis of the statistic we abstract from the final approximation which replaces the unknown θ_0 by the observed $\hat{\theta}$ .While clearly necessary for practical impllimentation of the statistic this final step introduces unnecessary elements and complexities for our theoretical analysis .The source of the problems with the Wald statistic with which we are concerned lies in the first two approximations, hence we shall consider below a Wald statistic of the form

$$W_{\theta_0} = g(\hat{\theta})[Dg(\theta_0)^T I_{\theta_0}^{-1} Dg(\theta_0)]^{-1} g(\hat{\theta}) \quad (5)$$

rather than the usual $W_{\hat{\theta}}$ which is defined as we have already stated as

$$W_{\hat{\theta}} = g(\hat{\theta})[Dg(\hat{\theta})^T I_{\hat{\theta}}^{-1} Dg(\hat{\theta})]^{-1} g(\hat{\theta}) \quad (6)$$

where the covariance matrix of $g(\hat{\theta})$ is evaluated at $\hat{\theta}$. Both of the statistics (5) and (6) imply a fixed metric and having conducted our theoretical argument in terms of statistic (5) it can be easily shown that precisely the same implications apply to the empirical Wald statistic (6). Our recommendations for a practical solution to the lack of invariance of the Wald Statistic apply in particular to the applied statistic (6).

3: The geometry of the Wald Statistic and the Geodesic Test.

3.1 An overview.

As discussed above the construction of a Wald statistic enables a standard test for hypotheses expressed on some parameter space indexing a family of distributions to be performed. If we parameterise this space, Θ, by $(\theta_1,....\theta_p)$ then at least locally we may, without loss of generality, write the null hypothesis as the zero set of the restriction function, g, where,

$$g: \Theta \to \mathbf{R}^r$$

is a smooth enough function. So the null hypothesis is the subset of Θ given by

$$H_0 = \{(\theta_1,....\theta_p) \mid g(\theta_1,....\theta_p)=0\}$$

For the generality we need in our analysis we take the space of distributions to be nonlinear or curved although even when it may be linear , as we shall see below, the effect of a nonlinear restriction is to introduce nonlinearity to the structure of the space.

The Wald test attempts to measure the probability of deviations from the null by constructing contours ,using the mathematical form of the statistic, around the null hypothesis . The Wald statistic then takes positive values as the estimated value of the parameter lies outside some chosen contour.

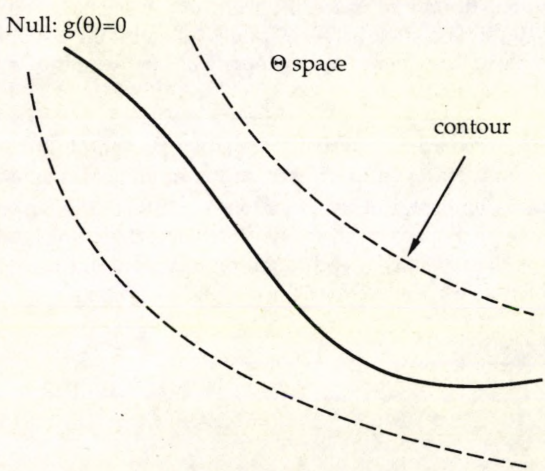


Fig.1

The problem with the Wald Test is that although many functions can equally be used to define the null hypothesis H_0 the approximation that is used to estimate the probability contours depends on which particular function is used. The reason loosely for this is although two different functions $g_1, g_2 : \Theta \rightarrow \mathbf{R}^r$ may agree on their zero set, i.e.

$$\{(g_1)^{-1}(0)\} = \{(g_2)^{-1}(0)\},$$

the sets that correspond to their other levels, $\{(g_1)^{-1}(c)\}$ and $\{(g_2)^{-1}(c)\}$ $(c \neq 0)$, can differ arbitarily. Moreover unlike the one dimensional case , considered by Væth (1985) , this problem cannot be resolved by a simple rescaling.

This implies that the reason for the inconsistency in the performance of the Wald test stem from this difference in behaviour of the level sets away from the null hypothesis which is not taken into account by the Wald statistic .

Geometrically we can see, given approximation (ii) above, that the Wald statistic has an interpretation as the squared length of a particular vector valued function , $(g(\hat{\theta})-0)$, on the curved manifold describing the family of potential distributions. The metric used to calculate the length of this vector is however taken, as will be shown below, from the tangent space to the manifold at θ_0. The Wald statistic for a nonlinear restriction is therefore a hybrid quantity measuring a vector corresponding to a point in a nonlinear, non-metric, space (the statistical manifold) with a metric taken from a linear tangent space. Notice that the Likelihood Ratio and Lagrange Multiplier statistics do not suffer from this inconsistency. The Likelihood Ratio statistic being simply a comparision of values taken by the likelihood function on the manifold and the Lagrange multiplier measuring the length of a vector in the tangent space with a consistent metric.

When the natural coordinate system defined by Θ is employed the lengths of tangent vectors are measured using the Fisher information metric, I_θ, at each point. The following diagram demonstrates the situation in general. The curved manifold corresponds to the nonlinear space of distributions indexed by the choice of θ or alternatively the value of $g(\theta)$ and for each point on this manifold there will be a tangent plane on which its associated metric is defined.

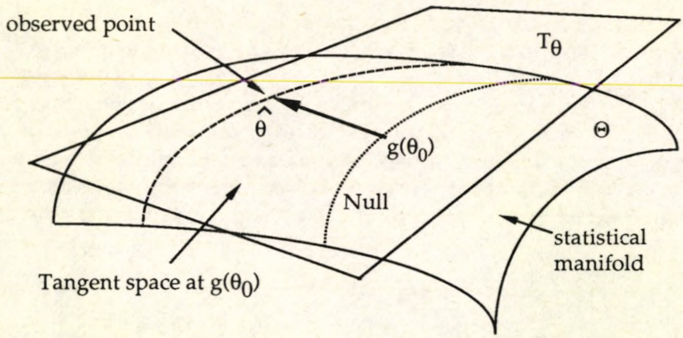


Fig. 2

Notice also that since we are free to choose any coordinate system to describe the statistical manifold and since the Wald statistic is itself indexed by the choice of restriction function $g(\theta)$ it is necessary to consider the manifold in terms of a new coordinate system that involves $g(\theta)$ rather than the natural coordinates $\{\theta_i\}$. This change in coordinate system however also induces a change in form of the metric used to measure distance in the tangent space which the Wald statistic exploits. Choosing a particular algebraic function, g_1 ,to express the restriction in effect then imposes a choice of (probably non-Euclidean) coordinates on the statistical manifold. Choosing a different function, say g_2 , therefore induces a change of coordinates, or a reparametrisation of the manifold and changes the form of the metric.

From this geometric point of view it is then possible to see how the Wald statistic does not transform in the correct way under the change of coordinates which corresponds to a different choice of restriction function, thus causing the inconsistencies in its behaviour. The Wald statistic is essentially a quadratic form on a linear space , the tangent space, which is appropriate to measuring the length of (linear) vectors in this space.The statistic transforms appropriately for linear transformations but inadequately for nonlinear transformations induced by nonlinear restrictions. In addition any nonlinear coordinate system on the manifold implies a different metric will be appropriate for every point on the manifold while the Wald statistic implicitly assumes that there is a single metric for the entire manifold. It is only under this constant metric assumption that the Wald statistic provides a well defined measure of length . Given this geometric insight we can see that Wald statistics computed for different nonlinear restriction functions are not comparable.

There are two main reasons why nonconstant metrics may come about in general. The first is that the underlying manifold has non-zero curvature and so there simply is no coordinate system which would give a constant metric. The variation of the metric may also be induced by the particular choice of coordinate system . This distinction corresponds with that made earlier by Bates and Watts (1980) between "intrinsic" and "parameter effects" curvature, where intrinsic curvature cannot be removed by a reparametrisation of the problem. Notice that even for a space with no curvature at all most coordinate systems will not give the constant representation of the metric that the Wald statistic requires.The property of a constant metric representation leads to what is known as an *Affine coordinate system*. An example of a non-affine coordinate system on a flat space is the use of polar coordinates ($r\cos\psi, r\sin\psi$) on the Euclidean plane. Here the standard metric will be given by the form.

$$\begin{bmatrix} r & 0 \\ 0 & 1 \end{bmatrix}$$

which is a nonconstant metric since it depends on the point (r,ψ).

In what follows we derive an alternative approach for calculating confidence regions in a space with such a varying metric. The resulting test statistic , the Geodesic statistic, has the advantage of behaving properly under changes of coordinates, and hence different choices of restriction function. This Geodesic statistic has a geometric interpretation as the length of a curve in the statistical manifold itself rather than in the tangent space and is invariant under coordinate transformation. We discuss cases in which it reduces to the standard Wald statistic and hence when the use of the Wald test will be free of its dependence on the form of restriction function.

Figure 3 below demonstrates our strategy. Initially the statistical problem is formulated in the $\{\theta_i\}$ coordinate system, and we change coordinates to a new coordinate system, (g,k), where g is the value of the restriction function and the remaining coordinates, k, are, without loss of generality, chosen orthogonally to g . Working in this particular , (g,k), coordinate system we can clearly see the geometric interpretation for the Wald Statistic as the length of some vector in a tangent space. Ideally this vector would correspond to the correct projection from the manifold to the tangent space of the point $\hat{\theta}$, the unrestricted parameter estimate. This projection is achieved by what is known as the exponential map which preserves the correct length of the implied vector. We use this map to show how the Wald statistic does not transform properly with respect to changes in coordinates.

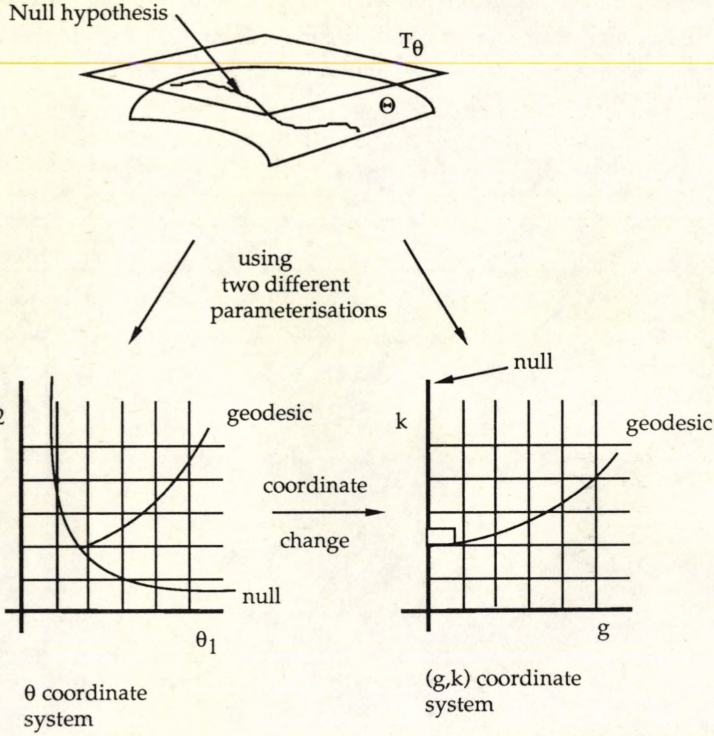


Fig. 3

3.2: Formal Analysis

The following lemma shows how the choice of restriction function , g , determines a choice of coordinates.

Lemma 3.2.1:

(a) If θ_0 is any point in Θ such that $Dg(\theta_0)$ has full rank, then in an open neighbourhood of θ_0 there exists a local coordinate system of the form,

$$(g_1(\theta),......g_r(\theta),k_1(\theta),....k_{p-r}(\theta))$$

where $g(\theta)=(g_1(\theta),......g_r(\theta))$ represents the r-vector of restrictions and $k_1,k_2,...,k_{p-r}$ are real valued smooth functions on $\mathbf{R}^p$. Furthermore if F(,) represents the Fisher information metric, then at any point on the level set of g to which θ_0 belongs we have that;

$$F\left(\frac{\partial}{\partial g_i},\frac{\partial}{\partial k_j}\right)=0 \qquad \forall i,j \tag{7}$$

where $\frac{\partial}{\partial g}$ and $\frac{\partial}{\partial k}$ form a basis for the coordinate system where the vector $k(\theta)$ is choosen such that the corresponding tangent vectors satisfies (7), as shown in the following diagram.

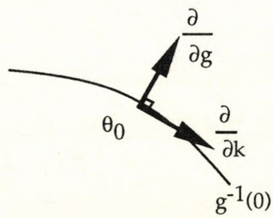


fig. 4

(b) In the case where the null hypothesis is given by just a single restriction we have that the above orthogonality condition is true for any point in an open region around the level set of g(.) to which θ_0 belongs and not just on the one level set.

Proof: All proofs are given in the appendix to the paper.

The implication of this lemma lies in that we must consider how the Fisher information metric is transformed as the coordinate system in which the statistical hypothesis is expressed itself varies with the change of form of g. Let I_θ be the Fisher information matrix and $(\theta_1,....,\theta_p)$ our original coordinate system on Θ then if we let $G=(\partial g_i/\partial\theta_j)$ we have the following result.

Lemma 3.2.2:

(a)In terms of the (g,k) coordinate system given above the matrix defining the Fisher Information metric at a point θ is given by

$$F_g = \left(\begin{pmatrix} G \\ K \end{pmatrix}^{-1} \right)^T I_\theta \left(\begin{pmatrix} G \\ K \end{pmatrix}^{-1} \right) \tag{8}$$

where $K=\{\partial k_i / \partial \theta_j\}$ is the $(p-r)\times p$ matrix that induces the change in coordinates for vectors parallel to the level sets of g, i.e. the vectors $\left\{\frac{\partial}{\partial k}\right\}$.

(b) In the single restriction case the formula holds in an open region of the g level set of θ_0.

Corollary 3.2.3:

If X is a vector field always orthogonal to the level sets of g, then working in our (g,k) coordinate system we see that the squared length of X at all points θ is given by $X^T(G^T I_\theta^{-1} G)^{-1} X$. This reduces to the either of our two forms of the Wald test statistic (5) or (6) , depending on where G and I_θ are evaluated given that X equals $g(\hat{\theta})$.

Notice in fact that the Wald test considers the length of the vector $(g_1, \ldots g_r, 0, \ldots, 0)$ which is orthogonal to the level sets of g and hence lies in the vector field X which is defined above. Any vector in X then has its length measured by the formula given in the corollary. This corollary shows the difficulty with the use of the Wald test lies in that instead of being a measure of a length in the curved manifold it is in fact measuring a length in the flat tangent space. It is this confusion between the manifold and its tangent space at a point which is causing the statistic to be dependent on the choice of coordinate system, and through the coordinate system the statistic ultimately depends on the particular algebraic form of the restriction. The difference between the two spaces is that while on the manifold the form of the Fisher metric changes from point to point, the tangent space is a linear space with a constant metric.

To understand the relationship between these two geometric structures we need to introduce the notion of the exponential map between the tangent space at a point θ and the manifold.

$$\exp: T_\theta \to M.$$

This exponential map is defined in the following way; $\exp_\theta(v)$ is the point which lies on the geodesic starting at θ in the direction v which is a geodesic distance $|v|$ from θ.

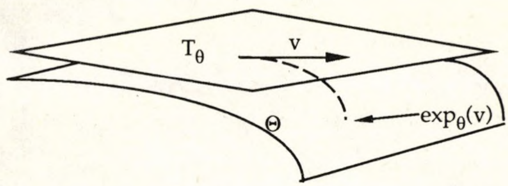


Fig. 5

Where,

Definition: *Path length on the manifold*

If $\gamma(t):[0,1]\to\Theta$ is the path starting at θ_1 and ending at θ_2 then the path length is given by,

$$\int_0^1 \sqrt{H(\tfrac{d}{dt}\gamma(t),\tfrac{d}{dt}\gamma(t))}\, dt \tag{9}$$

where H is the metric .

Definition: A *geodesic* in the manifold with a metric is a curve γ(s) which has the shortest path length between two points, where s is the arc length parameter. It can be characterised locally in our manifold as being the solution to the set of second degree differential equations given by:

$$\ddot{\gamma}^i + \Gamma^i_{jk}\dot{\gamma}^j\dot{\gamma}^k = 0 \qquad i,j,k = 1,\dots,p \tag{10}$$

where Γ^i_{jk} are the Christoffel symbols for the Levi-Civita connection (see for instance (Amari 1985)). These symbols are determined from the metric (h_{ij}), and its inverse (h^{ij}),by the equations:

$$\Gamma^i_{jk} = \tfrac{1}{2} h^{mi}\left(\frac{\partial h_{mk}}{\partial\theta_j} + \frac{\partial h_{jm}}{\partial\theta_k} - \frac{\partial h_{jk}}{\partial\theta_m}\right) \tag{11}$$

Applying these definitions to our problem we have,

Definition: Geodesic distance between θ *and* $\hat{\theta}$

If $\gamma(t):[0,1]\rightarrow\Theta$ is the geodesic starting at $\theta \in g^{-1}(0)$ and ending at $\hat{\theta}$ then the geodesic distance $(\theta,\hat{\theta})$ is given by,

$$\int_0^1 \sqrt{F_g(\tfrac{d}{dt}\gamma(t), \tfrac{d}{dt}\gamma(t))}\, dt \tag{12}$$

where F_g is the Fisher information metric and t is a parameter on the geodesic which is zero on the null hypothesis and 1 at the observed point in the manifold.

The Wald statistic defined in general as

$$(g(\hat{\theta}) - g(\theta))^T (G_\theta^T I_\theta^{-1} G_\theta)^{-1} (g(\hat{\theta}) - g(\theta))$$

then represents the squared length of $(g(\hat{\theta})-g(\theta))$ measured in the tangent space at a point θ. Whereas from the geometric point of view we would ideally wish to identify a point in the manifold corresponding to $\exp(g(\hat{\theta})-g(\theta))$. Notice in order to measure the length of $\exp(g(\hat{\theta})-g(\theta))$ we need to consider the sequence of metrics corresponding to the sequence of tangent planes which are based on those points on the geodesic line to $\exp(g(\hat{\theta})-g(\theta))$ from θ. The Wald statistic however is defined at the one tangent space based at θ and measures the corresponding length using the fixed metric from that one tangent space. The importance of this distinction arises when there is a change in coordinates such as that induced by a different choice of restriction function. The exponential map remains unaffected by this transformation since it considers the change of basis at each tangent space on the manifold. The Wald statistic is only determined by the change of basis in the one tangent space based at θ. Moreover the Wald statistic measures the length of a different vector and will not be comparable with that calculated using the original choice of restriction function.

Following this logic we are naturally led to introduce as an alternative to the Wald statistic the Geodesic statistic which follows the standard differential geometric construction for measuring the distance betwen two points on a manifold with a metric, a geodesic. As discussed above this geodesic statistic has the advantage of being coordinate free thus escaping the problem of being dependent on the choice of restriction function.

By changing the parametrisation on the geodesic to the value of the function g at each point we see that (12) is equivalent to

$$G = \int_0^{g(\hat{\theta})} \sqrt{F_g(\tfrac{d}{dg}\gamma(g), \tfrac{d}{dg}\gamma(g))}\, dg \, , \tag{13}$$

Where $\frac{d}{dg}\gamma(g)$ is the tangent field along the geodesic $\gamma(g)$ which we are using to measure the distance from the point θ.

We can therefore now formally define the Geodesic Statistic.

Definition: The geodesic statistic. For any point in our manifold , corresponding to $\hat{\theta}$ we can measure the geodesic distance (13) to any point on the null hypothesis. The minimum such length provides the value of the geodesic statistic.

The finite sample distribution of this statistic is uncertain but is considered in more detail in Critchley, Marriott and Salmon (1989b). However asymptotically at least we are assured that this statistic will be distributed χ^2_r and this forms the basis for the proposed Geodesic Test . As we show below this limiting distribution follows since asymptotically as $\hat{\theta}$ converges to the null the distinction between the Wald and Geodesic statistics vanishes and indeed the Wald statistic itself becomes immune to the problems of reparametrisation in this nonlinear environment. Although as shown by Philips and Park the speed of convergence to the limiting distribution may be critically determined by the choice of restriction function. These conclusions follow from the fundamental property of the geodesic which generates the geodesic statistic which is that it starts perpendicular to the null hypothesis before reaching the point $\hat{\theta}$.

One essential difference between the two statistics in finite samples lies in that the Wald statistic ignores the component of the total information held in the k-coordinates whereas the geodesic statistic exploits this ancillary information. More generally the following lemma establishes the conditions under which Wald and geodesic inference will coincide.

Lemma 3.2.4:

In the single restriction case the Wald test statistic will agree with the squared geodesic distance if

(i) F_g , the matrix representation of the metric is constant throughout the manifold,

(ii) the geodesics between any $\hat{\theta}$ and the null hypothesis are perpendicular to the level sets of g.

These conditions hold if we are working in Euclidean space and our restriction is just a linear function (as in the general linear model) so that all the level sets are parallel lines and all the geodesics are just orthogonal straight lines. Note that because the restriction function is linear the metric will stay constant in the (g,k) coordinate system. It is the second condition in this lemma that eliminates the dependence of the information in the k-coordinates.

4: An Error bound for the Wald statistic

In general it may be hard to calculate the statistic required for the geodesic test since it requires the solution to a set of second order quasi-linear differential equations (10), and then integrating along these curves. Both of these operations are , in general, difficult analytically although numerical methods are available to provide approximate solutions for a given example (Marriott and Salmon (1989)). In the general linear model with a constant metric in the natural coordinate system and nonlinear restrictions it is however possible to find explicit solutions to the geodesics which may then be evaluated numerically . An alternative and completely general approach that we follow in this section of the paper is to calculate a bound between the Wald and geodesic statistics . In this way we are able to determine whether the Wald statistic seriously deviates for a given form of restriction function and in addition we obtain a formal basis to compare different forms of restriction function. Thus we wish to consider the two statistics given by (13) and

$$W_{\theta_0} = g(\hat{\theta})[Dg(\theta_0)^T I_{\theta_0}^{-1} Dg(\theta_0)]^{-1} g(\hat{\theta})$$

The difference between these two statistics can be measured by

$$\left|G-\sqrt{W}\right| = \left| \int_0^{g(\hat{\theta})} \sqrt{F_g\left(\tfrac{d}{dg}\gamma(g), \tfrac{d}{dg}\gamma(g)\right)}\, dg - \sqrt{g(\hat{\theta})^T (Dg(\theta_0)^T I_{\theta_0}^{-1} Dg(\theta_0))^{-1} g(\hat{\theta})} \right| \tag{14}$$

where the two statistics are essentially of the same form , representing line integrals in the manifold expressed in the (g,k) coordinate system, except that the Wald Statistic reduces to its simpler form through its use of a constant metric at θ_0 and its independence of any information in the orthogonal direction given by the k coordinates. Hence the difference may be rewritten as

$$\left|G-\sqrt{W}\right| = \left| \int_0^{g(\hat{\theta})} \sqrt{F_g\left(\tfrac{d}{dg}\gamma(g), \tfrac{d}{dg}\gamma(g)\right)}\, dg - \int_0^{g(\hat{\theta})} \sqrt{F_{g=g(\theta_0)}\left(\tfrac{d}{dg}\gamma(g), \tfrac{d}{dg}\gamma(g)\right)}\, dg \right| \tag{15}$$

By applying the mean value theorem for differentiable functions we see that;

$$\left|G-\sqrt{W}\right| \leq \left| \int_{0}^{g(\hat{\theta})} \max_{g} \frac{d}{dg}\{\sqrt{F_g(\tfrac{d}{dg}\gamma(g), \tfrac{d}{dg}\gamma(g))}\}dg \right| = g(\hat{\theta}) \max_{g}\{\frac{d}{dg}\sqrt{F_g(\tfrac{d}{dg}\gamma(g), \tfrac{d}{dg}\gamma(g))}\} \tag{16}$$

Hence we can see that the two statistics will be the same if $F_g(\ ,\ )$ is constant for all values of g and so a choice of restriction function that induces the least possible variation in the metric is clearly seen to be preferred.

To make any further comparison we need to consider how the metric F_g varies with g and the rest of this section is devoted to providing an explicit bound on the difference between the two statistics based on the difference between the square of the geodesic length and the Wald statistic since the Wald statistic in fact represents a squared distance measure. For clarity we restrict our attention to the two dimensional case where Θ is a surface and the null hypothesis a curve although the analysis may easily be extended to higher dimensional manifolds, in particular if there is only one restriction function this is particularly easy. Working in the (g,k) coordinate system the Fisher metric is given by the matrix

$$\begin{bmatrix} f_{11} & 0 \\ 0 & f_{22} \end{bmatrix}$$

where by lemma 3.2.1(b) $f_{12}=f_{21}=F(\partial/\partial g,\partial/\partial k)=0$ by definition of the (g,k) coordinate system. The geodesic, $\gamma(g)$, is a curve parametrised by the value of the function g and therefore in these coordinates may be written $(g,\varphi(g))$, hence we have that

$$F_g(\tfrac{d}{dg}\gamma(g), \tfrac{d}{dg}\gamma(g)) = (f_{11} + (\varphi')^2 f_{22}). \tag{17}$$

Considering this expression in more detail it is clear that we need to understand both how the geodesic behaves with g and also how the form of the metric itself varies with g. We now take each of these questions in turn.

For a general analysis, without an explicit form for the geodesic, we are forced to consider its dependence on g using the projection of the geodesic on $\mathbf{R}^2$ Euclidean space . This projection is defined by the (g,k) coordinate system as shown in the diagram below . Notice that although the geodesic itself will have zero curvature in the manifold its image will have nonzero curvature and we can use this fact to establish a bound on the behaviour of the geodesic . In addition the coordinates of the geodesic and its image coincide although the relevant metric in each case will differ. We start by considering the curvature of the image of the geodesic using the angle ω as shown in the diagram.

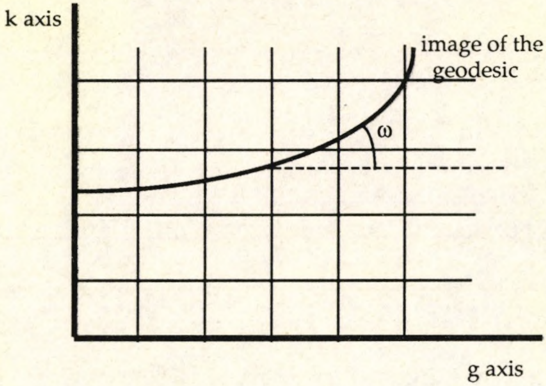


Fig 6.

We estimate the angle ω in the standard manner ; parameterising the image of the geodesic now by (α(s),β(s)), where s measures arc length in the manifold, the curvature is given by,

$$\kappa = \frac{\ddot{\alpha}.\dot{\beta} - \dot{\alpha}\ddot{\beta}}{(\dot{\alpha}^2 + \dot{\beta}^2)^{\frac{3}{2}}} \qquad (18)$$

Since γ(s) = (α(s),β(s)) is a geodesic of the surface it will satisfy by definition the differential equations (10) given earlier as

$$\ddot{\gamma}^i + \Gamma^i_{jk}\dot{\gamma}^j\dot{\gamma}^k = 0 \qquad i,j,k = 1,2$$

where Γ^i_{jk} are , as above, the Christoffel symbols for the Levi-Civita connection of the Fisher metric, with respect to the (g,k) coordinate system.We show how to calculate these symbols for a specific example later in the paper, but for the moment it is sufficient to know that they are determined entirely by the Fisher information metric . It should be noted that the Christoffel symbols are not themselves geometric objects and depend on the choice of coordinate system.

In order to calculate ω we write $\dot{\gamma}(s)$ as (rcosω, rsinω) and hence we find the differential equations (10) defining the geodesic to be,

$$\ddot{\alpha} = r^2(\cos\omega, \sin\omega)\begin{bmatrix}\Gamma^1_{11} & \Gamma^1_{12}\\ \Gamma^1_{21} & \Gamma^1_{22}\end{bmatrix}\begin{pmatrix}\cos\omega\\ \sin\omega\end{pmatrix}$$

$$\ddot{\beta} = r^2(\cos\omega, \sin\omega)\begin{bmatrix}\Gamma^2_{11} & \Gamma^2_{12}\\ \Gamma^2_{21} & \Gamma^2_{22}\end{bmatrix}\begin{pmatrix}\cos\omega\\ \sin\omega\end{pmatrix} \tag{19}$$

Hence the curvature of the image of the geodesic given by (18) is found to be

$$\kappa = \langle(\ddot{\alpha},\ddot{\beta}),(-\sin\omega,\cos\omega)\rangle = -\cos\omega\frac{\ddot{\beta}}{r^2} + \sin\omega\frac{\ddot{\alpha}}{r^2} \tag{20}$$

which with the explicit values of $\ddot{\alpha}$ and $\ddot{\beta}$ given in (19) substituted gives κ as a function of the angle ω and the Christoffel symbols , (<,> denotes the normal Euclidean product).

We are now in a position to consider how the geodesic varies with the choice of g, and this can be achieved even without an explicit form for the geodesic by considering bounds on the curvature of its image. A number of different bounds may be constructed and we present one simple and intuitive choice, others may in fact be tighter. The bound we propose exploits the maximum curvature of the image of the geodesic, however all such bounds are essentially determined by the values of the Christoffel symbols and so we show how these may usefully be regarded as criteria on which to base judgement about the particular choice of g function and its associated Wald test.

If we let λ_1, and λ_2 denote the maximum (in modulus) of the eigenvalues of the matrices

$$\Gamma^i = \begin{pmatrix}\Gamma^i_{11} & \Gamma^i_{12}\\ \Gamma^i_{21} & \Gamma^i_{22}\end{pmatrix} \quad i=1,2 \tag{21}$$

taken over the relevant region of Θ space. Then equations (19) and (20) give the maximum curvature κ_{max}, as

$$|\kappa_{max}| \le ((\lambda_1)^2+(\lambda_2)^2)^{1/2}. \tag{22}$$

We can now use κ_{max} to get an upper bound on ω and hence bound the behaviour of the geodesic. Figure 7 shows the situation , the angle ω which the tangent to the geodesic makes with the g-axis will be less than the angle made by any curve whose curvature is everywhere greater than that of the image of the geodesic. In particular the circle of radius $1/(\kappa_{max})$, whose curvature is everywhere κ_{max}.

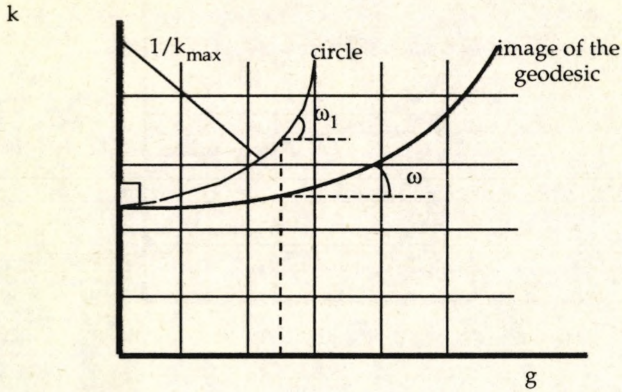


Fig 7.

Some simple geometry then tells us that

$$\omega \leq \omega_1 = \arcsin(g.\kappa_{max}). \tag{23}$$

and so ω_1 provides an upper bound for the variation of the geodesic with g. Notice that if κ_{max} was zero then the image of the geodesic would be a straight line in g-k space. From equation (22) it can be seen that this will come about if the Christoffel symbols are all zero.

We now turn to consider how the whole expression (17) varies with g, including the metric using this bound on ω that we have just established.

Since $(1,\varphi')=(r\cos\omega, r\sin\omega)$ we can see that $\varphi' = \tan\omega$ and so using (17)

$$F\left(\tfrac{d}{dg}\gamma(t), \tfrac{d}{dg}\gamma(t)\right) = f_{11} + f_{22}\tan^2\omega \tag{24}$$

Now as we are working in the g-k coordinate system;

$$\frac{d}{dg}F = \frac{\partial}{\partial g}F = \partial_1 F$$

and $\frac{\partial}{\partial k} = \partial_2$ is the (0,1) basis vector in the k direction

Hence,

$$\frac{d}{dg}F_g(\tfrac{d}{dg}\gamma(t),\tfrac{d}{dg}\gamma(t)) = \tfrac{1}{2}F(\nabla_{\partial_1}(\partial_1+\tan\omega.\partial_2),(\partial_1+\tan\omega.\partial_2))$$

$$= \tfrac{1}{2}F((\nabla_{\partial_1}\partial_1+\tan\omega.\nabla_{\partial_1}\partial_2+\partial_1(\tan\omega)\partial_2),(\partial_1+\tan\omega.\partial_2))$$

$$= \tfrac{1}{2}F((\nabla_{\partial_1}\partial_1,\partial_1)+\tfrac{1}{2}F(\tan\omega.\nabla_{\partial_1}\partial_2,\partial_1)+\tfrac{1}{2}F(\partial_1(\tan\omega)\partial_2,\partial_1)$$
$$+\tfrac{1}{2}\tan\omega F((\nabla_{\partial_1}\partial_1,\partial_2)+\tfrac{1}{2}\tan\omega F(\tan\omega.\nabla_{\partial_1}\partial_2,\partial_2)$$
$$+\tfrac{1}{2}\tan\omega F(\partial_1(\tan\omega)\partial_2,\partial_2)$$

$$= \tfrac{1}{2}F((\nabla_{\partial_1}\partial_1,\partial_1)+\tfrac{\tan\omega}{2}F(\nabla_{\partial_1}\partial_2,\partial_1)+\tfrac{\tan\omega}{2}F((\nabla_{\partial_1}\partial_1,\partial_2)$$
$$+\tfrac{(\tan\omega)^2}{2}F(\nabla_{\partial_1}\partial_2,\partial_2)+\tfrac{\tan\omega}{2}\tfrac{d}{dg}(\tan\omega)F(\partial_2,\partial_2)$$

and so;

$$2\frac{d}{dg}F_g(\tfrac{d}{dg}\gamma(t),\tfrac{d}{dg}\gamma(t)) = \Gamma^1_{11}+\tan\omega.\Gamma^2_{12}+\tan\omega.\Gamma^2_{11}+\tan^2\omega.\Gamma^2_{12}+\tfrac{1}{2}\tfrac{d}{dg}(\tan^2\omega).f_{22}$$
$$= \Gamma^1_{11}+(\tan\omega+\tan^2\omega)\Gamma^2_{12}+\tan\omega.\Gamma^2_{11}+\tan\omega.\sec^2\omega.\tfrac{d\omega}{dg}f_{22} \qquad (25)$$

Considering the calculation of $d\omega/dg$. If we let the unit tangent to the image of the geodesic in Euclidean space be $\underline{T} = (\cos\omega,\sin\omega)$ then if we let s now be the arc length parameter in Euclidean space, we have that the curvature of the geodesic is given by

$$\kappa = |(d\underline{T}/ds)|$$
$$= |(d\underline{T}/dg)|.|(dg/ds)|$$
$$= |(-\sin\omega,\cos\omega)d\omega/dg)|.|dg/ds|$$
$$= |(d\omega/dg)|.|(dg/ds)|$$

where $|.|$ represents the Euclidean norm.

Further since s parameterises by arclength, we have by definition that

$$|d\gamma/ds| = 1$$

So

$$1 = |d\gamma/dg|.|dg/ds|$$

hence

$$d\omega/dg = \kappa|d\gamma/dg|$$

and since $d\gamma/dg = (1,\tan\omega)$

$$d\omega/dg = \kappa.|\sec(\omega)|$$

Given that in the range we are interested sec is an increasing function we have finally that

$$d\omega/dg \le \kappa_{max}.\sec(\omega) \tag{26}$$

Taking this result together with (23), that

$$\omega \le \arcsin(g(\hat{\theta}).\kappa_{max})$$

and substituting into (25) we have the estimate;

$$\left|\frac{dF()}{dg}\right| \le \left|\Gamma_{11}^{1}\right| + \left|\frac{g\kappa_{max}}{\sqrt{1-g^2\kappa_{max}^2}} + \frac{(g\kappa_{max})^2}{1-g^2\kappa_{max}^2}\right|\left|\Gamma_{12}^{2}\right| + \left|\frac{g\kappa_{max}}{\sqrt{1-g^2\kappa_{max}^2}}\right|\left|\Gamma_{11}^{2}\right| \tag{27}$$
$$+ \left|g\kappa_{max}^2.(1-g^2\kappa_{max}^2)\right|.\left|f_{22}\right|$$

If we denote the right hand side of this expression by $\mathrm{Err}(\Gamma_{ij}^{k}, g, \kappa_{max})$ then from the mean value theorem we can see that $F_g(.)$ changes by less than $g(\hat{\theta}).\mathrm{Err}(\Gamma_{ij}^{k}, g, \kappa_{max})$ as we move from the null to the estimated value of θ.

Hence,

$$\left| F_g(\tfrac{d}{dg}\gamma(g), \tfrac{d}{dg}\gamma(g)) - F_g(\tfrac{d}{dg}\gamma(0), \tfrac{d}{dg}\gamma(0))\right| \le g(\hat{\theta}).\mathrm{Err}$$

or

$$F_g(\tfrac{d}{dg}\gamma(0), \tfrac{d}{dg}\gamma(0)) - g(\hat{\theta}).\mathrm{Err} \le \left| F_g(\tfrac{d}{dg}\gamma(g), \tfrac{d}{dg}\gamma(g)) \right| \le g(\hat{\theta}).\mathrm{Err} + F_g(\tfrac{d}{dg}\gamma(0), \tfrac{d}{dg}\gamma(0))$$

and so

$$\left(F_g(\tfrac{d}{dg}\gamma(0), \tfrac{d}{dg}\gamma(0)) - g(\hat{\theta}).\mathrm{Err}\right)^{\frac{1}{2}} \le \left| F_g(\tfrac{d}{dg}\gamma(g), \tfrac{d}{dg}\gamma(g)) \right|^{\frac{1}{2}} \le \left(g(\hat{\theta}).\mathrm{Err} + F_g(\tfrac{d}{dg}\gamma(0), \tfrac{d}{dg}\gamma(0))\right)^{\frac{1}{2}}$$

Thus integrating over the geodesic curve, we see that;

$$g(\hat{\theta}).(\, F_g(\tfrac{d}{dg}\gamma(0),\tfrac{d}{dg}\gamma(0)) - g(\hat{\theta}).\mathrm{Err})^{\frac{1}{2}} \leq G \leq g(\hat{\theta}).(g(\hat{\theta}).\mathrm{Err} + F_g(\tfrac{d}{dg}\gamma(0),\tfrac{d}{dg}\gamma(0)))^{\frac{1}{2}}$$

therefore;

$$\left|G^2 - W\right| \leq \mathrm{Err}(\Gamma_{ij}^{k}, g, \kappa_{max}). \tag{28}$$

which is our final inequality bounding the deviation between our two statistics. Notice that the asymptotic distribution of the geodesic statistic follows directly from this inequality since on the null the right hand side tends asymptotically to zero as $g(\hat{\theta})$ goes to zero , ensuring that the squared geodesic length is distributed as χ_r^2.

Despite its apparent complexity the right hand side of this inequality may easily be shown to be zero when the Christoffel symbols are zero and monotonically increasing in the eigenvalues of the matrices Γ^i i=1,2 given in (21). Notice that from (22) if the eigenvalues of these matrices are zero κ_{max} itself will be zero . Hence , as suggested above the Christoffel symbols may be used as powerful indicators of the degree of nonlinearity and hence the lack of invariance in the Wald statistic. It should also be remembered that the tightness of the bound given above is entirely determined by our use of the circle of maximum curvature to limit the curvature of the geodesic and as such is a very crude example of the bound between G^2 and W. In any example better bounds might easily be found, using particular aspects of the problem. The common thread to all such estimates however will be the use of the Christoffel symbols as a measure of nonaffineness of a coordinate system. The calculation of these symbols alone will often be enough to indicate the validity of treating nonlinear systems as if they are linear and Euclidean.

In a practical example, as in the one below, this analysis also indicates how, given several alternative forms for the restriction function the 'best' one may be chosen. Essentially to minimise the effects of a changing metric we should select the restriction function with the smallest Christoffel symbols.

5: The Gregory and Veall Example.

We illustrate the discussion of the previous section with a geometrical analysis of the problem considered by Gregory and Veall (1985) . As a first step we set up the (g,k) coordinate system as in lemma 3.2.1, and then using the error bounds arguments we explain the large differences between the performance of the Wald statistic with the two restriction functions considered by Gregory and Veall.This analysis , using the Christoffel symbols, also shows quite clearly why one of the choices of formulation for the null hypothesis is to be preferred.

The example is a linear regression model given by

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon_t \tag{29}$$

where $\{\varepsilon_t\}$ is i.i.d. $N(0,\sigma^2)$. The two formulations of the null hypothesis to be considered are given by;

(A) $$H_0^A\colon g^A(\beta_1,\beta_2) = \beta_1 - \frac{1}{\beta_2} = 0 \tag{30}$$

(B) $$H_0^B\colon g^B(\beta_1,\beta_2) = \beta_1.\beta_2 - 1 = 0 \tag{31}$$

and we assume the Fisher metric to be;

$$\begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$$

where λ is assumed to be independent of (β_1,β_2).

To apply the previous theory we need to construct the (g,k) coordinate system for both forms of restriction function and then calculate the Fisher metric and the Christoffel symbols for each case.

Following the proof of lemma 3.2.1 , we construct the integral curves to the vector field given by grad(g), this vector field is given by

$$\text{grad}\,(g) = \left(\frac{1}{\lambda}.\frac{\partial g}{\partial \beta_1}, \frac{1}{\lambda}.\frac{\partial g}{\partial \beta_2}\right) \tag{32}$$

So for case (A) we have $\quad \text{grad}(g^A) = (1/\lambda)(1,(1/\beta_2)^2)$,

and in case(B) $\quad \text{grad}(g^B) = (1/\lambda)(\beta_2,\beta_1)$.

To find the integral curves of these vector fields we have to solve a set of first order differential equations (this is usually much easier than the second order differential equations you would have to solve to find the geodesics) and in our case we may do it explicitly.

Case (A).

We want to find a curve $\gamma(t) = (X(t),Y(t))$ such that

$$\text{grad}(g^A) = (dX/dt, dY/dt)$$

given that $\quad \text{grad}(g^A) = (1/\lambda)(1,(1/\beta_2)^2)$

In other words we need to solve the set of differential equations given by;

$$dX/dt = 1/\lambda,$$

$$\text{and} \quad dY/dt = (1/\lambda).(1/Y)^2$$

Solving we find

$$\gamma(t) = (1/\lambda)(t+A,\sqrt[3]{(t+B^3)})$$

where A and B are arbitrary constants. Since $\gamma(0)$ lies on the null hypothesis we see that A and B are related by $A.B = \lambda^2$.

So to write a point (β_1,β_2) in the (g,k) coordinate system we need to know the value of $g^A(\beta_1,\beta_2)=(\beta_1-1/\beta_2)$ and which of the integral curves (β_1,β_2) lies on. In other words we must find A such that

$$\beta_1=(1/\lambda)(t+A)$$

$$\text{and} \quad \beta_2=(1/\lambda)(\sqrt[3]{(t+(\lambda^2/A)^3)}).$$

Solving implies that (β_1,β_2) corresponds to $(\beta_1-1/\beta_2, 3\beta_1-\beta_2{}^3)$ in the (g,k) coordinates.Thus using the formula of lemma 3.2. 2 gives us that the metric in the (g,k) coordinate system is:

$$F_{g^A} = \begin{bmatrix} 1 & 1/\beta_2^2 \\ -3 & 3\beta_2^2 \end{bmatrix} \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \begin{bmatrix} 1 & -3 \\ 1/\beta_2^2 & 3\beta_2^2 \end{bmatrix} = \lambda \begin{bmatrix} 1+1/\beta_2^4 & 0 \\ 0 & 9+9\beta_2^4 \end{bmatrix} \qquad (33)$$

We can see immediately that the large deviation from a constant metric for small values of β_2. This distortion shows up well in the Monte Carlo analysis reported by Gregory and Veal.

Case (B):

Now we need to solve the diferential equations given by

$$dX/dt = Y/\lambda,$$

$$\text{and} \quad dY/dt = X/\lambda$$

Solving gives $X(t)=Ae^{t/\lambda}+Be^{-t/\lambda}$, and $Y(t)=Ae^{t/\lambda}-Be^{-t/\lambda}$. Using the initial condition that at t=0 we are on the null hypothesis,we get $B=\sqrt{(A^2-1)}$. So in this case the (g,k)-coordinates of (β_1,β_2) are given by $(\beta_1\beta_2-1,\beta_1^2-\beta_2^2)$.

The metric is given by:

$$F_{g^B}=\begin{bmatrix}\beta_2 & \beta_1\\ 2\beta_1 & -2\beta_2\end{bmatrix}\begin{bmatrix}\lambda & 0\\ 0 & \lambda\end{bmatrix}\begin{bmatrix}1 & -3\\ 1/\beta_2^2 & 3\beta_2^2\end{bmatrix}=\lambda(\beta_1^2+\beta_2^2)\begin{bmatrix}1 & 0\\ 0 & 4\end{bmatrix} \qquad (34)$$

Next we need to calculate the Christoffel symbols in the two different coordinate systems. Using the formula which tells use the symbols once we know the metric given above as;

$$\Gamma^i_{jk}=\frac{1}{2}f^{hi}\left(\frac{\partial f_{hk}}{\partial\beta_j}+\frac{\partial f_{jh}}{\partial\beta_k}-\frac{\partial f_{jk}}{\partial\beta_h}\right) \qquad i,j,k,h=1,2$$

Where $\{f^{ij}\}$ is the inverse of the metric $\{f_{ij}\}$.

Case (A)

In this case we get the matrices;

$$\Gamma^1=\begin{pmatrix}0 & -\dfrac{2}{(1+\beta_2^4)\beta_2}\\ -\dfrac{2}{(1+\beta_2^4)\beta_2} & 0\end{pmatrix}$$

$$\Gamma^2=\begin{pmatrix}\dfrac{2}{9(1+\beta_2^4)\beta_2^5} & 0\\ 0 & \dfrac{2\beta_2^3}{(1+\beta_2^4)}\end{pmatrix} \qquad (35)$$

We can see that for small values of β_2 the eigenvalues will blow up thus giving the indication of a large potential deviation between the Wald and Geodesic statistics .

Case(B)

The Christoffel symbols are now given by

$$\Gamma^1 = \frac{1}{(\beta_1^2+\beta_2^2)}\begin{pmatrix}\beta_1 & \beta_2 \\ \beta_2 & -4\beta_1\end{pmatrix}$$

$$\Gamma^2 = \frac{1}{(\beta_1^2+\beta_2^2)}\begin{pmatrix}-\frac{1}{4}\beta_2 & \beta_1 \\ \beta_1 & \beta_2\end{pmatrix} \quad (36)$$

Again the eigenvalues will blow up as β_1 and β_2 get small, but not as fast as in case (A) . Once again this provides a clear explanation for the numerical results of Gregory and Veall.

6 Graphical analysis and tools

These observations on the effect of the different restriction functions may also be displayed graphically as shown in figures 8, 9 and 10 below. In figures 9 and 10 the level sets of the restriction functions for H_A and H_B are shown. In figure 8 we show the level sets for the "geodesic restriction function", by which we mean that algebraic formulation of the restriction that is equivalent under the null to those in cases A and B but whose level sets away from the null can be seen to be parallel in the sense that all points on a given level set are equidistant from the null throughout the parameter space. This "optimal" form of the restriction function for which the Wald and Geodesic statistics would coincide is in fact impossible to derive in closed form , in this case, but a general procedure for evaluating this function for the case of quadratic restrictions has been given in Critchley (1989). The converse of this argument for the optimality of the "geodesic restriction function " can be seen in the graphs for the other two forms of restriction function. In figure 9 the bunching of the level sets as β_2 becomes small provides visible support for our theoretical predictions about the performance of the Wald statistic in this case. In figure 10 we see that non parallel behaviour of the level sets is found as either β_1 or β_2 become large, as in fact is clear from (34) where the metric is found to be proportional to $(\beta_1^2+\beta_2^2)$.

Another observation at this point lies in this question of the optimal choice of restriction function. The form $H_c : \beta_2 - \frac{1}{\beta_1} = 0$ is also equivalent under the null to H_A and H_B and its level sets will be the reflection of the level sets of H_A around the line $\beta_1 = \beta_2$ and hence will be bunched together as β_1 gets small. The obvious question arises as to whether a better restriction function can be formed by taking an optimal linear combination of H_A and H_B . While this operation may indeed reduce the bunching of the level sets in various parts of the parameter space the critical issue turns on where the observed parameter estimate lies and whether the metric is constant between this point and the null. While this may be quickly assessed graphically the Christoffel symbols for this new restriction function will clearly always provide this information analytically.

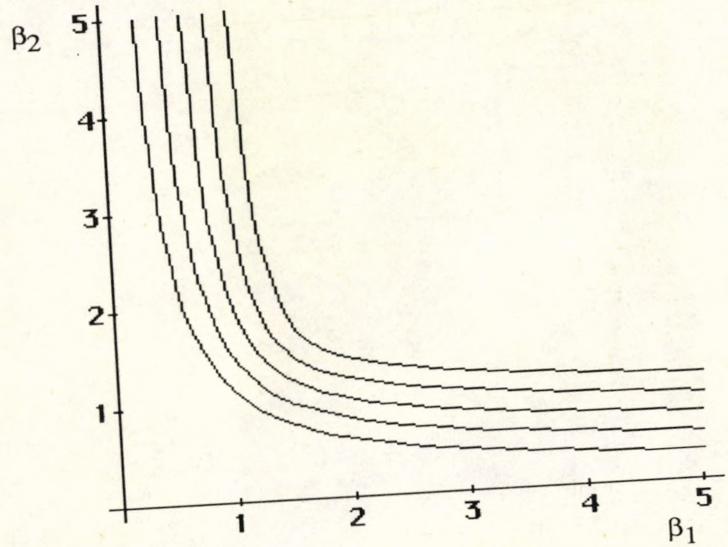


Fig. 8
The "Geodesic" restriction function

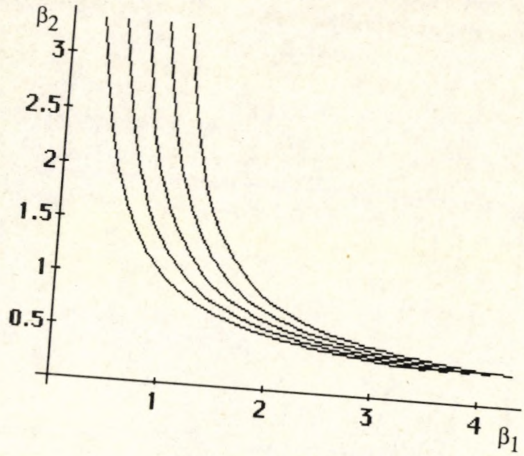


Fig. 9
The level sets of
H_A: $\beta_1 - \frac{1}{\beta_2} = 0$

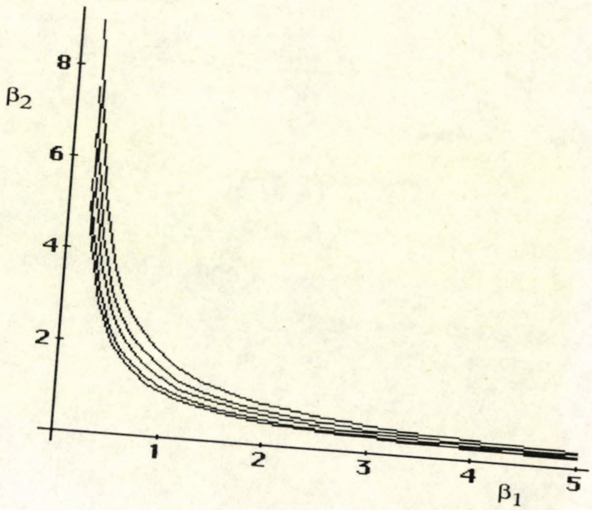


Fig. 10
The level sets of
H_B: $\beta_1\beta_2 - 1 = 0$

These graphical arguments can be rationalised formally and provide a general method of crudely examining whether the metric is changing in the relavant part of the parameter space. Notice that it will only be appropriate to use this simple approach if the metric in the natural coordinates is constant as in the linear regression case with nonlinear restriction functions. Consider how the variation of the form of the metric in the (g,k) coordinate system can be understood in the (β_1,β_2) coordinate system in a geometric fashion. Lemma 3.2.2 tells us that in the (g,k) coordinate system the metric is of the form

$$\begin{bmatrix} f_{11} & 0 \\ 0 & f_{22} \end{bmatrix}$$

Both of the terms in this matrix have readily observable geometric significance and for a constant metric we require both to be constant. First, by definition we have that $f_{11} = |\text{grad } g|$ and hence a constant value for f_{11} will imply that the level sets of the g function will be evenly distributed over the (β_1,β_2) space. For example compare the graph of the geodesic restriction function with thoses of either of the other two restriction functions above.

To understand the behaviour of f_{22} we need to look at the k-constant lines where f_{22} can be seen as a measure of how far they are apart from one another. This can be seen from the formula

$$(0 \quad 1)\begin{bmatrix} f_{11} & 0 \\ 0 & f_{22} \end{bmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = f_{22} \qquad (37)$$

Since (0,1) is the tangent vector to the g-constant lines with the parameterisation given by the value of k at each point. So we can use (37) to tell us the speed at which the value of k changes as we move along the g-constant lines. The length of the segment of g-constant line between k_1 and k_2 is given by

$$\int_{k_1}^{k_2} \sqrt{f_{22}}\,dk$$

Thus the smaller f_{22} the closer the $k=k_1$ line is to the $k=k_2$ line and the faster k changes as can be seen from the following diagram.

So f_{22} essentially provides a measure of the curvature of the level sets of g with a small value indictating a high curvature and this is indictated when the graphs for the two restriction functions are compared with their metrics (33) and (34).

Using this analysis we can see how a simple inspection of the level sets of the restriction function conveys information on how the metric changes through the parameter space and hence on the reliability of the Wald test. Obviously the ideal map for a restriction function that supports the use of the Wald statistic is a simple linear grid! More usefully, given the value of the estimated parameter, this sort of graphical analysis can indicate relevant regions of the parameter space in which particular restriction functions will imply regular behaviour. An example of this can be seen for the restriction function of H_A where good behaviour of the Wald test can be expected for small β_1 and large β_2 and this is confirmed by our Christoffel symbol analysis.

7 A Useful Inequality

So far in this paper we have provided a detailed discussion of how the Wald statistic behaves with different choices of restriction function and while the proposals we have made may be used to assess the sensitivity of the Wald statistic in the nonlinear case it can be seen from our analysis that there is fundamentally little that can be done to rescue the test in this situation. The introduction of the notion of a Geodesic statistic is one possible resolution and this is considered further in Critchley , Marriott and Salmon, (1989b). Another is to use the likelihood ratio test but this involves the calculation of the restricted maximum likelihood estimates which may in some cases prove troublesome. In general the Geodesic statistic may be difficult to compute, see Marriott and Salmon, (1989) but as we now show it is possible to establish an important inequality between the Wald and Geodesic statistics which will ensure reliable inference under certain conditions from the Wald test.

To derive the inequality we need to establish some technical conditions, in particular we must first clarifiy what we mean by a function "increasing" in some direction on a manifold.

Let $h(\beta_1,\beta_2)$ be a real valued function on our manifold. To talk about h increasing we really mean increasing along some regular path. In particular we require that we are in fact increasing along k constant lines i.e., the gradiant lines.

Definition: A real function h is *gradient increasing* if it is increasing along all the k constant lines which cut the null hypothesis.

We can now produce a very useful inequality between the standard Wald statistic, $W_{\hat{\theta}}$ evaluated at the unrestricted maximum likelihood estimate and the geodesic statistic.

Lemma 7.1:

If f_{11} is gradiently increasing towards the null, then $G^2 \leq W_{\hat{\theta}}$.

This lemma tells us that if the level sets of the restriction function are more dense closer to null then the standard Wald test gives us confidence regions inside those of the Geodesic test. Hence a non rejection inference with the Wald test would also imply non rejection under the geodesic test and this may re-establish some utility for the use of the standard Wald statistic in this situation. Notice that the condition underlying the inequality is quite weak and easily checked analytically and also from the graphical inspection of the restriction function . In addition this inequality applies for all sample sizes but as will be clear from our previous arguments becomes an equality asymptotically.

8: Conclusions

This paper has provided a geometric analysis of the Wald statistic and has shown why it is possible to obtain any value from the statistic by a suitable transformationn of the algebraic form used to express the nonlinear restriction being tested. We have shown that the essential problem with the Wald statistic is that it is not a true geometric quantity in that it is not invariant to a change in coordinates. Although there is little that can be done to retrieve the utility of the Wald test in the nonlinear environment and we have provided a number of tools , both analytic and graphical, that may be used to assess whether this problem with the Wald test is likely to be severe in any particular example.

Moreover the geometric approach that we have adopted suggests the use of a new test in the nonlinear context based on the Geodesic Statistic that tranforms properly when the nonlinear restriction is re-expressed since it is a true geometric quantity.

A bound has been established between the between the Wald and Geodesic Statistics that establishes the importance of the Christoffel symbols as indicators of the degree of nonlinearity in an inference problem and hence indicate the severity of the problem with the use of the Wald Statistic. Graphical methods are introduced to support this analysis that are particularly appropriate to the linear regression case.

Finally we have established a powerful inequality between the Wald and Geodesic statistics that enables unambiguous inference to be achieved with the Wald test even in a nonlinear environment.

Appendix:

Proofs of the various theorems and lemmas not given in the main text follow.

Lemma 3.2.1 : Proof

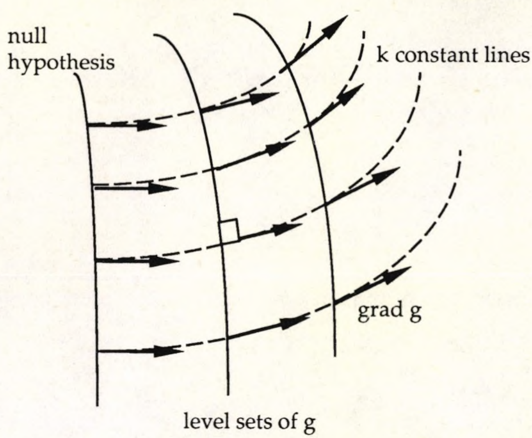
(a) A coordinate system such as (g,k) may be defined arbitrarily (see Thorpe(1979)), the only issue of concern is the smoothness of the functions which we assume.

(b) Since $\mathbf{D}g(q_0)$ has full rank we can use the implicit function theorem, locally around θ_0, $g^{-1}(0)$, the null hypothesis, is an p-1 dimensional submanifold as are all $g^{-1}(c)$ for $c\in(-\varepsilon,\varepsilon)$.

Since the null hypothesis is a submanifold we can put coordinates on it;parametrise by the coordinates $(k_1^0(\theta),....k_{p-1}^0(\theta))$.

Since $g:\Theta\rightarrow\mathbf{R}$ $(\theta\rightarrow g(\theta))$ is assumed to be a real valued smooth function on Θ we are assured that the gradient function, grad(g) exists.

The operation grad takes the function g to a vector field which has the property that each vector is perpendicular to the level set of g through which it passes as shown in the diagram below.



In Euclidean space grad g is given by the formula

$$\text{grad}(g) = (\frac{\partial g}{\partial \theta_1}, \frac{\partial g}{\partial \theta_2}, \ldots, \frac{\partial g}{\partial \theta_p})$$

and in a space with a metric given by (h_{ij}) it is given by

$$grad(g) = (h^{ij}\frac{\partial g}{\partial \theta_i}.\frac{\partial}{\partial \theta_j})$$

where (h^{ij}) is the inverse of the metric matrix.

Then because g is a C^1 function we see that grad(g) is a continuous vector field. Which means by the theorem on the existence of solutions to ordinary differential equations there exists $\{\gamma_i(s)\}$ the set of integral paths to our vector field,see (Arnol'd (1983))

By flowing along these integral paths we may construct a diffeomorphism between $g^{-1}(0)$ and $g^{-1}(c)$. So we define on $g^{-1}(c)$ the coordinate system $(k_1^c,\dots,k_{p-1}^c)$ by pushing forward the original coordinate system $(k_1^0,\dots,k_{p-1}^0)$ along the γ's. Hence we have local coordinates everywhere defined by $(g(\theta),k_1(\theta),\dots k_{p-1}(\theta))$ and since $\frac{\partial}{\partial g}$ is parallel to grad(g) which is tangent to $\gamma_i(s)$, we have that

$$F\left(\frac{\partial}{\partial g},\frac{\partial}{\partial k_i}\right) = 0 \qquad \forall i$$

Lemma 3.2.2 : Proof

(a)The change of basis going from the (g,k) to θ coordinate systems can easily be seen to be $(G,K)^T$ For vectors orthogonal to the level sets $\gamma_i(s)$ are the integral curves for $\frac{\partial}{\partial g_i}$ then

$$g(\gamma_i(s))= (0....0,s,0....0)$$

with the non zero element in the i'th position.Then differentiating with respect to s we find

$$\sum_{j=1}^{p}\frac{\partial g}{\partial \gamma_j}\frac{d\gamma_j}{dt} = (0,\dots,0,1,0,\dots,0)$$

Hence G takes $\frac{\partial}{\partial g_i}$ to (0...0,1,0....0) and so G is a change of basis matrix. For vectors parallel to the level sets K is the change of basis.

(b) The proof follows exactly the same from as in (a).

Lemma 3.2.4 : Proof

By Lemma 3.2.2 , if the geodesics are orthogonal to the level sets we have $F(\gamma(s),\gamma(s))$ is equal to the Wald Statistic at that point and since F is constant we see the geodesic distance

$$\int_0^1 \sqrt{F_g} dg = \sqrt{F_g} = \text{constant} = \sqrt{\text{Wald}}$$

Lemma 7.1:

If f_{11} is gradiently increasing towards the null then $G^2 \leq W_{\hat{\theta}}$.

Proof. Consider the length of the path $\nu(g(\theta))$ from $\hat{\theta}$ to the null hypothesis which is orthogonal to the level sets of g.

At each point of ν the length of its tangent vector is clearly given by f_{11}. Hence the total length $l(\nu)$ of the path is

$$\int_0^{g(\hat{\theta})} \sqrt{f_{11}(g)}\, dg$$

Now by definition

$$W_{\hat{\theta}} = g^2(\hat{\theta}) f_{11}(\hat{\theta})$$

Therefore if f_{11} is increasing towards the null we have that

$$W_{\hat{\theta}} \geq (l(\nu))^2.$$

However, by definition the distance G is the shortest path length from the estimate to the the null hypothesis. Therefore we have

$$W_{\hat{\theta}} \geq (l(\nu))^2 \geq G^2.$$

References

Arnol'd V.I.,(1983), *Geometric methods in the theory of ordinary differential equations*, Springer Verlag.

Amari Sun-ichi,(1985), *Differential geometrical methods in statistics*, Springer-Verlag Lecture notes in Statistics No.28.

Amari Sun-ichi et al,(1987),*Differential Geometry in Statistical inference*, Institute of Mathematical Statistics Lecture Notes- Monograph Series Volume 10.

Amari,S-I.,Barndorff-Nielsen O.,Kass R.E., Lauritzen S.L.,and C.R.Rao, (1987), *Differential Geometry in Statistical Inference*, Institute of Mathematical Statistics, Lecture Notes-Monograph Series Volume 10.

Barndorff-Nielsen O.,Cox D.R. and Reid N.,(1986), The role of Differential Geometry in Statistical Theory,*International Statistical Review*, 54,1,83-96.

Bates D.M. and Watts D.G. (1980), Relative curvature measures of nonlinearity, *Journal of the Royal Statistical Society, Series B* 423,No.1,1-25

Cox D. and Hinkley D.V., (1973), *Theoretical Statistics*, London Chapman and Hall, London

Critchley F.(1989), On the minimisation of a positive definite quadratic form under a quadratic constraint: analytical solution and statistical applications, *mimeo, Department of Statistics ,University of Warwick*.

Critchley F., P.Marriott and M.Salmon, (1989a), An Introduction to Differential Geometry for Econometrics, *mimeo, Department of Economics, University of Warwick*.

Critchley F., P.Marriott and M.Salmon, (1989b), Geodesic Inference,*mimeo, Department of Statistics, University of Warwick*.

Davidson R.and J.G. MacKinnon, (1987), Implicit alternatives and the local power of test statistics, *Econometrica*,Vol. 55,No.6,pp 1305-1329.

Gregory A. and Veall,(1985), Formulating Wald tests of nonlinear restrictions, *Econometrica*,53,549-563.

Gregory A. and Veall,(1986),Wald tests of common factor restrictions, *Economic Letters*,22,203-208.

Hauck W.W. and Donner A.,(1977),Wald's test as applied to hypotheses in logit analysis, Jnl. Am. Statist. Ass. 72,pps 851-853, Corrigendum (1980), 75, 482.

Kumon M. and Amari S.,(1983), Geometrical theory of higher order asymptotics of test, interval estimator and conditional inference,*Procedings of the Royal Society,London*,A387,429-458.

Lafontaine F.and K.J. White,(1986),Obtaining any Wald Statistic you want,*Economic Letters*,21,35-40.

Marriott P. and Salmon M. ,(1989), On the use and calculation of geodesics in statistics. *Mimeo, Department of Economics, European University Institute, Florence.*

Moolgavkar S.H. and Venzon D.J., (1987),Confidence regions in curved exponential families: application to matched case-control and survial studies with general relative risk function, *The Annals of Statistics*,Vol 15,No.1, pps 346-359.

Phillips P. and J.Y.Park, (1988),On the formulation of Wald tests of nonlinear restrictions,*Econometrica*, Vol56, No.5,pps 1065-1083.

Silvey D. (1975), *Statistical inference*, London Chapman and Hall, London

Spivak (1981), *A Comprehensive Introduction to Differential Geometry* , Publish or Perish, Berkley.

Thorpe J.A.,(1979),*Elementary Topics in Differential Geometry*, Springer Verlag,

Væth M.,(1985),On the use of Wald's test in exponential families, *International Statistical Review* ,53,2, pps 199-214.

Working Papers of the Department of Economics Published since 1989

89/370
B. BENSAID/ R.J. GARY-BOBO/
S. FEDERBUSCH
The Strategic Aspects of Profit Sharing in the Industry

89/372
Jean-Philippe ROBE
Countervailing Duties, State Protectionism and the Challenge of the Uruguay Round

89/374
Francisco S. TORRES
Small Countries and Exogenous Policy Shocks

89/375
Renzo DAVIDDI
Rouble Convertibility: A Realistic Target

89/377
Elettra AGLIARDI
On the Robustness of Contestability Theory

89/378
Stephen MARTIN
The Welfare Consequences of Transaction Costs in Financial Markets

89/381
Susan SENIOR NELLO
Recent Developments in Relations Between the EC and Eastern Europe

89/382
Jean GABSZEWICZ/ Paolo GARELLA/
Charles NOLLET
Spatial Price Competition With Uninformed Buyers

89/383
Benedetto GUI
Beneficiary and Dominant Roles in Organizations: The Case of Nonprofits

89/384
Agustín MARAVALL/ Daniel PEÑA
Missing Observations, Additive Outliers and Inverse Autocorrelation Function

89/385
Stephen MARTIN
Product Differentiation and Market Performance in Oligopoly

89/386
Dalia MARIN
Is the Export-Led Growth Hypothesis Valid for Industrialized Countries?

89/387
Stephen MARTIN
Modeling Oligopolistic Interaction

89/388
Jean-Claude CHOURAQUI
The Conduct of Monetary Policy: What have we Learned From Recent Experience

89/390
Corrado BENASSI
Imperfect Information and Financial Markets: A General Equilibrium Model

89/394
Serge-Christophe KOLM
Adequacy, Equity and Fundamental Dominance: Unanimous and Comparable Allocations in Rational Social Choice, with Applications to Marriage and Wages

89/395
Daniel HEYMANN/ Axel LEIJONHUFVUD
On the Use of Currency Reform in Inflation Stabilization

89/400
Robert J. GARY-BOBO
On the Existence of Equilibrium Configurations in a Class of Asymmetric Market Entry Games

89/402
Stephen MARTIN
Direct Foreign Investment in The United States

89/413
Francisco S. TORRES
Portugal, the EMS and 1992: Stabilization and Liberalization

89/416
Joerg MAYER
Reserve Switches and Exchange-Rate Variability: The Presumed Inherent Instability of the Multiple Reserve-Currency System

89/417
José P. ESPERANÇA/ Neil KAY
Foreign Direct Investment and Competition in the Advertising Sector: The Italian Case

89/418
Luigi BRIGHI/ Mario FORNI
Aggregation Across Agents in Demand Systems

89/420
Corrado BENASSI
A Competitive Model of Credit Intermediation

89/422
Marcus MILLER/ Mark SALMON
When does Coordination pay?

89/423
Marcus MILLER/ Mark SALMON/
Alan SUTHERLAND
Time Consistency, Discounting and the Returns to Cooperation

89/424
Frank CRITCHLEY/ Paul MARRIOTT/
Mark SALMON
On the Differential Geometry of the Wald Test with Nonlinear Restrictions

89/425
Peter J. HAMMOND
On the Impossibility of Perfect Capital Markets

89/426
Peter J. HAMMOND
Perfected Option Markets in Economies with Adverse Selection

89/427
Peter J. HAMMOND
Irreducibility, Resource Relatedness, and Survival with Individual Non-Convexities

EUI
WORKING
PAPERS

EUI Working Papers are published and distributed by the European University Institute, Florence.

Copies can be obtained free of charge - depending on the availability of stocks - from:

The Publications Officer
European University Institute
Badia Fiesolana
I - 50016 San Domenico di Fiesole (FI)
Italy

Please use order form overleaf

PUBLICATIONS OF THE EUROPEAN UNIVERSITY INSTITUTE

To The Publications Officer
European University Institute
Badia Fiesolana
I - 50016 San Domenico di Fiesole (FI)
Italy

From Name ....
Address ....
....
....
....
....

Please send me the following EUI Working Paper(s):

No. ....
Author, title: ....
....
....
....
....

Date ....

Signature

....

89/383
Benedetto GUI
Beneficiary and Dominant Roles in Organizations: The Case of Nonprofits

89/384
Agustín MARAVALL/
Daniel PEÑA
Missing Observations, Additive Outliers and Inverse Autocorrelation Function

89/385
Stephen MARTIN
Product Differentiation and Market Performance in Oligopoly

89/386
Dalia MARIN
Is the Export-Led Growth Hypothesis Valid for Industrialized Countries?

89/387
Stephen MARTIN
Modeling Oligopolistic Interaction

89/388
Jean-Claude CHOURAQUI
The Conduct of Monetary Policy: What have we Learned From Recent Experience

89/389
Léonce BEKEMANS
Economics in Culture vs. Culture in Economics

89/390
Corrado BENASSI
Imperfect Information and Financial Markets: A General Equilibrium Model

89/391
Patrick DEL DUCA
Italian Judicial Activism in Light of French and American Doctrines of Judicial Review and Administrative Decisionmaking: The Case of Air Pollution

89/392
Dieter ZIEGLER
The Bank of England in the Provinces: The Case of the Leicester Branch Closing, 1872

89/393
Gunther TEUBNER
How the Law Thinks: Toward a Constructivist Epistemology of Law

89/394
Serge-Christophe KOLM
Adequacy, Equity and Fundamental Dominance: Unanimous and Comparable Allocations in Rational Social Choice, with Applications to Marriage and Wages

89/395
Daniel HEYMANN/
Axel LEIJONHUFVUD
On the Use of Currency Reform in Inflation Stabilization

* Working Paper out of print

89/396
Gisela BOCK
Challenging Dichotomies: Theoretical and Historical Perspectives on Women's Studies in the Humanities and Social Sciences

89/397
Giovanna C. CIFOLETTI
Quaestio sive aequatio: la nozione di problema nelle *Regulae*

89/398
Michela NACCI
L'équilibre difficile. Georges Friedmann avant la sociologie du travail

89/399
Bruno WANROOIJ
Zefthe Akaira, o delle identità smarrite

89/400
Robert J. GARY-BOBO
On the Existence of Equilibrium Configurations in a Class of Asymmetric Market Entry Games

89/401
Federico ROMERO
The US and Western Europe: A Comparative Discussion of Labor Movements in the Postwar Economy

89/402
Stephen MARTIN
Direct Foreign Investment in The United States

89/403
Christine LAMARRE
La vie des enfants et des vieillards assistés à Dijon au 18e siècle

89/404
Christian JOERGES
Product liability and product safety in the European Community

89/405
Giandomenico MAJONE
Regulating Europe: Problems and Prospects

89/406
Fabio SDOGATI
Exchange Rate Fluctuations and the Patterns of International Trade: A Study of the Flow of Trade from Newly Industrialized Countries to the European Community at the Industry Level

89/407
Angela LIBERATORE
EC Environmental Research and EC Environmental Policy: A study in the utilization of knowledge for regulatory purposes

89/408
J. -Matthias Graf von der SCHULENBURG
Regulation and Deregulation of Insurance Markets in the Federal Republic of Germany

* Working Paper out of print

89/409
Greg KASER
Acceptable Nuclear Risk: Some Examples from Europe

89/410
Léonce BEKEMANS/ Manfred GLAGOW/ Jeremy MOON
Beyond Market and State Alternative Approaches to Meeting Societal Demands

89/411
Erich KAUFER
The Regulation of Drug Development: In Search of a Common European Approach

89/412
Gianna GIANNELLI/ Gøsta ESPING-ANDERSEN
Labor Costs and Employment in the Service Economy

89/413
Francisco S. TORRES
Portugal, the EMS and 1992: Stabilization and Liberalization

89/414
Gøsta ESPING-ANDERSEN/ Harald SONNBERGER
The Demographics of Age in Labor Market Management

89/415
Fritz von NORDHEIM NIELSEN
The Scandinavian Model: Reformist Road to Socialism or Dead End Street?

89/416
Joerg MAYER
Reserve Switches and Exchange-Rate Variability: The Presumed Inherent Instability of the Multiple Reserve-Currency System

89/417
José P. ESPERANÇA/ Neil KAY
Foreign Direct Investment and Competition in the Advertising Sector: The Italian Case

89/418
Luigi BRIGHI/ Mario FORNI
Aggregation Across Agents in Demand Systems

89/419
Hans Ulrich JESSURUN d'OLIVEIRA
Nationality and Apartheid: Some Reflections on the Use of Nationality Law as a Weapon against Violation of Fundamental Rights

89/420
Corrado BENASSI
A Competitive Model of Credit Intermediation

89/421
Ester STEVERS
Telecommunications Regulation in the European Community: The Commission of the European Communities as Regulatory Actor

89/422
Marcus MILLER/ Mark SALMON
When does Coordination pay?

* Working Paper out of print

89/423
Marcus MILLER/ Mark SALMON/
Alan SUTHERLAND
Time Consistency, Discounting and the Returns to Cooperation

89/424
Frank CRITCHLEY/ Paul MARRIOTT/
Mark SALMON
On the Differential Geometry of the Wald Test with Nonlinear Restrictions

89/425
Peter J. HAMMOND
On the Impossibility of Perfect Capital Markets

89/426
Peter J. HAMMOND
Perfected Option Markets in Economies with Adverse Selection

89/427
Peter J. HAMMOND
Irreducibility, Resource Relatedness, and Survival with Individual Non-Convexities

89/428
Joanna GOYDER
"Business Format" Franchising and EEC Competition Law

* Working Paper out of print