

A note on the local power of the LR, Wald, score and gradient tests

Artur J. Lemonte and Silvia L.P. Ferrari

Departamento de Estatística, Universidade de São Paulo, São Paulo/SP, Brazil

e-mail: arturlemonte@gmail.com; silviaferrari.usp@gmail.com

Abstract: This paper examines the local power of the likelihood ratio, Wald, score and gradient tests under the presence of a scalar parameter, ϕ say, that is orthogonal to the remaining parameters. We show that some of the coefficients that define the local powers remain unchanged regardless of whether ϕ is known or needs to be estimated, whereas the others can be written as the sum of two terms, the first of which being the corresponding term obtained as if ϕ were known, and the second, an additional term yielded by the fact that ϕ is unknown. The contribution of each set of parameters on the local powers of the tests can then be examined. Various implications of our main result are stated and discussed. Several examples are presented for illustrative purposes.

Keywords and phrases: asymptotic expansions, gradient test, likelihood ratio test, local power, score test, wald test.

Received February 2012.

Contents

1	Introduction	421
2	Main result	424
3	Examples	429
	Acknowledgments	433
	References	433

1. Introduction

The likelihood ratio (LR), Wald and Rao score tests are usually employed for testing hypotheses in parametric models. These tests have been widely used in economics, engineering and biology, among other fields. Recently, [Terrell \(2002\)](#) proposed a new criterion for testing hypotheses, referred to as the *gradient test*. As we will see below, the gradient statistic is very simple to compute when compared with the Wald and the score statistics. Due to its simplicity, [Rao \(2005\)](#) wrote: “The suggestion by Terrell is attractive as it is simple to compute. It would be of interest to investigate the performance of the [gradient] statistic.” An interesting result about the gradient statistic is that it shares the same first order asymptotic properties with the LR, Wald and score statistics. That is, to the first order of approximation, these statistics have the same asymptotic

distributional properties either under the null hypothesis or under a sequence of Pitman alternatives, i.e. a sequence of local alternatives that shrink to the null hypothesis at a convergence rate $n^{-1/2}$, n being the sample size. Additionally, it is known that, up to an error of order n^{-1} , the LR, Wald, score and gradient tests have the same size properties but their local powers differ in the $n^{-1/2}$ term. Therefore, a meaningful comparison among the criteria can be performed by comparing the nonnull asymptotic expansions to order $n^{-1/2}$ of the distribution functions of these statistics under a sequence of Pitman alternatives.

The nonnull asymptotic expansions up to order $n^{-1/2}$ for the distribution functions of the LR and Wald statistics for testing a composite hypothesis in the presence of nuisance parameters were derived by Hayakawa (1975). Harris and Peers (1980) obtained an analogous result for the score statistic. The asymptotic expansion up to order $n^{-1/2}$ for the distribution function of the gradient statistic was derived by Lemonte and Ferrari (2012). The null asymptotic expansion up to order n^{-1} for the distribution function of the likelihood ratio statistic for testing a composite hypothesis in the presence of nuisance parameters is given in Hayakawa (1977, 1987), while an analogous result for the score statistic was obtained by Harris (1985); see also Hayakawa and Puri (1985). The derivation of the null asymptotic expansion up to order n^{-1} for the distribution function of the gradient statistic is in progress and will be published in a future article.

Let $\pi(\boldsymbol{\theta})$ be a continuous parametric model and $\ell(\boldsymbol{\theta})$ be the corresponding total log-likelihood function, where $\boldsymbol{\theta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top, \phi)^\top$ is a $(p+1)$ -vector of unknown parameters. The dimensions of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are q and $p-q$, respectively, and ϕ is a scalar parameter. We focus on testing the composite null hypothesis $\mathcal{H}_0 : \boldsymbol{\beta}_2 = \boldsymbol{\beta}_{20}$ against the two-sided alternative hypothesis $\mathcal{H}_1 : \boldsymbol{\beta}_2 \neq \boldsymbol{\beta}_{20}$, where $\boldsymbol{\beta}_{20}$ is a specified vector and $\boldsymbol{\beta}_1$ and ϕ are nuisance parameters. Let $\mathbf{U}_\boldsymbol{\theta} = \partial\ell(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ and $\mathbf{K}_\boldsymbol{\theta} = \mathbb{E}(\mathbf{U}_\boldsymbol{\theta}\mathbf{U}_\boldsymbol{\theta}^\top)$ be the score function and the Fisher information matrix for $\boldsymbol{\theta}$, respectively. Also, let $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}_1^\top, \hat{\boldsymbol{\beta}}_2^\top, \hat{\phi})^\top$ and $\tilde{\boldsymbol{\theta}} = (\tilde{\boldsymbol{\beta}}_1^\top, \tilde{\boldsymbol{\beta}}_{20}^\top, \tilde{\phi})^\top$ denote the unrestricted (under \mathcal{H}_1) and restricted (under \mathcal{H}_0) maximum likelihood estimators of $\boldsymbol{\theta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top, \phi)^\top$. The likelihood ratio (S_1), Wald (S_2), score (S_3) and gradient (S_4) statistics for testing \mathcal{H}_0 versus \mathcal{H}_1 are given, respectively, by

$$S_1 = 2\{\ell(\hat{\boldsymbol{\theta}}) - \ell(\tilde{\boldsymbol{\theta}})\}, \quad S_2 = (\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}})^\top \widehat{\mathbf{K}}_\boldsymbol{\theta}(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}),$$

$$S_3 = \tilde{\mathbf{U}}_\boldsymbol{\theta}^\top \widetilde{\mathbf{K}}_\boldsymbol{\theta}^{-1} \tilde{\mathbf{U}}_\boldsymbol{\theta}, \quad S_4 = \tilde{\mathbf{U}}_\boldsymbol{\theta}^\top (\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}),$$

where $\widehat{\mathbf{K}}_\boldsymbol{\theta} = \mathbf{K}_\boldsymbol{\theta}(\hat{\boldsymbol{\theta}})$, $\widetilde{\mathbf{K}}_\boldsymbol{\theta} = \mathbf{K}_\boldsymbol{\theta}(\tilde{\boldsymbol{\theta}})$ and $\tilde{\mathbf{U}}_\boldsymbol{\theta} = \mathbf{U}_\boldsymbol{\theta}(\tilde{\boldsymbol{\theta}})$. The limiting distribution of S_1 , S_2 , S_3 and S_4 is χ_{p-q}^2 under \mathcal{H}_0 . Under \mathcal{H}_1 , these statistics have a $\chi_{p-q, \lambda}^2$, i.e. a noncentral chi-square distribution with $p-q$ degrees of freedom and an appropriate noncentrality parameter λ . The null hypothesis is rejected for a given nominal level, γ say, if the test statistic exceeds the upper $100(1-\gamma)\%$ quantile of the χ_{p-q}^2 distribution.

In this paper, we shall assume that $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$ is globally orthogonal to ϕ in the sense of Cox and Reid (1987). In other words, the Fisher information

matrix for θ and its inverse are block-diagonal,

$$\mathbf{K}_\theta = \text{diag}\{\mathbf{K}_\beta, K_\phi\}, \quad \mathbf{K}_\theta^{-1} = \text{diag}\{\mathbf{K}_\beta^{-1}, K_\phi^{-1}\},$$

say. Here, \mathbf{K}_β is the Fisher information matrix for β and K_ϕ is the information relative to ϕ . There are numerous statistical models for which global orthogonality holds; see Section 3.

The global orthogonality of the parameters will be exploited and we will show an interesting decomposition of the $n^{-1/2}$ term of the expansion of the nonnull distribution function of the four statistics. From the partition of θ and the global orthogonality between β and ϕ , we have the corresponding partitions:

$$\mathbf{U}_\theta = (\mathbf{U}_{\beta_1}^\top, \mathbf{U}_{\beta_2}^\top, U_\phi)^\top, \quad \mathbf{K}_\beta = \begin{bmatrix} \mathbf{K}_{\beta 11} & \mathbf{K}_{\beta 12} \\ \mathbf{K}_{\beta 21} & \mathbf{K}_{\beta 22} \end{bmatrix}, \quad \mathbf{K}_\beta^{-1} = \begin{bmatrix} \mathbf{K}^{11} & \mathbf{K}^{12} \\ \mathbf{K}^{21} & \mathbf{K}^{22} \end{bmatrix}.$$

Hence, the statistics S_2, S_3 and S_4 can be rewritten as

$$S_2 = (\hat{\beta}_2 - \beta_{20})^\top \widehat{\mathbf{K}}^{22^{-1}} (\hat{\beta}_2 - \beta_{20}), \quad S_3 = \widetilde{\mathbf{U}}_{\beta_2}^\top \widetilde{\mathbf{K}}^{22} \widetilde{\mathbf{U}}_{\beta_2}, \quad S_4 = \widetilde{\mathbf{U}}_{\beta_2}^\top (\hat{\beta}_2 - \beta_{20}),$$

where $\widehat{\mathbf{K}}^{22} = \mathbf{K}^{22}(\hat{\theta})$, $\widetilde{\mathbf{K}}^{22} = \mathbf{K}^{22}(\tilde{\theta})$ and $\widetilde{\mathbf{U}}_{\beta_2} = \mathbf{U}_{\beta_2}(\tilde{\theta})$. Notice that S_4 , the gradient statistic, has a very simple form and does not involve the information matrix, neither expected nor observed, unlike S_2 and S_3 . Terrell (2002) points out that the gradient statistic “is not transparently non-negative, even though it must be so asymptotically.” His Theorem 2 implies that if the log-likelihood function is concave and is differentiable at $\tilde{\theta}$, then $S_4 \geq 0$.

The subject matter of this note is the local power of the LR, Wald, score and gradient tests for testing the null hypothesis $\mathcal{H}_0 : \beta_2 = \beta_{20}$ under a sequence of Pitman alternatives, when global orthogonality between β and ϕ holds. The nonnull distribution function of the statistics S_1, S_2, S_3 and S_4 under Pitman alternatives for testing $\mathcal{H}_0 : \beta_2 = \beta_{20}$ takes the form

$$\Pr(S_i \leq x) = G_{f,\lambda}(x) + \sum_{k=0}^3 b_{ik} G_{f+2k,\lambda}(x) + O(n^{-1}), \quad (1)$$

for $i = 1, 2, 3, 4$, where $G_{f,\lambda}(x)$ is the cumulative distribution function of a non-central chi-square variate with f degrees of freedom and an appropriate non-centrality parameter λ . Here, $f = p - q$. Clearly, the local power (up to order $n^{-1/2}$) of the four corresponding tests are given by $1 - \Pr(S_i \leq x)$, where x is replaced by the appropriate quantile of the χ_{p-q}^2 distribution according to the chosen nominal level. The coefficients b_{ik} ($i = 1, 2, 3, 4$ and $k = 0, 1, 2, 3$) and λ are given in Hayakawa (1975), Harris and Peers (1980) and Lemonte and Ferrari (2012), and are reproduced in Section 2.

We will show that the coefficients b_{i2} and b_{i3} , for $i = 1, 2, 3, 4$, in (1) remain unchanged regardless of whether ϕ is known or needs to be estimated, whereas the coefficients b_{i1} , for $i = 1, 2, 3, 4$, can be written as the sum of two terms, the first of which being the corresponding term obtained as if ϕ were known, and the second, an additional term yielded by the fact that ϕ is unknown. A sufficient

condition under which this additional term is zero will be given. The general result derived in this paper allows one to explicitly verify the contribution of each parameter on the local power of the LR, Wald, score and gradient tests for testing the null hypothesis \mathcal{H}_0 . We also discuss on the local power of the tests for testing the null hypothesis $\mathcal{H}_0 : \phi = \phi_0$, where ϕ_0 is a specified scalar value, when β acts as a vector of nuisance parameters. Some examples which include probability density functions and regression models are considered to illustrate our result.

2. Main result

In order to derive our main result, we shall assume that the local alternative hypothesis is $\mathcal{H}_{1n} : \beta_2 = \beta_{20} + \epsilon$, where $\epsilon = (\epsilon_{q+1}, \dots, \epsilon_p)^\top$ is of order $n^{-1/2}$. We define the quantities

$$\epsilon^* = \begin{bmatrix} \mathbf{K}_{\beta_{11}}^{-1} \mathbf{K}_{\beta_{12}} \\ -\mathbf{I}_{p-q} \\ \mathbf{0} \end{bmatrix} \epsilon, \quad \mathbf{A} = ((a_{rs})) = \begin{bmatrix} \mathbf{A}_\beta & \mathbf{0} \\ \mathbf{0} & K_\phi^{-1} \end{bmatrix},$$

$$\mathbf{M} = ((m_{rs})) = \begin{bmatrix} \mathbf{M}_\beta & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix},$$

where \mathbf{I}_{p-q} is a $(p-q) \times (p-q)$ identity matrix, $r, s = 1, \dots, p+1$, and

$$\mathbf{M}_\beta = \mathbf{K}_\beta^{-1} - \mathbf{A}_\beta, \quad \mathbf{A}_\beta = \begin{bmatrix} \mathbf{K}_{\beta_{11}}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

Note that whenever an index equals $p+1$, it refers to ϕ , the last component of the parameter vector. To make the notation more intuitive, in many instances we will write ϕ instead of $p+1$. We have $m_{r\phi} = m_{\phi r} = m_{\phi\phi} = 0$, $a_{r\phi} = a_{\phi r} = 0$ ($r = 1, \dots, p$) and $a_{\phi\phi} = K_\phi^{-1}$. In the following, we use the standard notation for the cumulants of the log-likelihood derivatives (Lawley, 1956; Harris, 1985). Let $U_r = \partial\ell(\boldsymbol{\theta})/\partial\theta_r$, $U_{rs} = \partial^2\ell(\boldsymbol{\theta})/\partial\theta_r\partial\theta_s$, $U_{rst} = \partial^3\ell(\boldsymbol{\theta})/\partial\theta_r\partial\theta_s\partial\theta_t$ and so on, for $r, s, t = 1, \dots, p+1$. We have $\kappa_{rs} = \mathbb{E}(U_{rs})$, $\kappa_{rst} = \mathbb{E}(U_{rst})$, $\kappa_{r,st} = \mathbb{E}(U_r U_{rs})$, $\kappa_{r,s,t} = \mathbb{E}(U_r U_s U_t)$, etc. All cumulants κ 's refer to a total over the sample and are, in general, of order n .

The general expressions for the coefficients b_{ik} 's ($i = 1, 2, 3, 4$ and $k = 0, 1, 2, 3$) that define the nonnull expansions of the distribution function of the statistics S_1, S_2, S_3 and S_4 under Pitman alternatives for testing $\mathcal{H}_0 : \beta_2 = \beta_{20}$, which are given in (1), can be written as

$$b_{11} = -\frac{1}{6} \sum_{r,s,t=1}^{p+1} (\kappa_{rst} - 2\kappa_{r,s,t}) \epsilon_r^* \epsilon_s^* \epsilon_t^* - \frac{1}{2} \sum_{r,s,t=1}^{p+1} (\kappa_{rst} + 2\kappa_{r,st}) a_{rs} \epsilon_t^*$$

$$- \frac{1}{2} \sum_{r=q+1}^p \sum_{s,t=1}^{p+1} (\kappa_{rst} + \kappa_{r,st}) \epsilon_r \epsilon_s^* \epsilon_t^*,$$

$$\begin{aligned}
b_{21} &= -\frac{1}{2} \sum_{r,s,t=1}^{p+1} (\kappa_{rst} + 2\kappa_{r,st}) \epsilon_r^* \epsilon_s^* \epsilon_t^* + \sum_{r,s,t=1}^{p+1} \kappa_{r,st} m_{rs} \epsilon_t^* \\
&\quad - \frac{1}{2} \sum_{r,s,t=1}^p (\kappa_{rst} + 2\kappa_{r,st}) \kappa^{r,s} \epsilon_t^* - \frac{1}{2} \sum_{r=q+1}^p \sum_{s,t=1}^{p+1} (\kappa_{rst} + \kappa_{r,st}) \epsilon_r \epsilon_s^* \epsilon_t^*, \\
b_{31} &= -\frac{1}{6} \sum_{r,s,t=1}^{p+1} (\kappa_{rst} - 2\kappa_{r,s,t}) \epsilon_r^* \epsilon_s^* \epsilon_t^* + \frac{1}{2} \sum_{r,s,t=1}^{p+1} \kappa_{r,s,t} m_{rs} \epsilon_t^* \\
&\quad - \frac{1}{2} \sum_{r,s,t=1}^{p+1} (\kappa_{rst} + 2\kappa_{r,st}) a_{rs} \epsilon_t^* - \frac{1}{2} \sum_{r=q+1}^p \sum_{s,t=1}^{p+1} (\kappa_{rst} + \kappa_{r,st}) \epsilon_r \epsilon_s^* \epsilon_t^*, \\
b_{41} &= \frac{1}{4} \sum_{r,s,t=1}^{p+1} \kappa_{rst} \kappa^{r,s} \epsilon_t^* - \frac{1}{2} \sum_{r,s,t=1}^{p+1} (\kappa_{rst} + 2\kappa_{r,st}) \epsilon_r^* \epsilon_s^* \epsilon_t^* \\
&\quad - \frac{1}{4} \sum_{r,s,t=1}^{p+1} (4\kappa_{r,st} + 3\kappa_{rst}) a_{rs} \epsilon_t^* - \frac{1}{2} \sum_{r=q+1}^p \sum_{s,t=1}^{p+1} (\kappa_{rst} + \kappa_{r,st}) \epsilon_r \epsilon_s^* \epsilon_t^*, \\
b_{12} &= -\frac{1}{6} \sum_{r,s,t=1}^{p+1} \kappa_{r,s,t} \epsilon_r^* \epsilon_s^* \epsilon_t^*, \quad b_{13} = 0, \\
b_{22} &= \frac{1}{2} \sum_{r,s,t=1}^{p+1} \kappa_{r,st} \epsilon_r^* \epsilon_s^* \epsilon_t^* + \frac{1}{2} \sum_{r,s,t=1}^{p+1} \kappa_{rst} m_{rs} \epsilon_t^*, \\
b_{23} &= \frac{1}{6} \sum_{r,s,t=1}^{p+1} \kappa_{rst} \epsilon_r^* \epsilon_s^* \epsilon_t^*, \quad b_{32} = -\frac{1}{2} \sum_{r,s,t=1}^{p+1} \kappa_{r,s,t} m_{rs} \epsilon_t^*, \\
b_{33} &= -\frac{1}{6} \sum_{r,s,t=1}^{p+1} \kappa_{r,s,t} \epsilon_r^* \epsilon_s^* \epsilon_t^*, \\
b_{42} &= -\frac{1}{4} \sum_{r,s,t=1}^{p+1} \kappa_{rst} m_{rs} \epsilon_t^* + \frac{1}{4} \sum_{r,s,t=1}^{p+1} (\kappa_{rst} + 2\kappa_{r,st}) \epsilon_r^* \epsilon_s^* \epsilon_t^*, \\
b_{43} &= -\frac{1}{12} \sum_{r,s,t=1}^{p+1} \kappa_{r,s,t} \epsilon_r^* \epsilon_s^* \epsilon_t^*,
\end{aligned}$$

and $b_{i0} = -(b_{i1} + b_{i2} + b_{i3})$, for $i = 1, 2, 3, 4$, where $\kappa^{r,s}$ is the (r, s) th element of \mathbf{K}_θ^{-1} and ϵ_r^* is the r th element of the vector $\boldsymbol{\epsilon}^*$. Here, the non-centrality parameter is given by $\lambda = \boldsymbol{\epsilon}^{*\top} \mathbf{K}_\theta \boldsymbol{\epsilon}^*$. The coefficients b_{ik} 's are of order $n^{-1/2}$ and all quantities except $\boldsymbol{\epsilon}$ are evaluated under the null hypothesis.

Based on the general expressions of the coefficients presented above and exploiting the orthogonality between $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$, we arrive, after long and tedious algebraic manipulations, at the following general result.

Theorem 1 Let $\boldsymbol{\theta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top, \phi)^\top$ be the parameter vector of dimension $p+1$, where the dimensions of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are q and $p-q$, respectively, and ϕ is a scalar parameter. Assume that $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$ and ϕ are globally orthogonal. The nonnull asymptotic expansions of the distribution functions of the LR, Wald, score and gradient statistics for testing the null hypothesis $\mathcal{H}_0 : \boldsymbol{\beta}_2 = \boldsymbol{\beta}_{20}$ under a sequence of Pitman alternatives are given by (1) with $f = p-q$, $\lambda = \boldsymbol{\epsilon}^\top (\mathbf{K}_{\boldsymbol{\beta}22} - \mathbf{K}_{\boldsymbol{\beta}21} \mathbf{K}_{\boldsymbol{\beta}11}^{-1} \mathbf{K}_{\boldsymbol{\beta}12}) \boldsymbol{\epsilon}$, and

$$\begin{aligned} b_{11} &= b_{11}^0 + \xi, & b_{21} &= b_{21}^0 + \xi, & b_{31} &= b_{31}^0 + \xi, & b_{41} &= b_{41}^0 + \xi, \\ b_{12} &= b_{33} = 2b_{43} = -\frac{1}{6} \sum_{r,s,t=1}^p \kappa_{r,s,t} \epsilon_r^* \epsilon_s^* \epsilon_t^*, \\ b_{22} &= \frac{1}{2} \sum_{r,s,t=1}^p \kappa_{r,st} \epsilon_r^* \epsilon_s^* \epsilon_t^* + \frac{1}{2} \sum_{r,s,t=1}^p \kappa_{rst} m_{rs} \epsilon_t^*, \\ b_{13} &= 0, & b_{23} &= \frac{1}{6} \sum_{r,s,t=1}^p \kappa_{rst} \epsilon_r^* \epsilon_s^* \epsilon_t^*, & b_{32} &= -\frac{1}{2} \sum_{r,s,t=1}^p \kappa_{r,s,t} m_{rs} \epsilon_t^*, \\ b_{42} &= -\frac{1}{4} \sum_{r,s,t=1}^p \kappa_{rst} m_{rs} \epsilon_t^* + \frac{1}{4} \sum_{r,s,t=1}^p (\kappa_{rst} + 2\kappa_{r,st}) \epsilon_r^* \epsilon_s^* \epsilon_t^*, \end{aligned}$$

where

$$\xi = -\frac{K_\phi^{-1}}{2} \sum_{t=1}^p (\kappa_{\phi\phi t} + 2\kappa_{\phi,\phi t}) \epsilon_t^*,$$

with $\kappa_{\phi\phi t} = \mathbb{E}(\partial^3 \ell(\boldsymbol{\theta}) / \partial \phi^2 \partial \beta_t)$ and $\kappa_{\phi,\phi t} = \mathbb{E}\{(\partial \ell(\boldsymbol{\theta}) / \partial \phi)(\partial^2 \ell(\boldsymbol{\theta}) / \partial \phi \partial \beta_t)\}$,

$$\begin{aligned} b_{11}^0 &= -\frac{1}{6} \sum_{r,s,t=1}^p (\kappa_{rst} - 2\kappa_{r,s,t}) \epsilon_r^* \epsilon_s^* \epsilon_t^* - \frac{1}{2} \sum_{r,s,t=1}^p (\kappa_{rst} + 2\kappa_{r,st}) a_{rs} \epsilon_t^* \\ &\quad - \frac{1}{2} \sum_{r=q+1}^p \sum_{s,t=1}^p (\kappa_{rst} + \kappa_{r,st}) \epsilon_r \epsilon_s^* \epsilon_t^*, \\ b_{21}^0 &= -\frac{1}{2} \sum_{r,s,t=1}^p (\kappa_{rst} + 2\kappa_{r,st}) \epsilon_r^* \epsilon_s^* \epsilon_t^* + \sum_{r,s,t=1}^p \kappa_{r,st} m_{rs} \epsilon_t^* \\ &\quad - \frac{1}{2} \sum_{r,s,t=1}^p (\kappa_{rst} + 2\kappa_{r,st}) \kappa^{r,s} \epsilon_t^* - \frac{1}{2} \sum_{r=q+1}^p \sum_{s,t=1}^p (\kappa_{rst} + \kappa_{r,st}) \epsilon_r \epsilon_s^* \epsilon_t^*, \\ b_{31}^0 &= -\frac{1}{6} \sum_{r,s,t=1}^p (\kappa_{rst} - 2\kappa_{r,s,t}) \epsilon_r^* \epsilon_s^* \epsilon_t^* + \frac{1}{2} \sum_{r,s,t=1}^p \kappa_{r,s,t} m_{rs} \epsilon_t^* \\ &\quad - \frac{1}{2} \sum_{r,s,t=1}^p (\kappa_{rst} + 2\kappa_{r,st}) a_{rs} \epsilon_t^* - \frac{1}{2} \sum_{r=q+1}^p \sum_{s,t=1}^p (\kappa_{rst} + \kappa_{r,st}) \epsilon_r \epsilon_s^* \epsilon_t^*, \end{aligned}$$

$$\begin{aligned}
 b_{41}^0 &= \frac{1}{4} \sum_{r,s,t=1}^p \kappa_{rst} \kappa^{r,s} \epsilon_t^* - \frac{1}{2} \sum_{r,s,t=1}^p (\kappa_{rst} + 2\kappa_{r,st}) \epsilon_r^* \epsilon_s^* \epsilon_t^* \\
 &\quad - \frac{1}{4} \sum_{r,s,t=1}^p (4\kappa_{r,st} + 3\kappa_{rst}) a_{rs} \epsilon_t^* - \frac{1}{2} \sum_{r=q+1}^p \sum_{s,t=1}^p (\kappa_{rst} + \kappa_{r,st}) \epsilon_r \epsilon_s^* \epsilon_t^*,
 \end{aligned}$$

and $b_{i0} = -(b_{i1} + b_{i2} + b_{i3})$, for $i = 1, 2, 3, 4$.

Notice that b_{i1}^0 ($i = 1, 2, 3, 4$) and b_{ik} ($i = 1, 2, 3, 4$ and $k = 2, 3$) represent the contribution of the parameter vector β to the local power of the LR, Wald, score and gradient tests for testing the null hypothesis $\mathcal{H}_0 : \beta_2 = \beta_{20}$, since these expressions are only obtained over the components of β , i.e. as if ϕ were known. On the other hand, the quantity ξ , which depends on third-order mixed cumulants involving ϕ and β , can be regarded as the contribution of the parameter ϕ to the local power of the LR, Wald, score and gradient tests when it is unknown, that is, when it needs to be estimated. It is interesting to note that the contribution yielded by the fact that ϕ is unknown is the same for the four tests. Additionally, the contribution of the parameter ϕ to the local power of the tests only appears in the coefficient b_{i1} ($i = 1, 2, 3, 4$) and, of course, in b_{i0} ($i = 1, 2, 3, 4$).

Theorem 1 implies that the limiting distribution of the four statistics, namely a non-central chi-square distribution with non-centrality parameter λ , is the same regardless of whether ϕ is known or estimated from the data. Notice that ξ is the only term that involves cumulants of log-likelihood derivatives with respect to ϕ , and it decreases with the Fisher information for ϕ and vanishes if ϕ is known. By using the Bartlett identity $\kappa_{\phi,\phi t} = \kappa_{\phi t}^{(\phi)} - \kappa_{\phi\phi t}$, for $t = 1, \dots, p$, where $\kappa_{\phi t}^{(\phi)} = \partial \kappa_{\phi t} / \partial \phi$, we have $\kappa_{\phi,\phi t} = -\kappa_{\phi\phi t}$ since the orthogonality between β and ϕ implies that $\kappa_{\phi t} = 0$. Therefore, we can write

$$\xi = \frac{K_\phi^{-1}}{2} \sum_{t=1}^p \kappa_{\phi\phi t} \epsilon_t^*. \tag{2}$$

Theorem 1 has a practical application when the goal is to obtain explicit formulas to the nonnull distribution function of any of the four tests for special models in which orthogonality holds. It suggests that the coefficients b_{ik} 's should be obtained as if the scalar orthogonal parameter ϕ were known, and the extra contribution due to the estimation of ϕ should be obtained from (2).

Now, let Π_i^0 and Π_i , for $i = 1, 2, 3, 4$, be the local powers (ignoring terms of order smaller than $n^{-1/2}$) of the test that uses the statistic S_i when ϕ is known and when ϕ is unknown, respectively. It is well known that $G_{m,\lambda}(x) - G_{m+2,\lambda}(x) = 2g_{m+2,\lambda}(x)$, where $g_{f,\lambda}(x)$ is the probability density function of a non-central chi-square variate with f degrees of freedom and non-centrality parameter λ . We can then write $\Pi_i - \Pi_i^0 = c\xi$ for $i = 1, 2, 3, 4$, where $c = 2g_{p-q+2,\lambda}(x) > 0$ and x represents the appropriate quantile of the reference distribution for the chosen nominal level. Therefore, the difference between the local powers can be zero, or it can increase or decrease when ϕ needs to be

estimated, depending on the sign of the components of ϵ . If $\kappa_{\phi\phi t} = 0$, for $t = 1, \dots, p$, we have $\xi = 0$ and hence the local powers of the four tests do not change when a scalar parameter, which is globally orthogonal to the remaining parameters, is included in the model specification. In the next section we will present various examples for which this happens.

If $q = 0$, the null hypothesis takes the form $\mathcal{H}_0 : \beta = \beta_0$. Also, $\epsilon^* = -\epsilon$ and $\mathbf{A} = \mathbf{0}$. Therefore, an immediate consequence of Theorem 1 is the following corollary.

Corollary 1 *Let $\theta = (\beta, \phi)^\top$ be the parameter vector with β and ϕ being globally orthogonal parameters. The nonnull asymptotic expansions of the distribution functions of the LR, Wald, score and gradient statistics for testing the null hypothesis $\mathcal{H}_0 : \beta = \beta_0$ under a sequence of Pitman alternatives are given by (1) with $f = p$ and $\lambda = \epsilon^\top \mathbf{K}_\beta \epsilon$, and the coefficients are $b_{11} = b_{11}^0 + \xi$, $b_{21} = b_{21}^0 + \xi$, $b_{31} = b_{31}^0 + \xi$, $b_{41} = b_{41}^0 + \xi$, $\xi = \frac{1}{2} K_\phi^{-1} \sum_{t=1}^p (\kappa_{\phi\phi t} + 2\kappa_{\phi, \phi t}) \epsilon_t$, $b_{11}^0 = \frac{1}{6} \sum_{r,s,t=1}^p (\kappa_{rst} - 2\kappa_{r,s,t}) \epsilon_r \epsilon_s \epsilon_t - \frac{1}{2} \sum_{r=q+1}^p \sum_{s,t=1}^p (\kappa_{rst} + \kappa_{r,st}) \epsilon_r \epsilon_s \epsilon_t$,*

$$\begin{aligned} b_{21}^0 &= \frac{1}{2} \sum_{r,s,t=1}^p (\kappa_{rst} + 2\kappa_{r,st}) \epsilon_r \epsilon_s \epsilon_t + \frac{1}{2} \sum_{r,s,t=1}^p \kappa_{rst} \kappa^{r,s} \epsilon_t \\ &\quad - \frac{1}{2} \sum_{r=q+1}^p \sum_{s,t=1}^p (\kappa_{rst} + \kappa_{r,st}) \epsilon_r \epsilon_s \epsilon_t, \end{aligned}$$

$$\begin{aligned} b_{31}^0 &= \frac{1}{6} \sum_{r,s,t=1}^p (\kappa_{rst} - 2\kappa_{r,s,t}) \epsilon_r \epsilon_s \epsilon_t - \frac{1}{2} \sum_{r,s,t=1}^p \kappa_{r,s,t} \kappa^{r,s} \epsilon_t \\ &\quad - \frac{1}{2} \sum_{r=q+1}^p \sum_{s,t=1}^p (\kappa_{rst} + \kappa_{r,st}) \epsilon_r \epsilon_s \epsilon_t, \end{aligned}$$

$$\begin{aligned} b_{41}^0 &= -\frac{1}{4} \sum_{r,s,t=1}^p \kappa_{rst} \kappa^{r,s} \epsilon_t + \frac{1}{2} \sum_{r,s,t=1}^p (\kappa_{rst} + 2\kappa_{r,st}) \epsilon_r \epsilon_s \epsilon_t \\ &\quad - \frac{1}{2} \sum_{r=q+1}^p \sum_{s,t=1}^p (\kappa_{rst} + \kappa_{r,st}) \epsilon_r \epsilon_s \epsilon_t, \end{aligned}$$

$b_{12} = b_{33} = 2b_{43} = \frac{1}{6} \sum_{r,s,t=1}^p \kappa_{r,s,t} \epsilon_r \epsilon_s \epsilon_t$, $b_{13} = 0$, $b_{22} = -\frac{1}{2} \sum_{r,s,t=1}^p \kappa_{r,st} \epsilon_r \epsilon_s \epsilon_t - \frac{1}{2} \sum_{r,s,t=1}^p \kappa_{rst} \kappa^{r,s} \epsilon_t$, $b_{23} = -\frac{1}{6} \sum_{r,s,t=1}^p \kappa_{rst} \epsilon_r \epsilon_s \epsilon_t$, $b_{32} = \frac{1}{2} \sum_{r,s,t=1}^p \kappa_{r,s,t} \kappa^{r,s} \epsilon_t$, $b_{42} = \frac{1}{4} \sum_{r,s,t=1}^p \kappa_{rst} \kappa^{r,s} \epsilon_t - \frac{1}{4} \sum_{r,s,t=1}^p (\kappa_{rst} + 2\kappa_{r,st}) \epsilon_r \epsilon_s \epsilon_t$ and $b_{i0} = -(b_{i1} + b_{i2} + b_{i3})$, for $i = 1, 2, 3, 4$.

If $q = 0$ and $p = 1$, the null hypothesis is $\mathcal{H}_0 : \beta = \beta_0$, where β_0 is a specified scalar, and hence we have the corollary.

Corollary 2 *Let $\theta = (\beta, \phi)^\top$ be the parameter vector with β and ϕ being globally orthogonal scalar parameters. The nonnull asymptotic expansions of the*

distribution functions of the LR, Wald, score and gradient statistics for testing the null hypothesis $\mathcal{H}_0 : \beta = \beta_0$ under a sequence of Pitman alternatives are given by (1) with $f = 1$ and $\lambda = \epsilon^2 K_\beta$, and the coefficients are $b_{11} = b_{11}^0 + \xi$, $b_{21} = b_{21}^0 + \xi$, $b_{31} = b_{31}^0 + \xi$, $b_{41} = b_{41}^0 + \xi$, $\xi = \frac{1}{2} K_\phi^{-1} (\kappa_{\phi\phi\phi} + 2\kappa_{\phi,\phi\phi})\epsilon$, $b_{11}^0 = -\frac{1}{6}(\kappa_{\beta\beta\beta} + \kappa_{\beta,\beta,\beta})\epsilon^3$, $b_{21}^0 = \frac{1}{2}\kappa_{\beta\beta\beta}K_\beta^{-1}\epsilon + \frac{1}{2}\kappa_{\beta,\beta,\beta}\epsilon^3$, $b_{31}^0 = -\frac{1}{2}\kappa_{\beta,\beta,\beta}K_\beta^{-1}\epsilon - \frac{1}{6}(\kappa_{\beta\beta\beta} + \kappa_{\beta,\beta,\beta})\epsilon^3$, $b_{41}^0 = -\frac{1}{4}\kappa_{\beta\beta\beta}K_\beta^{-1}\epsilon + \frac{1}{2}\kappa_{\beta,\beta,\beta}\epsilon^3$, $b_{12} = b_{33} = 2b_{43} = \frac{1}{6}\kappa_{\beta,\beta,\beta}\epsilon^3$, $b_{13} = 0$, $b_{22} = -\frac{1}{2}\kappa_{\beta,\beta,\beta}\epsilon^3 - \frac{1}{2}\kappa_{\beta\beta\beta}K_\beta^{-1}\epsilon$, $b_{23} = -\frac{1}{6}\kappa_{\beta\beta\beta}\epsilon^3$, $b_{32} = \frac{1}{2}\kappa_{\beta,\beta,\beta}K_\beta^{-1}\epsilon$, $b_{42} = \frac{1}{4}\kappa_{\beta\beta\beta}K_\beta^{-1}\epsilon - \frac{1}{4}(\kappa_{\beta\beta\beta} + 2\kappa_{\beta,\beta,\beta})\epsilon^3$ and $b_{i0} = -(b_{i1} + b_{i2} + b_{i3})$, for $i = 1, 2, 3, 4$, where $\epsilon = \beta - \beta_0$ and K_β is the information of β . Here, $\kappa_{\phi\phi\phi} = \mathbb{E}(\partial^3 \ell(\boldsymbol{\theta}) / \partial \phi^3)$, $\kappa_{\beta,\beta,\beta} = \mathbb{E}\{(\partial \ell(\boldsymbol{\theta}) / \partial \beta)(\partial^2 \ell(\boldsymbol{\theta}) / \partial \beta^2)\}$, and so on.

If we wish to test $\mathcal{H}_0 : \phi = \phi_0$, where ϕ_0 is a specified scalar and β acts as a vector of nuisance parameters, then the expressions for the coefficients b_{ik} ($i = 1, 2, 3, 4$ and $k = 0, 1, 2, 3$) that define the nonnull asymptotic expansions of the distribution functions of the LR, Wald, score and gradient statistics for testing the null hypothesis $\mathcal{H}_0 : \phi = \phi_0$ under a sequence of Pitman alternatives $\mathcal{H}_{1n} : \phi = \phi_0 + \epsilon$, where $\epsilon = \phi - \phi_0$ is assumed to be of order $n^{-1/2}$, are given by $b_{11} = b_{11}^0 + \xi$, $b_{21} = b_{21}^0 + \xi$, $b_{31} = b_{31}^0 + \xi$, $b_{41} = b_{41}^0 + \xi$, $b_{11}^0 = -\frac{1}{6}(\kappa_{\phi\phi\phi} + \kappa_{\phi,\phi,\phi})\epsilon^3$, $b_{21}^0 = \frac{1}{2}\kappa_{\phi\phi\phi}K_\phi^{-1}\epsilon + \frac{1}{2}\kappa_{\phi,\phi,\phi}\epsilon^3$, $b_{31}^0 = -\frac{1}{2}\kappa_{\phi,\phi,\phi}K_\phi^{-1}\epsilon - \frac{1}{6}(\kappa_{\phi\phi\phi} + \kappa_{\phi,\phi,\phi})\epsilon^3$, $b_{41}^0 = -\frac{1}{4}\kappa_{\phi\phi\phi}K_\phi^{-1}\epsilon + \frac{1}{2}\kappa_{\phi,\phi,\phi}\epsilon^3$, $b_{12} = b_{33} = 2b_{43} = \frac{1}{6}\kappa_{\phi,\phi,\phi}\epsilon^3$, $b_{13} = 0$, $b_{22} = -\frac{1}{2}\kappa_{\phi,\phi,\phi}\epsilon^3 - \frac{1}{2}\kappa_{\phi\phi\phi}K_\phi^{-1}\epsilon$, $b_{23} = -\frac{1}{6}\kappa_{\phi\phi\phi}\epsilon^3$, $b_{32} = \frac{1}{2}\kappa_{\phi,\phi,\phi}K_\phi^{-1}\epsilon$, $b_{42} = \frac{1}{4}\kappa_{\phi\phi\phi}K_\phi^{-1}\epsilon - \frac{1}{4}(\kappa_{\phi\phi\phi} + 2\kappa_{\phi,\phi,\phi})\epsilon^3$ and $b_{i0} = -(b_{i1} + b_{i2} + b_{i3})$, for $i = 1, 2, 3, 4$. Here, the additional contribution on the local power of the tests when β needs to be estimated takes the form $\xi = \frac{1}{2} \sum_{r,s=1}^p (\kappa_{rs\phi} + 2\kappa_{r,s\phi})\kappa_{r,s}^{-1}\epsilon$. Also, if $p = 1$, then $\xi = \frac{1}{2} K_\beta^{-1} (\kappa_{\beta\beta\phi} + 2\kappa_{\beta,\beta,\phi})\epsilon$.

As a final remark, we shall point out that our results are also valid when the statistics are modified via a Bartlett or a Bartlett-type correction (see, for example, [Cribari–Neto and Cordeiro, 1996](#)) since the corrections have no effect on the $n^{-1/2}$ term of the local power of the corresponding tests. The advantage of the corrected tests is that they are less size distorted than the tests in the original form in small and moderate-sized samples.

3. Examples

In this section, we discuss some examples to illustrate our results. We focus on ξ , which determines whether the local power changes if a parameter that is globally orthogonal to the remaining parameters is introduced in the model. It is evident that several other special cases could be considered.

Normal distribution. Let y_1, \dots, y_n be independent and identically distributed (i.i.d.) random variables with probability density function (p.d.f.) in the form $\pi(y; \mu, \sigma^2) = (2\pi\sigma^2)^{-1/2} \exp\{-(y - \mu)^2 / (2\sigma^2)\}$, where $-\infty < y < \infty$, $\mu \in \mathbb{R}$ and $\sigma^2 > 0$. First, let $\mathcal{H}_0 : \mu = \mu_0$ be the null hypothesis of interest, where μ_0 is a specified scalar. In this case, $\xi = 0$. Here, $b_{ik} = 0$ ($i = 1, 2, 3, 4$

and $k = 0, 1, 2, 3$), which implies that $\Pi_1 = \Pi_2 = \Pi_3 = \Pi_4$ up to order $n^{-1/2}$, as expected. Now, consider the test of the null hypothesis $\mathcal{H}_0 : \sigma^2 = \sigma_0^2$ against $\mathcal{H}_1 : \sigma^2 \neq \sigma_0^2$, where σ_0^2 is a specified positive scalar. We have $\xi = -\epsilon/(2\sigma_0^2)$, where $\epsilon = \sigma^2 - \sigma_0^2 = O(n^{-1/2})$. Notice that the additional contribution on the local power of the LR, Wald, score and gradient tests for testing $\mathcal{H}_0 : \sigma^2 = \sigma_0^2$ by considering μ unknown does not depend on this parameter.

Birnbaum–Saunders distribution. The Birnbaum–Saunders (BS) distribution introduced by [Birnbaum and Saunders \(1969a,b\)](#) is also known as the fatigue life distribution. It describes the total time until the damage caused by the development and growth of a dominant crack reaches a threshold level and causes a failure. The BS distribution has been used quite effectively to model times to failure for materials subject to fatigue and for modeling lifetime data. Let y_1, \dots, y_n be i.i.d. random variables with a BS distribution with p.d.f. given by $\pi(y; \alpha, \eta) = \kappa(\alpha, \eta)y^{-3/2}(y + \eta) \exp\{-\tau(y/\eta)/(2\alpha^2)\}$, where $y > 0$, $\alpha > 0$ (shape parameter), $\eta > 0$ (scale parameter), $\kappa(\alpha, \eta) = \exp(\alpha^{-2})/(2\alpha\sqrt{2\pi\eta})$ and $\tau(z) = z + z^{-1}$. We first consider the null hypothesis $\mathcal{H}_0 : \alpha = \alpha_0$, where α_0 is a specified positive scalar. It is possible to show that $\xi = -\epsilon(2 + \alpha_0^2)/\{2[\alpha_0(2\pi)^{-1/2}h(\alpha_0) + 1]\}$, where $\epsilon = \alpha - \alpha_0 = O(n^{-1/2})$ and $h(\alpha_0) = \alpha_0(\pi/2)^{1/2} - \pi \exp(2/\alpha_0^2)[1 - \Phi(2/\alpha_0)]$, $\Phi(\cdot)$ being the cumulative distribution function of the standard normal distribution. It is interesting to note that ξ does not involve the parameter η . For testing the null hypothesis $\mathcal{H}_0 : \eta = \eta_0$ against $\mathcal{H}_1 : \eta \neq \eta_0$, where η_0 is a specified positive scalar, ξ reduces to $\xi = 0$ and hence the coefficients that define the nonnull asymptotic expansions of the distribution functions of the LR, Wald, score and gradient statistics do not change when the parameter α needs to be estimated.

von Mises distribution. Let y_1, \dots, y_n be i.i.d. random variables with a von Mises distribution with mean direction μ and concentration parameter ϕ and p.d.f. $\pi(y; \mu, \phi) = \{2\pi I_0(\phi)\}^{-1} \exp\{\phi \cos(y - \mu)\}$, where $0 \leq y < 2\pi$, $0 \leq \mu < 2\pi$, $\phi > 0$ and $I_0(\cdot)$ is the modified Bessel function of the first kind and order 0. The positive parameter ϕ measures the concentration of the distribution: as $\phi \rightarrow 0$ the von Mises distribution converges to the uniform distribution around the circumference, whereas for $\phi \rightarrow \infty$ the distribution tends to the point distribution concentrated in the mean direction. This distribution is particularly useful for the analysis of circular data. For testing $\mathcal{H}_0 : \mu = \mu_0$, where μ_0 is a specified scalar, we have $\xi = 0$ and hence when one introduces unknown concentration the coefficients that define the nonnull asymptotic expansions of the distribution functions of the LR, Wald, score and gradient statistics do not change. The additional contribution on the local powers of the LR, Wald, score and gradient tests (up to order $n^{-1/2}$) for testing $\mathcal{H}_0 : \phi = \phi_0$, where ϕ_0 is a specified positive scalar, when the parameter μ unknown, reduces to $\xi = \epsilon/(2\phi_0)$, where $\epsilon = \phi - \phi_0 = O(n^{-1/2})$.

Generalized linear models. In the following consider the problem of testing hypothesis in the class of generalized linear models ([McCullagh and Nelder](#),

1989). Suppose that the random variables y_1, \dots, y_n are independent and each y_l has a probability density function of the form $\pi(y; \zeta_l, \phi) = \exp[\phi\{y\zeta_l - b(\zeta_l)\} + c(y, \phi)]$, where $b(\cdot)$ and $c(\cdot, \cdot)$ are known appropriate functions. The function $c(\cdot, \cdot)$ admits a factorization of the form $c(y, \phi) = c_1(\phi) + c_2(y)$. The mean and the variance of y_l are $\mathbb{E}(y_l) = \mu_l = db(\zeta_l)/d\zeta_l$ and $\text{var}(y_l) = \phi^{-1}V_l$, where $V_l = d\mu_l/d\zeta_l$ is called the variance function and $\zeta_l = q(\mu_l) = \int V_l^{-1}d\mu_l$ is a known one-to-one function of μ_l . The choice of the variance function V_l as a function of μ_l determines $q(\mu_l)$. We have $V_l = 1 [q(\mu_l) = \mu_l]$, $V_l = \mu_l^2 [q(\mu_l) = -1/\mu_l]$ and $V_l = \mu_l^3 [q(\mu_l) = -1/(2\mu_l^2)]$ for the normal, gamma and inverse Gaussian models, respectively. The parameters ζ_l and $\phi > 0$ are called the canonical and precision parameters, respectively. The systematic part of the model is defined by $d(\mu_l) = \eta_l = \mathbf{x}_l^\top \boldsymbol{\beta}$ ($l = 1, \dots, n$), where $d(\cdot)$ is a known one-to-one differentiable link function, $\mathbf{x}_l^\top = (x_{l1}, \dots, x_{lp})$ is a vector of known variables associated with the l th observable response and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ is a set of unknown parameters to be estimated ($p < n$). Let $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^\top$ be the model matrix with full column rank, i.e. $\text{rank}(\mathbf{X}) = p$. The hypothesis of interest is $\mathcal{H}_0 : \boldsymbol{\beta}_2 = \boldsymbol{\beta}_{20}$, which will be tested against the alternative hypothesis $\mathcal{H}_1 : \boldsymbol{\beta}_2 \neq \boldsymbol{\beta}_{20}$, where $\boldsymbol{\beta}$ is partitioned as $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$, with $\boldsymbol{\beta}_1 = (\beta_1, \dots, \beta_q)^\top$ and $\boldsymbol{\beta}_2 = (\beta_{q+1}, \dots, \beta_p)^\top$. We have $\xi = 0$ and hence the coefficients that define the nonnull asymptotic expansions of the distribution functions of the LR, Wald, score and gradient statistics do not change when the parameter ϕ needs to be estimated. On the other hand, for testing $\mathcal{H}_0 : \phi = \phi_0$, where ϕ_0 is a specified positive scalar, we have $\xi = p\epsilon/(2\phi_0)$ with $\epsilon = \phi - \phi_0 = O(n^{-1/2})$. Notice that the additional contribution on the local powers of the LR, Wald, score and gradient tests (up to order $n^{-1/2}$) for testing \mathcal{H}_0 by considering the parameter vector $\boldsymbol{\beta}$ unknown depends on the rank of the matrix \mathbf{X} . Also, ξ does not involve the unknown $\boldsymbol{\beta}$. For the class of exponential family nonlinear models (Cordeiro and Paula, 1989), which is a natural extension of the generalized linear models for allowing nonlinear systematic component, we arrive exactly at the same conclusions.

Symmetric linear regression models. We say that the random variable y follows a symmetric distribution if its probability density function takes the form $\pi(y; \mu, \phi) = \phi^{-1}g([y - \mu]^2/\phi^2)$, $y \in \mathbb{R}$, where $\mu \in \mathbb{R}$ is a location parameter and $\phi > 0$ is a scale parameter. The function $g : \mathbb{R} \rightarrow [0, \infty)$ is such that $\int_0^\infty g(u)du < \infty$ and $\int_0^\infty u^{-1/2}g(u)du = 1$ to guarantee that $\pi(\cdot; \mu, \phi)$ is a density function. We then write $y \sim S(\mu, \phi^2)$. The function $g(\cdot)$, which is independent of y , μ and ϕ , is typically known as density generator. The probability density function of $z = (y - \mu)/\phi$ is $\pi(v; 0, 1) = g(v^2)$, $v \in \mathbb{R}$, i.e. $z \sim S(0, 1)$. The symmetrical family of distributions allows extensions of the normal distribution for statistical modeling of real data involving distributions with heavier and lighter tails than the ones of the normal distribution. Some special cases are the following: normal, Cauchy, Student- t , generalized Student- t , type I logistic, type II logistic, generalized logistic, Kotz distribution, generalized Kotz distribution, contaminated normal, double exponential, power exponential and

extended power family. The symmetric linear regression model is defined as $y_l = \mu_l + \varepsilon_l$ ($l = 1, \dots, n$), where $\mu_l = \mathbf{x}_l^\top \boldsymbol{\beta}$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ is a vector of unknown regression parameters, $\mathbf{x}_l^\top = (x_{l1}, \dots, x_{lp})$ contains the l th observation on p covariates and $\varepsilon_l \sim S(0, \phi^2)$; see Ferrari and Uribe-Opazo (2001). Let $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^\top$ be the model matrix with full column rank, i.e. $\text{rank}(\mathbf{X}) = p$. We have, when they exist, that $\mathbb{E}(y_l) = \mu_l$ and $\text{var}(y_l) = \eta\phi^2$, where $\eta > 0$ is a constant that may be obtained from the expected value of the radial variable or from the derivative of the characteristic function. Let $\mathcal{H}_0 : \boldsymbol{\beta}_2 = \boldsymbol{\beta}_{20}$ be the null hypothesis of interest, which will be tested against the alternative hypothesis $\mathcal{H}_1 : \boldsymbol{\beta}_2 \neq \boldsymbol{\beta}_{20}$, where $\boldsymbol{\beta}$ is partitioned as $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$, with $\boldsymbol{\beta}_1 = (\beta_1, \dots, \beta_q)^\top$ and $\boldsymbol{\beta}_2 = (\beta_{q+1}, \dots, \beta_p)^\top$. It can be shown that $\xi = 0$ and hence the coefficients that define the nonnull asymptotic expansions of the distribution functions of the LR, Wald, score and gradient statistics do not change for unknown ϕ . For the test of $\mathcal{H}_0 : \phi = \phi_0$, where ϕ_0 is a specified positive scalar, we have $\xi = -p\epsilon(\alpha_{3,1} + 2\alpha_{2,0})/(2\phi_0\alpha_{2,0})$ with $\epsilon = \phi - \phi_0 = O(n^{-1/2})$, where $\alpha_{r,s} = \mathbb{E}(t^{(r)}(z)z^s)$ with $r, s = 0, 1, 2, 3$, where $t(z) = \log g(z^2)$ and $t^{(r)}(z) = d^r t(z)/dz^r$. For example, for the Student- t distribution with ν degrees of freedom we have $\alpha_{2,0} = -(\nu + 1)/(\nu + 3)$ and $\alpha_{3,1} = 6(\nu + 1)\{(\nu + 3)(\nu + 5)\}$. Note that the additional contribution on the local powers of the LR, Wald, score and gradient tests (up to order $n^{-1/2}$) for testing \mathcal{H}_0 by considering the parameter vector $\boldsymbol{\beta}$ unknown depends on p , the rank of the matrix \mathbf{X} . For the Student- t model, for example, ξ depends also on the degrees of freedom ν .

Birnbaum–Saunders regression model. Rieck and Nedelman (1991) proposed a log-linear regression model based on the BS distribution by showing that y (the logarithm of a $\text{BS}(\alpha, \eta)$ variate) has a sinh-normal distribution with shape, location and scale parameters given by α , $\mu = \log(\eta)$ and $\sigma = 2$, respectively, say $y \sim \text{SN}(\alpha, \mu, 2)$. The regression model proposed by the authors is given by $y_l = \mu_l + \varepsilon_l$ ($l = 1, \dots, n$), where y_l is the logarithm of the l th observed lifetime, $\mu_l = \mathbf{x}_l^\top \boldsymbol{\beta}$, $\mathbf{x}_l^\top = (x_{l1}, \dots, x_{lp})$ contains the l th observation on p covariates ($p < n$), $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ is a vector of unknown regression parameters and $\varepsilon_l \sim \text{SN}(\alpha, 0, 2)$. The log-BS regression model is becoming increasingly popular in lifetime analyses and reliability studies. First, consider the null hypothesis $\mathcal{H}_0 : \boldsymbol{\beta}_2 = \boldsymbol{\beta}_{20}$, which will be tested against the alternative hypothesis $\mathcal{H}_1 : \boldsymbol{\beta}_2 \neq \boldsymbol{\beta}_{20}$, where $\boldsymbol{\beta}$ is partitioned as $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$, with $\boldsymbol{\beta}_1 = (\beta_1, \dots, \beta_q)^\top$ and $\boldsymbol{\beta}_2 = (\beta_{q+1}, \dots, \beta_p)^\top$. It can be shown that the coefficients that define the nonnull asymptotic expansions of the distribution functions of the LR, Wald, score and gradient statistics do not change when the parameter α needs to be estimated, since the additional contribution on the local powers of the LR, Wald, score and gradient tests (up to order $n^{-1/2}$) for testing \mathcal{H}_0 by considering the parameter α unknown is $\xi = 0$. Now, let $\mathcal{H}_0 : \alpha = \alpha_0$ be the null hypothesis of interest, where α_0 is a specified positive scalar. In this case, $\xi = -2p\epsilon(2 + \alpha_0^2)/\{\alpha_0^3\psi(\alpha_0)\}$ with $\epsilon = \alpha - \alpha_0 = O(n^{-1/2})$, where $\psi(\alpha_0) = 2 + 4/\alpha_0^2 - (\sqrt{2\pi}/\alpha_0)\{1 - \text{erf}(\sqrt{2}/\alpha_0)\} \exp(2/\alpha_0^2)$ and $\text{erf}(\cdot)$ is the error function (see, for instance, Gradshteyn and Ryzhik, 2007).

Acknowledgments

We gratefully acknowledge the financial support of FAPESP and CNPq (Brazil).

References

- BIRNBAUM, Z.W., SAUNDERS, S.C. (1969a). A new family of life distributions. *Journal of Applied Probability* 6, 319–327. [MR0253493](#)
- BIRNBAUM, Z.W., SAUNDERS, S.C. (1969b). Estimation for a family of life distributions with applications to fatigue. *Journal of Applied Probability* 6, 328–347. [MR0251807](#)
- CORDEIRO, G.M., PAULA, G.A. (1989). Improved likelihood ratio statistics for exponential family nonlinear models. *Biometrika* 76, 93–100. [MR0991426](#)
- COX, D.R., REID, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *Journal of the Royal Statistical Society B* 40, 1–39. [MR0893334](#)
- CRIBARI-NETO, F., CORDEIRO, G.M. (1996). On Bartlett and Bartlett-type corrections. *Econometric Reviews* 15, 339–367. [MR1423902](#)
- FERRARI, S.L.P., URIBE-OPAZO, M.A. (2001). Corrected likelihood ratio tests in a class of symmetric linear regression models. *Brazilian Journal of Probability and Statistics* 15, 49–67. [MR1907860](#)
- GRADSHTEYN, I.S., RYZHIK, I.M. (2007). *Table of Integrals, Series, and Products*. New York: Academic Press. [MR2360010](#)
- HARRIS, P. (1985). An asymptotic expansion for the null distribution of the efficient score statistic. *Biometrika* 72, 653–659. [MR0817580](#)
- HARRIS, P., PEERS, H.W. (1980). The local power of the efficient score test statistic. *Biometrika* 67, 525–529. [MR0601088](#)
- HAYAKAWA, T. (1975). The likelihood ratio criterion for a composite hypothesis under a local alternative. *Biometrika* 62, 451–460. [MR0383630](#)
- HAYAKAWA, T. (1977). The likelihood ratio criterion and the asymptotic expansion of its distribution. *Annals of the Institute of Statistical Mathematics* 29, 359–378. [MR0474588](#)
- HAYAKAWA, T. (1987). Correction to ‘The likelihood ratio criterion and the asymptotic expansion of its distribution’. *Annals of the Institute of Statistical Mathematics* 39, 681. [MR0930540](#)
- HAYAKAWA, T., PURI, M.L. (1985). Asymptotic expansions of the distributions of some test statistics. *Annals of the Institute of Statistical Mathematics* 37, 95–108. [MR0790378](#)
- LAWLEY, D. (1956). A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika* 43, 295–303. [MR0082237](#)
- LEMONTE, A.J., FERRARI, S.L.P. (2012). The local power of the gradient test. *Annals of the Institute of Statistical Mathematics* 64, 373–381.
- MCCULLAGH, P., NELDER, J.A. (1989). *Generalized Linear Models*. London: Chapman and Hall. 2nd. [MR0727836](#)
- RAO, C.R. (2005). Score test: historical review and recent developments. In *Advances in Ranking and Selection, Multiple Comparisons, and Reliability*,

- N. Balakrishnan, N. Kannan and H. N. Nagaraja, eds. Birkhuser, Boston.
[MR2111498](#)
- RIECK, J.R., NEDELMAN, J.R. (1991). A log-linear model for the Birnbaum–Saunders distribution. *Technometrics* 33, 51–60.
- TERRELL, G.R. (2002). The gradient statistic. *Computing Science and Statistics* 34, 206–215.