Local power of the LR, Wald, score and gradient tests in dispersion models

Artur J. Lemonte, Silvia L.P. Ferrari

Departamento de Estatística, Universidade de São Paulo, Brazil

Abstract

We derive asymptotic expansions up to order $n^{-1/2}$ for the nonnull distribution functions of the likelihood ratio, Wald, score and gradient test statistics in the class of dispersion models, under a sequence of Pitman alternatives. The asymptotic distributions of these statistics are obtained for testing a subset of regression parameters and for testing the precision parameter. Based on these nonnull asymptotic expansions it is shown that there is no uniform superiority of one test with respect to the others for testing a subset of regression parameters. Furthermore, in order to compare the finite-sample performance of these tests in this class of models, Monte Carlo simulations are presented. An empirical application to a real data set is considered for illustrative purposes.

Key words: Asymptotic expansions; Chi-square distribution; Dispersion models; Gradient test; Likelihood ratio test; Local power; Score test; Wald test.

1 Introduction

The paper by Nelder and Wedderburn (1972) introduced the class of generalised linear models (GLMs) and showed that a large variety of non-normal data may be analysed by a simple general technique (see, for example, McCullagh and Nelder, 1989; Dobson and Barnett, 2008). The GLMs were originally developed for the exponential family of distributions, but the main ideas were extended to a wider class of models called dispersion models (DMs) in such a way that most of their good properties were preserved. This class of models was introduced by Jørgensen (1987a) and studied in details in Jørgensen (1997a). Some recent references about DMs are Kokonendji et al. (2004), Jørgensen et al. (2010), Simas et al. (2010) and Rocha et al. (2010).

The class of DMs with position parameter θ (which vary in an interval of the real line) and precision parameter $\phi > 0$ has probability density function of the form

$$\pi(y;\theta,\phi) = \exp\{\phi t(y,\theta) + c(y,\phi)\}, \tag{1}$$

where $t(\cdot, \cdot)$ and $c(\cdot, \cdot)$ are known functions. If Y is continuous, π is assumed to be a density with respect to the Lebesgue measure, while if Y is discrete, π is assumed to be a density with respect to the counting measure. The parameter θ may be generally interpreted as a kind of location parameter, not necessarily the mean of the distribution. Several models of the form (1) are discussed by Jørgensen (1987a,b, 1997a), who also examined their statistical properties. It is evident that some special cases arise from (1). Exponential dispersion models (EDMs) represent a special case of DMs with $t(y, \theta) = y\theta - b(\theta)$, where $\mathbb{E}(Y) = \mathrm{d}b(\theta)/\mathrm{d}\theta$; see Jørgensen (1992). An important subclass of DMs of special interest, called proper dispersion models (PDMs), arise when $c(y, \phi)$ is additive, i.e. $c(y, \phi) = a_1(y) + a_2(\phi)$, where $a_1(\cdot)$ and $a_2(\cdot)$ are known functions (see, for instance, Jørgensen, 1997b). The class of PDMs covers important distributions which are not covered by the EDMs, such as the log-gamma distribution, the McCullagh distribution (McCullagh, 1989), the reciprocal inverse Gaussian distribution and the simplex distribution, which is suitable for modeling continuous proportions (Barndorff–Nielsen and Jørgensen, 1991). The von Mises distribution, which also belongs to the class of PDMs and does not belong to the EDMs, is particularly useful for the analysis of circular data; see Mardia and Jupp (2000). The PDMs have two important general properties. First, the distribution of the statistic $T = t(Y, \theta)$ does not depend on θ when ϕ is known, that is, T is a pivotal quantity for θ. Second, (1) is an exponential family with canonical statistic T when θ is known.

Large-sample tests, such as the likelihood ratio, Wald and Rao score tests, are usually employed for testing hypotheses in parametric models. A new criterion for testing hypotheses, referred to as the *gradient test*, was proposed in Terrell (2002). Its statistic is very simple to compute when compared with the other three classic statistics. Here, it is worthwhile to quote Rao (2005): "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." Also, Terrell's statistic shares the same first order asymptotic properties with the likelihood ratio, Wald and score statistics. That is, to the first order of approximation, the likelihood ratio, Wald, score and gradient 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 $n^{-1/2}$. Additionally, it is known that, up to an error of order n^{-1}, the likelihood ratio, 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.

In this paper, our main objective is to derive nonnull asymptotic expansions to order $n^{-1/2}$ of the distribution functions of the likelihood ratio, Wald, score and gradient statistics under a sequence of local alternatives and to compare the local power of the corresponding tests in the class of DMs. In order to compare the finite-sample performance of these tests in this class of models we also perform a Monte Carlo simulation study. As far as we know, there is no mention in the statistical literature on the use of the gradient test in DMs.

The nonnull asymptotic expansions up to order $n^{-1/2}$ for the distribution functions of the likelihood ratio and Wald statistics were derived by Hayakawa (1975), while an analogous result for the score statistic was obtained by Harris and Peers (1980). The asymptotic expansion up to order $n^{-1/2}$ for the distribution functions of the gradient statistic was derived by Lemonte and Ferrari (2010). The expansions are very general, although being difficult or even impossible to particularize their formulas for specific regression models. As we shall see below, we have been capable to apply their results for DMs. In particular, we derive closed-form expressions for the coefficients that define the nonnull asymptotic expansions of these statistics in this class of models and show that there is no uniform superiority of one test with respect to the others for testing a subset of regression parameters.

The rest of the paper is organized as follows. Section 2 briefly describes the likelihood ratio, Wald, score and gradient tests. We present the class of DMs in Section 3. In Section 4 we derive the nonnull asymptotic expansions of the likelihood ratio, Wald, score and gradient statistics for testing a subset of regression parameters in DMs. The local power of the likelihood ratio, Wald, score and gradient tests are compared in Section 5. In Section 6 we consider hypothesis testing on the precision parameter. Monte Carlo simulation results are addressed in Section 7. We consider an empirical application in Section 8 for illustrative purposes. Section 9 closes the paper with some concluding remarks.

2 Background

Let $\ell(\boldsymbol{\theta})$, $\boldsymbol{U}_{\boldsymbol{\theta}}$ and $\boldsymbol{K}_{\boldsymbol{\theta}}$ denote the total log-likelihood function, the score function and the information matrix for the parameter vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)^\top$ of dimension k, respectively. Let $\boldsymbol{K}_{\boldsymbol{\theta}}^{-1}$ denote the inverse of $\boldsymbol{K}_{\boldsymbol{\theta}}$. Consider the partition $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$, where the dimensions of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are q and $k-q$, respectively. Suppose the interest lies in testing the composite null hypothesis $\mathcal{H}_0 : \boldsymbol{\theta}_2 = \boldsymbol{\theta}_{20}$ against $\mathcal{H}_1 : \boldsymbol{\theta}_2 \neq \boldsymbol{\theta}_{20}$, where $\boldsymbol{\theta}_{20}$ is a specified vector. Hence, $\boldsymbol{\theta}_1$ acts as a vector of nuisance parameters. 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

$$S_1 = 2\big\{\ell(\widehat{\boldsymbol{\theta}}) - \ell(\widetilde{\boldsymbol{\theta}})\big\}, \qquad S_2 = (\widehat{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\theta}})^\top \widehat{\boldsymbol{K}_{\boldsymbol{\theta}}}(\widehat{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\theta}}),$$

$$S_3 = \widetilde{\boldsymbol{U}}_{\boldsymbol{\theta}}^\top \widetilde{\boldsymbol{K}_{\boldsymbol{\theta}}}^{-1} \widetilde{\boldsymbol{U}}_{\boldsymbol{\theta}}, \qquad S_4 = \widetilde{\boldsymbol{U}}_{\boldsymbol{\theta}}^\top (\widehat{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\theta}}),$$

where $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\theta}}_1^\top, \widehat{\boldsymbol{\theta}}_2^\top)^\top$ and $\widetilde{\boldsymbol{\theta}} = (\widetilde{\boldsymbol{\theta}}_1^\top, \boldsymbol{\theta}_{20}^\top)^\top$ denote the maximum likelihood estimators of $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$ under $\mathcal{H}_1$ and $\mathcal{H}_0$, respectively, $\widehat{\boldsymbol{K}_{\boldsymbol{\theta}}} = \boldsymbol{K}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}})$, $\widetilde{\boldsymbol{K}_{\boldsymbol{\theta}}} = \boldsymbol{K}_{\boldsymbol{\theta}}(\widetilde{\boldsymbol{\theta}})$ and $\widetilde{\boldsymbol{U}}_{\boldsymbol{\theta}} = \boldsymbol{U}_{\boldsymbol{\theta}}(\widetilde{\boldsymbol{\theta}})$. The limiting distribution of S_1, S_2, S_3 and S_4 is χ^2_{k-q} under $\mathcal{H}_0$ and $\chi^2_{k-q,\lambda}$, i.e. a noncentral chi-square distribution with $k-q$ degrees of freedom and an appropriate noncentrality parameter λ, under $\mathcal{H}_1$. 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^2_{k-q} distribution.

From the partition of $\boldsymbol{\theta}$, we have the corresponding partitions

$$\boldsymbol{U}_{\boldsymbol{\theta}} = (\boldsymbol{U}_{\boldsymbol{\theta}_1}^\top, \boldsymbol{U}_{\boldsymbol{\theta}_2}^\top)^\top, \quad \boldsymbol{K}_{\boldsymbol{\theta}} = \begin{bmatrix} \boldsymbol{K}_{11} & \boldsymbol{K}_{12} \\ \boldsymbol{K}_{21} & \boldsymbol{K}_{22} \end{bmatrix}, \quad \boldsymbol{K}_{\boldsymbol{\theta}}^{-1} = \begin{bmatrix} \boldsymbol{K}^{11} & \boldsymbol{K}^{12} \\ \boldsymbol{K}^{21} & \boldsymbol{K}^{22} \end{bmatrix}.$$

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

$$S_2 = (\widehat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_{20})^\top \widehat{\boldsymbol{K}}^{22^{-1}} (\widehat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_{20}), \quad S_3 = \widetilde{\boldsymbol{U}}_{\boldsymbol{\theta}_2}^\top \widetilde{\boldsymbol{K}}^{22} \widetilde{\boldsymbol{U}}_{\boldsymbol{\theta}_2}, \quad S_4 = \widetilde{\boldsymbol{U}}_{\boldsymbol{\theta}_2}^\top (\widehat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_{20}),$$

where $\widehat{\boldsymbol{K}}^{22} = \boldsymbol{K}^{22}(\widehat{\boldsymbol{\theta}})$, $\widetilde{\boldsymbol{K}}^{22} = \boldsymbol{K}^{22}(\widetilde{\boldsymbol{\theta}})$ and $\widetilde{\boldsymbol{U}}_{\boldsymbol{\theta}_2} = \boldsymbol{U}_{\boldsymbol{\theta}_2}(\widetilde{\boldsymbol{\theta}})$.

Noticed that S_4 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 $\widetilde{\boldsymbol{\theta}}$, then $S_4 \geq 0$.

Recently, Lemonte and Ferrari (2011) obtained the nonnull asymptotic expansions of the likelihood ratio, Wald, score and gradient statistics in Birnbaum–Saunders regression models (Rieck and Nedelman, 1991). An interesting finding is that, up to an error of order n^{-1}, the four tests have the same local power in this class of models. Their simulation study evidenced that the score and the gradient tests perform better than the likelihood ratio and Wald tests in small and moderate-sized samples and hence they concluded that the gradient test is an appealing alternative to the three classic asymptotic tests in Birnbaum–Saunders regressions.

3 Dispersion models

We assume that the random variables $y_1, \ldots, y_n$ are independent and each y_l has a probability density function of the form

$$\pi(y_l; \theta_l, \phi) = \exp\{\phi t(y_l, \theta_l) + c(y_l, \phi)\}, \quad l = 1, \ldots, n. \tag{2}$$

The mean of Y_l will be denoted by μ_l, and is not necessary equal to θ_l, the parameter of interest. In order to introduce a regression structure in the class of models in (2), we assume that

$$d(\theta_l) = \eta_l = f(\boldsymbol{x}_l; \boldsymbol{\beta}), \quad l = 1, \ldots, n, \tag{3}$$

where $d(\cdot)$ is a known one-to-one differentiable link function, $\boldsymbol{x}_l = (x_{l1}, \ldots, x_{lm})^\top$ is an m-vector of nonstocastic variables associated with the l-th response, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ is a set of unknown parameters to be estimated ($m \leq p < n$), and $f(\cdot;\cdot)$ is a possible nonlinear twice continuous differenciable function with respect to $\boldsymbol{\beta}$. The regression structure links the covariates $\boldsymbol{x}_l$ to the parameter of interest θ_l. The $n \times p$ matrix of derivatives of $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)^\top$ with respect to $\boldsymbol{\beta}$, specified by $\boldsymbol{X}^* = \partial\boldsymbol{\eta}/\partial\boldsymbol{\beta}^\top$, is assumed to be of full rank, i.e. rank$(\boldsymbol{X}^*) = p$ for all $\boldsymbol{\beta}$. Further, it is assumed that the precision parameter is unknown and it is the same for all observations. It is also assumed that the usual regularity conditions for maximum likelihood estimation and large sample inference hold; see Cox and Hinkley (1974, Ch. 9).

The class of regression models defined by (2) and (3) extends the class of generalised linear models discussed by McCullagh and Nelder (1989) in two directions. First and as noted before, it includes important distributions which are not exponential family models. Second, it allows for a nonlinear structure in $\boldsymbol{\eta}$. The class of models in (2)-(3) is also a natural extension of the exponential family nonlinear models (EFNLMs) introduced by Cordeiro and Paula (1989), which in turn extends the well-known GLMs by allowing the regression structure to be nonlinear. The EFNLMs are defined by equations (2) and (3), with $t(y_l, \theta_l) = y_l\theta_l - b(\theta_l)$ and $c(y_l, \phi) = a_1(y_l) + a_2(\phi)$ in (2).

Let $\ell = \ell(\boldsymbol{\beta}, \phi) = \sum_{l=1}^{n}\{\phi t(y_l, \theta_l) + c(y_l, \phi)\}$ be the total log-likelihood function for $\boldsymbol{\beta}$ and ϕ, where θ_l is related to $\boldsymbol{\beta}$ by (3). We define $D_{il} = D_{il}(\theta_l, \phi) = \mathbb{E}\{\partial^i t(Y_l, \phi)/\partial\theta_l^i\}$, for $i = 1, 2, 3$ and $l = 1, \ldots, n$. From regularity conditions we have that $D_{1l} = 0$, for $l = 1, \ldots, n$. Table 1 lists D_{2l} and D_{3l} for some dispersion models. The total score function and the total Fisher information matrix for $\boldsymbol{\beta}$ are given, respectively, by $\boldsymbol{U_\beta} = \phi \boldsymbol{X}^{*\top}\dot{\boldsymbol{t}}$ and $\boldsymbol{K_\beta} = \phi\boldsymbol{X}^{*\top}\boldsymbol{W}\boldsymbol{X}^*$, where $\dot{\boldsymbol{t}} = \dot{\boldsymbol{t}}(\boldsymbol{y}, \boldsymbol{\theta}) = (\dot{t}_1, \ldots, \dot{t}_n)^\top$ is an $n \times 1$ vector with $\dot{t}_l = \partial t(y_l, \theta_l)/\partial\theta_l$, $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)^\top$ and $\boldsymbol{W} = \mathrm{diag}\{w_1, \ldots, w_n\}$ with $w_l = -D_{2l}(\mathrm{d}\theta_l/\mathrm{d}\eta_l)^2$. A simple calculation shows that $\mathbb{E}(\partial^2\ell/\partial\boldsymbol{\beta}\partial\phi) = \mathbf{0}$ and then the parameters $\boldsymbol{\beta}$ and ϕ are globally orthogonal (Cox and Reid, 1987). Let $\alpha_i = \sum_{l=1}^{n}\mathbb{E}\{\partial^i c(Y_l, \phi)/\partial\phi^i\} = \sum_{l=1}^{n}\mathbb{E}\{c^{(i)}(Y_l, \phi)\}$, for $i = 1, 2, 3$. The derivatives of the α_i's with respect to ϕ are written with primes, i.e. $\alpha_i' = \mathrm{d}\alpha_i/\mathrm{d}\phi$ and so on. We have that the joint information matrix for $(\boldsymbol{\beta}^\top, \phi)^\top$ is given by $\mathrm{diag}\{\boldsymbol{K_\beta}, -\alpha_2\}$.

Table 1: Expressions of D_{2l} and D_{3l} ($l = 1, \ldots, n$) for some dispersion models.†

Model	D_{2l}	D_{3l}
Normal	-1	0
Inverse Gaussian	$-(-2\theta_l)^{-3/2}$	$-3(-2\theta_l)^{-5/2}$
Reciprocal inverse Gaussian	$-1/\theta_l$	0
Gamma	$-1/\theta_l^2$	$2/\theta_l^3$
Reciprocal gamma	$-1/\theta_l^2$	$2/\theta_l^3$
Log-gamma	-1	1
von Mises	$-I_1(\phi)/I_0(\phi)$	0
generalised hyperbolic secant	$2/(\theta_l^2 + 1)^3$	$(2\theta_l^3 + 10\theta_l)/(\theta_l^2 + 1)^3$

† $I_j(\phi)$ is the modified Bessel function of the first kind and order j.

The maximum likelihood estimate (MLE) $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ can be obtained iteratively using standard reweighted least squares method (Jørgensen, 1983, 1984):

$$\boldsymbol{X}^{*(m)\top}\boldsymbol{W}^{(m)}\boldsymbol{X}^{*(m)}\boldsymbol{\beta}^{(m+1)} = \boldsymbol{X}^{*(m)\top}\boldsymbol{W}^{(m)}\boldsymbol{y}^{*(m)}, \quad m = 0, 1, \ldots,$$

where $\boldsymbol{y}^{*(m)} = \boldsymbol{X}^{*(m)}\boldsymbol{\beta}^{(m)} + \boldsymbol{N}^{(m)}\dot{\boldsymbol{t}}^{(m)}$ is an adjusted dependent variable and $\boldsymbol{N}$ is a diagonal matrix given by $\boldsymbol{N} = -\mathrm{diag}\{D_{21}^{-1}(\mathrm{d}\theta_1/\mathrm{d}\eta_1)^{-1}, \ldots, D_{2n}^{-1}(\mathrm{d}\theta_n/\mathrm{d}\eta_n)^{-1}\}$. The estimate $\widehat{\boldsymbol{\beta}}$ depends directly on the distribution only through the function D_{2l} and does not depend on the parameter ϕ. The maximum likelihood estimate $\widehat{\phi}$ of ϕ is the solution of

$$\sum_{l=1}^{n}\{t(y_l, \widehat{\theta}_l) + c^{(1)}(y_l, \widehat{\phi})\} = 0. \tag{4}$$

The maximum likelihood estimators $\widehat{\boldsymbol{\beta}}$ and $\widehat{\phi}$ are asymptotically independent due to their asymptotic normality and the block diagonal structure of the joint information matrix. If the model is a PDM the α_i's can be expressed as functions of ϕ only, namely $\alpha_i = na_2^{(i)}(\phi)$ for $i = 1, 2, 3$, where $a_2^{(i)}(\phi)$ is the i-th derivative of $a_2(\phi)$ with respect to ϕ. In this case, the $(p+1, p+1)$-th element of the joint information matrix is simply $-na_2^{(2)}(\phi)$ and equation (4) reduces to $a_2^{(1)}(\widehat{\phi}) = -\sum_{l=1}^{n} t(y_l, \widehat{\theta}_l)/n$.

In what follows, we shall consider tests based on the likelihood ratio (S_1), Wald (S_2), Rao score (S_3) and gradient (S_4) statistics in the class of DMs for testing a composite null hypothesis $\mathcal{H}_0 : \boldsymbol{\beta}_2 = \boldsymbol{\beta}_{20}$. This hypothesis 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 $\boldsymbol{\beta}_1 = (\beta_1, \ldots, \beta_q)^\top$ and $\boldsymbol{\beta}_2 = (\beta_{q+1}, \ldots, \beta_p)^\top$. Here, $\boldsymbol{\beta}_{20}$ is a fixed column vector of dimension $p - q$. The partition of the parameter vector $\boldsymbol{\beta}$ induces the corresponding partitions $\boldsymbol{U}_{\boldsymbol{\beta}} = (\boldsymbol{U}_{\boldsymbol{\beta}_1}^\top, \boldsymbol{U}_{\boldsymbol{\beta}_2}^\top)^\top$, with $\boldsymbol{U}_{\boldsymbol{\beta}_1} = \phi \boldsymbol{X}_1^{*\top} \dot{\boldsymbol{t}}$ and $\boldsymbol{U}_{\boldsymbol{\beta}_2} = \phi \boldsymbol{X}_2^{*\top} \dot{\boldsymbol{t}}$,

$$\boldsymbol{K}_{\boldsymbol{\beta}} = \begin{bmatrix} \boldsymbol{K}_{\boldsymbol{\beta}11} & \boldsymbol{K}_{\boldsymbol{\beta}12} \\ \boldsymbol{K}_{\boldsymbol{\beta}21} & \boldsymbol{K}_{\boldsymbol{\beta}22} \end{bmatrix} = \phi \begin{bmatrix} \boldsymbol{X}_1^{*\top} \boldsymbol{W} \boldsymbol{X}_1^* & \boldsymbol{X}_1^{*\top} \boldsymbol{W} \boldsymbol{X}_2^* \\ \boldsymbol{X}_2^{*\top} \boldsymbol{W} \boldsymbol{X}_1^* & \boldsymbol{X}_2^{*\top} \boldsymbol{W} \boldsymbol{X}_2^* \end{bmatrix},$$

with the matrix $\boldsymbol{X}^*$ partitioned as $\boldsymbol{X}^* = \begin{bmatrix} \boldsymbol{X}_1^* & \boldsymbol{X}_2^* \end{bmatrix}$, $\boldsymbol{X}_1^*$ being $n \times q$ and $\boldsymbol{X}_2^*$ being $n \times (p - q)$. Let $(\widehat{\boldsymbol{\beta}}_1, \widehat{\boldsymbol{\beta}}_2, \widehat{\phi})$ and $(\widetilde{\boldsymbol{\beta}}_1, \boldsymbol{\beta}_{20}, \widetilde{\phi})$ be the unrestricted and restricted MLEs of $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \phi)$, respectively. The likelihood ratio, Wald, score and gradient statistics for testing $\mathcal{H}_0$ can be expressed, respectively, as

$$S_1 = 2\big\{\ell(\widehat{\boldsymbol{\beta}}_1, \widehat{\boldsymbol{\beta}}_2, \widehat{\phi}) - \ell(\widetilde{\boldsymbol{\beta}}_1, \boldsymbol{\beta}_{20}, \widetilde{\phi})\big\}, \qquad S_2 = \widehat{\phi}(\widehat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_{20})^\top (\widehat{\boldsymbol{R}}^\top \widehat{\boldsymbol{W}} \widehat{\boldsymbol{R}})(\widehat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_{20}),$$

$$S_3 = \widetilde{\boldsymbol{s}}^\top \widetilde{\boldsymbol{W}}^{1/2} \widetilde{\boldsymbol{X}}_2^* (\widetilde{\boldsymbol{R}}^\top \widetilde{\boldsymbol{W}} \widetilde{\boldsymbol{R}})^{-1} \widetilde{\boldsymbol{X}}_2^{*\top} \widetilde{\boldsymbol{W}}^{1/2} \widetilde{\boldsymbol{s}}, \qquad S_4 = \widetilde{\phi}^{1/2} \widetilde{\boldsymbol{s}}^\top \widetilde{\boldsymbol{W}}^{1/2} \widetilde{\boldsymbol{X}}_2^* (\widehat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_{20}),$$

where $\boldsymbol{s} = (s_1, \ldots, s_n)^\top$ with $s_l = \phi^{1/2} \dot{t}_l (-D_{2l})^{-1/2}$ and $\boldsymbol{R} = \boldsymbol{X}_2^* - \boldsymbol{X}_1^* (\boldsymbol{X}_1^{*\top} \boldsymbol{W} \boldsymbol{X}_1^*)^{-1} \boldsymbol{X}_1^{*\top} \boldsymbol{W} \boldsymbol{X}_2^*$. Here, tildes and hats indicate evaluation at the restricted and unrestricted MLEs, respectively. The limiting distribution of all these statistics under $\mathcal{H}_0$ is χ^2_{p-q}. Note that, unlike the Wald and score statistics, the gradient statistic does not involve any matrix inversion.

4 Nonnull asymptotic distributions in DMs

We present in this section expressions for the nonnull asymptotic expansions up to order $n^{-1/2}$ for the nonnull distribution of the likelihood ratio, Wald, score and gradient statistics for testing a subset of regression parameters in DMs. It should be mentioned that the general nonnull asymptotic expansions derived in Hayakawa (1975), Harris and Peers (1980) and Lemonte and Ferrari (2010) were developed for continuous distributions. It implies that the results derived in this section are only valid for continuous DMs. Here, we shall assume the following local alternative hypothesis $\mathcal{H}_{1n} : \boldsymbol{\beta}_2 = \boldsymbol{\beta}_{20} + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} = (\epsilon_{q+1}, \ldots, \epsilon_p)^\top$ with $\epsilon_r = O(n^{-1/2})$ for $r = q + 1, \ldots, p$.

We introduce the following quantities:

$$\boldsymbol{\epsilon}^* = \begin{bmatrix} \boldsymbol{K}_{\boldsymbol{\beta}11}^{-1} \boldsymbol{K}_{\boldsymbol{\beta}12} \\ -\boldsymbol{I}_{p-q} \end{bmatrix} \boldsymbol{\epsilon}, \quad \boldsymbol{A} = \begin{bmatrix} \boldsymbol{K}_{\boldsymbol{\beta}11}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}, \quad \boldsymbol{M} = \boldsymbol{K}_{\boldsymbol{\beta}}^{-1} - \boldsymbol{A},$$

where $\boldsymbol{I}_{p-q}$ is a $(p-q)\times(p-q)$ identity matrix. Additionally, let $\boldsymbol{Z} = \boldsymbol{X}^*(\boldsymbol{X}^{*\top}\boldsymbol{W}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*\top} = \{z_{lm}\}$, $\boldsymbol{Z}_1 = \boldsymbol{X}_1^*(\boldsymbol{X}_1^{*\top}\boldsymbol{W}\boldsymbol{X}_1^*)^{-1}\boldsymbol{X}_1^{*\top} = \{z_{1lm}\}$,

$$\boldsymbol{X}_l^* = \left\{\frac{\partial^2\eta_l}{\partial\beta_r\partial\beta_s}\right\} = \begin{bmatrix}\boldsymbol{X}_{11l}^* & \boldsymbol{X}_{12l}^*\\ \boldsymbol{X}_{21l}^* & \boldsymbol{X}_{22l}^*\end{bmatrix}, \quad r,s=1,\dots,p, \quad l=1,\dots,n,$$

$\boldsymbol{Z}_d = \mathrm{diag}\{z_{11},\dots,z_{nn}\}$, $\boldsymbol{Z}_{1d} = \mathrm{diag}\{z_{111},\dots,z_{1nn}\}$, $\boldsymbol{F} = \mathrm{diag}\{f_1,\dots,f_n\}$, $\boldsymbol{G} = \mathrm{diag}\{g_1,\dots,g_n\}$, $\boldsymbol{E} = \mathrm{diag}\{e_1,\dots,e_n\}$, $\boldsymbol{t} = (t_1,\dots,t_n)^\top = \boldsymbol{X}^*\boldsymbol{\epsilon}^*$, $\boldsymbol{b} = (b_1,\dots,b_n)^\top = \boldsymbol{X}_2^*\boldsymbol{\epsilon}$, $\boldsymbol{T} = \mathrm{diag}\{t_1,\dots,t_n\}$, $\boldsymbol{T}^{(2)} = \boldsymbol{T}\odot\boldsymbol{T}$, $\boldsymbol{T}^{(3)} = \boldsymbol{T}^{(2)}\odot\boldsymbol{T}$ and $\boldsymbol{B} = \mathrm{diag}\{b_1,\dots,b_n\}$, where "$\odot$" denotes the Hadamard (direct) product of matrices, and

$$f_l = -\frac{\mathrm{d}\theta_l}{\mathrm{d}\eta_l}\frac{\mathrm{d}^2\theta_l}{\mathrm{d}\eta_l^2}D_{2l} - \left(\frac{\mathrm{d}\theta_l}{\mathrm{d}\eta_l}\right)^3 D_{3l}, \quad g_l = -\frac{\mathrm{d}\theta_l}{\mathrm{d}\eta_l}\frac{\mathrm{d}^2\theta_l}{\mathrm{d}\eta_l^2}D_{2l}, \quad e_l = -\left(\frac{\mathrm{d}\theta_l}{\mathrm{d}\eta_l}\right)^3 D_{2l}', \quad l=1,\dots,n,$$

where D_{2l}' denotes the first derivative of D_{2l} with respect to θ_l, for $l=1,\dots,n$.

The nonnull distributions of S_1, S_2, S_3 and S_4 under Pitman alternatives for testing $\mathcal{H}_0:\boldsymbol{\beta}_2=\boldsymbol{\beta}_{20}$ in DMs can be expressed as

$$\Pr(S_i\le x) = G_{p-q,\lambda}(x) + \sum_{k=0}^{3} b_{ik}G_{p-q+2k,\lambda}(x) + O(n^{-1}), \quad i=1,2,3,4,$$

where $G_{m,\lambda}(x)$ is the cumulative distribution function of a non-central chi-square variate with m degrees of freedom and non-centrality parameter λ. Here, $\lambda = \phi\mathrm{tr}\{\boldsymbol{K}_{22.1}\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top\}/2$, where $\boldsymbol{K}_{22.1} = \boldsymbol{K}_{\boldsymbol{\beta}22} - \boldsymbol{K}_{\boldsymbol{\beta}21}\boldsymbol{K}_{\boldsymbol{\beta}11}^{-1}\boldsymbol{K}_{\boldsymbol{\beta}12}$ and $\mathrm{tr}(\cdot)$ denotes the trace operator. The coefficients b_{ik}'s ($i=1,2,3,4$ and $k=0,1,2,3$) can be written in matrix notation, after extensive algebra, as

$$\begin{aligned}b_{11} &= \frac{\phi}{2}\mathrm{tr}\{(\boldsymbol{E}+2\boldsymbol{G})\boldsymbol{B}\boldsymbol{T}^{(2)} + (2\boldsymbol{E}-\boldsymbol{F}+2\boldsymbol{G})\boldsymbol{T}^{(3)} + \boldsymbol{W}\boldsymbol{T}(\boldsymbol{C}+2\boldsymbol{P})\}\\ &\quad+\frac{1}{2}\mathrm{tr}\{(2\boldsymbol{E}-\boldsymbol{F}+2\boldsymbol{G})\boldsymbol{Z}_{1d}\boldsymbol{T} + \boldsymbol{W}\boldsymbol{J}\boldsymbol{T}\},\end{aligned}$$

$$b_{12} = -\frac{\phi}{6}\mathrm{tr}\{(3\boldsymbol{E}-2\boldsymbol{F}+2\boldsymbol{G})\boldsymbol{T}^{(3)}\}, \quad b_{13}=0,$$

$$\begin{aligned}b_{21} &= \frac{\phi}{2}\mathrm{tr}\{(\boldsymbol{E}+2\boldsymbol{G})\boldsymbol{B}\boldsymbol{T}^{(2)} + (2\boldsymbol{E}-\boldsymbol{F}+2\boldsymbol{G})\boldsymbol{T}^{(3)} + \boldsymbol{W}\boldsymbol{T}(\boldsymbol{C}+2\boldsymbol{P})\}\\ &\quad+\frac{1}{2}\mathrm{tr}\{(2\boldsymbol{E}-\boldsymbol{F}+2\boldsymbol{G})\boldsymbol{Z}_d\boldsymbol{T} + 2(\boldsymbol{F}-\boldsymbol{E})(\boldsymbol{Z}_d-\boldsymbol{Z}_{1d})\boldsymbol{T} + \boldsymbol{W}(\boldsymbol{U}\boldsymbol{T}+2\boldsymbol{H})\},\end{aligned}$$

$$b_{22} = \frac{\phi}{2}\mathrm{tr}\{(\boldsymbol{F}-\boldsymbol{E})\boldsymbol{T}^{(3)} + \boldsymbol{W}\boldsymbol{T}\boldsymbol{C}\} - \frac{1}{2}\mathrm{tr}\{(\boldsymbol{F}+2\boldsymbol{G})(\boldsymbol{Z}_d-\boldsymbol{Z}_{1d})\boldsymbol{T} + \boldsymbol{W}\boldsymbol{T}(\boldsymbol{U}-\boldsymbol{J}) + 2\boldsymbol{W}\boldsymbol{H}\},$$

$$b_{23} = -\frac{\phi}{6}\mathrm{tr}\{(\boldsymbol{F}+2\boldsymbol{G})\boldsymbol{T}^{(3)} + 3\boldsymbol{W}\boldsymbol{T}\boldsymbol{C}\},$$

$$b_{31} = \frac{\phi}{2}\text{tr}\{(\boldsymbol{E}+2\boldsymbol{G})\boldsymbol{B}\boldsymbol{T}^{(2)} + (2\boldsymbol{E}-\boldsymbol{F}+2\boldsymbol{G})\boldsymbol{T}^{(3)} + \boldsymbol{W}\boldsymbol{T}(\boldsymbol{C}+2\boldsymbol{P})\}$$
$$+ \frac{1}{2}\text{tr}\{(2\boldsymbol{E}-\boldsymbol{F}+2\boldsymbol{G})\boldsymbol{Z}_{1d}\boldsymbol{T} + (3\boldsymbol{E}-2\boldsymbol{F}+2\boldsymbol{G})(\boldsymbol{Z}_d-\boldsymbol{Z}_{1d})\boldsymbol{T} + \boldsymbol{W}\boldsymbol{T}\boldsymbol{J}\},$$

$$b_{32} = -\frac{1}{2}\text{tr}\{(3\boldsymbol{E}-2\boldsymbol{F}+2\boldsymbol{G})(\boldsymbol{Z}_d-\boldsymbol{Z}_{1d})\boldsymbol{T}\}, \quad b_{33} = -\frac{\phi}{6}\text{tr}\{(3\boldsymbol{E}-2\boldsymbol{F}+2\boldsymbol{G})\boldsymbol{T}^{(3)}\},$$

$$b_{41} = \frac{\phi}{2}\text{tr}\{(\boldsymbol{E}+2\boldsymbol{G})\boldsymbol{B}\boldsymbol{T}^{(2)} + (2\boldsymbol{E}-\boldsymbol{F}+2\boldsymbol{G})\boldsymbol{T}^{(3)} + \boldsymbol{W}\boldsymbol{T}(\boldsymbol{C}+2\boldsymbol{P})\}$$
$$+ \frac{1}{4}\text{tr}\{(6\boldsymbol{G}-\boldsymbol{F}+4\boldsymbol{E})\boldsymbol{Z}_{1d}\boldsymbol{T} - (\boldsymbol{F}+2\boldsymbol{G})\boldsymbol{Z}_d\boldsymbol{T} + \boldsymbol{W}\boldsymbol{T}(3\boldsymbol{J}-\boldsymbol{U}) - 2\boldsymbol{W}\boldsymbol{H}\},$$

$$b_{42} = -\frac{\phi}{4}\text{tr}\{(2\boldsymbol{E}-\boldsymbol{F}+2\boldsymbol{G})\boldsymbol{T}^{(3)} + \boldsymbol{W}\boldsymbol{T}\boldsymbol{C}\}$$
$$+ \frac{1}{4}\text{tr}\{(\boldsymbol{F}+2\boldsymbol{G})(\boldsymbol{Z}_d-\boldsymbol{Z}_{1d})\boldsymbol{T} + \boldsymbol{W}\boldsymbol{T}(\boldsymbol{U}-\boldsymbol{J}) + 2\boldsymbol{W}\boldsymbol{H}\},$$

$$b_{43} = \frac{\phi}{12}\text{tr}\{(\boldsymbol{F}+2\boldsymbol{G})\boldsymbol{T}^{(3)} + 3\boldsymbol{W}\boldsymbol{T}\boldsymbol{C}\},$$

where $\boldsymbol{U} = \text{diag}\{u_1,\ldots,u_n\}$ with $u_l = \text{tr}\{\boldsymbol{X}_l^*(\boldsymbol{X}^{*\top}\boldsymbol{W}\boldsymbol{X}^*)^{-1}\}$, $\boldsymbol{J} = \text{diag}\{j_1,\ldots,j_n\}$ with $j_l = \text{tr}\{\boldsymbol{X}_{11l}^*(\boldsymbol{X}_1^{*\top}\boldsymbol{W}\boldsymbol{X}_1^*)^{-1}\}$, $\boldsymbol{C} = \text{diag}\{c_1,\ldots,c_n\}$ with $c_l = \text{tr}\{\boldsymbol{X}_l^*\boldsymbol{\epsilon}^*\boldsymbol{\epsilon}^{*\top}\}$, $\boldsymbol{P} = \text{diag}\{p_1,\ldots,p_n\}$ with $p_l = \text{tr}\{\boldsymbol{X}_l^*\boldsymbol{\epsilon}^*\boldsymbol{\delta}^\top\}$, $\boldsymbol{H} = \text{diag}\{h_1,\ldots,h_n\}$ with $h_l = \phi\text{tr}\{\boldsymbol{M}\boldsymbol{X}_l^*\boldsymbol{\epsilon}^*\boldsymbol{x}_l^{*\top}\}$, $\boldsymbol{\delta}^\top = (\boldsymbol{0}^\top, \boldsymbol{\epsilon}^\top)$ and $\boldsymbol{x}_l^{*\top}$ is the lth line of $\boldsymbol{X}^*$. The coefficients b_{i0} are obtained from $b_{i0} = -(b_{i1}+b_{i2}+b_{i3})$, for $i = 1,2,3,4$. The b_{ik}'s are of order $n^{-1/2}$ and all quantities except $\boldsymbol{\epsilon}$ are evaluated under the null hypothesis $\mathcal{H}_0$. The detailed derivation of these expressions is long and extremely tedious but may be obtained from the authors upon request.

It is interesting to note that the b_{ik}'s are functions of the local derivative matrix and of the (possibly unknown) precision parameter. These coefficients depend on the second derivative of the (possibly nonlinear) function $f(\boldsymbol{x}_l;\boldsymbol{\beta})$ and involve the link function and its first and second derivatives. Unfortunately, they are very difficult to interpret. The matrices $\boldsymbol{C}$, $\boldsymbol{H}$, $\boldsymbol{J}$, $\boldsymbol{P}$ and $\boldsymbol{U}$ may be considered the nonlinear contribution of the dispersion model since they vanish if the regression model is linear. Obviously, these coefficients depend heavily on the particular dispersion model under consideration. In particular, these coefficients do not change for the class of PDMs, since the only difference between PDMs and DMs is the form of the function $c(\cdot,\cdot)$, which can be decomposed as $c(y,\phi) = a_1(y) + a_2(\phi)$ for PDMs. By replacing $\boldsymbol{E}$ by $\boldsymbol{F}-\boldsymbol{G}$ in these coefficients, we obtain the nonnull asymptotic distributions of the four statistics in the class of EFNLMs (see Lemonte, 2011).

Some simplifications in the coefficients b_{ik} ($i = 1,2,3,4$ and $k = 0,1,2,3$) can be achieved by examining special cases. For example, consider the null hypothesis $\mathcal{H}_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ (i.e. $q = 0$) and an identity link function ($d(\theta_l) = \theta_l$), which implies that $f_l = -D_{3l}$, $g_l = 0$ and $e_l = -D'_{2l}$ ($l = 1,\ldots,n$). Therefore, the b_{ik}'s can be written as

$$b_{11} = \frac{\phi}{2}\text{tr}\{\boldsymbol{E}\boldsymbol{B}\boldsymbol{T}^{(2)} + (2\boldsymbol{E}-\boldsymbol{F})\boldsymbol{T}^{(3)} + \boldsymbol{W}\boldsymbol{T}(\boldsymbol{C}+2\boldsymbol{P})\} + \frac{1}{2}\text{tr}\{\boldsymbol{W}\boldsymbol{J}\boldsymbol{T}\},$$

$$b_{12} = b_{33} = -\frac{\phi}{6}\text{tr}\{(3\boldsymbol{E} - 2\boldsymbol{F})\boldsymbol{T}^{(3)}\}, \quad b_{13} = 0, \quad b_{32} = -\frac{1}{2}\text{tr}\{(3\boldsymbol{E} - 2\boldsymbol{F})\boldsymbol{Z}_d\boldsymbol{T}\},$$

$$b_{21} = \frac{\phi}{2}\text{tr}\{\boldsymbol{E}\boldsymbol{B}\boldsymbol{T}^{(2)} + (2\boldsymbol{E} - \boldsymbol{F})\boldsymbol{T}^{(3)} + \boldsymbol{W}\boldsymbol{T}(\boldsymbol{C} + 2\boldsymbol{P})\} + \frac{1}{2}\text{tr}\{\boldsymbol{F}\boldsymbol{Z}_d\boldsymbol{T} + \boldsymbol{W}(\boldsymbol{U}\boldsymbol{T} + 2\boldsymbol{H})\},$$

$$b_{22} = \frac{\phi}{2}\text{tr}\{(\boldsymbol{F} - \boldsymbol{E})\boldsymbol{T}^{(3)} + \boldsymbol{W}\boldsymbol{T}\boldsymbol{C}\} - \frac{1}{2}\text{tr}\{\boldsymbol{F}\boldsymbol{Z}_d\boldsymbol{T} + \boldsymbol{W}\boldsymbol{T}(\boldsymbol{U} - \boldsymbol{J}) + 2\boldsymbol{W}\boldsymbol{H}\},$$

$$b_{23} = -2b_{43} = -\frac{\phi}{6}\text{tr}\{\boldsymbol{F}\boldsymbol{T}^{(3)} + 3\boldsymbol{W}\boldsymbol{T}\boldsymbol{C}\},$$

$$b_{31} = \frac{\phi}{2}\text{tr}\{\boldsymbol{E}\boldsymbol{B}\boldsymbol{T}^{(2)} + (2\boldsymbol{E} - \boldsymbol{F})\boldsymbol{T}^{(3)} + \boldsymbol{W}\boldsymbol{T}(\boldsymbol{C} + 2\boldsymbol{P})\} + \frac{1}{2}\text{tr}\{(3\boldsymbol{E} - 2\boldsymbol{F})\boldsymbol{Z}_d\boldsymbol{T} + \boldsymbol{W}\boldsymbol{T}\boldsymbol{J}\},$$

$$b_{41} = \frac{\phi}{2}\text{tr}\{\boldsymbol{E}\boldsymbol{B}\boldsymbol{T}^{(2)} + (2\boldsymbol{E} - \boldsymbol{F})\boldsymbol{T}^{(3)} + \boldsymbol{W}\boldsymbol{T}(\boldsymbol{C} + 2\boldsymbol{P})\} + \frac{1}{4}\text{tr}\{-\boldsymbol{F}\boldsymbol{Z}_d\boldsymbol{T} + \boldsymbol{W}\boldsymbol{T}(3\boldsymbol{J} - \boldsymbol{U}) - 2\boldsymbol{W}\boldsymbol{H}\},$$

$$b_{42} = -\frac{\phi}{4}\text{tr}\{(2\boldsymbol{E} - \boldsymbol{F})\boldsymbol{T}^{(3)} + \boldsymbol{W}\boldsymbol{T}\boldsymbol{C}\} + \frac{1}{4}\text{tr}\{\boldsymbol{F}\boldsymbol{Z}_d\boldsymbol{T} + \boldsymbol{W}\boldsymbol{T}(\boldsymbol{U} - \boldsymbol{J}) + 2\boldsymbol{W}\boldsymbol{H}\},$$

and $b_{i0} = -(b_{i1} + b_{i2} + b_{i3})$, for $i = 1, 2, 3, 4$. For the log-gamma model, the above coefficients reduce to

$$b_{11} = \frac{\phi}{2}\text{tr}\{-\boldsymbol{F}\boldsymbol{T}^{(3)} + \boldsymbol{W}\boldsymbol{T}(\boldsymbol{C} + 2\boldsymbol{P})\} + \frac{1}{2}\text{tr}\{\boldsymbol{W}\boldsymbol{J}\boldsymbol{T}\}, \quad b_{12} = b_{33} = \frac{\phi}{3}\text{tr}\{\boldsymbol{F}\boldsymbol{T}^{(3)}\}, \quad b_{13} = 0,$$

$$b_{21} = \frac{\phi}{2}\text{tr}\{-\boldsymbol{F}\boldsymbol{T}^{(3)} + \boldsymbol{W}\boldsymbol{T}(\boldsymbol{C} + 2\boldsymbol{P})\} + \frac{1}{2}\text{tr}\{\boldsymbol{F}\boldsymbol{Z}_d\boldsymbol{T} + \boldsymbol{W}(\boldsymbol{U}\boldsymbol{T} + 2\boldsymbol{H})\},$$

$$b_{22} = \frac{\phi}{2}\text{tr}\{\boldsymbol{F}\boldsymbol{T}^{(3)} + \boldsymbol{W}\boldsymbol{T}\boldsymbol{C}\} - \frac{1}{2}\text{tr}\{\boldsymbol{F}\boldsymbol{Z}_d\boldsymbol{T} + \boldsymbol{W}\boldsymbol{T}(\boldsymbol{U} - \boldsymbol{J}) + 2\boldsymbol{W}\boldsymbol{H}\},$$

$$b_{23} = -2b_{43} = -\frac{\phi}{6}\text{tr}\{\boldsymbol{F}\boldsymbol{T}^{(3)} + 3\boldsymbol{W}\boldsymbol{T}\boldsymbol{C}\}, \quad b_{32} = \text{tr}\{\boldsymbol{F}\boldsymbol{Z}_d\boldsymbol{T}\},$$

$$b_{31} = \frac{\phi}{2}\text{tr}\{-\boldsymbol{F}\boldsymbol{T}^{(3)} + \boldsymbol{W}\boldsymbol{T}(\boldsymbol{C} + 2\boldsymbol{P})\} + \frac{1}{2}\text{tr}\{-2\boldsymbol{F}\boldsymbol{Z}_d\boldsymbol{T} + \boldsymbol{W}\boldsymbol{T}\boldsymbol{J}\},$$

$$b_{41} = \frac{\phi}{2}\text{tr}\{-\boldsymbol{F}\boldsymbol{T}^{(3)} + \boldsymbol{W}\boldsymbol{T}(\boldsymbol{C} + 2\boldsymbol{P})\} + \frac{1}{4}\text{tr}\{-\boldsymbol{F}\boldsymbol{Z}_d\boldsymbol{T} + \boldsymbol{W}\boldsymbol{T}(3\boldsymbol{J} - \boldsymbol{U}) - 2\boldsymbol{W}\boldsymbol{H}\},$$

$$b_{42} = -\frac{\phi}{4}\text{tr}\{-\boldsymbol{F}\boldsymbol{T}^{(3)} + \boldsymbol{W}\boldsymbol{T}\boldsymbol{C}\} + \frac{1}{4}\text{tr}\{\boldsymbol{F}\boldsymbol{Z}_d\boldsymbol{T} + \boldsymbol{W}\boldsymbol{T}(\boldsymbol{U} - \boldsymbol{J}) + 2\boldsymbol{W}\boldsymbol{H}\},$$

Also, for the von Mises model we have

$$b_{11} = b_{31} = \frac{\phi}{2}\text{tr}\{\boldsymbol{W}\boldsymbol{T}(\boldsymbol{C} + 2\boldsymbol{P})\} + \frac{1}{2}\text{tr}\{\boldsymbol{W}\boldsymbol{J}\boldsymbol{T}\}, \quad b_{12} = b_{13} = b_{32} = b_{33} = 0,$$

$$b_{21} = \frac{\phi}{2}\text{tr}\{\boldsymbol{W}\boldsymbol{T}(\boldsymbol{C} + 2\boldsymbol{P})\} + \frac{1}{2}\text{tr}\{\boldsymbol{W}(\boldsymbol{U}\boldsymbol{T} + 2\boldsymbol{H})\}, \quad b_{23} = -2b_{43} = -\frac{\phi}{2}\text{tr}\{\boldsymbol{W}\boldsymbol{T}\boldsymbol{C}\},$$

$$b_{22} = -2b_{42} = \frac{\phi}{2}\text{tr}\{\boldsymbol{W}\boldsymbol{T}\boldsymbol{C}\} - \frac{1}{2}\text{tr}\{\boldsymbol{W}\boldsymbol{T}(\boldsymbol{U} - \boldsymbol{J}) + 2\boldsymbol{W}\boldsymbol{H}\},$$

$$b_{41} = \frac{\phi}{2}\text{tr}\{\boldsymbol{W}\boldsymbol{T}(\boldsymbol{C} + 2\boldsymbol{P})\} + \frac{1}{4}\text{tr}\{\boldsymbol{W}\boldsymbol{T}(3\boldsymbol{J} - \boldsymbol{U}) - 2\boldsymbol{W}\boldsymbol{H}\},$$

Note that for the von Mises linear regression model, the b_{ij}'s above vanish and hence we can write

$$\Pr(S_i \leq x) = G_{p,\lambda}(x) + O(n^{-1}), \qquad i = 1, 2, 3, 4.$$

This is a very interesting result, which implies that the likelihood ratio, score, Wald and gradient tests for testing the null hypothesis $\mathcal{H}_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ have exactly the same local power up to an error of order n^{-1} when we consider an identity link function. It should be noticed that this result also happens for testing the composite null hypothesis $\mathcal{H}_0 : \boldsymbol{\beta}_2 = \boldsymbol{\beta}_{20}$, i.e $\Pr(S_i \leq x) = G_{p-q,\lambda}(x) + O(n^{-1})$, for $i = 1, 2, 3, 4$.

Now, we present the coefficients that define the nonnull asymptotic distributions of the likelihood ratio, Wald, score and gradient statistics for testing the composite null hypothesis $\mathcal{H}_0 : \boldsymbol{\beta}_2 = \boldsymbol{\beta}_{20}$ in GLMs. We have $t(y_l, \theta_l) = y_l\theta_l - b(\theta_l)$ and $\mu_l = \mathbb{E}(Y_l) = \text{d}b(\theta_l)/\text{d}\theta_l$. The class of GLMs is characterized by its variance function $V_l = \text{d}\mu_l/\text{d}\theta_l$, which plays a key role in the study of its mathematical properties and estimation. The variance of Y_l can be written as $\text{var}(Y_l) = \phi^{-1}V_l$. For the GLMs, we have $D_{2l} = -V_l^{-1}$ and $D_{3l} = 2V_l^{-1}(\text{d}V_l/\text{d}\mu_l)$ and hence we can rewrite

$$f_l = \frac{1}{V_l}\frac{\text{d}\mu_l}{\text{d}\eta_l}\frac{\text{d}^2\mu_l}{\text{d}\eta_l^2}, \quad g_l = \frac{1}{V_l}\frac{\text{d}\mu_l}{\text{d}\eta_l}\frac{\text{d}^2\mu_l}{\text{d}\eta_l^2} - \frac{1}{V_l^2}\frac{\text{d}V_l}{\text{d}\mu_l}\left(\frac{\text{d}\mu_l}{\text{d}\eta_l}\right)^3, \quad l = 1, \ldots, n,$$

and redefine the matrices $\boldsymbol{F}$ and $\boldsymbol{G}$ given before. Additionally, the link function is $d(\mu_l) = \eta_l = \boldsymbol{x}_l^\top\boldsymbol{\beta}$ with $m = p$. Also, $\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta}$ with $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^\top$, i.e. here $\boldsymbol{X}^* = \boldsymbol{X}$. Hence, in this class of models we have

$$b_{11} = \frac{\phi}{2}\text{tr}\{(\boldsymbol{F} + \boldsymbol{G})\boldsymbol{B}\boldsymbol{T}^{(2)} + \boldsymbol{F}\boldsymbol{T}^{(3)}\} + \frac{1}{2}\text{tr}\{\boldsymbol{F}\boldsymbol{Z}_{1d}\boldsymbol{T}\}, \quad b_{12} = b_{33} = -\frac{\phi}{6}\text{tr}\{(\boldsymbol{F} - \boldsymbol{G})\boldsymbol{T}^{(3)}\},$$

$$b_{21} = \frac{\phi}{2}\text{tr}\{(\boldsymbol{F} + \boldsymbol{G})\boldsymbol{B}\boldsymbol{T}^{(2)} + \boldsymbol{F}\boldsymbol{T}^{(3)}\} + \frac{1}{2}\text{tr}\{\boldsymbol{F}\boldsymbol{Z}_d\boldsymbol{T} + 2\boldsymbol{G}(\boldsymbol{Z}_d - \boldsymbol{Z}_{1d})\boldsymbol{T}\},$$

$$b_{22} = \frac{\phi}{2}\text{tr}\{\boldsymbol{G}\boldsymbol{T}^{(3)}\} - \frac{1}{2}\text{tr}\{(\boldsymbol{F} + 2\boldsymbol{G})(\boldsymbol{Z}_d - \boldsymbol{Z}_{1d})\boldsymbol{T}\}, \quad b_{13} = 0,$$

$$b_{23} = -2b_{43} = -\frac{\phi}{6}\text{tr}\{(\boldsymbol{F} + 2\boldsymbol{G})\boldsymbol{T}^{(3)}\}, \quad b_{32} = -\frac{1}{2}\text{tr}\{(\boldsymbol{F} - \boldsymbol{G})(\boldsymbol{Z}_d - \boldsymbol{Z}_{1d})\boldsymbol{T}\},$$

$$b_{31} = \frac{\phi}{2}\text{tr}\{(\boldsymbol{F} + \boldsymbol{G})\boldsymbol{B}\boldsymbol{T}^{(2)} + \boldsymbol{F}\boldsymbol{T}^{(3)}\} + \frac{1}{2}\text{tr}\{\boldsymbol{F}\boldsymbol{Z}_{1d}\boldsymbol{T} + (\boldsymbol{F} - \boldsymbol{G})(\boldsymbol{Z}_d - \boldsymbol{Z}_{1d})\boldsymbol{T}\},$$

$$b_{41} = \frac{\phi}{2}\text{tr}\{(\boldsymbol{F} + \boldsymbol{G})\boldsymbol{B}\boldsymbol{T}^{(2)} + \boldsymbol{F}\boldsymbol{T}^{(3)}\} + \frac{1}{4}\text{tr}\{(3\boldsymbol{F} + 2\boldsymbol{G})\boldsymbol{Z}_{1d}\boldsymbol{T} - (\boldsymbol{F} + 2\boldsymbol{G})\boldsymbol{Z}_d\boldsymbol{T}\},$$

$$b_{42} = -\frac{\phi}{4}\text{tr}\{\boldsymbol{F}\boldsymbol{T}^{(3)}\} + \frac{1}{4}\text{tr}\{(\boldsymbol{F} + 2\boldsymbol{G})(\boldsymbol{Z}_d - \boldsymbol{Z}_{1d})\boldsymbol{T}\},$$

By considering the identity link function, these coefficients reduce to

$$b_{11} = \frac{\phi}{2}\text{tr}\{\boldsymbol{G}\boldsymbol{B}\boldsymbol{T}^{(2)}\}, \quad b_{12} = b_{33} = \frac{\phi}{6}\text{tr}\{\boldsymbol{G}\boldsymbol{T}^{(3)}\}, \quad b_{32} = b_{42} = \frac{1}{2}\text{tr}\{\boldsymbol{G}(\boldsymbol{Z}_d - \boldsymbol{Z}_{1d})\boldsymbol{T}\},$$

$$b_{13} = 0, \quad b_{23} = -2b_{43} = -2b_{12}, \quad b_{21} = b_{11} + 2b_{32}, \quad b_{22} = 3b_{12} - 2b_{32}, \quad b_{31} = b_{41} = b_{11} - b_{32}.$$

As expected, the above coefficients vanish for the normal model since the nonnull distributions of all the four criteria agree with the $\chi^2_{p-q,\lambda}$ distribution.

5 Power comparisons

It is known that, to the first order of approximation, the likelihood ratio, Wald, score and gradient statistics have the same asymptotic distributional properties either under the null hypothesis or under a sequence of local alternatives. On the other hand, up to an error of order n^{-1} the corresponding criteria have the same size properties but their local powers differ in the $n^{-1/2}$ term. A meaningful comparison among the criteria can then be performed by comparing the nonnull asymptotic expansions to order $n^{-1/2}$, i.e. ignoring terms or order less than $n^{-1/2}$.

In what follows, we shall compare the local powers of the rival tests based on the general nonnull asymptotic expansions derived in Section 4 for testing the null hypothesis $\mathcal{H}_0 : \boldsymbol{\beta}_2 = \boldsymbol{\beta}_{20}$ in the class of DMs. Let Π_i be the power function, up to order $n^{-1/2}$, of the test that uses the statistic S_i, for $i = 1, 2, 3, 4$. We have

$$\Pi_i - \Pi_j = \sum_{k=0}^{3}(b_{jk} - b_{ik})G_{p-q+2k,\lambda}(x), \tag{5}$$

for $i \neq j$. It is well known that

$$G_{m,\lambda}(x) - G_{m+2,\lambda}(x) = 2g_{m+2,\lambda}(x), \tag{6}$$

where $g_{\nu,\lambda}(x)$ is the probability density function of a non-central chi-square random variable with ν degrees of freedom and non-centrality parameter λ. From (5) and (6) we have after some algebra

$$\begin{aligned}
&\Pi_1 - \Pi_4 = k_1 g_{p-q+4,\lambda}(x) + k_2 g_{p-q+6,\lambda}(x), \quad \Pi_2 - \Pi_4 = k_3 g_{p-q+4,\lambda}(x) + k_4 g_{p-q+6,\lambda}(x),\\
&\Pi_3 - \Pi_4 = k_5 g_{p-q+4,\lambda}(x) + k_6 g_{p-q+6,\lambda}(x), \quad \Pi_1 - \Pi_2 = k_7 g_{p-q+4,\lambda}(x) + k_8 g_{p-q+6,\lambda}(x),\\
&\Pi_1 - \Pi_3 = k_9 g_{p-q+4,\lambda}(x) + k_{10} g_{p-q+6,\lambda}(x), \quad \Pi_2 - \Pi_3 = k_{11} g_{p-q+4,\lambda}(x) + k_{12} g_{p-q+6,\lambda}(x),
\end{aligned} \tag{7}$$

where

$$k_1 = -\frac{1}{2}\text{tr}\{(\boldsymbol{F} + 2\boldsymbol{G})(\boldsymbol{Z}_d - \boldsymbol{Z}_{1d})\boldsymbol{T}\} + \frac{1}{2}\text{tr}\{\boldsymbol{W}\boldsymbol{T}(\boldsymbol{J} - \boldsymbol{U}) - 2\boldsymbol{W}\boldsymbol{H}\},$$

$$k_2 = -\frac{\phi}{6}\text{tr}\{(\boldsymbol{F} + 2\boldsymbol{G})\boldsymbol{T}^{(3)}\} - \frac{\phi}{2}\text{tr}\{\boldsymbol{W}\boldsymbol{T}\boldsymbol{C}\}, \quad k_3 = 3k_1, \quad k_4 = 3k_2,$$

$$k_5 = k_1 - \text{tr}\{(3\boldsymbol{E} - 2\boldsymbol{F} + 2\boldsymbol{G})(\boldsymbol{Z}_d - \boldsymbol{Z}_{1d})\boldsymbol{T}\},$$

$$k_6 = -\frac{\phi}{2}\text{tr}\{(2\boldsymbol{E} - \boldsymbol{F} + 2\boldsymbol{G})\boldsymbol{T}^{(3)}\} - \frac{\phi}{2}\text{tr}\{\boldsymbol{W}\boldsymbol{T}\boldsymbol{C}\},$$

$$k_7 = -2k_1, \quad k_8 = -2k_2, \quad k_9 = k_1 - k_5, \quad k_{10} = \frac{\phi}{3}\mathrm{tr}\{(3\boldsymbol{E} - 2\boldsymbol{F} + 2\boldsymbol{G})\boldsymbol{T}^{(3)}\},$$
$$k_{11} = -3\mathrm{tr}\{(\boldsymbol{F} - \boldsymbol{E})(\boldsymbol{Z}_d - \boldsymbol{Z}_{1d})\boldsymbol{T}\} - \mathrm{tr}\{\boldsymbol{W}\boldsymbol{T}(\boldsymbol{U} - \boldsymbol{J}) + 2\boldsymbol{W}\boldsymbol{H}\},$$
$$k_{12} = -\phi\mathrm{tr}\{(\boldsymbol{F} - \boldsymbol{E})\boldsymbol{T}^{(3)}\} - \phi\mathrm{tr}\{\boldsymbol{W}\boldsymbol{T}\boldsymbol{C}\}.$$

For proper dispersion models, the above expressions are the same. Replacing $\boldsymbol{E}$ by $\boldsymbol{F} - \boldsymbol{G}$ we obtain these quantities for exponential family nonlinear models. From equations (7) we have $\Pi_1 > \Pi_3$ if $k_9 \geq 0$ and $k_{10} \geq 0$ with $k_9 + k_{10} > 0$, and if $k_9 \leq 0$ and $k_{10} \leq 0$ with $k_9 + k_{10} < 0$, we have $\Pi_1 < \Pi_3$. Also, $\Pi_1 = \Pi_3$ if $k_9 = k_{10} = 0$, i.e. $\boldsymbol{F} = \boldsymbol{G}$ and $\boldsymbol{E} = \boldsymbol{0}$, which occurs only for von Mises and normal models with any link function. Additionally, equations (7) show that with the exception of the likelihood ratio and score tests, is not possible to have any other equality among the power functions in the class of DMs for testing the null hypothesis $\mathcal{H}_0 : \boldsymbol{\beta}_2 = \boldsymbol{\beta}_{20}$. The reason is that $\boldsymbol{C}$, $\boldsymbol{H}$, $\boldsymbol{J}$ and $\boldsymbol{U}$, which may be considered as the nonlinear contribution of the dispersion model, vanish only for linear regression models. It implies that only strict inequality holds for any other power comparison among the power functions of the tests that are based on the statistics S_1, S_2, S_3 and S_4. For example, from (7) we have $\Pi_1 > \Pi_4$ ($\Pi_1 < \Pi_4$) if $k_1 \geq 0$ and $k_2 \geq 0$ with $k_1 + k_2 > 0$ (if $k_1 \leq 0$ and $k_2 \leq 0$ with $k_1 + k_2 < 0$), and so on.

We now move to the class of GLMs, in which $\boldsymbol{C} = \boldsymbol{H} = \boldsymbol{J} = \boldsymbol{P} = \boldsymbol{U} = \boldsymbol{0}$. By using the coefficients derived for this class of models in Section 4, the quantities that define equation (7) reduce to

$$k_1 = -\frac{1}{2}\mathrm{tr}\{(\boldsymbol{F} + 2\boldsymbol{G})(\boldsymbol{Z}_d - \boldsymbol{Z}_{1d})\boldsymbol{T}\}, \quad k_2 = -\frac{\phi}{6}\mathrm{tr}\{(\boldsymbol{F} + 2\boldsymbol{G})\boldsymbol{T}^{(3)}\}, \quad k_3 = 3k_1,$$
$$k_5 = k_1 - \mathrm{tr}\{(\boldsymbol{F} - \boldsymbol{G})(\boldsymbol{Z}_d - \boldsymbol{Z}_{1d})\boldsymbol{T}\}, \quad k_6 = -\frac{\phi}{2}\mathrm{tr}\{\boldsymbol{F}\boldsymbol{T}^{(3)}\}, \quad k_4 = 3k_2,$$
$$k_7 = -2k_1, \quad k_8 = -2k_2, \quad k_9 = k_1 - k_5, \quad k_{10} = \frac{\phi}{3}\mathrm{tr}\{(\boldsymbol{F} - \boldsymbol{G})\boldsymbol{T}^{(3)}\},$$
$$k_{11} = -3\mathrm{tr}\{\boldsymbol{G}(\boldsymbol{Z}_d - \boldsymbol{Z}_{1d})\boldsymbol{T}\}, \quad k_{12} = -\phi\mathrm{tr}\{\boldsymbol{G}\boldsymbol{T}^{(3)}\}.$$

For GLMs with canonical link ($\boldsymbol{G} = \boldsymbol{0}$), we have $k_{11} = k_{12} = 0$ and hence $\Pi_2 = \Pi_3$. It is possible to show that $\Pi_1 = \Pi_2 = \Pi_4$ if $\boldsymbol{F} = -2\boldsymbol{G}$, that is

$$\frac{\mathrm{d}^2\mu_l}{\mathrm{d}\eta_l^2} = \frac{2}{3V_l}\left(\frac{\mathrm{d}\mu_l}{\mathrm{d}\eta_l}\right)^2, \quad l = 1, \ldots, n.$$

The GLMs for which this equality holds have the link function defined by $\eta_l = \int V_l^{-3/2}\mathrm{d}\mu_l$ ($l = 1, \ldots, n$). For the gamma model this function is $\eta_l = \mu_l^{-1/3}$ ($l = 1, \ldots, n$). Additionally, we have that $\Pi_3 = \Pi_4$ for any GLM with identity link function, i.e. $\boldsymbol{F} = \boldsymbol{0}$. Also, $\Pi_1 = \Pi_3$ if $k_9 = k_{10} = 0$, i.e. $\boldsymbol{F} = \boldsymbol{G}$, which occurs only for normal models with any link. Finally, the equality $\Pi_1 = \Pi_2 = \Pi_3 = \Pi_4$ holds only for normal models with identity link function.

We can conclude that there is no uniform superiority of one test with respect to the others for testing the null hypothesis $\mathcal{H}_0 : \boldsymbol{\beta}_2 = \boldsymbol{\beta}_{20}$ in the class of DMs. Hence, if the sample size is large, all tests could be recommended, since their type I error probabilities do not significantly deviate from the true nominal level and their local powers are approximately equal. The natural question is how these tests perform when the sample size is small or of moderate size, and which one is the most reliable. In Section 7, we shall use Monte Carlo simulations to shed some light on this issue.

6 Tests for the precision parameter

In this section we derive asymptotic expansions for the nonnull distribution of the four statistics for testing the precision parameter ϕ in DMs. We are interested in testing the null hypothesis $\mathcal{H}_0 : \phi = \phi_0$ against a two-sided alternative hypothesis $\mathcal{H}_1 : \phi \neq \phi_0$, where ϕ_0 is a positive specified value for ϕ. Here, $\boldsymbol{\beta}$ acts as a nuisance parameter. The likelihood ratio, Wald, score and gradient statistics are expressed as follows:

$$S_1 = \sum_{l=1}^{n}\{(\widehat{\phi}-\phi_0)t(y_l,\widehat{\theta}_l) + c(y_l,\widehat{\phi}) - c(y_l,\phi_0)\}, \quad S_2 = (\widehat{\phi}-\phi_0)^2\{-\alpha_2(\widehat{\phi})\},$$

$$S_3 = \{-\alpha_2(\phi_0)\}^{-1}\left[\sum_{l=1}^{n}\{t(y_l,\widehat{\theta}_l) + c^{(1)}(y_l,\phi_0)\}\right]^2, \quad S_4 = (\widehat{\phi}-\phi_0)\sum_{l=1}^{n}\{t(y_l,\widehat{\theta}_l) + c^{(1)}(y_l,\phi_0)\}.$$

For PDMs, these statistics can be expressed as

$$S_1 = 2n\{a_2(\widehat{\phi}) - a_2(\phi_0) - (\widehat{\phi}-\phi_0)a_2^{(1)}(\widehat{\phi})\}, \quad S_2 = -n(\widehat{\phi}-\phi_0)^2 a_2^{(2)}(\widehat{\phi}),$$

$$S_3 = -\frac{n\{a_2^{(1)}(\widehat{\phi}) - a_2^{(1)}(\phi_0)\}^2}{a_2^{(2)}(\phi_0)}, \quad S_4 = n\{a_2^{(1)}(\phi_0) - a_2^{(1)}(\widehat{\phi})\}(\widehat{\phi}-\phi_0).$$

For example, for the von Mises model $a_2(\phi) = -\log\{I_0(\phi)\}$. Also, $a_2^{(1)}(\phi) = -r(\phi)$ and $a_2^{(2)}(\phi) = r(\phi)^2 + r(\phi)/\phi - 1$, where $r(\phi) = I_1(\phi)/I_0(\phi)$. Thus, we can write

$$S_1 = 2n[\log\{I_0(\phi_0)/I_0(\widehat{\phi})\} + (\widehat{\phi}-\phi_0)r(\widehat{\phi})], \quad S_2 = -n(\widehat{\phi}-\phi_0)^2\{r(\widehat{\phi})^2 + r(\widehat{\phi})/\widehat{\phi} - 1\},$$

$$S_3 = -\frac{n\{r(\phi_0) - r(\widehat{\phi})\}^2}{r(\phi_0)^2 + r(\phi_0)/\widehat{\phi}_0 - 1}, \quad S_4 = n\{r(\widehat{\phi}) - r(\phi_0)\}(\widehat{\phi}-\phi_0).$$

Also, for normal and inverse Gaussian models we have $a_2(\phi) = \log(\phi)/2$. Hence

$$S_1 = 2n\left\{\log\left(\frac{\widehat{\phi}}{\phi_0}\right) - \left(\frac{\widehat{\phi}-\phi_0}{\widehat{\phi}}\right)\right\}, \quad S_2 = S_3 = \frac{n}{2}\left\{\frac{\widehat{\phi}-\phi_0}{\widehat{\phi}}\right\}^2, \quad S_4 = \frac{n}{2}\left\{\frac{\widehat{\phi}-\phi_0}{\phi_0} - \frac{\widehat{\phi}-\phi_0}{\widehat{\phi}}\right\}.$$

We have $a_2(\phi) = \phi\log(\phi) - \log\{\Gamma(\phi)\}$ for the gamma model and therefore these statistics reduce to

$$S_1 = 2n\left\{\phi_0\log\left(\frac{\widehat{\phi}}{\phi_0}\right) - \log\left(\frac{\Gamma(\widehat{\phi})}{\Gamma(\phi_0)}\right) - (\widehat{\phi}-\phi_0)(1-\psi(\widehat{\phi}))\right\},$$

$$S_2 = n\{\widehat{\phi}\psi'(\widehat{\phi}) - 1\}\frac{(\widehat{\phi}-\phi_0)^2}{\widehat{\phi}}, \quad S_3 = \frac{n\phi_0\{\log(\widehat{\phi}/\phi_0) - (\psi(\widehat{\phi}) - \psi(\phi_0))\}}{\phi_0\psi'(\phi_0) - 1}$$

and

$$S_4 = n(\widehat{\phi} - \phi_0)\left\{\log\left(\frac{\widehat{\phi}}{\phi_0}\right) + \psi(\widehat{\phi}) - \psi(\phi_0)\right\},$$

where $\Gamma(\cdot)$, $\psi(\cdot)$ and $\psi'(\cdot)$ are the gamma, digamma and trigamma functions, respectively.

The nonnull asymptotic distributions of S_1, S_2, S_3 and S_4 for testing $\mathcal{H}_0 : \phi = \phi_0$ in DMs under the local alternative $\mathcal{H}_{1n} : \phi = \phi_0 + \epsilon$, where $\epsilon = \phi - \phi_0$ is assumed to be $O(n^{-1/2})$, is

$$\Pr(S_i \leq x) = G_{1,\lambda}(x) + \sum_{k=0}^{3} b_{ik} G_{1+2k,\lambda}(x) + O(n^{-1}), \qquad i = 1, 2, 3, 4.$$

The noncentrality parameter is given by $\lambda = -\alpha_2\epsilon^2$ and the the coefficients b_{ik}'s can be written as

$$b_{11} = \frac{(\alpha_2' - \alpha_3)\epsilon^3}{2} + \frac{p\epsilon}{2\phi}, \quad b_{12} = \frac{(2\alpha_3 - 3\alpha_2')\epsilon^3}{6}, \quad b_{13} = 0,$$

$$b_{21} = \frac{(\alpha_2' - \alpha_3)\epsilon^3}{2} - \frac{\alpha_3\epsilon}{2\alpha_2} + \frac{p\epsilon}{2\phi}, \quad b_{22} = -\frac{(\alpha_2' - \alpha_3)\epsilon^3}{2} + \frac{\alpha_3\epsilon}{2\alpha_2}, \quad b_{23} = -\frac{\alpha_3\epsilon^3}{6},$$

$$b_{31} = \frac{(\alpha_2' - \alpha_3)\epsilon^3}{2} + \frac{(2\alpha_3 - 3\alpha_2')\epsilon}{2\alpha_2} + \frac{p\epsilon}{2\phi}, \quad b_{32} = -\frac{(2\alpha_3 - 3\alpha_2')\epsilon}{2\alpha_2}, \quad b_{33} = \frac{(2\alpha_3 - 3\alpha_2')\epsilon^3}{6},$$

$$b_{41} = \frac{(\alpha_2' - \alpha_3)\epsilon^3}{2} + \frac{\alpha_3\epsilon}{4\alpha_2} + \frac{p\epsilon}{2\phi}, \quad b_{42} = -\frac{(2\alpha_2' - \alpha_3)\epsilon^3}{4} - \frac{\alpha_3\epsilon}{4\alpha_2}, \quad b_{43} = \frac{\alpha_3\epsilon^3}{12},$$

with $b_{i0} = -(b_{i1} + b_{i2} + b_{i3})$, for $i = 1, 2, 3, 4$. It should be noticed that the above expressions depend on the parameter ϕ and depend on the local derivative matrix $\boldsymbol{X}^*$ only through its rank p. Since $\alpha_2' = \alpha_3 = na_2^{(3)}(\phi)$ for PDMs, these coefficients reduce to

$$b_{11} = \frac{p\epsilon}{2\phi}, \quad b_{12} = b_{23} = b_{33} = -\frac{na_2^{(3)}(\phi)\epsilon^3}{6}, \quad b_{13} = 0, \quad b_{21} = b_{31} = \frac{p\epsilon}{2\phi} - \frac{a_2^{(3)}(\phi)\epsilon}{2a_2^{(2)}(\phi)},$$

$$b_{22} = b_{32} = b_{11} - b_{21}, \quad b_{41} = b_{11} + \frac{1}{2}(b_{11} - b_{21}), \quad b_{42} = -\frac{1}{2}(b_{11} - b_{21} - 3b_{12}), \quad b_{43} = -\frac{b_{12}}{2},$$

with $b_{i0} = -(b_{i1} + b_{i2} + b_{i3})$, for $i = 1, 2, 3, 4$. These coefficients do not change for the class of GLMs.

In what follows, we present an analytical comparison among the local powers of the four tests for testing the null hypothesis $\mathcal{H}_0 : \phi = \phi_0$. We have

$$\Pi_i - \Pi_j = \sum_{k=0}^{3} (b_{jk} - b_{ik}) G_{1+2k,\lambda}(x).$$

After some algebra, we can write

$$\Pi_1 - \Pi_2 = -\frac{\alpha_3\epsilon}{\alpha_2} g_{5,\lambda}(x) + \frac{\alpha_3\epsilon^3}{3} g_{7,\lambda}(x),$$

$$\Pi_1 - \Pi_3 = \frac{(2\alpha_3 - 3\alpha_2')\epsilon}{\alpha_2} g_{5,\lambda}(x) - \frac{(2\alpha_3 - 3\alpha_2')\epsilon^3}{3} g_{7,\lambda}(x),$$

$$\Pi_1 - \Pi_4 = \frac{\alpha_3\epsilon}{2\alpha_2}g_{5,\lambda}(x) - \frac{\alpha_3\epsilon^3}{6}g_{7,\lambda}(x),$$
$$\Pi_2 - \Pi_3 = \frac{3(\alpha_3 - \alpha_2')\epsilon}{\alpha_2}g_{5,\lambda}(x) - (\alpha_3 - \alpha_2')\epsilon^3 g_{7,\lambda}(x),$$
$$\Pi_2 - \Pi_4 = \frac{3\alpha_3\epsilon}{2\alpha_2}g_{5,\lambda}(x) - \frac{\alpha_3\epsilon^3}{2}g_{7,\lambda}(x),$$
$$\Pi_3 - \Pi_4 = -\frac{3(\alpha_3 - 2\alpha_2')\epsilon}{\alpha_2}g_{5,\lambda}(x) + \frac{(\alpha_3 - 2\alpha_2')}{2}\epsilon^3 g_{7,\lambda}(x).$$

From the above expressions, we can obtain the following general conclusions. By assuming $\phi > \phi_0$ (opposite inequalities hold if $\phi < \phi_0$), we have that $\Pi_3 < \Pi_2 < \Pi_1 < \Pi_4$ if $\alpha_3 > 0$ with $\alpha_2' > \alpha_3$. Also, $\Pi_2 = \Pi_3 < \Pi_1 < \Pi_4$ if $\alpha_2' = \alpha_3 > 0$. For example, for normal and inverse Gaussian models we have $a_2(\phi) = \log(\phi)/2$, which implies that $a_2^{(1)}(\phi) = 1/(2\phi)$, $a_2^{(2)}(\phi) = -1/(2\phi^2)$ and $a_2^{(3)}(\phi) = 1/\phi^3$. Since $\alpha_2' = \alpha_3 = n/\phi^3 > 0$, we arrive at the following inequalities: $\Pi_2 = \Pi_3 < \Pi_1 < \Pi_4$ if $\phi > \phi_0$, and $\Pi_2 = \Pi_3 > \Pi_1 > \Pi_4$ if $\phi < \phi_0$.

7 Monte Carlo simulation

In this section we conduct Monte Carlo simulations in order to compare the performance of the likelihood ratio, Wald, score and gradient tests in small- and moderate-sized samples.

We consider the von Mises regression model, which is quite useful for modeling circular data; see Fisher (1993) and Mardia and Jupp (2000). Here,

$$\pi(y; \theta, \phi) = \frac{\exp\{\phi\cos(y - \theta)\}}{2\pi I_0(\phi)}, \quad y \in (-\pi, \pi),$$

where $\theta \in (-\pi, \pi)$ and $\phi > 0$. This density function is symmetric around $y = \theta$, which is the mode and the circular mean of the distribution. Also, ϕ is a precision parameter in the sense that the larger the value of ϕ the more concentrated the density function around θ. It is evident the density function above is a special case of (1) with $t(y, \theta) = \cos(y - \theta)$ and $c(y, \phi) = -\log(I_0(\phi))$.

We assume that

$$\tan(\theta_l/2) = \eta_l = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip},$$

where $x_{i1} = 1$ and $\theta_l = 2\arctan(\eta_l)$, $l = 1, \ldots, n$. The covariate values were selected as random draws from the $\mathcal{U}(0, 1)$ distribution and for fixed n those values were kept constant throughout the experiment. The number of Monte Carlo replications was 10,000, the nominal levels of the tests were γ = 10%, 5% and 1%, and all simulations were carried out using the `Ox` matrix programming language (Doornik, 2007). `Ox` is freely distributed for academic purposes and available at http://www.doornik.com.

First, the null hypothesis is $\mathcal{H}_0 : \beta_{p-1} = \beta_p = 0$, which is tested against a two-sided alternative. The sample size is $n = 50$, $\phi = 1.5, 2.5, 4$ and $p = 3, 4, \ldots, 8$. The values of the response were generated using $\beta_1 = \cdots = \beta_{p-2} = 1$. The null rejection rates of the four tests are presented in Table 1. It is clear that the likelihood ratio (S_1) and Wald (S_2) tests are markedly liberal, more so as the number of regressors increases. The score (S_3) and gradient (S_4) tests are also liberal in most of the cases, but much less size distorted than the likelihood ratio and Wald tests in all cases. For instance, when $\phi = 2.5$, $p = 4$ and $\gamma = 5\%$, the rejection rates are 7.05% (S_1), 8.28% (S_2), 5.15% (S_3) and 6.30% (S_4). We note that the score test is much less liberal than the likelihood ratio and Wald tests and slightly less liberal than the gradient test. Additionally, the Wald test is much more liberal than the other tests. Note that as ϕ increases the tests become less size distorted, as expected, since the von Mises distribution approaches a normal distribution as ϕ increases.

Table 2: Null rejection rates (%); ϕ = 1.5, 2.5 and 4, with $n = 50$.

	$\phi = 1.5$											
	$\gamma = 10\%$				$\gamma = 5\%$				$\gamma = 1\%$			
p	S_1	S_2	S_3	S_4	S_1	S_2	S_3	S_4	S_1	S_2	S_3	S_4
3	13.31	15.42	10.12	10.42	6.90	9.93	4.65	5.04	1.75	4.13	0.79	1.20
4	14.48	16.31	10.26	12.49	7.75	10.86	4.83	6.83	1.93	4.62	0.59	2.08
5	16.65	19.34	10.92	12.46	9.55	12.36	5.05	6.62	2.67	4.87	0.84	1.83
6	19.04	21.93	11.94	14.81	11.78	15.00	5.90	8.26	3.62	6.50	1.03	2.40
7	22.09	26.39	12.44	15.94	13.71	18.12	6.12	8.87	4.27	7.67	1.27	2.21
8	24.16	26.58	13.03	17.66	15.87	17.42	6.63	9.82	5.23	6.82	1.39	2.76
	$\phi = 2.5$											
	$\gamma = 10\%$				$\gamma = 5\%$				$\gamma = 1\%$			
p	S_1	S_2	S_3	S_4	S_1	S_2	S_3	S_4	S_1	S_2	S_3	S_4
3	12.02	12.96	10.56	10.50	6.21	7.35	5.17	5.29	1.39	2.31	0.78	1.04
4	12.97	13.66	11.05	11.77	7.05	8.28	5.15	6.30	1.73	3.05	0.90	1.52
5	14.28	16.38	10.97	11.68	7.96	10.31	4.94	6.25	2.11	4.28	0.85	1.65
6	14.83	15.33	11.90	13.02	8.36	9.82	5.71	7.27	2.09	3.85	1.01	1.80
7	15.93	18.00	12.60	13.87	9.20	11.30	6.66	7.60	2.72	3.71	1.53	1.87
8	18.12	19.53	13.45	16.12	11.16	12.29	7.02	9.38	3.31	4.79	1.55	2.68
	$\phi = 4$											
	$\gamma = 10\%$				$\gamma = 5\%$				$\gamma = 1\%$			
p	S_1	S_2	S_3	S_4	S_1	S_2	S_3	S_4	S_1	S_2	S_3	S_4
3	11.99	12.59	10.72	10.81	6.32	7.19	5.02	5.25	1.37	2.20	0.82	1.12
4	13.15	14.48	11.49	11.74	7.19	8.66	5.50	5.83	1.67	2.89	0.84	1.13
5	13.59	13.67	11.87	12.26	7.21	7.64	5.72	6.25	1.68	2.50	0.96	1.35
6	14.08	15.60	11.85	12.65	7.57	9.04	5.88	6.30	1.73	2.88	1.00	1.21
7	15.16	16.42	12.79	13.52	8.34	9.55	6.42	7.03	2.28	3.16	1.43	1.71
8	16.14	17.36	13.53	14.57	9.28	10.31	7.13	7.84	2.42	2.96	1.28	1.61

Table 3 reports results for $\phi = 3$, $p = 4$ and sample sizes ranging from 20 to 150. As expected, the null rejection rates of all the tests approach the corresponding nominal levels as the sample size grows. Again, the score and gradient tests present the best performances. In Table 4 we present the first two moments of S_1, S_2, S_3 and S_4 and the corresponding moments of the limiting χ^2 distribution. Note that the gradient and score statistics present a good agreement between the true moments (obtained by simulation) and the moments of the limiting distribution.

Table 3: Null rejection rates (%); $\phi = 3$, $p = 4$ and different sample sizes.

	$\gamma = 10\%$				$\gamma = 5\%$				$\gamma = 1\%$			
n	S_1	S_2	S_3	S_4	S_1	S_2	S_3	S_4	S_1	S_2	S_3	S_4
20	17.33	19.18	13.71	13.89	10.50	11.95	6.92	7.04	3.33	4.38	1.16	1.14
30	15.04	16.33	11.65	12.76	8.29	10.19	5.10	6.66	2.05	4.14	0.75	1.50
40	13.49	15.23	11.44	11.44	7.56	9.43	5.72	5.96	1.81	3.07	0.92	1.18
50	12.51	13.78	10.77	11.05	6.65	7.79	5.40	5.59	1.66	2.31	1.02	1.25
70	12.01	12.46	11.00	11.17	6.20	6.90	5.41	5.58	1.48	2.18	1.12	1.28
100	11.30	12.13	10.74	10.69	5.86	6.65	4.92	5.44	1.22	2.04	0.94	1.07
150	10.51	11.01	10.02	10.10	5.05	6.03	4.59	4.63	1.08	1.66	0.94	0.95

Table 4: Moments; $\phi = 2$, $n = 35$, $p = 4$.

	S_1	S_2	S_3	S_4	χ^2_2
Mean	2.50	2.68	2.16	2.23	2.0
Variance	6.23	8.73	4.14	4.63	4.0

We also performed Monte Carlo simulations considering hypothesis testing on ϕ. To save space, the results are not shown. The score and gradient tests exhibited superior behaviour than the likelihood ratio and Wald tests. For example, when $n = 35$, $p = 3$, $\gamma = 10\%$ and $\mathcal{H}_0 : \phi = 2$, we obtained the following null rejection rates: 13.23% (S_1), 14.75% (S_2), 10.61% (S_3) and 9.97% (S_4). Again, the best performing tests are the score and gradient tests.

Overall, in small to moderate-sized samples the best performing tests are the score and the gradient tests. They are less size distorted than the other two. Hence, these tests may be recommended for testing hypotheses on the regression parameters in the von Mises regression model. The gradient test has a slight advantage over the score test because the gradient statistic is simpler to calculate than the score statistic for testing a subset of regression parameters. In particular, no matrix needs to be inverted; see Section 3.

8 Application

In this section we shall illustrate an application of the likelihood ratio, Wald, score and gradient tests in a real data set. We consider the data described in Fisher and Lee (1992) regarding the distance traveled by 31 small blue periwinkles (*Nodilittorina unifasciata*) after they have moved down-shore from the height at which they normally live. Following Fisher and Lee (1992) we assume a von Mises distribution for the animals' path, but with the assumption of constant dispersion and link function

$$\tan(\theta_l/2) = \beta_1 + \beta_2 x_l, \quad l = 1, \ldots, 31,$$

where $\theta_l = 2\arctan(\beta_1 + \beta_2 x_l)$ denotes the mean direction for a given distance moved x_l (cm). These data have been previously analysed by Paula (1996) and Souza and Paula (2002) with emphasis on local influence and residual analysis, respectively. The angular responses were transformed to the range $(-\pi, \pi)$. The maximum likelihood estimates of the parameters (asymptotic standard errors in parentheses) are: $\widehat{\beta}_1 = -0.323\,(0.151)$, $\widehat{\beta}_2 = -0.013\,(0.004)$ and $\widehat{\phi} = 3.265\,(0.726)$. The values of the likelihood ratio (S_1), Wald (S_2), score (S_3) and gradient (S_4) statistics for testing the null hypothesis $\mathcal{H}_0 : \beta_2 = 0$ are 9.526 (p-value: 0.002), 11.031 (p-value: 0.001), 7.126 (p-value: 0.008) and 8.280 (p-value: 0.004), respectively. At any usual significance level, all tests lead to the same conclusion, i.e. the null hypothesis should be rejected.

Now, we consider different values for β_{20} and we wish to test $\mathcal{H}_0 : \beta_2 = \beta_{20}$ against $\mathcal{H}_1 : \beta_2 \neq \beta_{20}$. Table 5 lists the observed values of the different test statistics and the corresponding p-values for $\beta_{20} = -0.026, -0.024, -0.022, -0.020$ and -0.018. The asterisks indicate that the null hypothesis is rejected at respectively the 1% (***), the 5% (**) or at the 10% (*) significance level. Notice that the same decision is reached by all the tests when $\beta_{20} = -0.018$ but not when $\beta_{20} = -0.026, -0.024, -0.022$ and -0.020. In all cases considered here, the score and gradient tests lead to the same conclusion. Additionally, the likelihood ratio and Wald tests display the smallest p-values in all cases, in accordance with their liberal behaviours observed in our simulation study.

Table 5: Test statistics for $\mathcal{H}_0 : \beta_2 = \beta_{20}$ against $\mathcal{H}_1 : \beta_2 \neq \beta_{20}$ (p-values between parentheses).

	β_{20}				
statistic	-0.026	-0.024	-0.022	-0.020	-0.018
S_1	7.314 (0.007)***	5.606 (0.018)**	4.011 (0.045)**	2.591 (0.107)	1.411 (0.235)
S_2	11.409 (0.001)***	8.193 (0.004)***	5.509 (0.019)**	3.355 (0.067)*	1.733 (0.188)
S_3	5.872 (0.015)**	4.636 (0.031)**	3.407 (0.065)*	2.251 (0.134)	1.249 (0.264)
S_4	5.728 (0.017)**	4.611 (0.032)**	3.458 (0.063)*	2.332 (0.127)	1.321 (0.250)

Notice that the sample size is $n = 31$, but if n were smaller, the tests could lead to different conclusions. To illustrate this, a randomly chosen subset of the data set with $n = 10$ was drawn. The null hypothesis to be tested is $\mathcal{H}_0 : \beta_2 = 0$. The observed value of the test statistics are $S_1 = 2.939$ (p-value: 0.086), $S_2 = 2.980$ (p-value: 0.084), $S_3 = 2.491$ (p-value: 0.114) and $S_4 = 2.682$ (p-value = 0.101). Hence, at the 10% significance level, the score and gradient tests do not reject the null hypothesis unlike the likelihood ratio and Wald tests, which are much more oversized than the score and gradient tests as evidenced by our simulation results.

9 Concluding remarks

The dispersion models (DMs) extend the well-known generalised linear models (Nelder and Wedderburn, 1972) and also the exponential family nonlinear models (Cordeiro and Paula, 1989). Additionally, the class of DMs covers a comprehensive range of non-normal distributions. In this paper, we dealt with the issue of performing hypothesis testing in DMs. We considered the three classic tests, likelihood ratio, Wald and score tests, and a recently proposed test, the gradient test. We have derived formulae for the asymptotic expansions up to order $n^{-1/2}$ of the distribution functions of the likelihood ratio, Wald, score and gradient statistics, under a sequence of Pitman alternatives, for testing a subset of regression parameters and for testing the dispersion parameter. The formulae derived are simple to be used analytically to obtain closed-form expressions for these expansions in special models. Also, the power of all four criteria, which are equivalent to first order, were compared under specific conditions based on second order approximations. Additionally, we present Monte Carlo simulations in order to compare the finite-sample performance of these tests. From the simulation results we can conclude that the score and gradient tests should be preferred. Finally, we present an empirical application for illustrative purposes.

Acknowledgments

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

References

Barndorff–Nielsen, O.E., Jørgensen, B. (1991). Some parametric models on the simplex. *Journal of Multivariate Analysis* 38, 106–116.

Cordeiro, G.M., Paula, G.A. (1989). Improved likelihood ratio statistics for exponential family nonlinear models. *Biometrika* 76, 93–100.

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

Cox, D.R., Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *Journal of the Royal Statistical Society B* 40, 1–39.

Dobson, A.J., Barnett, A.G. (2008). *Introduction to generalised Linear Models*, 3rd ed. Chapman & Hall/CRC.

Doornik, J.A. (2007). *An Object-Oriented Matrix Language*, 5th ed. London: Timberlake Consultants Press.

Fisher, N.I. (1993). *Statistical Analysis of Circular Data*. Cambridge University Press, New York.

Fisher, N.I., Lee, A.J. (1992). Regression models for an angular response. *Biometrics* 48, 665–677

Harris, P., Peers, H.W. (1980). The local power of the efficient score test statistic. *Biometrika* 67, 525–529.

Hayakawa, T. (1975). The likelihood ratio criterion for a composite hypothesis under a local alternative. *Biometrika* 62, 451–460.

Jørgensen, B. (1983). Maximum likelihood estimation ans large-sample inference for generalised linear and nonlinear models. *Biometrika* 70, 19–28.

Jørgensen, B. (1984). The delta algorithm and GLIM. *International Statistical Review* 52, 283–300.

Jørgensen, B. (1987a). Exponential dispersion models (with discussion). *Journal of the Royal Statistical Society B* 49, 127–162.

Jørgensen, B. (1987b). Small dispersion asymptotics. *Brazilian Journal of Probability and Statistics* 1, 59–90.

Jørgensen, B. (1992). Exponential dispersion models and extensions: A review. *International Statistical Review* 60, 5–20.

Jørgensen, B. (1997a). *The Theory of Dispersion Models*. Chapman & Hall, London.

Jørgensen, B. (1997b). Proper dispersion models (with discussion). *Brazilian Journal of Probability and Statistics* 11, 89–140.

Jørgensen, B., Goegebeur, Y., Martínez, J.R. (2010). Dispersion models for extremes. *Extremes* **13**, 399–437.

Kokonendji, C.C, Dossou-Gbété, S., Demétrio, C.G.B. (2004). Some discrete exponential dispersion models: Poisson-Tweedie and Hinde-Demétrio classes. *SORT* 28, 201–214.

Lemonte, A.J. (2011). Local power of some tests in exponential family nonlinear models. *Journal of Statistical Planning and Inference* 141, 1981–1989.

Lemonte, A.J., Ferrari, S.L.P. (2010). The local power of the gradient test. *Annals of the Institute of Statistical Mathematics*. DOI:10.1007/s10463-010-0315-4.

Lemonte, A.J., Ferrari, S.L.P. (2011). Size and power properties of some tests in the Birnbaum–Saunders regression model. *Computational Statistics and Data Analysis* 55, 1109–1117.

McCullagh, P. (1989). Some statistical properties of a family of continuous univariate distributions. *Journal of the American Statistical Association* 84, 125–129.

McCullagh, P., Nelder, J.A. (1989). *generalised Linear Models*. London: Chapman and Hall. 2nd.

Mardia, K.V., Jupp, P.E. (2000). *Directional Statistics*. John Wiley & Sons.

Nelder, J., Wedderburn, R. (1972). generalised linear model. *Journal of the Royal Statistical Society A* 135, 370–384.

Paula, G.A. (1996). Influence diagnostic in proper dispersion models. *The Australian Journal of Statistics* 38, 307–316.

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.

Rieck, J.R., Nedelman, J.R. (1991). A log-linear model for the Birnbaum–Saunders distribution. *Technometrics* 33, 51–60.

Rocha, A.V., Simas, A.B., Cordeiro, G.M. (2010). Second-order asymptotic expressions for the covariance matrix of maximum likelihood estimators in dispersion models. *Statistics and Probability Letters* 80, 718–725.

Simas, A.B., Rocha, A.V., Cordeiro, G.M. (2010). Skewness of maximum likelihood estimators in dispersion models. *Journal of Statistical Planning and Inference* 140, 2111–2121.

Souza, F.A.M., Paula, G.A. (2002). Deviance residuals for an angular response. *Australian and New Zealand Journal of Statistics* 44, 345–356.

Terrell, G.R. (2002). The gradient statistic. *Computing Science and Statistics* 34, 206–215.