RT-MAE 2001-14

THREE CORRECTED SCORE TESTS FOR GENERALIZED LINERAR MODELS WITH DISPERSION COVARIATES

by

Gauss M. Cordeiro, Denise A. Botter, Lúcia P. Barroso
and
Silvia L. P. Ferrari

Palavras-Chave: Bartlett-type correction; Chi-squared distribution; Dispersion parameter; Generalized linear model; Link function; Precision parameter; Score test.
Classificação AMS: 62J02, 62F04.
(AMS Classification)

- Julho de 2001 -

Three Corrected Score Tests for Generalized Linear Models with Dispersion Covariates[1]

Gauss M. Cordeiro

Departamento de Estatística, Universidade Federal da Bahia, Brazil

Denise A. Botter

Lúcia P. Barroso

Silvia L. P. Ferrari

Departamento de Estatística, Universidade de São Paulo, Brazil

Abstract

We develop three corrected score tests for generalized linear models with dispersion covariates, thus generalizing the results of Cordeiro, Ferrari and Paula (1993) and Cribari-Neto and Ferrari (1995). We present, in matrix notation, general formulae for the coefficients which define three different corrected score statistics. These statistics are asymptotically equivalent to order n^{-1}, where n is the sample size. The formulae only require simple operations on matrices and can be used to obtain analytically closed-form corrections for score test statistics in a variety of special generalized linear models with dispersion covariates. They also have advantages for numerical purposes since our formulae are readily computable using a language supporting numerical linear algebra. Several examples, namely, iid sampling without covariates on the mean or dispersion parameter, two-parameter regression models for both linear predictors and one-way classification models, are given. We also present some simulations where the three corrected statistics perform better than the usual score test. Finally, we present a numerical example for a data set discussed by Simonoff and Tsai (1994).

Key Words: Bartlett-type correction; Chi-squared distribution; Dispersion parameter; Generalized linear model; Link function; Precision parameter; Score test.

1. Introduction

Score tests are commonly used in statistics and econometrics to test a wide range of hypotheses in many different regression models. These tests rely upon an asymptotic approximation: the chi-squared distribution is used as a large-sample approximation to the true (unknown) null distribution of S_R, Rao's score statistic (Rao, 1947). However, the chi-squared distribution may be a poor approximation to the null distribution of S_R when the sample size n is not sufficiently large. It is thus important to obtain refinements for inference based on score tests from second-order asymptotic theory. For generalized linear models with known dispersion parameter, these corrections were obtained by Cordeiro, Ferrari and Paula (1993) for testing that a subset of the regression coefficients equals a given vector of constants. When the dispersion parameter is unknown, similar corrections were developed by Cribari-Neto and Ferrari (1995).

This paper derives finite-sample corrections for score tests in generalized linear models with dispersion covariates, thus generalizing both the results of Cordeiro, Ferrari and

[1] *Correspondence to*: Silvia L. P. Ferrari, Departamento de Estatística, Universidade de São Paulo, Caixa Postal 66281, 05315-970, São Paulo, Brazil. email: sferrari@ime.usp.br.

Paula (1993) and Cribari-Neto and Ferrari (1995). Similar corrections for the likelihood ratio statistics were obtained by Botter and Cordeiro (1997). The mean and the dispersion of the dependent variable are nonlinear functions of linear combinations of some explanatory variables. The null hypothesis that specify fixed values for some of the coefficients of those linear combinations is considered. The score test with a Bartlett-type correction is used to test this hypothesis. We develop the specific calculations that are needed to implement this correction in the class of models under study.

The basic idea of transforming the score test statistic in such a way that it becomes better approximated by the reference chi-squared distribution is due to Cordeiro and Ferrari (1991). The corrected score statistic S_R^* proposed by these authors is obtained by multiplying the original score statistic S_R by a second-degree polynomial in the statistic S_R itself, producing a modified score test statistic whose null distribution has its asymptotic chi-squared approximation error reduced from $O(n^{-1})$ to $O(n^{-3/2})$. Thus, improved score tests may be based on S_R^* which are expected to deliver more accurate inferences with samples of typical sizes encountered by applied practitioners.

Recently, Bartlett-type corrections leading to a better approximation of the null distribution of the score statistic by the chi-squared distribution received a considerable attention in the literature. The recent developments have focused on two main problems, namely: (i) establishing various formulae for the Bartlett-type correction in general as well as for special classes of regression models; and (ii) showing that the distribution function of the corrected score statistic is closer to the χ^2 distribution than the distribution function of the usual score statistic.

From the corrected score test introduced in Cordeiro and Ferrari (1991), many results have appeared in the literature for a great variety of regression models. Cordeiro, Ferrari and Paula (1993) obtained improved score tests in generalized linear models and Cribari-Neto and Ferrari (1995) extended their results for the case of unknown dispersion. Cordeiro and Ferrari (1996) derived corrected score test statistics in proper dispersion models. Ferrari and Cordeiro (1996) and also Ferrari, Uribe-Opazo and Cribari-Neto (1997) derived this type of correction for exponential family nonlinear models. Some of the results in these previous papers are extended in Ferrari, Cordeiro and Cribari-Neto (2001). For other important applications of Bartlett-type corrections, see the references mentioned in Cribari-Neto and Cordeiro (1996).

The paper is organized in the following form. In Section 2 we define an extended class of generalized linear models (GLMs), where the dispersion parameter of the response is a function of extra covariates. Our class of models allows the simultaneous modeling of the mean and the dispersion by making use of the GLM framework. We also give the score statistic for testing a composite hypothesis that some components in both mean and dispersion vectors of parameters have specified values. In Section 3, we obtain some joint cumulants of log-likelihood derivatives for the regression model defined in Section 2, and present general formulae for Bartlett-type corrections, in matrix notation, to improve score statistics to simultaneously test that some of the mean effects and some of the dispersion effects have specified values. In Section 4, we consider two special cases: one case where only mean effects are being tested and another case where only dispersion effects are being tested. In Section 5 we deal with the Bartlett-type correction for some special model structures of the covariates. Examples of iid sampling without covariates on the location or dispersion, two-parameter regression models for both linear predictors and one-way classification models are given. Section 6 presents some simulation results to illustrate the merits of the corrections.

The final section of the paper illustrates the computation of the Bartlett-type corrections in a heteroscedastic normal regression model using a dataset (Simonoff and Tsai, 1994) from a study on the monthly excess returns over the riskless rate for a market. Finally, for the sake of readability all proofs are deferred to the Appendix.

2. Model Specification

We consider a class of models with more complex systematic components than those provided by standard generalized linear models (McCullagh and Nelder, 1989) for which the dispersion parameter is constant over the observations. Suppose that the random variables $Y_1, \ldots, Y_n$ are independent and each Y_l has a density (or probability function) in the following family of distributions

$$\pi(y; \theta_l, \phi_l) = \exp[\phi_l\{y\theta_l - b(\theta_l) + c(y)\} + a(y) + d(\phi_l)], \tag{1}$$

where $b(\cdot)$, $c(\cdot)$, $a(\cdot)$ and $d(\cdot)$ are known functions. The mean and the variance of Y_l are $\mathrm{E}(Y_l) = \mu_l = \mathrm{d}b(\theta_l)/\mathrm{d}\theta_l$ and $\mathrm{Var}(Y_l) = \phi_l^{-1}V_l$, where $V_l = \mathrm{d}\mu_l/\mathrm{d}\theta_l$ and $\theta_l = \int V_l^{-1}\mathrm{d}\mu_l = q(\mu_l)$ is a known one-to-one function of μ_l. The parameters θ_l and ϕ_l in (1) are called the canonical and the precision parameters, respectively. The inverse of ϕ_l is the dispersion parameter of the distribution. The choice of the variance function V_l determines the interpretation of ϕ_l. For normal, gamma and inverse Gaussian distributions, the means and variances are θ_l^{-1}, $-\theta_l^{-1}$ and $(-2\theta_l)^{-1/2}$ and ϕ_l^{-1}, $\phi_l^{-1}\mu_l^2$ and $\phi_l^{-1}\mu_l^3$, respectively.

We now assume that both parameters μ_l and ϕ_l vary across observations through regression models which are parameterized as $\mu_l = \mu_l(\beta)$ and $\phi_l = \phi_l(\gamma)$. In classical GLMs the precision parameter is constant although possibly unknown. The usual systematic component for the mean is $d(\mu_l) = \eta_l = x_l^\top\beta$, where $d(\cdot)$ is the mean link function, $\eta = (\eta_1, \ldots, \eta_n)^\top$ is the linear predictor, $x_l^\top = (x_{l1}, \ldots, x_{lp})$ is the l-th row of X, a specified $n \times p$ matrix of full rank $p < n$, and $\beta = (\beta_1, \ldots, \beta_p)^\top$ is a set of unknown parameters to be estimated. Analogously, the precision parameter ϕ_l is assumed to vary in the following systematic way: $g(\phi_l) = \tau_l = s_l^\top\gamma$, where $\phi = (\phi_1, \ldots, \phi_n)^\top$, $\tau = (\tau_1, \ldots, \tau_n)^\top$ is the dispersion linear predictor, $g(\cdot)$ is the dispersion link function, $s_l^\top = (s_{l1}, \ldots, s_{lq})$ is the l-th row of S, a specified $n \times q$ matrix of full rank $q < n$, and $\gamma = (\gamma_1, \ldots, \gamma_q)^\top$ is also a vector of unknown parameters. Both $d(\cdot)$ and $g(\cdot)$ are known one-to-one continuously twice differentiable functions. The linear structure $\tau_l = s_l^\top\gamma$ measures the dispersion for the l-th observation. The dispersion link function $g(\cdot)$ is a positive valued function. A simple choice for $g(\cdot)$ is $g(\phi_l) = \log\phi_l$. The idea of constructing and fitting formal models for the dependence of both μ_l and ϕ_l on several covariates was first put forward by Pregibon (1984). The dispersion covariates in S constitute, in general, although not necessarily, a subset of the covariates in X. The parameters β and γ are assumed to be functionally independent, which leaves us with $p + q$ parameters to be estimated.

We are interested in simultaneously estimating β and γ which represent the effects of the explanatory variables on the mean response and dispersion parameter, respectively. Let $\ell(\beta, \gamma)$ be the total log-likelihood function for a given GLM with dispersion covariates, which is assumed to be regular (Cox and Hinkley, 1974; Chapter 9) with respect to all β and γ derivatives up to the fourth order. From now on, we use the following notation

$d_{il} = \mathrm{d}^i d(\phi_l)/\mathrm{d}\phi_l^i$ and $\phi_{il} = \mathrm{d}^i \phi_l/\mathrm{d}\tau_l^i$ for $i = 1, \ldots, 4$ and $l = 1, \ldots, n$. Denote the efficient score by

$$U = U(\beta, \gamma) = \begin{pmatrix} \partial\ell(\beta,\gamma)/\partial\beta \\ \partial\ell(\beta,\gamma)/\partial\gamma \end{pmatrix}$$

whose components are

$$\frac{\partial\ell(\beta,\gamma)}{\partial\beta} = X^\top \Phi T V^{-1}(y-\mu) \quad \text{and} \quad \frac{\partial\ell(\beta,\gamma)}{\partial\gamma} = S^\top \Phi_1 v,$$

where $y = (y_1, \ldots, y_n)^\top$, $\mu = (\mu_1, \ldots, \mu_n)^\top$, $\Phi = \mathrm{diag}\{\phi_1, \ldots, \phi_n\}$, $\Phi_1 = \mathrm{diag}\{\phi_{11}, \ldots, \phi_{1n}\}$, $V = \mathrm{diag}\{V_1, \ldots, V_n\}$, $T = \mathrm{diag}\{\mathrm{d}\mu_1/\mathrm{d}\eta_1, \ldots, \mathrm{d}\mu_n/\mathrm{d}\eta_n\}$ and $v = (v_1, \ldots, v_n)^\top$ with $v_l = y_l\theta_l - b(\theta_l) + c(y_l) + d_{1l}$. The partition $(\beta^\top, \gamma^\top)^\top$ induces a corresponding partition on the total information matrix for these parameters. The information matrix $K = K(\beta,\gamma) = \mathrm{diag}\{K_{\beta,\beta}, K_{\gamma,\gamma}\}$ is block-diagonal, where $K_{\beta,\beta} = X^\top W \Phi X$ and $K_{\gamma,\gamma} = -S^\top D_2 \Phi_1^2 S$, with $W = \mathrm{diag}\{w_1, \ldots, w_n\}$, $w_l = V_l^{-1}(\mathrm{d}\mu_l/\mathrm{d}\eta_l)^2$ and $D_2 = \mathrm{diag}\{d_{21}, \ldots, d_{2n}\}$, are the information matrices for β and γ, respectively. Thus, the parameters β and γ are globally orthogonal (Cox and Reid, 1987), and their maximum likelihood estimates (MLEs) $\hat\beta$ and $\hat\gamma$ are asymptotically independent. These estimates satisfy equations $U(\hat\beta, \hat\gamma) = 0$, which are in general non-linear but can be solved by Fisher's scoring method or, equivalently, by the iterative reweighted least squares method. In fact, the Fisher's scoring method can be used to estimate β and γ by iteratively solving the following equations:

$$\begin{aligned} X^\top W^{(m)}\Phi^{(m)}X(\beta^{(m+1)} - \beta^{(m)}) &= X^\top W^{(m)}\Phi^{(m)} z_\beta^{(m)}, \\ (-S^\top D_2^{(m)}\Phi_1^{(m)2}S)(\gamma^{(m+1)} - \gamma^{(m)}) &= S^\top(-D_2^{(m)}\Phi_1^{(m)2}) z_\gamma^{(m)}, \end{aligned} \tag{2}$$

where $z_\beta = T^{-1}(y-\mu)$ and $z_\gamma = -(D_2\Phi_1)^{-1}v$ are $n \times 1$ vectors.

Equations (2) show that any software with a weighted regression routine can be used to evaluate the MLEs $\hat\beta$ and $\hat\gamma$. In general terms we have to regress the $2n \times 1$ modified dependent variable $(z_\beta^\top, z_\gamma^\top)^\top$ on the $n \times (p+q)$ model matrix (X, S) with the $2n \times 2n$ modified weight matrix defined by $\mathrm{diag}\{W\Phi, -D_2\Phi_1^2\}$. The process is iterative because both the adjusted dependent variable and the weights depend on the current estimates of β and γ. The iterative procedure (2) is easy to perform in the GLIM algorithm following the same lines described in Cordeiro and Paula (1989) and Cordeiro and Demétrio (1989). We define the model through the own directive and the starting procedure is done by choosing suitable initial values of β and γ. The cycle to obtain the increments of the current estimates of β and γ is repeated until convergence.

We consider regular problems in which as n, the dimension of y, increases, the null distribution of the score statistic tends to a chi-squared distribution. For model (1) we are first interested in testing a subset of the β and γ parameters. Partitioning the parameters as $\beta = (\beta_1^\top, \beta_2^\top)^\top$ and $\gamma = (\gamma_1^\top, \gamma_2^\top)^\top$, where $\beta_1 = (\beta_1, \ldots, \beta_{p_1})^\top, \beta_2 = (\beta_{p_1+1}, \ldots, \beta_p)^\top, \gamma_1 = (\gamma_1, \ldots, \gamma_{q_1})^\top$ and $\gamma_2 = (\gamma_{q_1+1}, \ldots, \gamma_q)^\top$, we are interested in testing the null hypothesis $H_0^1 : \beta_1 = \beta_1^{(0)}, \gamma_1 = \gamma_1^{(0)}$ against H_1^1: violation of at least one equality, where $\beta_1^{(0)}$ and $\gamma_1^{(0)}$ are specified vectors of dimensions p_1 and q_1, respectively. We assume that $0 \le p_1 \le p$ and $0 \le q_1 \le q$ but the trivial case $p_1 = q_1 = 0$ is excluded because there are no parameters left under the null hypothesis. Evidently, the simplest case $p_1 = p$ and $q_1 = q$ has little interest in practical applications since the null hypothesis becomes simple. The cases $p_1 = 0$ and $q_1 = 0$

are treated separately in Section 4. Following the partition induced by H_0^1, let $X = (X_1, X_2)$ and $S = (S_1, S_2)$ be the corresponding partitioned model matrices, where X_1, X_2, S_1 and S_2 are, respectively, $n \times p_1, n \times (p - p_1), n \times q_1$ and $n \times (q - q_1)$ matrices of full ranks. We denote the unrestricted MLEs of β and γ by $\hat{\beta}$ and $\hat{\gamma}$, while the restricted MLEs of the nuisance parameters β_2 and γ_2 under H_0^1 are denoted by $\tilde{\beta}_2$ and $\tilde{\gamma}_2$. Similarly, all quantities evaluated at the unrestricted MLEs will be denoted with the addition of a circumflex, while those evaluated at the restricted MLEs will be denoted by adding a tilde.

The score statistic for the hypothesis $H_0^1 : \beta_1 = \beta_1^{(0)}, \gamma_1 = \gamma_1^{(0)}$ is $S_R = \widetilde{U}^\top \widetilde{K}^{-1} \widetilde{U}$ which can be written as a sum of two quadratic forms

$$S_R = \tilde{z}_\beta^\top \widetilde{W} \widetilde{\Phi} \widetilde{Z}_\beta \widetilde{\Phi} \widetilde{W} \tilde{z}_\beta + \tilde{z}_\gamma^\top \widetilde{D}_2 \widetilde{\Phi}_1^2 \widetilde{Z}_\gamma \widetilde{\Phi}_1^2 \widetilde{D}_2 \tilde{z}_\gamma, \tag{3}$$

where the $n \times n$ matrices $Z_\beta = X(X^\top W \Phi X)^{-1} X^\top$ and $Z_\gamma = S(-S^\top D_2 \Phi_1 \; S)^{-1} S^\top$ are interpreted as the asymptotic covariance matrices of $\widehat{\eta} = X\widehat{\beta}$ and $\widehat{\tau} = S\widehat{\gamma}$, respectively. Equation (3) also holds for testing a null hypothesis of mean effects $H_0^2 : \beta_1 = \beta_1^{(0)}$ against the alternative hypothesis $H_1^2 : \beta_1 \neq \beta_1^{(0)}$ and for testing a null hypothesis of dispersion effects $H_0^3 : \gamma_1 = \gamma_1^{(0)}$ against $H_1^3 : \gamma_1 \neq \gamma_1^{(0)}$.

3. The main results

Finite-sample corrections for score tests can be found in full generality in Cordeiro and Ferrari (1991). They showed that, in regular problems, the score statistic S_R can be improved by a Bartlett-type correction which is not exactly a Bartlett correction because it involves a polynomial of second degree in the original statistic. Let S_R be a continuous score statistic having a chi-squared distribution with ν degrees of freedom asymptotically. Cordeiro and Ferrari proposed a modified score statistic given by

$$S_R^* = S_R\{1 - (c + bS_R + aS_R^2)\}, \tag{4}$$

where the coefficients a, b and c of order $O(n^{-1})$ make the modified score statistic in (4) have a χ_ν^2 distribution under the null hypothesis when terms of order smaller than n^{-1} are neglected. These coefficients are functions of some joint cumulants of log-likelihood derivatives and they depend upon the hypotheses being tested. When the coefficients a, b and c involve unknown parameters they should be replaced by their MLEs under the null hypothesis but this does not affect the order of approximation of the correction. Cordeiro and Ferrari (1998) expressed these coefficients as functions of the $O(n^{-1})$ terms of the expansions to the first three moments of the unmodified statistic S_R thus giving a simple "method of moments" for obtaining the Bartlett-type correction in (4). However, for complex regression models such as GLMs with dispersion covariates, it is not possible to obtain the first three moments of S_R to $O(n^{-1})$ directly. Then, we have to compute a, b and c from the three quantities A_1, A_2 and A_3, all of order n^{-1}, given in Harris's (1985) asymptotic expansion for the null distribution of S_R to order n^{-1}. Those quantities are obtainable from the cumulants of the total log-likelihood derivatives up to the fourth order, and will, in general, depend on unknown parameters. The coefficients in (4) are given by

$$a = \frac{A_3}{12\nu(\nu+2)(\nu+4)}, \quad b = \frac{(A_2 - 2A_3)}{12\nu(\nu+2)}, \quad c = \frac{(A_1 - A_2 + A_3)}{12\nu}. \tag{5}$$

One difficulty encountered with the use of S_R^* rather than S_R is the fact that the required coefficients a, b and c may be difficult to compute. For models where the joint cumulants of log-likelihood derivatives are invariant under permutation of parameters, we can obtain these coefficients using Harris's formulae for the quantities A_1, A_2 and A_3. Equation (4) is a general result which can be given to improve many important tests in econometrics and statistics. However, the improved statistic S_R^* is not always a monotone transformation. To overcome this, Kakizawa (1996) suggested a monotone transformation $K(S_R) = S_R^* + P(S_R)$ involving the statistic S_R itself and the coefficients a, b and c, where $P(S_R)$ is a polynomial of 5th degree in the original score statistic S_R and is of order $O_p(n^{-2})$. $P(S_R)$ is given by

$$P(S_R) = \frac{1}{4}\left\{c^2 S_R + 2bcS_R^2 + \left(2ac + \frac{4}{3}b^2\right) S_R^3 + 3abS_R^4 + \frac{9}{5}a^2 S_R^5\right\}.$$

Also, Cordeiro, Ferrari and Cysneiros (1998) presented an alternative formula for the improved score statistic which is a monotone transformation of S_R. The alternative statistic $\tilde{S}_R$ is expressed in terms of the normal distribution function $\Phi(\cdot)$ by

$$\tilde{S}_R = \sqrt{\frac{\pi}{3a}}\exp\left(\frac{b^2}{3a} - c\right)\left\{\Phi\left(\sqrt{6a}S_R + \sqrt{\frac{2}{3a}}b\right) - \Phi\left(\sqrt{\frac{2}{3a}}b\right)\right\}$$

if $a > 0$ (a is always non negative) and

$$\tilde{S}_R = \frac{1}{2b}\exp(-c)\{1 - \exp(-2bS)\}$$

if $a = 0$ and $b \neq 0$. Note that, if $a = b = 0$, S_R is a monotone transformation of S and there is no need to define an alternative corrected statistic. The three statistics S_R^*, $K(S_R)$ and $\tilde{S}_R$ are equivalent to the second order, i.e. they typically differ by $O_p(n^{-3/2})$.

The goal here is to apply formulae (4) and (5) to improve the score statistic S_R used in (3) to test the hypothesis $H_0^1 : \beta_1 = \beta_1^{(0)}, \gamma_1 = \gamma_1^{(0)}$ in model (1). From the log-likelihood function $\ell = \ell(\beta, \gamma)$ for the parameters β and γ defined in Section 2, we now introduce the notation used throughout the paper. The total log-likelihood derivatives with respect to the unknown parameters are indicated by subscripts, where lower-case letters r, s, $t, \ldots$ correspond to derivatives with respect to the β parameters, while upper-case letters R, S, $T, \ldots$ correspond to derivatives with respect to the γ parameters. Thus, $U_r = \partial\ell/\partial\beta_r$, $U_R = \partial\ell/\partial\gamma_R$, $U_{Rs} = \partial^2\ell/\partial\gamma_R\partial\beta_s$, $U_{rsT} = \partial^3\ell/\partial\beta_r\partial\beta_s\partial\gamma_T$ and so on. The standard notation for joint cumulants of these derivatives is used (Lawley, 1956; Cordeiro, 1993): $\kappa_{rs} = E(U_{rs})$, $\kappa_{R,S} = E(U_R U_S)$, $\kappa_{rs,T} = E(U_{rs}U_T)$, $\kappa_{rst} = E(U_{rst})$, $\kappa_{rs,TU} = E(U_{rs}U_{TU}) - \kappa_{rs}\kappa_{TU}$, $\kappa_{r,s,TU} = E(U_r U_s U_{TU}) - \kappa_{r,s}\kappa_{TU}$, etc., where all κ's refer to a total over the sample, and are, in general, of order n. Also, their derivatives are denoted by $\kappa_{rs}^{(t)} = \partial\kappa_{rs}/\partial\beta_t$, $\kappa_{rS}^{(T)} = \partial\kappa_{rS}/\partial\gamma_T$, etc. These cumulants satisfy standard regularity equations which usually facilitate their calculations, such as $\kappa_{r,s} = -\kappa_{rs}$, $\kappa_{r,s,t} = -\kappa_{rst} - \kappa_{rs,t} - \kappa_{rt,s} - \kappa_{st,r}$, $\kappa_{rs,t} = -\kappa_{rst} + \kappa_{rs}^{(t)}$, etc. Moreover, the information matrices for β and γ are denoted by K_β and K_γ with their typical elements given by $\kappa_{r,s}$ and $\kappa_{R,S}$, respectively. We assume that K_β and K_γ are nonsingular and denote the typical elements of their inverses by $\kappa^{r,s}$ and $\kappa^{R,S}$, respectively.

Some joint cumulants for the model under study are $\kappa_{rs} = -\sum_l \phi_l w_l x_{lr} x_{ls}$, $\kappa_{RS} = \sum_l \phi_{1l}^2 d_{2l} s_{lR} s_{lS}$, $\kappa_{Rst} = -\sum \phi_{1l} w_l x_{ls} x_{lt} s_{lR}$, $\kappa_{rST} = \kappa_{R,st} = 0$ and $\kappa_{RST} = \sum_l (d_{3l}\phi_{1l}^3 +$

$3d_{2l}\phi_{1l}\phi_{2l})s_{lR}s_{lS}s_{lT}$. The other cumulants needed for the derivation of the corrected score statistic S_R^* for testing H_0^1 are given in the Appendix.

Let θ_1 and θ_2 be defined as $\theta_1^\top = (\beta_1^\top, \gamma_1^\top)$ and $\theta_2^\top = (\beta_2^\top, \gamma_2^\top)$. Hence, θ_1 represents a $(p_1 + q_1)$-dimensional vector with the parameters of interest, while θ_2 is a $(p_2 + q_2)$-dimensional vector with the nuisance parameters, $p_2 = p - p_1$ and $q_2 = q - q_1$. We write, correspondingly, the total score function U as $U^\top = (U_1^\top, U_2^\top)$, where $U_1 = \partial\ell(\beta,\gamma)/\partial\theta_1$ and $U_2 = \partial\ell(\beta,\gamma)/\partial\theta_2$. Accordingly, consider the total Fisher information matrix $K = E(UU^\top)$ and its inverse K^{-1} partitioned as

$$K = \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix} \quad \text{and} \quad K^{-1} = \begin{pmatrix} K^{11} & K^{12} \\ K^{21} & K^{22} \end{pmatrix}. \tag{6}$$

From the global orthogonality between β and γ, it is easy to show that all the matrices K_{ij} defined above are block-diagonal, and they are given by $K_{ij} = \mathrm{diag}\{X_i^\top W\Phi X_j, -S_i^\top D_2\Phi_1^2 S_j\}$, for $i, j = 1, 2$. Following the partition considered here, we define

$$A = \begin{pmatrix} 0 & 0 \\ 0 & K_{22}^{-1} \end{pmatrix}$$

and let $M = K^{-1} - A$. Then, if a_{ij} and m_{ij} denote the (i,j)th elements of A and M, respectively, we obtain from Harris (1985) and Cordeiro, Ferrari and Paula (1993) the coefficients A_1, A_2 and A_3 that define the null Edgeworth expansion of S_R up to an error of order $n^{-3/2}$ and also the Bartlett-type correction in (4) to the score statistic. We can write from these authors the formulae for A_1, A_2 and A_3 as

$$\begin{aligned} A_1 &= 3A_{11} - 6A_{12} + 6A_{13} - 6A_{14}, \\ A_2 &= -3A_{21} + 6A_{22} - 6A_{23} + 3A_{24}, \\ A_3 &= 3A_{31} + 2A_{32}, \end{aligned} \tag{7}$$

where the A_{ij}'s are given by the corresponding summations in the expressions of the A_i's.

From the partitioned information matrix K given in (6) we can obtain the submatrices in K^{-1} as $K^{ij} = \mathrm{diag}\{K_\beta^{ij}, K_\gamma^{ij}\}$, where

$$\begin{aligned} &K_\beta^{11} = \{X_1^\top W\Phi(X_1 - X_2\xi_{12})\}^{-1}, \quad K_\gamma^{11} = \{-S_1^\top D_2\Phi_1^2(S_1 - S_2\rho_{12})\}^{-1}, \\ &K_\beta^{12} = \{X_1^\top W\Phi(X_1 - X_2\xi_{12})\}^{-1}\xi_{12}^\top, \quad K_\beta^{21} = K_\beta^{12T}, \\ &K_\gamma^{12} = -\{-S_1^\top D_2\Phi_1^2(S_1 - S_2\rho_{12})\}^{-1}\rho_{12}^\top, \quad K_\gamma^{21} = K_\gamma^{12T}. \end{aligned}$$

Here, $\xi_{12} = (X_2^\top W\Phi X_2)^{-1}X_2^\top W\Phi X_1$ and $\rho_{12} = (-S_2^\top D_2\Phi_1^2 S_2)^{-1}S_2^\top D_2\Phi_1^2 S_1$ have the forms of linear weighted least-squares equations with dependent variables X_1 and S_1 and weights $W\Phi$ and $-D_2\Phi_1^2$, respectively. Using these submatrices we can obtain the inverse of the information matrix K algebraically and then the asymptotic covariances matrices of the unrestricted estimates $\widehat{\beta}_1, \widehat{\beta}_2, \widehat{\gamma}_1$ and $\widehat{\gamma}_2$, explicitly. The asymptotic covariance matrix of the restricted estimates $\widetilde{\beta}_2$ and $\widetilde{\gamma}_2$ is simply $\mathrm{diag}\{(X_2^\top W\Phi X_2)^{-1}, (-S_2^\top D_2\Phi_1^2\, S_2)^{-1}\}$.

Equation (3), which uses the matrices Z_β and Z_γ to obtain the score statistic S_R for testing $H_0^1 : \beta_1 = \beta_1^{(0)}, \gamma_1 = \gamma_1^{(0)}$, has an disadvantage since it depends on inverting matrices of orders p and q. Large values of these numbers can give rise to numerical instability from rounding errors in the calculations. To overcome this problem, we give an alternative formula for S_R which requires the inversion of matrices of orders p_1, p_2, q_1 and q_2. Using the inverses given before we can obtain

$$\begin{aligned} S_R &= \tilde{z}_\beta^\top \widetilde{W}\widetilde{\Phi}X_1\{X_1^\top \widetilde{W}\widetilde{\Phi}(X_1 - X_2\tilde{\xi}_{12})\}^{-1}X_1^\top \widetilde{\Phi}\widetilde{W}\tilde{z}_\beta \\ &+ \tilde{z}_\gamma^\top \widetilde{D}_2\widetilde{\Phi}_1 S_1\{-S_1^\top \widetilde{D}_2\widetilde{\Phi}_1(S_1 - S_2\tilde{\rho}_{12})\}^{-1}S_1^\top \widetilde{\Phi}_1\widetilde{D}_2\tilde{z}_\gamma. \end{aligned} \tag{8}$$

Under the null hypothesis H_0^1 the score statistic (8) converges in distribution, as n increases, to a chi-squared distribution with $p_1 + q_1$ degrees of freedom, the error of this approximation being $O(n^{-1})$. However, this asymptotic approximation may not work well if n is not sufficiently large. It is then important to obtain an improved test which can be based on the modified statistic S_R^* in equation (4) and the $\chi^2_{p_1+q_1}$ reference distribution.

In fact, when the null hypothesis is true and under some regularity conditions, $P(S_R^* > x_\alpha) = \alpha + O(n^{-3/2})$ whereas $P(S_R > x_\alpha) = \alpha + O(n^{-1})$, where x_α is such that $P(\chi^2_{p_1+q_1} > x_\alpha) = \alpha$ and α is the desired significance level for the test. Then, improved score tests based on S_R^* have an error of order $O(n^{-3/2})$ under the null hypothesis H_0^1, whereas inference based on the original statistic S_R has an error of order $O(n^{-1})$. We now obtain matrix formulae for the A_{ij}'s in the specific test of $H_0^1 : \beta_1 = \beta_1^{(0)}, \gamma_1 = \gamma_1^{(0)}$ in order to define S_R^* from (4), (5) and (7). Improved likelihood ratio tests for GLMs with dispersion covariates were obtained by Botter and Cordeiro (1997) using the Lawley's (1956) general formulae.

Some additional notation is in order. Similarly to Z_β and Z_γ, we define the following $n \times n$ matrices: $Z_{\beta_2} = X_2(X_2^\top W\Phi X_2)^{-1}X_2^\top$ and $Z_{\gamma_2} = S_2(-S_2^\top D_2\Phi_1 S_2)^{-1}S_2^\top$. The subscript d indicates that the off-diagonal elements of the matrix were set equal to zero, 1 is an n-vector of ones, $\odot$ denotes the Hadamard (direct) product with further notation $Z^{(2)} = Z \odot Z, Z^{(3)} = Z^{(2)} \odot Z$, and $\mathrm{tr}(A)$ denotes the trace of the matrix A. Next, we define

$$f_l = \frac{1}{V_l}\frac{\mathrm{d}\mu_l}{\mathrm{d}\eta_l}\frac{\mathrm{d}^2\mu_l}{\mathrm{d}\eta_l^2}, \quad g_l = f_l - \frac{1}{V_l^2}\frac{\mathrm{d}V_l}{\mathrm{d}\mu_l}\left(\frac{\mathrm{d}\mu_l}{\mathrm{d}\eta_l}\right)^3, \quad b_l = \frac{1}{V_l^3}\left(\frac{\mathrm{d}\mu_l}{\mathrm{d}\eta_l}\right)^4\left\{\left(\frac{\mathrm{d}V_l}{\mathrm{d}\mu_l}\right)^2 + V_l\frac{\mathrm{d}^2V_l}{\mathrm{d}\mu_l^2}\right\},$$

$$h_l = \frac{1}{V_l^2}\frac{\mathrm{d}V_l}{\mathrm{d}\mu_l}\left(\frac{\mathrm{d}\mu_l}{\mathrm{d}\eta_l}\right)^2\frac{\mathrm{d}^2\mu_l}{\mathrm{d}\eta_l^2} + \frac{1}{V_l^2}\frac{\mathrm{d}^2V_l}{\mathrm{d}\mu_l^2}\left(\frac{\mathrm{d}\mu_l}{\mathrm{d}\eta_l}\right)^4$$

and the diagonal matrices $F = \mathrm{diag}\{f_1, \ldots, f_n\}$, $G = \mathrm{diag}\{g_1, \ldots, g_n\}$, $B = \mathrm{diag}\{b_1, \ldots, b_n\}$, $H = \mathrm{diag}\{h_1, \ldots, h_n\}$, $D_i = \mathrm{diag}\{d_{i1}, \ldots, d_{in}\}$, for $i = 3, 4$, and $\Phi_2 = \mathrm{diag}\{\phi_{21}, \ldots, \phi_{2n}\}$. These matrices are functions of the first two derivatives of the variance and link functions and the function $d(\cdot)$ through its second, third and fourth derivatives.

Plugging the cumulants given in the Appendix into the expressions for the A_{ij}'s and using the definitions above, we are able to write the quantities A_{ij}'s in matrix notation. From the A_{ij}'s, we can obtain A_1, A_2 and A_3 using (7), and, finally, we arrive at the coefficients a, b and c in (5). The algebraic details of the derivation of the A_{ij}'s are too lengthy and tedious to be given here but they follow similar algebraic developments of Cordeiro, Ferrari and Paula (1993). These mathematical details can be obtained from the authors upon request. After a lengthy calculation, we arrive at the following results:

$$\begin{aligned} A_{11} &= 1^\top \Phi F Z_{\beta_2 d}(Z_\beta - Z_{\beta_2})Z_{\beta_2 d}F\Phi 1 \\ &\quad + 1^\top \Phi_1\{WZ_{\beta_2 d} + (D_3 + \Phi_2 D_2)Z_{\gamma_2 d}\}(Z_\gamma - Z_{\gamma_2})\{WZ_{\beta_2 d} + (D_3 + \Phi_2 D_2)Z_{\gamma_2 d}\}\Phi_1 1, \quad (9) \\ A_{12} &= -1^\top \Phi F Z_{\beta_2 d}Z_{\beta_2}(Z_{\beta d} - Z_{\beta_2 d})(F - G)\Phi 1 \end{aligned}$$

$$- 1^\top \Phi_1\{W Z_{\beta_2 d} + (D_3 + \Phi_2 D_2) Z_{\gamma_2 d}\} Z_{\gamma_2} \{W(Z_{\beta d} - Z_{2\beta d}) + D_3(Z_{\gamma d} - Z_{\gamma_2 d})\}\Phi_1 1, \quad (10)$$

$$A_{13} = -1^\top \Phi(2G - F) Z_{\beta_2}^{(2)} \odot (Z_\beta - Z_{\beta_2}) F \Phi 1 + 1^\top \Phi_1 W Z_{\beta_2}^{(2)} \odot (Z_\gamma - Z_{\gamma_2}) W \Phi_1 1$$
$$+ 1^\top \Phi_1 (D_3 - \Phi_2 D_2) Z_{\gamma_2}^{(2)} \odot (Z_\gamma - Z_{\gamma_2})(D_3 + \Phi_2 D_2)\Phi_1 1, \quad (11)$$

$$A_{14} = \mathrm{tr}\{\Phi H Z_{\beta_2 d}(Z_{\beta d} - Z_{\beta_2 d})\} + 2\,\mathrm{tr}\{\Phi^2 \Phi_1^2 W Z_{\beta_2 d}(Z_{\gamma d} - Z_{\gamma_2 d})\}$$
$$+ \mathrm{tr}\{(2\Phi^2\Phi_1^2 - \Phi_2) W Z_{\gamma_2 d}(Z_{\beta d} - Z_{\beta_2 d})\} - \mathrm{tr}\{\Phi_1^2(\Phi_1^2 D_4 + \Phi_2 D_3) Z_{\gamma_2 d}(Z_{\gamma d} - Z_{\gamma_2 d})\}, \quad (12)$$

$$A_{21} = 1^\top \Phi(F - G)(Z_{\beta d} - Z_{\beta_2 d}) Z_{\beta_2} (Z_{\beta d} - Z_{\beta_2 d})(F - G)\Phi 1$$
$$+ 1^\top \Phi_1\{W(Z_{\beta d} - Z_{\beta_2 d}) + D_3(Z_{\gamma d} - Z_{\gamma_2 d})\} Z_{\gamma_2} \{W(Z_{\beta d} - Z_{\beta_2 d})$$
$$+ D_3(Z_{\gamma d} - Z_{\gamma_2 d})\}\Phi_1 1, \quad (13)$$

$$A_{22} = -1^\top \Phi F Z_{\beta_2 d}(Z_\beta - Z_{\beta_2})(Z_{\beta d} - Z_{\beta_2 d})(F - G)\Phi 1$$
$$- 1^\top \Phi_1\{W Z_{\beta_2 d} + (D_3 + \Phi_2 D_2) Z_{\gamma_2 d}\}(Z_\gamma - Z_{\gamma_2})\{W(Z_{\beta d} - Z_{\beta_2 d})$$
$$+ D_3(Z_{\gamma d} - Z_{\gamma_2 d})\}\Phi_1 1, \quad (14)$$

$$A_{23} = 1^\top \Phi(F - G)(Z_\beta - Z_{\beta_2})^{(2)} \odot Z_{\beta_2}(F - G)\Phi 1$$
$$+ 2\, 1^\top \Phi_1 W (Z_\beta - Z_{\beta_2}) \odot Z_{\beta_2} \odot (Z_\gamma - Z_{\gamma_2}) W \Phi_1 1$$
$$+ 1^\top \Phi_1 W (Z_\beta - Z_{\beta_2})^{(2)} \odot Z_{\gamma_2} W \Phi_1 1 + 1^\top \Phi_1 D_3 (Z_\gamma - Z_{\gamma_2})^{(2)} \odot Z_{\gamma_2} D_3 \Phi_1 1, \quad (15)$$

$$A_{24} = \mathrm{tr}\{\Phi B (Z_{\beta d} - Z_{\beta_2 d})^2\} + 4\,\mathrm{tr}\{\Phi^2\Phi_1^2 W (Z_{\beta d} - Z_{\beta_2 d})(Z_{\gamma d} - Z_{\gamma_2 d})\}$$
$$- \mathrm{tr}\{\Phi_1^2 D_4 (Z_{\gamma d} - Z_{\gamma_2 d})^2\}, \quad (16)$$

$$A_{31} = 1^\top \Phi(F - G)(Z_{\beta d} - Z_{\beta_2 d})(Z_\beta - Z_{\beta_2})(Z_{\beta d} - Z_{\beta_2 d})(F - G)\Phi 1$$
$$+ 1^\top \Phi_1\{W(Z_{\beta d} - Z_{\beta_2 d}) + D_3(Z_{\gamma d} - Z_{\gamma_2 d})\}(Z_\gamma - Z_{\gamma_2})\{W(Z_{\beta d} - Z_{\beta_2 d})$$
$$+ D_3(Z_{\gamma d} - Z_{\gamma_2 d})\}\Phi_1 1 \quad (17)$$

and

$$A_{32} = 1^\top \Phi(F - G)(Z_\beta - Z_{\beta_2})^{(3)}(F - G)\Phi 1$$
$$+ 3\, 1^\top \Phi_1 W (Z_\beta - Z_{\beta_2})^{(2)} \odot (Z_\gamma - Z_{\gamma_2}) W \Phi_1 1 + 1^\top \Phi_1 D_3 (Z_\gamma - Z_{\gamma_2})^{(3)} D_3 \Phi_1 1. \quad (18)$$

Equations (9)–(18) represent the main result of our paper. These equations together with (7) and (5) yield the three coefficients a, b and c that define the corrected statistic S_R^* in (4) and also the asymptotically equivalent versions $K(S_R)$ and $\widetilde{S}_R$. Using these equations, one can design improved inference in GLMs with dispersion covariates based on these corrected statistics. Formulae (9)–(18) are functions of the matrices $Z_\beta, Z_{\beta_2}, Z_\gamma$ and Z_{γ_2}, of the unknown means μ_l's and of the unknown precision parameters ϕ_l's. They only involve simple operations on matrices and vectors, and can be easily implemented into a computer algebra system such as MAPLE or MATHEMATICA, or into a programming language with support for matrix operations, such as GAUSS, Ox or S-PLUS. This will be done in future work. Our formulae are general enough to include a number of published results as special cases, such as those in Cordeiro, Ferrari and Paula (1993) and Cribari-Neto and Ferrari (1995), as will be shown in Sections 4 and 5. For special forms of the matrices X and S we can obtain closed-form expressions for the A_{ij}'s (see Section 5). Finally, equations (9)–(18) are not easy to interpret since their individual terms are not invariant under reparameterization. Moreover, the A_{ij}'s provide no indication as to what structural aspects of the model contribute significantly to their magnitude.

4. Testing mean and dispersion effects separately

In this section we obtain the A_i's to improve two tests: first, the score test on mean parameters and, second, the score test on a subset of the γ parameters.

4.1. Testing mean effects

Now we are interested in testing a subset of the β parameters. In this situation, the null hypothesis is $H_0^2 : \beta_1 = \beta_1^{(0)}$ to be tested against $H_1^2 : \beta_1 \neq \beta_1^{(0)}$, where $\beta_1^{(0)}$ is a specified vector of dimension p_1, and β_2 and $\gamma = (\gamma_1^\top, \gamma_2^\top)^\top$ are the vectors of nuisance parameters. For testing H_0^2 the score statistic reduces to the first term of (8) evaluated at $(\beta_1^{(0)T}, \tilde{\beta}_2^\top, \tilde{\gamma}^\top)$. In this case, $Z_\gamma = Z_{\gamma_2}$ and equations (9)–(18) have some reduction leading to

$$\begin{aligned}
A_1 &= 3\,1^\top \Phi F Z_{\beta_2 d}(Z_\beta - Z_{\beta_2}) Z_{\beta_2 d} F \Phi 1 + 6\,1^\top \Phi F Z_{\beta_2 d} Z_{\beta_2}(Z_{\beta d} - Z_{\beta_2 d})(F - G)\Phi 1 \\
&+ 6\,1^\top \Phi_1 \{W Z_{2\beta d} + (D_3 + \Phi_2 D_2) Z_{\gamma d}\} Z_\gamma (Z_{\beta d} - Z_{\beta_2 d}) W \Phi_1 1 \\
&- 6\,1^\top \Phi(2G - F) Z_{\beta_2}^{(2)} \odot (Z_\beta - Z_{\beta_2}) F \Phi 1 - 6\,\mathrm{tr}\{\Phi H Z_{\beta_2 d}(Z_{\beta d} - Z_{\beta_2 d})\} \\
&- 6\,\mathrm{tr}\{(2\Phi^2 \Phi_1^2 - \Phi_2) W Z_{\gamma d}(Z_{\beta d} - Z_{\beta_2 d})\}, \qquad (19)\\
A_2 &= -3\,1^\top \Phi(F - G)(Z_{\beta d} - Z_{\beta_2 d}) Z_{\beta_2}(Z_{\beta d} - Z_{\beta_2 d})(F - G)\Phi 1 \\
&- 3\,1^\top \Phi_1 W (Z_{\beta d} - Z_{\beta_2 d}) Z_\gamma (Z_{\beta d} - Z_{\beta_2 d}) W \Phi_1 1 \\
&- 6\,1^\top \Phi F Z_{\beta_2 d}(Z_\beta - Z_{\beta_2})(Z_{\beta d} - Z_{\beta_2 d})(F - G)\Phi 1 \\
&- 6\,1^\top \Phi(F - G)(Z_\beta - Z_{\beta_2})^{(2)} \odot Z_{\beta_2}(F - G)\Phi 1 \\
&- 6\,1^\top \Phi_1 W (Z_\beta - Z_{\beta_2})^{(2)} \odot Z_\gamma W \Phi_1 1 + 3\,\mathrm{tr}\{\Phi B (Z_{\beta d} - Z_{\beta_2 d})^{(2)}\}, \qquad (20)\\
A_3 &= 3\,1^\top \Phi(F - G)(Z_{\beta d} - Z_{\beta_2 d})(Z_\beta - Z_{\beta_2})(Z_{\beta d} - Z_{\beta_2 d})(F - G)\Phi 1 \\
&+ 2\,1^\top \Phi(F - G)(Z_\beta - Z_{\beta d})^{(3)}(F - G)\Phi 1. \qquad (21)
\end{aligned}$$

When we plug expressions (19)–(21) evaluated at $(\beta_1^{(0)T}, \tilde{\beta}_2^\top, \tilde{\gamma}^\top)$ into equations (5) and (4) we obtain the corrected statistic S_R^*. The distribution of S_R^* can be approximated, under the null hypothesis, by a $\chi^2_{p_1}$ distribution with an error of order $O(n^{-3/2})$. Equations (19)–(21) can be checked in some special cases. For standard GLMs with common precision parameter ϕ we have $q = 1$ and these equations reduce to the A's given by Cribari-Neto and Ferrari (1995; equations (5a)-(5c), (7a)-(7b)).

We can consider a test on all elements of β, i.e. $H_0^2 : \beta = \beta^{(0)}$. In this case, $Z_{\beta_2} = 0$ and equations (19)–(21) lead to

$$\begin{aligned}
A_1 &= 6\,1^\top \Phi_1 (D_3 + \Phi_2 D_2) Z_{\gamma d} Z_\gamma Z_{\beta d} W \Phi_1 1 - 6\mathrm{tr}\{(\Phi^2 \Phi_1^2 - \Phi_2) W Z_{\gamma d} Z_{\beta d}\}, \\
A_2 &= -3\,1^\top \Phi_1 W Z_{\beta d} Z_\gamma Z_{\beta d} W \Phi_1 1 - 6\,1^\top \Phi_1 W Z_\beta^{(2)} \odot Z_\gamma W \Phi_1 1 + 3\,\mathrm{tr}\{\Phi B Z_{\beta d}^2\}, \\
A_3 &= 1^\top \Phi(F - G)\{3 Z_{\beta d} Z_\beta Z_{\beta d} + 2(Z_\beta - Z_{\beta d})^{(3)}\}(F - G)\Phi 1.
\end{aligned}$$

Finally, for the test of $H_0^2 : \beta = \beta^{(0)}$ in the class of GLMs ($q = 1$), the preceding three expressions reduce to the A's given by Cordeiro, Ferrari and Paula (1993; eq. (6.2)).

4.2. Testing dispersion effects

We now consider the test of a subset of parameters whose covariates model only the precision parameter ϕ. The null hypothesis is $H_0^3 : \gamma_1 = \gamma_1^{(0)}$ to be tested against $H_1^3 : \gamma_1 \neq \gamma_1^{(0)}$, where $\gamma_1^{(0)}$ is a specified vector of dimension q_1. For this test, the score statistic is just given by the second term of (8) with the functions being evaluated at $(\tilde{\beta}^{\top}, \gamma_1^{(0)T}, \tilde{\gamma}_2^{\top})$. Using $Z_\beta = Z_{\beta_2}$ the general expressions for the A_i's come from (9)–(18) and simplify to

$$\begin{aligned}
A_1 &= 3\,1^{\top}\Phi_1\{WZ_{\beta d} + (D_3 + \Phi_2 D_2)Z_{\gamma_2 d}\}(Z_\gamma - Z_{\gamma_2})\{WZ_{\beta d} + (D_3 + \Phi_2 D_2)Z_{\gamma_2 d}\}\Phi_1 1 \\
&+ 6\,1^{\top}\Phi_1\{WZ_{\beta d} + (D_3 + \Phi_2 D_2)Z_{\gamma_2 d}\}Z_{\gamma_2}\{D_3(Z_{\gamma d} - Z_{\gamma_2 d})\}\Phi_1 1 \\
&+ 6\,1^{\top}\Phi_1 W Z_\beta^{(2)} \odot (Z_\gamma - Z_{\gamma_2})W\Phi_1 1 \\
&+ 6\,1^{\top}\Phi_1(D_3 - \Phi_2 D_2)Z_{\gamma_2}^{(2)} \odot (Z_\gamma - Z_{\gamma_2})(D_3 + \Phi_2 D_2)\Phi_1 1 \\
&- 12\,\mathrm{tr}\{\Phi^2\Phi_1^2 W Z_{\beta d}(Z_{\gamma d} - Z_{\gamma_2 d})\} + 6\,\mathrm{tr}\{\Phi_1^2(\Phi_1^2 D_4 + \Phi_2 D_3)Z_{\gamma_2 d}(Z_{\gamma d} - Z_{\gamma_2 d})\}, \qquad (22) \\
A_2 &= -3\,1^{\top}\Phi_1 D_3(Z_{\gamma d} - Z_{\gamma_2 d})Z_{\gamma_2}(Z_{\gamma d} - Z_{\gamma_2 d})D_3\Phi_1 1 \\
&- 6\,1^{\top}\Phi_1\{WZ_{\beta d} + (D_3 + \Phi_2 D_2)Z_{\gamma_2 d}\}(Z_\gamma - Z_{\gamma_2})(Z_{\gamma d} - Z_{\gamma_2 d})D_3\Phi_1 1 \\
&- 6\,1^{\top}\Phi_1 D_3(Z_\gamma - Z_{\gamma_2})^{(2)} \odot Z_{\gamma_2} D_3\Phi_1 1 - 3\,\mathrm{tr}\{\Phi_1^4 D_4(Z_{\gamma d} - Z_{\gamma_2 d})^2\}, \qquad (23) \\
A_3 &= 1^{\top}\Phi_1 D_3\{3(Z_{\gamma d} - Z_{\gamma_2 d})(Z_\gamma - Z_{\gamma_2})(Z_{\gamma d} - Z_{2\gamma d}) + 2(Z_\gamma - Z_{\gamma_2})^{(3)}\}D_3\Phi_1 1. \qquad (24)
\end{aligned}$$

If the null hypothesis specifies the complete set of dispersion parameters, i.e., $H_0^3 : \gamma = \gamma^{(0)}$, which implies $q_1 = q$ and $Z_{\gamma_2} = 0$, equations (22)–(24) become

$$\begin{aligned}
A_1 &= 3\,1^{\top}\Phi_1 W Z_{\beta d} Z_\gamma Z_{\beta d} W\Phi_1 1 + 6\,1^{\top}\Phi_1 W Z_\beta^{(2)} \odot Z_\gamma W \Phi_1 1 - 12\,\mathrm{tr}\{\Phi^2\Phi_1^2 W Z_{\beta d} Z_{\gamma d}\}, \\
A_2 &= -6\,1^{\top}\Phi_1 W Z_{\beta d} Z_\gamma Z_{\gamma d} D_3 \Phi_1 1 - 3\,\mathrm{tr}\{\Phi_1^4 D_4 Z_{\gamma d}^2\}, \\
A_3 &= 1^{\top}\Phi_1 D_3\{3Z_{\gamma d} Z_\gamma Z_{\gamma d} + 2Z_\gamma^{(3)}\}D_3\Phi_1 1.
\end{aligned}$$

For $q = 1$, these A_i's above are in agreement with equations (5.3) given by Cordeiro, Ferrari and Paula (1993). The improved statistic S_R^* for testing $H_0^3 : \gamma_1 = \gamma_1^{(0)}$ is obtained from (22)–(24), (5) and (4). S_R^* is distributed as $\chi^2_{q_1}$, under the null hypothesis, with an error of order $O(n^{-3/2})$.

5. Special Models

Equations (9)–(18) involve only simple operations on matrices and vectors and can be used to derive Bartlett-type corrections for several GLMs with dispersion covariates. In this section we briefly discuss three special models: the simple model, the one-way classification structures for both linear predictors η and τ and the model defined by simple linear regression structures for these predictors. For all these cases, the mean and the precision parameters, under the null hypothesis, are constant over the observations. We then use the following notation: $V_l = V$, $\mu_l = \mu$, $\eta_l = \eta$, $\phi_l = \phi$, $\phi_{1l} = \phi_1$, $\phi_{2l} = \phi_2$, $d_{il} = d_i$, for $l = 1, \ldots, n$ and $i = 2, 3, 4$.

5.1. The simple model

We begin with the simple model where the observations are independent and identically distributed (iid). For this case, $p = q = 1$ and X and S are n-vectors of ones. We consider a test of the null hypothesis $H_0^1 : \beta = \beta^{(0)}, \gamma = \gamma^{(0)}$ against $H_1^1 : H_0^1$ is false. We have $Z_\beta = (\mathrm{d}\eta/\mathrm{d}\mu)^2 V (n\phi)^{-1} 1 1^{\mathsf{T}}$ and $Z_\gamma = -(\mathrm{d}\tau/\mathrm{d}\phi)^2 (n d_2)^{-1} 1 1^{\mathsf{T}}$. After some algebra, we obtain from (9)–(18) and (5)

$$
\begin{aligned}
A_1 &= 0, && (25)\\
A_2 &= \frac{3}{n\phi}\left\{\frac{\mathrm{d}^2 V}{\mathrm{d}\mu^2} + \frac{1}{V}\left(\frac{\mathrm{d}V}{\mathrm{d}\mu}\right)^2\right\} - \frac{12}{n\phi^2 d_2} - \frac{3d_4}{n d_2^2}, && (26)\\
A_3 &= \frac{5}{n\phi V}\left(\frac{\mathrm{d}V}{\mathrm{d}\mu}\right)^2 - \frac{9}{n\phi^2 d_2} + \frac{6d_3}{n\phi d_2^2} - \frac{5d_3^2}{n d_2^3}. && (27)
\end{aligned}
$$

The Bartlett-type correction obtained from (25)–(27) does not involve any estimation since the null hypothesis specifies the model completely. For the normal model, we have $V = 1$ and $d(\phi) = (\log\phi)/2$ and equations (25)–(27) and (5) yield $a = 41/(288n)$, $b = -13/(12n)$ and $c = 11/(12n)$. The Bartlett–type correction here does not involve any estimation since the null hypothesis specifies the model completely.

5.2. One-way classification models

We now consider the analysis of variance model. Each observation is classified into one of p different categories, where observations in the ith category have mean μ_i and precision parameter ϕ_i and we express the linear predictors η_i and τ_i as $\eta_i = \beta + \beta_i$ and $\tau_i = \gamma + \gamma_i$, for each $i = 1, \ldots, p$, where we fix $\beta_p = \gamma_p = 0$. Then, we have the one-way classification structure for both predictors η and τ. For this model, we consider the test of $H_0^1 : \beta_1 = \ldots = \beta_{p-1} = \gamma_1 = \ldots = \gamma_{p-1} = 0$ against H_1^1 : violation of at least one equality. Let n_i be the number of observations in category i, which means that $n_1 + \ldots + n_p = n$. Under the null hypothesis H_0^1, the matrix Z_β has a typical element given by $\zeta_{lm}/(n_l w\phi)$, where $\zeta_{lm} = 1$ if l and m index observations in the same population and zero otherwise. Analogously, the matrix Z_γ under H_0^1 has a typical element given by $-\zeta_{lm}/(n_l d_2 \phi_1^2)$. After some algebra we can obtain from (9)–(18) and (5)

$$
A_1 = \frac{6(p-1)}{n\phi^2 d_2^3}\left\{\phi^2(d_2 d_4 - d_3^2) + \phi d_2\left(-d_2^2\frac{\mathrm{d}^2 V}{\mathrm{d}\mu^2} + 2d_3\right) + 3d_2^2\right\}, \tag{28}
$$

$$
\begin{aligned}
A_2 = {} & \frac{3(p-1)}{n\phi^2 d_2^3 V}\left\{(p-1)\phi d_2\left[d_2^2\left(\frac{\mathrm{d}V}{\mathrm{d}\mu}\right)^2 - 2d_3 V\right] + V\left[(p+5)d_2^2 - (p-1)\phi^2 d_3^2\right]\right\}\\
& + \frac{3}{\phi^2 d_2^3 V}\left\{\phi d_2^3\left[\left(\frac{\mathrm{d}V}{\mathrm{d}\mu}\right)^2 + V\frac{\mathrm{d}^2 V}{\mathrm{d}\mu^2}\right] - V(4d_2 + \phi^2 d_4)\right\}\left\{\sum_{i=1}^{p}\frac{1}{n_i} + \frac{1-2p}{n}\right\},
\end{aligned} \tag{29}
$$

$$
\begin{aligned}
A_3 = {} & \frac{1}{\phi^2 d_2^3 V}\left\{\phi d_2^3\left(\frac{\mathrm{d}V}{\mathrm{d}\mu}\right)^2 - V(d_2^2 + \phi^2 d_3^2)\right\}\left\{5\sum_{i=1}^{p}\frac{1}{n_i} - \frac{3p^2 + 6p - 4}{n}\right\}\\
& + \frac{3d_3}{\phi d_2^2}\left(\sum_{i=1}^{p}\frac{1}{n_i} - \frac{p^2}{n}\right).
\end{aligned} \tag{30}
$$

We note from equations (28)–(30) that the Bartlett-type correction for this test does not depend on the link functions as expected, since the restricted MLEs of μ and ϕ are just sample moments which do not involve these functions. For the normal model we have $d_2(\phi) = (\log\phi)/2$ and $V = 1$ and equations (28)–(30) become

$$A_1 = -\frac{12(p-1)}{n},$$
$$A_2 = -\frac{6(p-1)(p+5)}{n} + 60\left(\sum_{i=1}^{p}\frac{1}{n_i} + \frac{1-2p}{n}\right),$$
$$A_3 = 12\left(\sum_{i=1}^{p}\frac{1}{n_i} - \frac{p^2}{n}\right) + 10\left(5\sum_{i=1}^{p}\frac{1}{n_i} - \frac{3p^2+6p-4}{n}\right).$$

5.3. Simple linear regression model

We now consider two-parameter regression models for both predictors η and τ. The systematic components are expressed by $d(\mu_l) = \eta_l = \beta_2 + \beta_1 x_l$ and $g(\phi_l) = \tau_l = \gamma_2 + \gamma_1 s_l$, where x_l and s_l denote the values of the explanatory variables x and s, for $l = 1, \ldots, n$. We consider the test of $H_0^1 : \beta_1 = \gamma_1 = 0$ against $H_1^1 : H_0^1$ is false. We use the notation $\delta_{ij} = n^{-1}\sum_l (x_l - \bar{x})^i (s_l - \bar{s})^j$, for $i, j = 0, \ldots, 4$, where $\bar{x} = n^{-1}\sum_l x_l$ and $\bar{s} = n^{-1}\sum_l s_l$. Under the unrestricted model, $X = (1\ x)$ and $S = (1\ s)$ with $x = (x_1, \ldots, x_n)^\top$ and $s = (s_1, \ldots, s_n)^\top$, the (l, m)th elements of the matrices Z_β and Z_γ, evaluated at $\beta_1 = \gamma_1 = 0$, are given by $z_{\beta_{lm}} = (nw\phi\delta_{20})^{-1}\{\delta_{20} + (x_l - \bar{x})(x_m - \bar{x})\}$ and $z_{\gamma_{lm}} = (nwd_2\phi_1\delta_{02})^{-1}\{\delta_{02} + (s_l - \bar{s})(s_m - \bar{s})\}$, respectively.

After some algebraic development we find the A_i's from (9)–(18) and (5) as functions of n, μ and ϕ as

$$A_1 = \frac{6}{n\phi^3 w d_2}\left\{d_3\phi\phi_1^2 + \left(\phi\phi_2 - \phi_1^2 - \frac{d^2V}{d\mu^2}\phi^2 w\right)d_2\right\}$$
$$+ \frac{6}{n\phi^2\phi_1^2 d_2^3}\left\{-d_3^2\phi^2\phi_1^2 + (3d_3\phi\phi_1^2 - 2d_3\phi^2\phi_2 + d_4\phi^2\phi_1^2)d_2 + (\phi\phi_2 + 2\phi_1^2)d_2^2\right\}, \quad (31)$$

$$A_2 = \frac{3}{n\phi^2 V d_2^3}\left\{3\phi d_2^3\left(\frac{dV}{d\mu}\right)^2 + 3Vd_2^2 - 2\phi V d_3 d_2 + 3\phi^2 V d_3^2\right\}$$
$$+ \frac{3}{n\phi^2 d_2^2\delta_{20}^2\delta_{02}^2}\left[4d_2\delta_{20}\delta_{02}(\delta_{11}^2 - \delta_{22}) + \left\{\frac{1}{V}\left(\frac{dV}{d\mu}\right)^2 + \frac{d^2V}{d\mu^2}\right\}\phi d_2^2\delta_{40}\delta_{02}^2 - \phi^2 d_4\delta_{04}\delta_{20}^2\right] \quad (32)$$

$$A_3 = \frac{1}{n\phi^2 d_2^3\delta_{20}^3\delta_{02}^3}\left\{\frac{5}{V}\left(\frac{dV}{d\mu}\right)^2\phi d_2^3\delta_{30}^2\delta_{02}^3 - 9d_2^2\delta_{21}^2\delta_{20}\delta_{02}^2 + 6\phi d_2 d_3\delta_{21}\delta_{20}^2\delta_{02}\delta_{03}\right.$$
$$\left. - 5\phi^2 d_3^2\delta_{20}^3\delta_{03}^2\right\}. \quad (33)$$

It is clear from A_2 and A_3 that the χ_2^2 approximation to the score test is particularly sensitive to changes in the sample joint moments δ_{lm} of x and s. For the normal model with identity link function for the mean and log link function for the precision parameter, equations (31)–(33) reduce to

$$A_1 = -\frac{12}{n}\left(\frac{1}{\phi}+3\right), \quad A_2 = \frac{6}{n}\left\{\frac{4(\delta_{22}-\delta_{11}^2)}{\delta_{20}\delta_{02}}+\frac{6\delta_{04}}{\delta_{02}^2}-19\right\},$$
$$A_3 = \frac{2}{n\delta_{20}^2\delta_{02}^3}\{9\delta_{21}^2\delta_{02}^2+12\delta_{21}\delta_{03}\delta_{20}\delta_{02}+20\delta_{03}^2\delta_{20}^2\}.$$

6. Simulation Results

In order to check our theoretical results, simulations are performed to study the finite-sample distributions of the likelihood ratio and score statistics and their corrected versions. The Bartlett correction for the likelihood ratio test was taken from Botter and Cordeiro (1997). The simulation results are based on a gamma regression model with $\mu_i = \eta_i^{-1}$ and $\log\phi_i = \tau_i$, where $\eta = X\beta$ and $\tau = S\gamma$. In what follows, we report the null rejection rates of some tests of the null hypothesis $H_0^2 : \beta_1 = 0$ against a two sided alternative, i.e., the percentage of times that they exceed the appropriate 10% and 5% upper points of a $\chi^2_{p_1}$ distribution, where $p_1 = \dim(\beta_1)$.

We considered the likelihood ratio test statistic (LR), its Bartlett-corrected version LR^*, the score statistic (S_R) and three corrected versions: the corrected score statistic S_R^* given in (4) and the monotone corrected score statistics $\widetilde{S}_R$ and $K(S_R)$ proposed by Cordeiro, Ferrari and Cysneiros (1998) and Kakizawa (1996), respectively. The covariate values for the second column of X and the first column of S were set equal 1, and all the other covariates were obtained as independent draws from a standard uniform distribution $\mathcal{U}(0,1)$. We set $p_1 = 1$, $p = 4, 6$, $q = 3, 5$, $n = 20, 30, 40, 50$ and $\beta_2 = \ldots = \beta_p = \gamma_1 = \ldots = \gamma_q = 1$ and report the results for two different nominal significance levels, namely: $\alpha = 0.10, 0.05$. The results are presented in Tables 1–4. Entries are percentages. All simulations were performed using the Ox matrix programming language (Doornik, 1999). The nonlinear maximizations of the relevant log-likelihood were carried out using a quasi-Newton algorithm known as 'BFGS' and all results are based on 10,000 replications.

The figures in Tables 1–4 reveal important information. First, the likelihood ratio and score tests are largely liberal, overrejecting the null hypothesis more frequently than expected based on the selected nominal levels. In fact, the likelihood ratio and score tests tend to overestimate considerably the magnitude of the upper tails of the χ_1^2 approximation. The asymptotic χ_1^2 approximation for the unmodified statistics only works well for very large n in agreement with the large sample asymptotic results. We also note that the unmodified score statistic yields a more accurate test when compared with the χ_1^2 distribution than the unmodified likelihood ratio statistic. Second, the corrected likelihood ratio and score statistics, which are deflations of the raw statistics, are really necessary even for relatively large sample sizes. The corrected likelihood ratio and score statistics do well, even for very small sample sizes. For all cases reported in Tables 1–4, the Bartlett and Bartlett-type corrections for the likelihood ratio and score statistics are very effective in pushing the rejection rates of the modified statistics toward to the nominal levels. Clearly, the corrections have less impact as the sample size increases. Third, for fixed p (q), the χ_1^2 approximation for the likelihood ratio and score statistics deteriorates when q (p) increases. For example, for $n = 30$ and

$(p,q) = (4,3)$ and $(p,q) = (4,5)$, the sizes of the likelihood ratio (score) test corresponding to $\alpha = 5\%$ are 10.6% (7.2%) and 17.3% (8.0%), respectively. Thus, the tendency of overrejecting the unmodified likelihood ratio and score tests becomes more evident as the number of nuisance parameters increases. Fourth, the modified score statistic S_R^* outperforms slightly the monotonic corrected statistics $\tilde{S}_R$ and $K(S_R)$ as far as size distortions are concerned, and then it delivers the most accurate of the four score tests. Finally, from the simulation results presented here, it seems reasonable to judge the χ^2 tests based on the modified statistics as much better tests than those based on the unadjusted statistics.

Table 1: Estimated sizes of the likelihood ratio, corrected likelihood ratio, score and three corrected score tests; $p = 4$, $q = 3$

Sample size	Nominal size	LR	LR^*	S_R	S_R^*	$\tilde{S}_R$	$K(S_R)$
20	10.0	21.8	15.1	12.1	10.0	9.8	10.1
	5.0	14.4	8.4	5.7	4.8	4.8	4.9
30	10.0	17.7	12.5	13.5	10.2	10.1	10.4
	50	10.6	6.9	7.2	5.2	4.9	5.3
40	10.0	14.2	10.9	11.4	10.1	10.2	10.1
	5.0	7.6	5.6	5.7	5.1	5.1	5.1
50	10.0	13.2	10.4	11.2	9.8	9.9	9.8
	5.0	7.5	5.5	5.8	5.1	5.1	5.1

Table 2: Estimated sizes of the likelihood ratio, corrected likelihood ratio, score and three corrected score tests; $p = 4$, $q = 5$

Sample size	Nominal size	LR	LR^*	S_R	S_R^*	$\tilde{S}_R$	$K(S_R)$
20	10.0	40.5	27.1	17.3	11.0	12.3	11.2
	5.0	31.1	18.0	8.9	5.3	6.0	5.5
30	10.0	25.6	17.1	15.3	10.7	11.5	10.8
	5.0	17.3	10.1	8.0	5.4	5.8	5.4
40	10.0	17.1	12.3	12.4	10.3	10.6	10.3
	5.0	10.2	6.5	6.2	5.2	5.2	5.2
50	10.0	15.6	11.2	11.3	9.8	10.0	9.8
	5.0	8.7	6.0	5.9	5.1	5.2	5.1

Table 3: Estimated sizes of the likelihood ratio, corrected likelihood ratio, score and three corrected score tests; $p=6$, $q=3$

Sample size	Nominal size	LR	LR^*	S_R	S_R^*	$\widetilde{S}_R$	$K(S_R)$
20	10.0	26.6	18.2	17.4	10.9	12.9	11.3
	5.0	18.6	11.3	9.9	5.3	6.6	5.4
30	10.0	21.7	14.9	16.0	10.7	10.1	11.3
	5.0	14.0	8.4	8.5	5.6	5.2	5.7
40	10.0	16.9	12.5	13.5	10.6	11.6	10.6
	5.0	10.2	6.8	6.9	5.3	5.8	5.3
50	10.0	15.0	10.9	12.8	10.0	10.1	10.0
	5.0	8.3	5.5	6.3	4.9	4.9	5.0

Table 4: Estimated sizes of the likelihood ratio, corrected likelihood ratio, score and three corrected score tests; $p=6$, $q=5$

Sample size	Nominal size	LR	LR^*	S_R	S_R^*	$\widetilde{S}_R$	$K(S_R)$
20	10.0	38.8	26.6	17.7	10.2	10.1	11.0
	5.0	30.0	17.8	9.9	5.4	5.3	5.8
30	10.0	29.2	19.4	17.0	10.8	11.5	11.1
	5.0	20.4	11.9	9.7	5.5	5.8	5.6
40	10.0	22.4	14.9	15.2	10.2	9.8	10.6
	5.0	14.4	8.2	8.1	5.1	4.8	5.2
50	10.0	17.0	12.1	12.5	10.2	10.9	10.2
	5.0	10.2	6.2	6.5	5.0	5.3	5.0

7. An Example

In modern portfolio theory, an important concept is the relative volatility of the return of a security compared to the return of the market. The regression coefficient of the return of the security (Y) on the return of the market (x) is an index of the systematic risk of the security. Quite often, the response variable that represents the return of a security displays heteroscedasticity. We consider the data given in Simonoff and Tsai (1994). These data represent the monthly excess returns over the riskless rate for the Acme Cleveland Corporation (Y) and for the market (x) for the period January 1986 - December 1990.

In the original analysis of these data, they suggested a heteroscedastic normal regression model $\mathrm{N}(\mu_l, \sigma_l^2)$ to predict Y_l in terms of a linear function of the return of the market x_l by considering two systematic components for the mean μ_l and the precision parameter ($\phi_l = \sigma_l^{-2}$): $\mu_l = \beta_2 + \beta_1 x_l$ and $-\log\phi_l = \gamma_2 + \gamma_1 x_l$, for $l = 1, \ldots, 59$, where the unusual observation 22 corresponding to October 1987 has been removed from the analysis. We are interested in testing the null hypothesis $H_0^3 : \gamma_1 = 0$ of homoscedasticity.

Under the alternative hypothesis $H_1^3 : \gamma_1 \neq 0$ of heteroscedasticity, we obtain $\widehat{\beta}_1 = 1.253$, $\widehat{\beta}_2 = -0.005$, $\widehat{\gamma}_1 = 8.092$ and $\widehat{\gamma}_2 = -4.410$. Under H_0^3, the restricted estimates are $\widetilde{\beta}_1 = 1.172$,

$\tilde{\beta}_2 = -0.010$ and $\tilde{\gamma}_2 = -4.739$. The large value of $\hat{\gamma}_2$ indicates potential, but not definite, heteroscedasticity. The likelihood ratio and score statistics are $LR = 7.33$ and $S_R = 2.69$ on one degree of freedom yielding p-values of 6.8% and 10.1%, respectively. For the test of $H_0^3 : \gamma_1 = 0$ of homoscedasticity in the normal model, equations (22)-(24) give $A_1 = 6(3\gamma_{1x}^2 - 4\gamma_{2x} - 8)/n$, $A_2 = 12(-2\gamma_{1x}^2 + 3\gamma_{2x} + 2)/n$ and $A_3 = 40\gamma_{1x}^2/n$, where $\gamma_{1x} = \delta_{30}/\delta_{20}^{3/2}$ and $\gamma_{2x} = \delta_{40}/\delta_{20}^2$, with the δ's defined in Section 5.3, are the sample measures of skewness and kurtosis of x. Thus, the corrected score statistics are $S_R^* = \tilde{S}_R = 2.89$ and $K(S_R) = 2.88$. The p-values for all the modified score statistics are now reduced to 8.91%. Using Botter and Cordeiro's (1997) results, we obtain $LR^* = 3.12$ yielding a p-value of 7.73%. The absolute difference between the p-values for the corrected likelihood ratio and score statistics is smaller than that for the unmodified statistics. Moreover, the reduction in the p-values for the corrected score statistics is in agreement with the conclusion expected based on the 10% nominal size of both modified and unmodified likelihood ratio tests. Then, the Bartlett-type correction for the score test gives further evidence to reject the null hypothesis of the homoscedasticity at least at a 10% nominal level.

Appendix: Joint cumulants of log-likelihood derivatives in GLMs with dispersion covariates

Differentiating the log-likelihood $\ell(\beta, \gamma)$ and using some regularity conditions and the orthogonality between β and γ, we can obtain the following third-order cumulants

$$\kappa_{rst} = -\sum_l \phi_l(f_l + 2g_l)x_{lr}x_{ls}x_{lt}, \quad \kappa_{rs,t} = \sum_l \phi_l g_l x_{lr}x_{ls}x_{lt}, \quad \kappa_{r,s,t} = \sum_l \phi_l(f_l - g_l)x_{lr}x_{ls}x_{lt},$$

$$\kappa_{Rs,t} = \sum_l \phi_{1l} w_l x_{ls}x_{lt}s_{lR}, \quad \kappa_{R,s,t} = -\sum_l \phi_{1l} w_l x_{ls}x_{lt}s_{lR}, \kappa_{RS,T} = -\sum_l d_{2l}\phi_{1l}\phi_{2l}s_{lR}s_{lS}s_{lT},$$

$$\kappa_{r,S,T} = \kappa_{R,st} = \kappa_{rS,T} = \kappa_{RS,t} = 0, \quad \kappa_{R,S,T} = -\sum_l d_{3l}\phi_{1l}^3 s_{lR}s_{lS}s_{lT}.$$

The fourth-order cumulants are given by

$$\begin{aligned}\kappa_{rstu} &= \sum_l \phi_l \left\{ -\frac{6}{V^3}\left(\frac{dV}{d\mu}\right)^2 \left(\frac{d\mu}{d\eta}\right)^4 + \frac{3}{V^2}\frac{d^2V}{d\mu^2}\left(\frac{d\mu}{d\eta}\right)^4 \right. \\ &+ \left. \frac{12}{V^2}\frac{dV}{d\eta}\left(\frac{d\mu}{d\eta}\right)^2 \frac{d^2\mu}{d\eta^2} - \frac{3}{V}\left(\frac{d^2\mu}{d\eta^2}\right)^2 - \frac{4}{V}\frac{d\mu}{d\eta}\frac{d^3\mu}{d\eta^3} \right\}_l x_{lr}x_{ls}x_{lt}x_{lu},\end{aligned}$$

$$\begin{aligned}\kappa_{rs,tu} &= \sum_l \phi_l \left\{ \frac{1}{V^3}\left(\frac{dV}{d\mu}\right)^2 \left(\frac{d\mu}{d\eta}\right)^4 - \frac{2}{V^2}\frac{dV}{d\mu}\left(\frac{d\mu}{d\eta}\right)^2 \frac{d^2\mu}{d\eta^2} \right. \\ &+ \left. \frac{1}{V}\left(\frac{d^2\mu}{d\eta^2}\right)^2 \right\}_l x_{lr}x_{ls}x_{lt}x_{lu},\end{aligned}$$

$$\kappa_{r,s,tu} = \sum_l \phi_l \left\{ \frac{1}{V^2}\frac{\mathrm{d}V}{\mathrm{d}\mu}\left(\frac{\mathrm{d}\mu}{\mathrm{d}\eta}\right)^2 \frac{\mathrm{d}^2\mu}{\mathrm{d}\eta^2} - \frac{1}{V^3}\left(\frac{\mathrm{d}V}{\mathrm{d}\mu}\right)^2 \left(\frac{\mathrm{d}\mu}{\mathrm{d}\eta}\right)^4 \right\}_l x_{lr}x_{ls}x_{lt}x_{lu},$$

$$\kappa_{r,s,t,u} = \sum_l \phi_l \left\{ \frac{1}{V^3}\left(\frac{\mathrm{d}V}{\mathrm{d}\mu}\right)^2 \left(\frac{\mathrm{d}\mu}{\mathrm{d}\eta}\right)^4 + \frac{1}{V^2}\frac{\mathrm{d}^2 V}{\mathrm{d}\mu^2}\left(\frac{\mathrm{d}\mu}{\mathrm{d}\eta}\right)^4 \right\}_l x_{lr}x_{ls}x_{lt}x_{lu},$$

$$\kappa_{R,S,tu} = \kappa_{RS,tu} = 0, \quad \kappa_{Rt,Su} = \sum_l \frac{\phi_{1l}^2 w_l}{\phi_l} x_{lt}x_{lu}s_{lR}s_{lS}, \quad \kappa_{r,s,TU} = -\sum_l \phi_{2l} w_l x_{lr} x_{ls} s_{lT} s_{lU},$$

$$\kappa_{R,S,TU} = -\sum_l d_{3l}\phi_{1l}^2\phi_{2l}s_{lR}s_{lS}s_{lT}s_{lU}, \quad \kappa_{R,S,t,u} = 2\sum_l \frac{\phi_{1l}^2 w_l}{\phi_l} x_{lt}x_{lu}s_{lR}s_{lS},$$

$$\kappa_{RS,TU} = -\sum_l \phi_{2l}^2 d_{2l} s_{lR}s_{lS}s_{lT}s_{lU}, \quad \kappa_{R,S,T,U} = -\sum_l d_{4l}\phi_{1l}^4 s_{lR}s_{lS}s_{lT}s_{lU}.$$

REFERENCES

Botter, D.A. and Cordeiro, G.M. (1997). Bartlett corrections for generalized linear models with dispersion covariates. *Communications in Statistics, Theory and Methods*, **26**, 279–307.

Cordeiro, G.M. (1993). Bartlett corrections and bias correction for two heteroscedastic regression models. *Communications in Statistics, Theory and Methods*, **22**, 169–188.

Cordeiro, G.M. and Demétrio, C.G.B. (1989). Estimation and tests in a quasi–likelihood model with a non–constant dispersion parameter. *Lecture Notes in Statistics*, **57**, 95–104. Eds. A. Decarli, B.J. Francis, R. Gilchrist and G.U.H. Seeber. *Proceedings of GLIM 89 and the 4th Workshop on Statistical Modelling*. Springer–Verlag, New York.

Cordeiro, G.M. and Ferrari, S.L.P. (1991). A modified score test statistic having chi-squared distribution to order n^{-1}. *Biometrika*, **78**, 573–582.

Cordeiro, G.M. and Ferrari, S.L.P. (1996). Bartlett-type corrections for some score tests in proper dispersion models. *Communications in Statistics – Theory and Methods*, **25**, 29–48.

Cordeiro, G.M. and Ferrari, S.L.P. (1998). A note on Bartlett-type corrections for the first few moments of test statistics. *Journal of Statistical Planning and Inference*, **71**, 261–269.

Cordeiro, G.M., Ferrari, S.L.P. and Cysneiros, A.H.M.A. (1998). A formula to improve score test statistics. *Journal of Statistical Computation and Simulation*, **62**, 123–136.

Cordeiro, G.M., Ferrari, S.L.P. and Paula, G.A. (1993). Improved score tests for generalized linear models. *Journal of the Royal Statistical Society* B, **55**, 661–674.

Cordeiro, G.M. and Paula, G.A. (1989). Fitting non–exponential family nonlinear models in GLIM by using the offset facility. *Lecture Notes in Statistics*, **57**, 105-114. Eds. A. Decarli, B. J. Francis, *Proceedings of GLIM 89 and the 4th Workshop on Statistical Modelling*. Springer–Verlag, New York.

Cox, D.R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society B*, **49**, 1–39.

Cox, D.R. and Hinkley, D.V. (1974). *Theoretical Statistics*. 2nd ed. Chapman and Hall, London.

Cribari–Neto, F. and G.M. Cordeiro (1996). On Bartlett and Bartlett-type corrections. *Econometric Reviews*, **15**, 339–367.

Cribari–Neto, F. and Ferrari, S.L.P. (1995). Second order asymptotics for score tests in generalised linear models. *Biometrika*, **82**, 426–432.

Doornik, J.A. (1999). *Object-oriented Matrix Programming Using Ox*, 3rd ed. Timberlake Consultants, London, and `http://www.nuff.ox.ac.uk/Users/Doornik`.

Ferrari, S.L.P. and Cordeiro, G.M. (1996). Corrected score tests for exponential family nonlinear models. *Statistics and Probability Letters*, **26**, 7–12.

Ferrari, S.L.P., Cordeiro, G.M. and Cribari–Neto, F. (2001). Higher order asymptotic refinements for score tests in proper dispersion models. *Journal of Statistical Planning and Inference*, **97**, 177-190.

Ferrari, S.L.P., Uribe–Opazo, M.A. and Cribari–Neto, F. (1997). Second order asymptotics for score tests in exponential family nonlinear models. *Journal of Statistical Computation and Simulation*, **59**, 179–194.

Harris, P. (1985). An asymptotic expansion for the null distribution of the efficient score statistic. *Biometrika*, **72**, 653–659.

Kakizawa, Y. (1996). Higher order monotone Bartlett-type adjustment for some multivariate test statistics. *Biometrika*, **83**, 923–927.

Lawley, D.N. (1956). A generalized method for approximating to the distribution of the likelihood ratio criteria. *Biometrika*, **43**, 295–303.

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

Pregibon, D. (1984). Review of generalized linear models. *The Annals of Statistics*, **12**, 1589–1596.

Rao, C.R. (1947). Large sample tests on statistical hypotheses concerning several parameters with applications to problems of estimation. *Proccedings of the Cambridge Philosophical Society*, **44**, 50–57.

Simonoff, J.S. and Tsai, C.–H. (1994). Use of modified profile likelihood for improved tests of constancy of variance in regression. *Applied Statistics*, **43**, 357–370.

ÚLTIMOS RELATÓRIOS TÉCNICOS PUBLICADOS

2001-01 - KOTTAS, A., BRANCO, M.D., GELFAND, A.E., A Nonparametric bayesian modeling approach for cytogenetic dosimetry. 2001. 19p. (RT-MAE-2001-01)

2001-02 - AOKI, R., BOLFARINE, H., SINGER, J.M., Null Intercept measurement error regression models. 2001. 18p.(RT-MAE-2001-02)

2001-03 - AOKI, R., BOLFARINE, H., SINGER, J.M., Asymptotic Efficiency of null intercept measurement error regression models. 2001. 11p. (RT-MAE-2001-03).

2001-04 - ALMEIDA, S.S., LIMA, C.R.O.P., SANDOVAL, M.C., Linear Calibration in functional models without the normality assumption. 2001. 18p. (RT-MAE-2001-04).

2001-05 - GARCIA-ALFARO, K.E., BOLFARINE, H. Comparative calibration with subgroups. 2001. 18p. (RT-MAE-2001-05)

2001-06 - DUNLOP, F.M., FERRARI, P.A., FONTES, L.R.G. A dynamic one-dimensional interface interacting with a wall. 2001. 21p. (RT-MAE-2001-06).

2001-07 - FONTES, L.R., ISOPI, M., NEWMAN, C.M., STEIN, D.L. 1*D* Aging. 2001. 10p. (RT-MAE-2001-07).

2001-08 - BARROS, S.R.M., FERRARI, P.A., GARCIA, N.L., MARTÍNEZ, S. Asymptotic behavior of a stationary silo with absorbing walls. 2001. 20p. (RT-MAE-2001-08).

2001-09 - GONZALEZ-LOPEZ, V.A., TANAKA, N.Y. Characterization of copula and its relationships with TP_2 (RR_2) association. 2001. 24p. (RT-MAE-2001-09).

2001-10 - ORTEGA, E.M.M., BOLFARINE, H., PAULA, G.A. Influence Diagnostics in Generalized Log-Gamma Regression Models. 2001. 24p. (RT-MAE-2001-10).

2001-11 - M.D, BRANCO, BOLFARINE, H., IGLESIAS, P., ARELLANO-VALLE, R.B. Bayesian and classical solutions for binomial cytogenetic dosimetry problem. 2001. 16p. (RT-MAE-2001-11).

2001-12 - GONÇALEZ-LOPES, V.A., TANAKA, N.I. Dependence structures and a posteriori distributions. 2001. 41p. (RT-MAE-2001-12).

2001-13 - MUTAFCHIEV, L.; KOLEV, N. The Number of Empty Cells in an Allocation Scheme Generated by a Zero-Inflated Distribution: Exact Results and Poisson Convergence. 2001. 11p. (RT-MAE-2001-13).

The complete list of "Relatórios do Departamento de Estatística", IME-USP, will be sent upon request.

Departamento de Estatística
IME-USP
Caixa Postal 66.281
05315-970 - São Paulo, Brasil