Brazilian Journal of Probability and Statistics
2021, Vol. 35, No. 1, 37–49
https://doi.org/10.1214/20-BJPS489

A general expression for second-order covariance matrices—an application to dispersion models

Tiago M. Magalhães[a], **Denise A. Botter**[b] and **Mônica C. Sandoval**[b]

[a]*Federal University of Juiz de Fora*
[b]*University of São Paulo*

Abstract. We present a general expression that allows the calculation of both the n^{-2} asymptotic covariance matrices of the maximum likelihood estimator (MLE) and the first-order bias corrected MLE, where n is the sample size. The formula is presented in a matrix notation which has numerical advantages since it requires only simple operations on matrices and vectors. The usefulness of the formula is to construct better Wald statistics. We apply our findings to dispersion models and develop simulation studies which show that modification in the Wald statistic effectively removes size distortions of the type I error probability with no power loss. For illustrative purposes, a real data application is considered to support our theoretical results.

1 Introduction

Regression models attempt to explain the behavior of a variable of interest (or response) from covariables (explanatory variables). In general, a function called a link function, links a characteristic of the response variable, usually the mean, to the explanatory variables through parameters to be estimated from observed data.

In general, in the frequentist context, the maximum likelihood method is used to estimate the parameters of the regression models. The inferences depend strongly on asymptotic properties of the maximum likelihood estimators. Among these properties, we have that the MLE is unbiased and follows a normal distribution.

However, likelihood inferences based on an asymptotic approach may not be reliable, when sample sizes are small or moderate. In this context, several papers have been developed to improve the likelihood of inference procedures. Cox and Snell (1968) obtained a general expression for the first-order bias of MLE, which allows us to obtain MLE with reduced bias (BCE). Shenton and Bowman (1977) and Peers and Iqbal (1985) derived a general formula for the MLE second-order covariance matrix. Based on Peers and Iqbal (1985) several studies have been published, for example, Cordeiro, Barroso and Botter (2006) for generalized linear models (GLM), Cordeiro and Santana (2008) for exponential family nonlinear models, Rocha, Simas and Cordeiro (2010) for dispersion models, Lemonte (2011) and Lemonte (2020) for Birnbaum–Saunders and censored exponential regression models, respectively and Barroso, Botter and Cordeiro (2013) for heteroskedastic GLM. Until now, Cordeiro et al. (2014) is the only work obtaining a BCE second-order covariance matrix. However, Magalhães, Botter and Sandoval (2017) showed that the second-order covariance matrix of the MLE expression presented in Peers and Iqbal (1985) was incorrect and that the correct expression is the one in Shenton and Bowman (1977).

Hypotheses testing is an essential step in statistical inference in order to help investigators identify and understand the effect of covariates on the response variable. The main hypotheses tests are the likelihood ratio test (LR; Wilks, 1938), the score test (SR; Rao, 1948) and the Wald test (Wald, 1943). Under regularity conditions, these statistics have asymptotically

Key words and phrases. Bias estimator, covariance matrix, dispersion models, Wald test.
Received October 2019; accepted September 2020.

a chi-squared distribution with the number of degrees of freedom defined by the number of constraints imposed by the null hypothesis. However, the chi-squared distribution may not be a good approximation to the distribution of these statistics in small or moderate sample sizes. As an alternative for small sample sizes, Bartlett or Bartlett-type corrections were suggested for the LR and SR statistics, making the chi-squared distribution approximation more reliable. In a general setting, there is no Bartlett-type correction factor to improve the Wald test. Cordeiro et al. (2014) proposed a modified version of the Wald statistic for the GLM, replacing the MLE by the BCE and the inverse of the information matrix by the second-order covariance matrix of the bias-corrected maximum likelihood estimator.

Based on Shenton and Bowman (1977) and the ideas of Cordeiro et al. (2014), the chief goal in this paper is to obtain a single general expression for calculating both the MLE and BCE second-order covariance matrices. An important result is that our expression is written in matrix notation instead of tensorial notation. Our matrix formulation has numerical advantages since it requires only simple operations on matrices and vectors. This result allows an easy way to simultaneously obtain the MLE and BCE second-order covariance matrices in any class of regression models. Recently, Kosmidis, Kenne Pagui and Sartori (2020) presented an unified method to obtain bias reduced estimators for GLM, thus, the idea of integrating methods is being well received in the literature.

All results are applied to dispersion models (DM; Jørgensen, 1997b). The DM is an extension of the well-known generalized linear models. It is a wide class of models that allows new choices for the distribution of the response variable. Circular and unit intervals regression models are integrated with the DM. Some authors have attempted to develop a second-order asymptotic theory in dispersion models. For example, Simas, Cordeiro and Rocha (2010) and Simas, Rocha and Barreto-Souza (2011) obtained a matrix expression of order $n^{-1/2}$ for the skewness coefficient of the distribution of the MLE and the bias-corrected estimators, respectively. Lemonte and Ferrari (2012) derived asymptotic expansions for the nonnull distribution functions of the likelihood ratio, Wald, score and gradient test statistics. Cordeiro, Paula and Botter (1994), Cordeiro and Ferrari (1996) and Medeiros, Ferrari and Lemonte (2017) presented, respectively, the Bartlett and the Bartlett-type correction factor to the likelihood ratio, score and gradient statistics. Therefore, the DM has been widely studied and efforts to improve the inference are of great importance.

The paper unfolds as follows. In Section 2, we present in matrix notation a single general expression for the MLE and BCE second-order covariance matrices. In Section 3, we propose two Wald test statistics based on the matrix presented in Section 2. In Sections 4 and 5, we describe the dispersion models and obtain a general formula for the second-order covariance matrices of the MLE and BCE in DM, respectively. Monte Carlo simulation results are presented and discussed in Section 6. An empirical application that uses real data is presented and discussed in Section 7. This paper concludes with a brief discussion in Section 8.

2 General expression for second-order covariance matrices

Let $Y_1, \ldots, Y_n$ be n random variables with Y_i having a probability density function that satisfies the usual regularity conditions for large sample inference based on likelihood estimation (Cox and Hinkley, 1974). Let $\ell(\boldsymbol{\theta})$ be the log-likelihood function for an unknown p-vector parameter $\boldsymbol{\theta}$. The $\ell(\boldsymbol{\theta})$ derivatives concerning components $a, b, c, \ldots$ of $\boldsymbol{\theta}$ are denoted by

$$U_a = \partial l(\boldsymbol{\theta})/\partial\theta_a, \qquad U_{ab} = \partial^2 l(\boldsymbol{\theta})/\partial\theta_a\partial\theta_b, \qquad U_{abc} = \partial^3 l(\boldsymbol{\theta})/\partial\theta_a\partial\theta_b\partial\theta_c, \qquad \ldots$$

We use the notation introduced by Lawley (1956) to define the joint cumulants and their derivatives of $\ell(\boldsymbol{\theta})$:

$$\kappa_{ab} = \mathbb{E}(U_{ab}), \qquad \kappa_{abc} = \mathbb{E}(U_{abc}), \qquad \kappa_{a,bc} = \mathbb{E}(U_a U_{bc}), \qquad \kappa_{bc}^{(a)} = \partial\kappa_{bc}/\partial\theta_a.$$

All κ's refer to a total of the sample and are, in general, of n order. The Fisher information matrix, $\boldsymbol{K} = \boldsymbol{K}(\boldsymbol{\theta})$, has elements $\kappa_{a,b} = -\kappa_{ab}$. Also consider $\kappa^{a,b} = -\kappa^{ab}$ as the corresponding elements of its inverse, $\boldsymbol{K}^{-1} = \boldsymbol{K}^{-1}(\boldsymbol{\theta})$.

Let $\widehat{\boldsymbol{\theta}}$ be the maximum likelihood estimator of $\boldsymbol{\theta}$ and $\widetilde{\boldsymbol{\theta}}$ the MLE of $\boldsymbol{\theta}$ corrected by the bias up to order n^{-1}.

From Shenton and Bowman (1977) and Cordeiro et al. (2014) and following the ideas of Cordeiro (1993), Cordeiro and Klein (1994) and Patriota and Cordeiro (2011), we propose a single general matrix expression for the MLE and BCE second-order covariance matrices given by

$$\mathbf{Cov_2^{\boldsymbol{\tau}}}(\boldsymbol{\theta}^\star) = \boldsymbol{K}^{-1} + \boldsymbol{K}^{-1}\{\boldsymbol{\Delta} + \boldsymbol{\Delta}^\top\}\boldsymbol{K}^{-1} + \mathcal{O}(n^{-3}), \tag{2.1}$$

where the (a, b)th element, $a, b = 1, \ldots, p$, of the $\boldsymbol{\Delta}$ matrix, not necessarily a symmetric matrix, is given by δ_{ab}, with $\delta_{ab} = -\frac{1}{2}\delta_{ab}^{(1)} + \frac{1}{4}\delta_{ab}^{(2)} + \frac{\tau_2}{2}\delta_{ab}^{(3)}$, where

$$\delta_{ab}^{(1)} = \sum_{r,s=1}^{p} \kappa^{rs}\{\tau_1[2\kappa_{br}^{(as)} - \kappa_{brs}^{(a)}] + \kappa_{ar,bs}\},$$

$$\delta_{ab}^{(2)} = \sum_{r,s,t,u=1}^{p} \kappa^{rs}\kappa^{tu}\{\kappa_{aru}[3\kappa_{bst} + 2\kappa_{b,st} + 8\kappa_{bs,t}] + 2\kappa_{ar,u}[2\kappa_{b,st} + \kappa_{bt,s}]\},$$

$$\delta_{ab}^{(3)} = \sum_{r,s,t,u=1}^{p} \kappa^{rs}\kappa^{tu}\{\kappa_{bu}^{(a)}(\kappa_{st}^{(r)} + \kappa_{r,st})\},$$

with $\boldsymbol{\tau} = (\tau_1, \tau_2) = (1, 1)$ indicating the second-order covariance matrix of the MLE $\boldsymbol{\theta}^\star = \widehat{\boldsymbol{\theta}}$ denoted by $\mathbf{Cov_2}(\widehat{\boldsymbol{\theta}})$ and $\boldsymbol{\tau} = (0, -1)$ indicating the second-order covariance matrix of the BCE $\boldsymbol{\theta}^\star = \widetilde{\boldsymbol{\theta}}$ denoted by $\mathbf{Cov_2}(\widetilde{\boldsymbol{\theta}})$. The proof is presented in the Supplementary Material (see Magalhães, Botter and Sandoval, 2020). Note that for the second-order covariance matrix of the bias corrected maximum likelihood estimator, we do not need to derive the log-likelihood function of $\boldsymbol{\theta}$ four times and there are fewer cumulants to compute, which makes this matrix easier to calculate.

3 Wald test

One of the most used hypotheses tests for practitioners is the Wald test. However, for small sample sizes the conclusion based on this test can be unreliable, as well as for the LR and score tests. In a general setting, there is no Bartlett or Bartlett-type correction factor to improve the Wald test. Here, we describe a modification in the Wald statistic proposed by Cordeiro et al. (2014) which showed good performance when it was applied to GLM.

Consider the partition $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$, $\boldsymbol{\theta}_1$ being a q-dimensional vector and $\boldsymbol{\theta}_2$ containing the remaining $p - q$ parameters. We have an interest in testing the composite null hypothesis $\mathcal{H} : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^{(0)}$ against a composite alternative hypothesis $\mathcal{A}$: $\mathcal{H}$ is false, where $\boldsymbol{\theta}_1^{(0)}$ is a specified vector. This partition induces the corresponding partitions

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

The classic Wald test statistic is

$$W_0 = (\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^{(0)})^\top \{\widehat{\boldsymbol{K}}^{11}\}^{-1}(\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^{(0)}), \tag{3.1}$$

where $\widehat{\boldsymbol{K}}^{11}$ is the matrix $\boldsymbol{K}^{11}$ evaluated at $\widehat{\boldsymbol{\theta}}$. Under $\mathcal{H}$, W_0 has a χ_q^2.

In the first modification, we replaced $\widehat{\boldsymbol{K}}^{11}$ by the second-order covariance matrix $\mathbf{Cov_2^{11}}(\widehat{\boldsymbol{\theta}})$, obtained from

$$\mathbf{Cov_2}(\widehat{\boldsymbol{\theta}}) = \begin{pmatrix} \mathbf{Cov_2}(\widehat{\boldsymbol{\theta}})_{11} & \mathbf{Cov_2}(\widehat{\boldsymbol{\theta}})_{12} \\ \mathbf{Cov_2}(\widehat{\boldsymbol{\theta}})_{21} & \mathbf{Cov_2}(\widehat{\boldsymbol{\theta}})_{22} \end{pmatrix}$$

and

$$\mathbf{Cov_2^{-1}}(\widehat{\boldsymbol{\theta}}) = \begin{pmatrix} \mathbf{Cov_2^{11}}(\widehat{\boldsymbol{\theta}}) & \mathbf{Cov_2^{12}}(\widehat{\boldsymbol{\theta}}) \\ \mathbf{Cov_2^{21}}(\widehat{\boldsymbol{\theta}}) & \mathbf{Cov_2^{22}}(\widehat{\boldsymbol{\theta}}) \end{pmatrix},$$

which implies

$$W_1 = \big(\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^{(0)}\big)^\top \big\{\widehat{\mathbf{Cov}}\mathbf{_2^{11}}(\widehat{\boldsymbol{\theta}})\big\}^{-1} \big(\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^{(0)}\big), \tag{3.2}$$

where $\widehat{\mathbf{Cov}}\mathbf{_2^{11}}(\widehat{\boldsymbol{\theta}})$ is the matrix $\mathbf{Cov_2^{11}}(\widehat{\boldsymbol{\theta}})$ evaluated at $\widehat{\boldsymbol{\theta}}$.

The second modification is made when we replace $\widehat{\boldsymbol{\theta}}_1$ by $\widetilde{\boldsymbol{\theta}}_1$ and $\widehat{\boldsymbol{K}}^{11}$ by the second-order covariance matrix of the BCE, $\mathbf{Cov_2^{11}}(\widetilde{\boldsymbol{\theta}})$, obtained from partitions

$$\mathbf{Cov_2}(\widetilde{\boldsymbol{\theta}}) = \begin{pmatrix} \mathbf{Cov_2}(\widetilde{\boldsymbol{\theta}})_{11} & \mathbf{Cov_2}(\widetilde{\boldsymbol{\theta}})_{12} \\ \mathbf{Cov_2}(\widetilde{\boldsymbol{\theta}})_{21} & \mathbf{Cov_2}(\widetilde{\boldsymbol{\theta}})_{22} \end{pmatrix}$$

and

$$\mathbf{Cov_2^{-1}}(\widetilde{\boldsymbol{\theta}}) = \begin{pmatrix} \mathbf{Cov_2^{11}}(\widetilde{\boldsymbol{\theta}}) & \mathbf{Cov_2^{12}}(\widetilde{\boldsymbol{\theta}}) \\ \mathbf{Cov_2^{21}}(\widetilde{\boldsymbol{\theta}}) & \mathbf{Cov_2^{22}}(\widetilde{\boldsymbol{\theta}}) \end{pmatrix},$$

resulting in

$$W_2 = \big(\widetilde{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^{(0)}\big)^\top \big\{\widetilde{\mathbf{Cov}}\mathbf{_2^{11}}(\widetilde{\boldsymbol{\theta}})\big\}^{-1} \big(\widetilde{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^{(0)}\big), \tag{3.3}$$

where $\widetilde{\mathbf{Cov}}\mathbf{_2^{11}}(\widetilde{\boldsymbol{\theta}})$ is the matrix $\mathbf{Cov_2^{11}}(\widetilde{\boldsymbol{\theta}})$ evaluated at $\widetilde{\boldsymbol{\theta}}$.

Unlike the Bartlett and Bartlett-type corrections, the statistics (3.2) and (3.3) do not change the convergence order of the Wald test statistic.

4 Dispersion models

A DM is defined as follows. Consider $Y_1, \dots, Y_n$ independent random variables with a probability density function (or probability function) of the form (Jørgensen, 1997b):

$$\pi(y_i; \mu_i, \phi) = \exp\big\{\phi t(y_i, \mu_i) + a(y_i, \phi)\big\}, \quad y_i \in C, \tag{4.1}$$

where C is a convex support, μ_i varies in a subset of C, $\phi > 0$, $t(\cdot,\cdot)$ and $a(\cdot,\cdot)$ are known functions. Sometimes it is convenient to write $t(y_i, \mu_i) = -\mathcal{D}(y_i, \mu_i)/2$, where $\mathcal{D}(y_i, \mu_i)$ is the unit deviance. If Y_i is continuous, $\pi(\cdot)$ is assumed to be a density concerning Lebesgue measure, while if Y_i is discrete, $\pi(\cdot)$ is assumed to be a density with respect to the counting measure. We call ϕ the precision parameter and the inverse of ϕ the dispersion parameter. The parameter μ_i may be interpreted as a location parameter. In some cases, it may be the expectation of the distribution.

If $t(y_i, \mu_i) = \theta_i y_i - b(\theta_i)$, where $\mu_i = b'(\theta_i)$, (4.1) is a probability density of an exponential dispersion models (EDM; Jørgensen, 1987). If $a(y_i, \phi)$ in (4.1) can be rewritten in the form $a(y_i, \phi) = a_1(\phi) + a_2(y_i)$, where $a_1(\cdot)$ and $a_2(\cdot)$ are suitable functions, (4.1) is a probability density of a proper dispersion models (PDM; Jørgensen, 1997a). Examples of EDM are the normal, gamma, inverse Gaussian and Poisson distributions. The normal, gamma, inverse Gaussian, reciprocal gamma, simplex and von Mises distributions are examples of PDM.

We define the DM regression by the random component (4.1) and the systematic component $g(\mu_i) = \eta_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}$, where $g(\cdot)$ is a known one-to-one differentiable link function, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$, $p < n$, is a vector of an unknown parameter to be estimated and $\boldsymbol{x}_i$ is a $p \times 1$ vector of known explanatory variables associated with the ith observable response. Note that the generalized linear models are particular cases of dispersion models.

The log-likelihood function for $\boldsymbol{\beta}$ and ϕ, denoted by $\ell(\boldsymbol{\beta}, \phi)$, is given by

$$\ell(\boldsymbol{\beta}, \phi) = \sum_{i=1}^{n} \{\phi t(y_i, \mu_i) + a(y_i, \phi)\}. \tag{4.2}$$

The score function obtained by the differentiation of (4.2) with respect to $\boldsymbol{\beta}$ is given by

$$\boldsymbol{U}_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \phi) = \phi \boldsymbol{X}^\top \boldsymbol{t}'(\boldsymbol{y}, \boldsymbol{\mu}),$$

where $\boldsymbol{y} = (y_1, \dots, y_n)^\top$, $\boldsymbol{X} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_n)^\top$ is a specified $n \times p$ matrix of full rank $p < n$, $\boldsymbol{t}'(\boldsymbol{y}, \boldsymbol{\mu}) = \partial \boldsymbol{t}(\boldsymbol{y}, \boldsymbol{\mu})/\partial \boldsymbol{\mu}$ and $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^\top$. The MLE of $\boldsymbol{\beta}$ is obtained solving $\boldsymbol{U}_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \phi) = \boldsymbol{0}$.

The MLE of ϕ can be obtained solving

$$\sum_{i=1}^{n} a'(\phi, y_i) = -\sum_{i=1}^{n} t(y_i, \mu_i),$$

where $a'(\phi, y_i) = \partial a(\phi, y_i)/\partial \phi$.

We define $d_{ri} = d_r(\mu_i, \phi) = \mathbb{E}[\partial^r t(Y_i, \mu_i)/\partial \mu_i^r]$ for $r = 1, 2, 3$. Under the usual regularity conditions, we have $d_{1i} = 0$ and $d_{2i} = -\phi \mathbb{E}\{[\partial t(Y_i, \mu_i)/\partial \mu_i]^2\}$. The Fisher's information matrix for $\boldsymbol{\beta}$ has the form

$$\boldsymbol{K} = \boldsymbol{K}(\boldsymbol{\beta}) = \mathbb{E}\left[-\frac{\partial^2 \ell(\boldsymbol{\beta}, \phi)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top}\right] = \phi \boldsymbol{X}^\top \boldsymbol{W} \boldsymbol{X}, \tag{4.3}$$

where $\boldsymbol{W} = \mathrm{diag}\{w_1, \dots, w_n\}$ and $w_i = -d_{2i}(d\mu_i/d\eta_i)^2$. The MLEs of $\boldsymbol{\beta}$ and ϕ are asymptotically independent.

5 n^{-2} covariance matrix for $\boldsymbol{\beta}$ in a DM regression

The expression (2.1) is very general and in some cases, it is difficult or even impossible to obtain it for specific regression models. In the following, we will apply the results in (2.1) for dispersion models.

In order to express (2.1) for $\boldsymbol{\beta}$ in a DM regression considering the parameter ϕ is fixed, it is helpful to define the following matrices: $\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{K}^{-1}\boldsymbol{X}^\top$, $\boldsymbol{Z}_d = \mathrm{diag}\{z_{11}, \dots, z_{nn}\}$, $\boldsymbol{Z}^{(2)} = \boldsymbol{Z} \odot \boldsymbol{Z}$, with $\odot$ representing a direct product, $\boldsymbol{C} = \mathrm{diag}\{\boldsymbol{Z}(2\boldsymbol{F}_1 - \boldsymbol{F}_2 + 2\boldsymbol{F}_3)\boldsymbol{Z}_d\boldsymbol{1}\}$, $\boldsymbol{F}_j = \mathrm{diag}\{f_{j1}, \dots, f_{jn}\}$, $\boldsymbol{G}_j = \mathrm{diag}\{g_{j1}, \dots, g_{jn}\}$, for sake of brevity, the quantities f_{ji} and g_{ji}, $i = 1, \dots, n$, $j = 1, 2, 3$ are presented in the Appendix and $\boldsymbol{1}$ is a n-dimensional vector of ones.

In dispersion models, the general expression for the second-order covariance matrices is (2.1), where $\boldsymbol{K}^{-1}$ is given by (4.3) and $\boldsymbol{\Delta} = -0.5\boldsymbol{\Delta}^{(1)} + 0.25\boldsymbol{\Delta}^{(2)} + 0.5\tau_2\boldsymbol{\Delta}^{(3)}$ with

$$\begin{aligned}
\Delta^{(1)} &= -\boldsymbol{X}^\top[\tau_1(2\boldsymbol{G}_1 - \boldsymbol{G}_3) + \boldsymbol{G}_2]\boldsymbol{Z}_d\boldsymbol{X}, \\
\Delta^{(2)} &= \boldsymbol{X}^\top[(\boldsymbol{F}_2 + 2\boldsymbol{F}_3)\boldsymbol{Z}^{(2)}(10\boldsymbol{F}_1 - 7\boldsymbol{F}_2 + 6\boldsymbol{F}_3) \\
&\quad + 6(\boldsymbol{F}_2 - \boldsymbol{F}_1)\boldsymbol{Z}^{(2)}(\boldsymbol{F}_2 - \boldsymbol{F}_1)]\boldsymbol{X}, \\
\Delta^{(3)} &= \boldsymbol{X}^\top(\boldsymbol{F}_1 + 2\boldsymbol{F}_3)\boldsymbol{C}\boldsymbol{X}.
\end{aligned} \tag{5.1}$$

Table 1 $\widehat{\boldsymbol{K}}^{-1}$, $\widehat{\mathbf{Cov}}_{\mathbf{2}}(\widehat{\boldsymbol{\beta}})$ *and* $\widehat{\boldsymbol{\beta}}$ *sample covariance matrix, for* $n = 15$ *and* 25

	$n=15$				$n=25$			
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
$\hat{\beta}_1$	1.958	−1.689	−1.316	−0.475	1.392	−0.979	−0.645	−0.844
	2.256	−1.966	−1.511	−0.528	1.538	−1.095	−0.707	−0.929
	3.676	−3.270	−2.334	−0.897	2.090	−1.465	−0.908	−1.248
$\hat{\beta}_2$		2.905	0.597	−0.528		1.708	0.125	0.006
		3.377	0.668	−0.605		1.917	0.131	−0.003
		5.805	0.934	−0.905		2.655	0.161	0.000
$\hat{\beta}_3$			2.232	−0.137			1.307	0.097
			2.567	−0.173			1.450	0.093
			4.282	−0.297			2.007	0.103
$\hat{\beta}_4$				1.943				1.390
				2.216				1.556
				3.613				2.145

Although the expression (2.1) entails a great deal of algebra, the final expression of the second-order covariance matrix for a DM regression has a very nice form only involving simple operations on diagonal matrices and can be easily implemented into statistical software such as R (R Core Team, 2017). Additionally, the expression (5.1) generalizes Cordeiro et al. (2014) and Rocha, Simas and Cordeiro (2010). The detailed derivation of the expression (5.1) is presented in the Supplementary Material (see Magalhães, Botter and Sandoval, 2020).

6 Simulation study

To examine the performance of $\widehat{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\beta}}$ second-order covariance matrices, we conducted three simulation studies. In all studies, we considered the reciprocal gamma distribution, a PDM model that is commonly used in survival analysis, that is,

$$\pi(y_i; \mu_i, \phi) = \frac{\phi^{\phi} e^{-\phi}}{y\Gamma(\phi)} \exp\left\{-\phi\left(\frac{\mu}{y} - \log\frac{\mu}{y} - 1\right)\right\}, \quad y > 0,$$

where $\mu > 0$ and $\phi > 0$. We assume,

$$\sqrt{\mu}_i = \eta_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i}, \quad i = 1, \ldots, n.$$

The true values of the parameters were fixed as $\beta_1 = 3$, $\beta_2 = 1.5$, $\beta_3 = 2$, $\beta_4 = 0$ and $\phi = 1.5$. The covariates x_2, x_3, x_4 were obtained from a uniform distribution in the interval (0, 1) for each $n (n = 15, 25, 35, 45)$ and were held constant in all 10,000 simulations. All simulations are performed using the R software (R Core Team, 2017).

Up to this moment, we considered ϕ as a fixed value and derived the second-order covariance matrix for $\boldsymbol{\beta}$. However, in practice ϕ is unknown and we also need to estimate it. We adopt the alternative of taking $\phi = \hat{\phi}$ or $\phi = \tilde{\phi}$ in order to apply the discussed methodology, where $\hat{\phi}$ and $\tilde{\phi}$ denotes, respectively the MLE and the BCE (Simas, Rocha and Barreto-Souza, 2011) of ϕ.

In the first simulation study, we compared the covariance matrices of order n^{-1} and n^{-2} of the MLE $\widehat{\boldsymbol{\beta}}$ with the sample covariance matrix of $\widehat{\boldsymbol{\beta}}$. The results of this comparative analysis are shown in Table 1, for $n = 15$ and 25, and Table 2, for $n = 35$ and 45. In these tables, the first and second entries are the sample means of $\widehat{\boldsymbol{K}}^{-1}$ and $\widehat{\mathbf{Cov}}_{\mathbf{2}}(\widehat{\boldsymbol{\beta}})$, respectively, based on

Table 2 $\widehat{\boldsymbol{K}}^{-1}$, $\widehat{\mathbf{C}}\boldsymbol{ov}_{\mathbf{2}}(\widehat{\boldsymbol{\beta}})$ *and* $\widehat{\boldsymbol{\beta}}$ *sample covariance matrix, for* $n = 35$ *and* 45

	$n = 35$				$n = 45$			
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
$\hat{\beta}_1$	0.844	−0.612	−0.243	−0.709	0.902	−0.526	−0.726	−0.428
	0.933	−0.683	−0.264	−0.786	0.974	−0.569	−0.784	−0.462
	1.229	−0.844	−0.249	−0.984	1.221	−0.669	−0.895	−0.537
$\hat{\beta}_2$		1.220	−0.345	0.386		0.898	0.248	−0.071
		1.352	−0.392	0.441		0.969	0.268	−0.082
		1.782	−0.523	0.610		1.175	0.353	−0.082
$\hat{\beta}_3$			1.144	−0.230			1.086	0.117
			1.273	−0.273			1.164	0.128
			1.674	−0.391			1.407	0.148
$\hat{\beta}_4$				1.265				0.841
				1.407				0.912
				1.810				1.097

Table 3 $\widetilde{\boldsymbol{K}}^{-1}$, $\widetilde{\mathbf{C}}\boldsymbol{ov}_{\mathbf{2}}(\widetilde{\boldsymbol{\beta}})$ *and* $\widetilde{\boldsymbol{\beta}}$ *sample covariance matrix, for* $n = 15$ *and* 25

	$n = 15$				$n = 25$			
	$\tilde{\beta}_1$	$\tilde{\beta}_2$	$\tilde{\beta}_3$	$\tilde{\beta}_4$	$\tilde{\beta}_1$	$\tilde{\beta}_2$	$\tilde{\beta}_3$	$\tilde{\beta}_4$
$\tilde{\beta}_1$	2.312	−1.996	−1.554	−0.553	1.534	−1.080	−0.711	−0.929
	2.689	−2.348	−1.803	−0.622	1.694	−1.206	−0.779	−1.023
	3.684	−3.261	−2.349	−0.886	2.083	−1.453	−0.906	−1.242
$\tilde{\beta}_2$		3.429	0.706	−0.626		1.885	0.135	0.005
		4.029	0.796	−0.723		2.110	0.142	−0.004
		5.784	0.958	−0.924		2.633	0.158	0.002
$\tilde{\beta}_3$			2.634	−0.170			1.442	0.106
			3.060	−0.218			1.599	0.103
			4.291	−0.314			2.002	0.108
$\tilde{\beta}_4$				2.293				1.534
				2.643				1.714
				3.638				2.131

the 10,000 replications, where both matrices are evaluated at $\widehat{\boldsymbol{\beta}}$. The third entry in Tables 1 and 2 refers to the $\widehat{\boldsymbol{\beta}}$ sample covariance matrix. In the second simulation study, we performed the same analysis for $\widetilde{\boldsymbol{\beta}}$ and the results are shown in Table 3, for $n = 15$ and 25, and Table 4, for $n = 35$ and 45.

In the third simulation study, we showed the performance of the Wald statistics (3.1) to (3.3) and the likelihood ratio statistic through the estimated size of the four tests, where the null hypothesis was $\mathcal{H} : \beta_4^{(0)} = 0$ against $\mathcal{A} : \beta_4^{(0)} \neq 0$. Assuming that $\mathcal{H}$ is true, the empirical size of the Wald and LR tests is calculated as the proportion of the times that the test statistic exceeds the asymptotic critical value in the 10,000 Monte Carlo replicates for the nominal levels $\alpha = 10\%$, 5% and 1%. The power of the four tests was calculated under the alternative hypothesis $\mathcal{A} : \beta_4^{(0)} = \varepsilon$ for a grid of values for ε. The results are shown in Table 5 and Figures 1 and 2.

Based on Tables 1 to 4, we can point out some observations. The second-order covariances of the estimators $\widehat{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\beta}}$ are closer to the first-order covariances than the sample covariances.

Table 4 $\widetilde{\boldsymbol{K}}^{-1}$, $\widetilde{\mathbf{C}}\boldsymbol{ov}_{\mathbf{2}}(\widetilde{\boldsymbol{\beta}})$ *and* $\widetilde{\boldsymbol{\beta}}$ *sample covariance matrix, for* $n = 35$ *and* 45

	$n=35$				$n=45$			
	$\tilde{\beta}_1$	$\tilde{\beta}_2$	$\tilde{\beta}_3$	$\tilde{\beta}_4$	$\tilde{\beta}_1$	$\tilde{\beta}_2$	$\tilde{\beta}_3$	$\tilde{\beta}_4$
$\tilde{\beta}_1$	0.908	−0.658	−0.262	−0.761	0.953	−0.557	−0.767	−0.451
	0.999	−0.730	−0.284	−0.840	1.024	−0.598	−0.823	−0.485
	1.221	−0.833	−0.253	−0.972	1.215	−0.666	−0.889	−0.531
$\tilde{\beta}_2$		1.311	−0.371	0.414		0.949	0.262	−0.076
		1.445	−0.418	0.469		1.020	0.281	−0.086
		1.765	−0.511	0.596		1.170	0.351	−0.082
$\tilde{\beta}_3$			1.228	−0.247			1.146	0.123
			1.359	−0.288			1.223	0.133
			1.651	−0.374			1.400	0.144
$\tilde{\beta}_4$				1.358				0.889
				1.503				0.959
				1.785				1.089

Table 5 *Estimated sizes of the three Wald and LR tests*

n	$\alpha(\%)$	W_0	W_1	W_2	LR
15	1.0	7.50	6.05	4.41	4.51
	5.0	15.85	13.63	11.06	12.21
	10.0	22.97	20.42	17.03	19.06
25	1.0	4.62	3.73	3.02	2.64
	5.0	11.99	10.22	8.75	9.46
	10.0	19.18	16.92	14.73	15.45
35	1.0	3.79	2.84	2.43	2.10
	5.0	11.06	9.31	8.03	8.15
	10.0	17.34	15.44	14.01	14.25
45	1.0	2.94	2.43	2.02	1.91
	5.0	8.81	7.74	6.79	6.96
	10.0	14.81	13.27	12.17	12.81

In absolute values, both the first-order and the second-order covariances underestimated the sample covariances but the second-order covariances reduce this discrepancy. When the sample size increases, the variances of $\widehat{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\beta}}$ decrease.

Table 5 and Figure 1 show that for $n = 15$, the empirical sizes of the original Wald test are very distant from the corresponding nominal levels. This table and figure also show that W_0 has the worst performance for all n. The statistic W_2 given in (3.3) presents empirical sizes closer to nominal levels, showing that the simultaneous use of the corrected estimator $\widetilde{\boldsymbol{\beta}}$ and covariance matrix $\mathbf{Cov_2}(\widetilde{\boldsymbol{\beta}})$ greatly improves the performance of the Wald test. Its performance is comparable to the LR statistic. When n increases, the empirical sizes of the four statistics converge to the true nominal levels. It is worth mentioning that all four tests are liberal, that is, they wrongly reject the null hypothesis more frequently than expected based on the selected nominal level.

Since all tests have different sizes when one uses their asymptotic chi-squared distribution, we previously ran Monte Carlo replicates to estimate the exact critical value for each test. So, the exact critical value guarantees that all tests have the same type I error, which allows us to compare their powers, as presented in Figure 2, for $n = 45$. The power of all tests, as expected, as $|\varepsilon|$ increases, tends to be 1. Nevertheless, the power of all tests is similar.

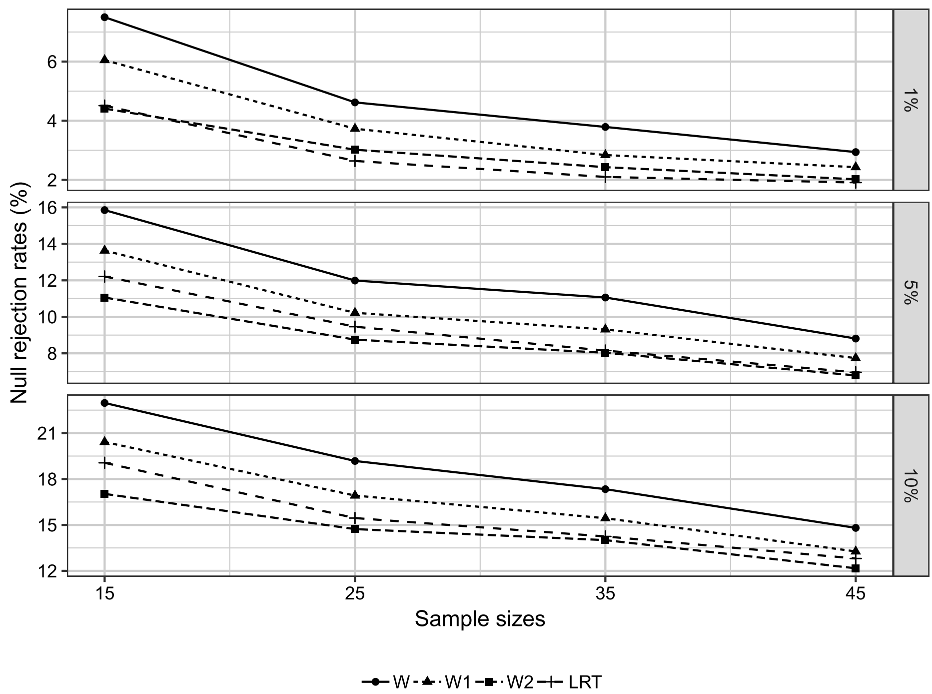

Figure 1 *Estimated sizes of the three Wald and LR tests.*

We must say that the simulation findings for different values of $\boldsymbol{\beta}$, ϕ, p, number of parameters in $\mathcal{H}$ are qualitatively similar to those reported here. As earlier remarked, there is no general Bartlett–type correction available for the Wald test and hence, the W_2 statistic is a suitable option to correct its liberal behavior in small and moderate-sized samples. The choice of W_2 statistic instead of a bootstrap statistic seems to be more attractive because it is not computationally costly.

7 Application

In this section, we present an application based on real data. The data set is presented in McCullagh and Nelder (1989, p. 300) and, most recently, in Rasch, Verdooren and Pilz (2020, p. 421). The response variable (y) is the clotting time of blood in seconds for normal plasma diluted to nine different percentage concentrations with prothrombin-free plasma (x_2). The clotting time was also induced by two lots of thromboplastin (x_3). The data set is shown in Table 6. The sample size is $n = 18$ observations of clotting time.

To illustrate an application of the results presented in the Sections 2 and 3 to DM, we fitted a reciprocal gamma model regression, with $\eta_i = \log(\mu_i) = \beta_1 + \beta_2 \log(x_{2i}) + \beta_3 x_{3i}$, $i = 1, \dots, 18$ and $x_3 = 0$, if Lot 2 and $x_3 = 1$, if Lot 1.

Table 7 presents the estimation results. This table shows that MLE and BCE for $\boldsymbol{\beta}$ have close values, although this not happens for ϕ. As expected through the simulation studies results, the n^{-1} standard errors are lower than the corresponding n^{-2} ones.

Table 8 displays the values and associated p-value of the three Wald statistics to test the hypotheses $\mathcal{H}: \beta_j = 0$ against $\mathcal{A}: \beta_j \neq 0,\ j = 1, 2, 3$. All tests reject the null hypothesis $\mathcal{H}: \beta_j = 0$ for $j = 1, 2$. However, for $\mathcal{H}: \beta_3 = 0$, the tests based on the statistics W_0 and W_1 reject $\mathcal{H}$ considering a significance level of 10% while the test based on W_2 does not reject this hypothesis.

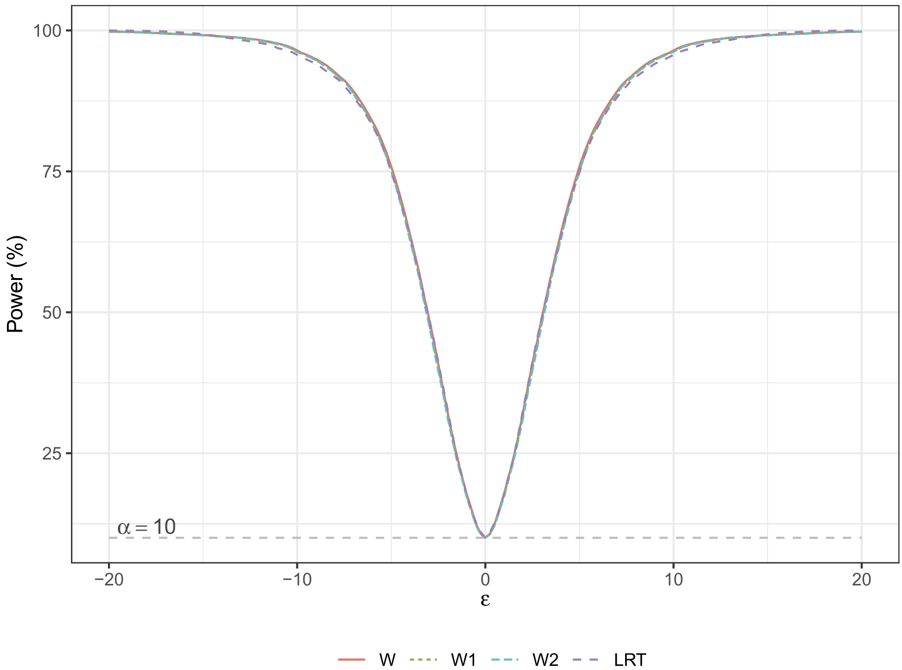

Figure 2 *Power of the three Wald tests and LRT.*

Table 6 *Clotting time of blood (y) for nine percentage concentrations of normal plasma (x_2) and two lots of clotting agents*

	Clotting time (y)	
x_2	Lot 1	Lot 2
5	118	69
10	58	35
15	42	26
20	35	21
30	27	18
40	25	16
60	21	12
80	19	12
100	18	12

Table 7 *Parameter estimates, n^{-1} and n^{-2} standard errors (SE)*

	MLE	n^{-1} SE	n^{-2} SE	BCE	n^{-1} SE	n^{-2} SE
β_1	5.849	0.669	0.676	5.890	0.707	0.717
β_2	−0.575	0.150	0.152	−0.578	0.159	0.161
β_3	−0.469	0.283	0.285	−0.469	0.299	0.302
ϕ	2.781			2.491		

As shown in the simulation study, the W_0 statistic is too liberal while the W_2 presents null rejection rates closer to the nominal levels with no power loss, then the W_2 statistic should be used in practice.

Table 8 *Wald test statistics and p-values*

	W_0	p-value	W_1	p-value	W_2	p-value
β_1	76.522	<0.001	74.775	<0.001	67.541	<0.001
β_2	14.644	<0.001	14.273	<0.001	12.848	<0.001
β_3	2.756	0.097	2.705	0.100	2.412	0.120

8 Concluding remarks

Since the last decade, there has been considerable interest in finding closed-form expressions for the second-order covariance matrix of maximum likelihood estimators in many classes of regression models. Although MLE bias reduction studies are older, up to now, the work of Cordeiro et al. (2014) is the only one that obtains the BCE second-order covariance matrix. The reason is that the expression for obtaining the second-order matrix for the BCE probably requires even more difficult algebra. In this paper, we obtain a general expression for the second-order covariance matrices which, at the same time, gives the second-order covariance matrix for the maximum likelihood estimator and its bias-corrected version. Surprisingly, the expression for the BCE second-order covariance matrix requires less effort to be derived than for the MLE. The second-order covariance matrices presented in this work correct previously published papers based on the results from Peers and Iqbal (1985) instead of those from Shenton and Bowman (1977). We also apply the second-order covariances to construct improved Wald statistics. Next, the general expression of the second-order covariances matrices expression is derived for the dispersion models, an interesting class of models that extends the well-known generalized linear models. We show that the matrix expression for DM is very simple and can be easily implemented in a programming language with support for matrix operations, such as `R`. Our simulation studies indicate that the discrepancy between the sample covariances and the second-order covariances is less than the discrepancy between the sample covariances and the first-order ones. Additionally, they show that the proposed modification to the Wald statistic removes size distortions of the type I error probability without power loss. Finally, we also present an empirical application which illustrates that traditional Wald statistic may lead to misleading conclusions and that W_2 statistic should be preferred.

Appendix

The quantities f_{ji} and g_{ji}, $i = 1, \dots, n$, $j = 1, 2, 3$, in (5.1), are given by

$$f_{1i} = -\left(\frac{d\mu_i}{d\eta_i}\right)^3 d'_{2i}, \qquad f_{2i} = -\frac{d\mu_i}{d\eta_i}\frac{d^2\mu_i}{d\eta_i^2} d_{2i} - \left(\frac{d\mu_i}{d\eta_i}\right)^3 d_{3i},$$

$$f_{3i} = -\frac{d\mu_i}{d\eta_i}\frac{d^2\mu_i}{d\eta_i^2} d_{2i}, \qquad w_i = -\left(\frac{d\mu_i}{d\eta_i}\right)^2 d_{2i},$$

$$g_{1i} = \left(\frac{d\mu_i}{d\eta_i}\right)^4 d''_{2i} + 5\left(\frac{d\mu_i}{d\eta_i}\right)^2 \frac{d^2\mu_i}{d\eta_i^2} d'_{2i} + 2\frac{d\mu_i}{d\eta_i}\frac{d^3\mu_i}{d\eta_i^3} d_{2i} + 2\left(\frac{d^2\mu_i}{d\eta_i^2}\right)^2 d_{2i},$$

$$g_{2i} = \phi\left(\frac{d\mu_i}{d\eta_i}\right)^4 d_{2i}^{(2)} + 2\left(\frac{d\mu_i}{d\eta_i}\right)^2 \frac{d^2\mu_i}{d\eta_i^2}(d'_{2i} - d_{3i}) - \left(\frac{d^2\mu_i}{d\eta_i^2}\right)^2 d_{2i}$$
$$- \phi\left(\frac{d\mu_i}{d\eta_i}\right)^4 d_{2i}^2,$$

$$g_{3i} = 3\left(\frac{d\mu_i}{d\eta_i}\right)^2 \frac{d^2\mu_i}{d\eta_i^2} d'_{2i} + 3\frac{d\mu_i}{d\eta_i}\frac{d^3\mu_i}{d\eta_i^3} d_{2i} + 3\left(\frac{d^2\mu_i}{d\eta_i^2}\right)^2 d_{2i} + \left(\frac{d\mu_i}{d\eta_i}\right)^4 d'_{3i}$$
$$+ 3\left(\frac{d\mu_i}{d\eta_i}\right)^2 \frac{d^2\mu_i}{d\eta_i^2} d_{3i},$$

where d'_{ri} e d''_{ri} are the first and second partial derivatives of d_{ri} with respect to μ_i, respectively and $d_{2i}^{(2)} = \mathbb{E}[\{\partial^2 t(Y_i, \mu_i)/\partial \mu_i^2\}^2]$.

Acknowledgments

This paper is based in part on the Ph.D. dissertation of Dr. Tiago M. Magalhães submitted to the Institute of Mathematics and Statistics, University of São Paulo, under the supervision of the two other authors. We are also grateful to the National Council for Scientific and Technological Development (CNPq) and the National Council for the Improvement of Higher Education (CAPES) for financial support. We also acknowledge the referees' suggestions that helped us improve this work.

Supplementary Material

Supplement to "A general expression for second-order covariance matrices—an application to dispersion models" (DOI: 10.1214/20-BJPS489SUPP; .pdf). Supplementary material related to this article can be found online at https://doi.org/10.1214/20-BJPS489SUPP.

References

Barroso, L. P., Botter, D. A. and Cordeiro, G. M. (2013). Second-order covariance matrix formula for heteroskedastic generalized linear models. *Communications in Statistics Theory and Methods* **42**, 1618–1627. MR3041489 https://doi.org/10.1080/03610926.2011.594543

Cordeiro, G. M. (1993). General matrix formulae for computing bartlett corrections. *Statistics & Probability Letters* **16**, 11–18. MR1208493 https://doi.org/10.1016/0167-7152(93)90115-Y

Cordeiro, G. M., Barroso, L. P. and Botter, D. A. (2006). Covariance matrix formula for generalized linear models with unknown dispersion. *Communications in Statistics Theory and Methods* **35**, 113–120. MR2274004 https://doi.org/10.1080/03610920500439687

Cordeiro, G. M., Botter, D. A., Cavalcanti, A. B. and Barroso, L. P. (2014). Covariance matrix of the bias-corrected maximum likelihood estimator in generalized linear models. *Statistical Papers* **55**, 643–652. MR3227544 https://doi.org/10.1007/s00362-013-0514-1

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. MR1378945 https://doi.org/10.1080/03610929608831678

Cordeiro, G. M. and Klein, R. (1994). Bias correction in arma models. *Statistics & Probability Letters* **30**, 1317–1334. MR1278646 https://doi.org/10.1016/0167-7152(94)90100-7

Cordeiro, G. M., Paula, G. A. and Botter, D. A. (1994). Improved likelihood ratio tests for dispersion models. *International Statistical Review* **62**, 257–274.

Cordeiro, G. M. and Santana, R. G. (2008). Covariance matrix formula for exponential nonlinear models. *Communications in Statistics Theory and Methods* **37**, 2724–2734. MR2458951 https://doi.org/10.1080/03610920802040407

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*. London: Chapman & Hall. MR0370837

Cox, D. R. and Snell, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society, Series B, Methodological* **30**, 248–275. MR0237052

Jørgensen, B. (1987). Exponential dispersion models (with discussion). *Journal of the Royal Statistical Society, Series B* **49**, 127–162. MR0905186

Jørgensen, B. (1997a). Proper dispersion models (with discussion). *Brazilian Journal of Probability and Statistics* **11**, 89–140. MR1619709

Jørgensen, B. (1997b). *The Theory of Dispersion Models*. London: Chapman & Hall. MR1462891

Kosmidis, I., Kenne Pagui, E. C. and Sartori, N. (2020). Mean and median bias reduction in generalized linear models. *Statistics and Computing* **30**, 43–59. MR4057470 https://doi.org/10.1007/s11222-019-09860-6

Lawley, D. N. (1956). A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika* **43**, 295–303. MR0082237 https://doi.org/10.1093/biomet/43.3-4.295

Lemonte, A. J. (2011). Covariance matrix formula for Birnbaum-Saunders regression models. *Journal of Statistical Computation and Simulation* **81**, 899–908. MR2806933 https://doi.org/10.1080/00949650903555288

Lemonte, A. J. (2020). Covariance matrix of maximum likelihood estimators in censored exponential regression models. *Communications in Statistics—Theory and Methods*. https://doi.org/10.1080/03610926.2020.1767142

Lemonte, A. J. and Ferrari, S. L. P. (2012). Local power and size properties of the lr, Wald, score and gradient tests in dispersion models. *Statistical Methodology* **9**, 537–554. MR2925985 https://doi.org/10.1016/j.stamet.2012.03.001

Magalhães, T. M., Botter, D. A. and Sandoval, M. C. (2017). Corrigendum to: "covariance matrix formula for generalized linear models with unknown dispersion" by G. M. Cordeiro, L. P. Barroso and D. A. Botter [Communications in Statistics—Theory and Methods (2006) 35(1), 113–120]. *Communications in Statistics Theory and Methods* **46**, 10702–10703. MR3740831 https://doi.org/10.1080/03610926.2016.1242745

Magalhães, T. M., Botter, D. A. and Sandoval, M. C. (2020). Supplement to "A general expression for second-order covariance matrices—an application to dispersion models." https://doi.org/10.1214/20-BJPS489SUPP

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. London: Chapman & Hall. MR3223057 https://doi.org/10.1007/978-1-4899-3242-6

Medeiros, F. M. C., Ferrari, S. L. P. and Lemonte, A. J. (2017). Improved inference in dispersion models. *Applied Mathematical Modelling* **51**, 317–328. MR3694537 https://doi.org/10.1016/j.apm.2017.06.045

Patriota, A. G. and Cordeiro, G. M. (2011). A matrix formula for the skewness of maximum likelihood estimators. *Statistics & Probability Letters* **81**, 529–537. MR2765174 https://doi.org/10.1016/j.spl.2010.12.009

Peers, H. W. and Iqbal, M. (1985). Asymptotic expansions for confidence limits in the presence of nuisance parameters, with applications. *Journal of the Royal Statistical Society, Series B, Methodological* **47**, 547–554. MR0844486

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rao, C. R. (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Proceedings of the Cambridge Philosophical Society* **44**, 50–57.

Rasch, D., Verdooren, R. and Pilz, J. (2020). *Applied Statistics—Theory and Problem Solutions with R*. New Jersey: Wiley.

Rocha, A. V., Simas, A. B. and Cordeiro, G. M. (2010). Second-order asymptotic expressions for the covariance matrix of maximum likelihood estimators in dispersion models. *Statistics & Probability Letters* **80**, 718–725. MR2595152 https://doi.org/10.1016/j.spl.2009.12.030

Shenton, L. R. and Bowman, K. O. (1977). *Maximum Likelihood Estimation in Small Samples*. London: Griffin.

Simas, A. B., Cordeiro, G. M. and Rocha, A. V. (2010). Skewness of maximum likelihood estimators in dispersion models. *Journal of Statistical Planning and Inference* **140**, 2111–2121. MR2606745 https://doi.org/10.1016/j.jspi.2010.02.007

Simas, A. B., Rocha, A. V. and Barreto-Souza, W. (2011). Bias-corrected estimators for dispersion models with dispersion covariates. *Journal of Statistical Planning and Inference* **141**, 3063–3074. MR2796012 https://doi.org/10.1016/j.jspi.2011.03.028

Wald, A. (1943). Test of statistical hypotheses concerning several parameter when the number of observations is large. *Transactions of the American Mathematical Society* **54**, 426–482. MR0012401 https://doi.org/10.2307/1990256

Wilks, S. S. (1938). The large-sample distribution of likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics* **9**, 60–62.

T. M. Magalhães
Department of Statistics
Institute of Exact Sciences
University of Juiz de Fora
Rua José Lourenço Kelmer, s/n, 36036–900
Juiz de Fora MG
Brazil
E-mail: tiago.magalhaes@ice.ufjf.br

D. A. Botter
M. C. Sandoval
Department of Statistics
Institute of Mathematics and Statistics
University of São Paulo
Rua do Matão 1010, 05508–090
São Paulo SP
Brazil
E-mail: botter@ime.usp.br
sandoval@ime.usp.br