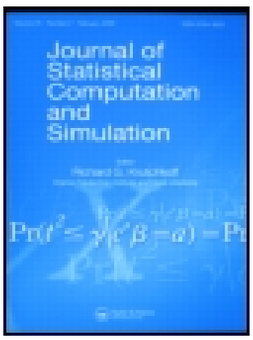

Journal of Statistical Computation and Simulation

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/gscs20

Improved gradient statistic in heteroskedastic generalized linear models

Fabiana U. Barros, Denise A. Botter, Mônica C. Sandoval & Tiago M. Magalhães

To cite this article: Fabiana U. Barros, Denise A. Botter, Mônica C. Sandoval & Tiago M. Magalhães (2023): Improved gradient statistic in heteroskedastic generalized linear models, Journal of Statistical Computation and Simulation, DOI: 10.1080/00949655.2023.2172170

To link to this article: https://doi.org/10.1080/00949655.2023.2172170

JOURNAL OF STATISTICAL COMPUTATION AND SIMULATION
https://doi.org/10.1080/00949655.2023.2172170

Improved gradient statistic in heteroskedastic generalized linear models

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

[a]Department of Statistics, University of São Paulo, São Paulo, Brazil; [b]Department of Statistics, Federal University of Juiz de Fora, Juiz de Fora, Brazil

ABSTRACT

In this paper, we obtain the Bartlett-type correction factor for the gradient test statistic in heteroskedastic generalized linear models (HGLMs). We present an extensive Monte Carlo study to evaluate the performance of the corrected gradient test in small sample sizes. We also compare its performance with the usual gradient, likelihood ratio (LR), score, Wald, improved LR and improved score statistics along with bootstrap tests.

ARTICLE HISTORY
Received 22 June 2022
Accepted 19 January 2023

KEYWORDS
Gradient test; heteroskedastic generalized linear models; maximum likelihood estimators

MSC:
Primary 62Fxx; secondary 62F12

1. Introduction

Hypothesis testing is an important stage of statistical inference. Without loss of generality, let $\boldsymbol{\theta}$ be the unknown p-dimensional parameter vector. Additionally, let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$, $\boldsymbol{\theta}_1$ be a q-dimensional vector and $\boldsymbol{\theta}_2$ a vector containing the remaining $p-q$ parameters. In a hypothesis test, the interest lies in $\mathcal{H}_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^{(0)}$, the null hypothesis, where $\boldsymbol{\theta}_1^{(0)}$ is a known q-vector. In other words, the null hypothesis imposes q restrictions on the parameter vector. Hence, $\boldsymbol{\theta}_2$ is the vector of nuisance parameters. The partition $\boldsymbol{\theta}$ induces the following partition: $\boldsymbol{U}(\boldsymbol{\theta}) = (\boldsymbol{U_1}(\boldsymbol{\theta})^\top, \boldsymbol{U_2}(\boldsymbol{\theta})^\top)^\top$, where $\boldsymbol{U}(\boldsymbol{\theta})$ is the score function. Consider $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\theta}}_1^\top, \widehat{\boldsymbol{\theta}}_2^\top)^\top$ as the unrestricted maximum likelihood estimator of $\boldsymbol{\theta}$ and $\widetilde{\boldsymbol{\theta}} = (\boldsymbol{\theta}_1^{(0)\top}, \tilde{\boldsymbol{\theta}}_2^\top)^\top$ as the restricted maximum likelihood estimator $\boldsymbol{\theta}$ under the null hypothesis.

The gradient test [1] is an alternative to the classical likelihood ratio [2], Wald [3] and score tests [4], and has been employed to test hypotheses. Rao [5] emphasized the following: 'The suggestion by Terrell is attractive as it is simple to compute. It would be of interest to investigate the performance of the [gradient] statistic.' To compute this statistic, it is not necessary to know the information matrix, either expected or observed. The

CONTACT Tiago M. Magalhães tiago.magalhaes@ufjf.br

Supplemental data for this article can be accessed here. https://doi.org/10.1080/00949655.2023.2172170

gradient statistic for testing $\mathcal{H}_0$ against $\mathcal{H}_1$ is defined as:

$$T = \boldsymbol{U}(\widetilde{\boldsymbol{\theta}})^\top(\widehat{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\theta}}),$$

and can be expressed as $T = \boldsymbol{U_1}(\widetilde{\boldsymbol{\theta}})^\top(\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^{(0)})$, since $\boldsymbol{U_2}(\widetilde{\boldsymbol{\theta}}) = \mathbf{0}$. Otherwise, the gradient statistic is not invariant under reparameterizations of the model that preserve the parameter of interest. Montoril and Souza [6] obtained some properties of the gradient statistic and Lemonte [7] presented a detailed survey on the gradient test.

Under the null hypothesis with large sample size, the gradient test statistic is asymptotically χ^2_q distributed. On the other hand, the chi-squared distribution may not be a good approximation of the null distribution of the gradient test statistic in samples with small or moderate sizes. Vargas et al. [8] derived a general Bartlett-type correction factor for the gradient statistic. Under the null hypothesis, the corrected gradient statistic is distributed as chi-squared up to an error of order $o(n^{-1})$, while the uncorrected gradient statistic has a chi-squared distribution up to an error of order $o(n^{-1/2})$. Based on Vargas et al. [8], various studies have been published: Vargas et al. [9] for generalized linear models; Medeiros et al. [10] for dispersion models; and Medeiros and Lemonte [11] and Magalhães and Gallardo [12] for exponential and Weibull regression models, respectively.

Our chief goal in this paper is to derive the Bartlett-type correction factor for the gradient statistic in heteroskedastic generalized linear models (HGLMs). Some attempts have been made to develop a second-order asymptotic theory for HGLMs to obtain better likelihood inference procedures. Botter and Cordeiro [13] derived nearly unbiased maximum likelihood estimators. Barroso et al. [14] provided a matrix formula for second-order covariances of maximum likelihood estimates. Botter and Cordeiro [15] and Cordeiro et al. [16] derived the Bartlett and the Bartlett-type correction factors in matrix notation for the likelihood ratios (LRs) and score tests, respectively, which reduce the size distortion of these tests.

Another goal in this paper is to perform Monte Carlo simulation experiments to evaluate and compare the finite-sample performance of the corrected gradient statistic with the usual gradient, LR, score, Wald, improved LR and improved score statistics. Bootstrap-based tests are also included in the Monte Carlo experiments. Since we have not found any comprehensive simulation study in the statistical literature performing this type of numerical process in HGLMs, we believe our paper fills this gap.

The paper is organized as follows. In Section 2, we describe the class of HGLMs and discuss estimation and hypothesis testing inference for the mean and precision regression parameters. In Section 3, we derive the specific Bartlett-type correction factor of the gradient statistic for the HGLMs. The Monte Carlo simulation results are presented and discussed in Section 4. An empirical application that uses real data is presented and discussed in Section 5. The paper ends with a discussion in Section 6.

2. Heteroskedastic generalized linear models

In the standard formulation of the generalized linear models (GLMs) proposed by Nelder and Wedderburn [17], the mean of the response variable is a function of covariates, but the precision parameter is assumed constant across the observations. However, in many practical situations, this assumption does not hold; see, for instance, Brooks et al. [18], Hussain et al. [19] and Smyth et al. [20]. It is important to formulate the precision parameter

also a function of covariates. For this reason, Smyth [21] introduced the generalized linear models with varying dispersion, also known as heteroskedastic generalized linear models or double generalized linear models.

Let $Y_1, \ldots, Y_n$ be independent random variables, with each Y_ℓ having probability density function (or probability function) in the exponential family given by:

$$\pi(y_\ell; \theta_\ell, \phi_\ell) = \exp\{\phi_\ell[y_\ell\theta_\ell - b(\theta_\ell) + c(y_\ell)] + a(y_\ell) + d(\phi_\ell)\}, \tag{1}$$

where $a(\cdot)$, $b(\cdot)$, $c(\cdot)$ and $d(\cdot)$ are known functions. The mean and variance of Y_ℓ are, respectively, $\mathbb{E}(Y_\ell) = \mu_\ell = \mathrm{d}b(\theta_\ell)/\mathrm{d}\theta_\ell$ and $\mathrm{Var}(Y_\ell) = \phi_\ell^{-1}V_\ell$, where $V_\ell = \mathrm{d}\mu_\ell/\mathrm{d}\theta_\ell$ is the variance function and $\theta_\ell = \int V_\ell{}^{-1}\mathrm{d}\mu_\ell = q(\mu_\ell)$, $q(\mu_\ell)$ is a known one-to-one function of the mean μ_ℓ. The parameters θ_ℓ and $\phi_\ell > 0$ are called the canonical and the precision parameters, respectively.

The HGLMs are defined by (1) and the systematic components

$$t_1(\boldsymbol{\mu}) = \boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta} \quad \text{and} \quad t_2(\boldsymbol{\phi}) = \boldsymbol{\tau} = \boldsymbol{S}\boldsymbol{\lambda}, \tag{2}$$

where $t_1(\cdot)$ is a known, continuous, invertible and twice differentiable function, called the mean link function, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^\top$; $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)^\top$ is the mean linear predictor, $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^\top$ is a specified $n \times p$ matrix of full rank $p < n$, with ℓth row $\boldsymbol{x}_\ell = (x_{\ell 1}, \ldots, x_{\ell p})^\top$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ is a vector of unknown parameters to be estimated, $t_2(\cdot)$ is a known, continuous, invertible and twice differentiable function, called the dispersion link function, $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_n)^\top$, $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_n)^\top$ is the dispersion linear predictor, $\boldsymbol{S} = (\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n)^\top$ is a specified $n \times q$ matrix of full rank $q < n$, with ℓth row $\boldsymbol{s}_\ell = (s_{\ell 1}, \ldots, s_{\ell q})^\top$, and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_q)^\top$ is also a vector of unknown parameters to be estimated. Occasionally, some columns of $\boldsymbol{X}$ and $\boldsymbol{S}$ can be equal. In others words, $\boldsymbol{X}$ and $\boldsymbol{S}$ can overlap.

Let $\ell(\boldsymbol{\beta}, \boldsymbol{\lambda})$ be the logarithm of the likelihood function of the model described by (1) and (2) given a set of observations $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$. Thus

$$\ell(\boldsymbol{\beta}, \boldsymbol{\lambda}) = \sum_{\ell=1}^{n} \phi_\ell[y_\ell\theta_\ell - b(\theta_\ell) + c(y_\ell)] + \sum_{\ell=1}^{n}[a(y_\ell) + d(\phi_\ell)]. \tag{3}$$

We assume that the function in (3) satisfies the usual regularity conditions [22, Chapter 9] concerning all derivatives for $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ to the fourth-order. The following notation will be used from now on: $d_{k\ell} = \mathrm{d}^k d(\phi_\ell)/\mathrm{d}\phi_\ell^k$ and $\phi_{k\ell} = \mathrm{d}^k\phi_\ell/\mathrm{d}\tau_\ell^k$ for $k = 1, \ldots, 4$ and $\ell = 1, \ldots, n$.

Table 1 shows some quantities for distributions belonging to the exponential family. The quantity $\Gamma(\phi) = \int_0^\infty t^{\phi-1}e^{-t}dt$ is the gamma function, $\psi(\phi) = \Gamma'(\phi)/\Gamma(\phi)$ is the digamma function, and $\psi^{(r)}(\phi) = \mathrm{d}^r\psi(\phi)/\mathrm{d}\phi^r$ for $r = 1, 2$ and 3 is th rth digamma function's derivative.

The score vector, obtained by differentiating (3) for the unknown parameters, is given by $\boldsymbol{U}(\boldsymbol{\beta}, \boldsymbol{\lambda}) = (\boldsymbol{U}_\beta(\boldsymbol{\beta}, \boldsymbol{\lambda})^\top, \boldsymbol{U}_\lambda(\boldsymbol{\beta}, \boldsymbol{\lambda})^\top)^\top$, where

$$\boldsymbol{U}_\beta(\boldsymbol{\beta}, \boldsymbol{\lambda}) = \boldsymbol{X}^\top\boldsymbol{\Phi}\boldsymbol{W}^{1/2}\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}) \quad \text{and} \quad \boldsymbol{U}_\lambda(\boldsymbol{\beta}, \boldsymbol{\lambda}) = \boldsymbol{S}^\top\boldsymbol{\Phi}_1\boldsymbol{\nu},$$

with $\boldsymbol{W} = \mathrm{diag}\{w_1, \ldots, w_n\}$, $w_\ell = V_\ell^{-1}(\mathrm{d}\mu_\ell/\mathrm{d}\eta_\ell)^2$, $\boldsymbol{V} = \mathrm{diag}\{V_1, \ldots, V_n\}$, $\boldsymbol{\Phi} = \mathrm{diag}\{\phi_1, \ldots, \phi_n\}$, $\boldsymbol{\Phi}_1 = \mathrm{diag}\{\phi_{11}, \ldots, \phi_{1n}\}$ and $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_n)^\top$, $\nu_\ell = y_\ell\theta_\ell - b(\theta_\ell) + c(y_\ell) + d_{1\ell}$.

Table 1. Quantities and their derivatives for some distributions in the exponential family.

Quantity	Normal: N(μ,σ^2)	Inverse Gaussian: IG(μ,ϕ)	Gamma: G(μ,ϕ)
θ	μ	$-1/2\mu^2$	$-1/\mu$
$b(\theta)$	$\theta^2/2$	$-\sqrt{-2\theta}$	$-\log(-\theta)$
ϕ	σ^{-2}	ϕ	$\mathbb{E}^2(Y)/\mathrm{Var}(Y)$
$V(\mu)$	1	μ^3	μ^2
$d(\phi)$	$1/2\log\phi$	$1/2\log\phi$	$\phi\log\phi-\log\Gamma(\phi)$
d_1	$1/2\phi$	$1/2\phi$	$\log\phi+1-\psi(\phi)$
d_2	$-1/2\phi^2$	$-1/2\phi^2$	$(1/\phi)-\psi^{(1)}(\phi)$
d_3	$1/\phi^3$	$1/\phi^3$	$(-1/\phi^2)-\psi^{(2)}(\phi)$
d_4	$-3/\phi^4$	$-3/\phi^4$	$(2/\phi^3)-\psi^{(3)}(\phi)$

The Fisher information matrix is given by:

$$\boldsymbol{K}=\boldsymbol{K}(\boldsymbol{\beta},\boldsymbol{\lambda})=\begin{pmatrix}\boldsymbol{K}_\beta & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{K}_\lambda\end{pmatrix},$$

where $\boldsymbol{K}_\beta=\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{\Phi}\boldsymbol{X}$, $\boldsymbol{K}_\lambda=-\boldsymbol{S}^\top\boldsymbol{D}_2\boldsymbol{\Phi}_1^2\boldsymbol{S}$ and $\boldsymbol{D}_2=\mathrm{diag}\{d_{21},\ldots,d_{2n}\}$.

Note that the Fisher information matrix is a diagonal block matrix, so the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ are globally orthogonal [23]. Under the regularity conditions, referenced above, when the sample size is large, $(\hat{\boldsymbol{\beta}}^\top,\hat{\boldsymbol{\lambda}}^\top)^\top$ is approximately distributed as $N_{p+q}((\boldsymbol{\beta}^\top,\boldsymbol{\lambda}^\top)^\top,\boldsymbol{K}^{-1}(\boldsymbol{\beta},\boldsymbol{\lambda}))$. It follows from these facts that $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\lambda}}$ are asymptotically independent. These estimators satisfy equations $\boldsymbol{U}(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\lambda}})=0$, and to solve them it is necessary to apply iterative methods [24] such as Fisher's scoring, or equivalently, re-weighted least squares.

3. Corrected gradient statistic

In HGLMs we have $\boldsymbol{\theta}=(\boldsymbol{\beta}^\top,\boldsymbol{\lambda}^\top)^\top$. Assuming we are interested in simultaneously testing the mean and dispersion effects, the null hypothesis is written as $\mathcal{H}_0:\boldsymbol{\beta_1}=\boldsymbol{\beta_1^{(0)}},\boldsymbol{\lambda_1}=\boldsymbol{\lambda_1^{(0)}}$, which will be tested against the alternative hypothesis $\mathcal{H}_1$: at least one violation, where $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ are partitioned as $\boldsymbol{\beta}=(\boldsymbol{\beta_1^\top},\boldsymbol{\beta_2^\top})^\top$, with $\boldsymbol{\beta_1}=(\beta_1,\ldots,\beta_{p_1})^\top$ and $\boldsymbol{\beta_2}=(\beta_{p_1+1},\ldots,\beta_p)^\top$, $\boldsymbol{\lambda}=(\boldsymbol{\lambda_1^\top},\boldsymbol{\lambda_2^\top})^\top$, with $\boldsymbol{\lambda_1}=(\lambda_1,\ldots,\lambda_{q_1})^\top$ and $\boldsymbol{\lambda_2}=(\lambda_{q_1+1},\ldots,\lambda_q)^\top$. The dimensions of the specified vectors $\boldsymbol{\beta_1^{(0)}}$ and $\boldsymbol{\lambda_1^{(0)}}$ are p_1 and q_1, respectively. The vectors $\boldsymbol{\beta_2}$ and $\boldsymbol{\lambda_2}$ are nuisance parameter vectors. Additionally, these partitions induce:

$$\boldsymbol{K}_\beta=\begin{bmatrix}\boldsymbol{K}_{\beta 11} & \boldsymbol{K}_{\beta 12}\\ \boldsymbol{K}_{\beta 21} & \boldsymbol{K}_{\beta 22}\end{bmatrix},\quad \boldsymbol{K}_\beta^{-1}=\begin{bmatrix}\boldsymbol{K}_\beta^{11} & \boldsymbol{K}_\beta^{12}\\ \boldsymbol{K}_\beta^{21} & \boldsymbol{K}_\beta^{22}\end{bmatrix},$$

$$\boldsymbol{K}_\lambda=\begin{bmatrix}\boldsymbol{K}_{\lambda 11} & \boldsymbol{K}_{\lambda 12}\\ \boldsymbol{K}_{\lambda 21} & \boldsymbol{K}_{\lambda 22}\end{bmatrix},\quad \boldsymbol{K}_\lambda^{-1}=\begin{bmatrix}\boldsymbol{K}_\lambda^{11} & \boldsymbol{K}_\lambda^{12}\\ \boldsymbol{K}_\lambda^{21} & \boldsymbol{K}_\lambda^{22}\end{bmatrix},$$

where $\boldsymbol{K_\beta}$ and $\boldsymbol{K_\lambda}$ are the Fisher information matrices of $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$, respectively, with inverses $\boldsymbol{K}_\beta^{-1}$ and $\boldsymbol{K}_\lambda^{-1}$.

Let T be the gradient statistic [1] for testing a general null hypothesis $\mathcal{H}_0$. Vargas et al. [8] improved the gradient test by defining a statistic, say T_b, such that $\mathbb{P}_{\mathcal{H}_0}(T_b\leq w)=\mathbb{P}(\chi_q^2\leq w)+\mathcal{O}(n^{-2})$, under the null hypothesis, while $\mathbb{P}_{\mathcal{H}_0}(T\leq w)=\mathbb{P}(\chi_q^2\leq$

$w) + \mathcal{O}(n^{-1})$. More accurate inferences are expected with this error reduction. The expansion proposed by Vargas et al. [8] is very general and there is no closed-form for all models. For instance, although Vargas et al. [9] found the improved gradient statistic for the dispersion models, there is no closed-form for the simplex distribution.

Using the general result of Vargas et al. [8] and after some tedious algebra, we obtain a closed-form for the Bartlett-type corrected gradient statistic for testing $\mathcal{H}_0$ in heteroskedastic generalized linear models, as follows:

$$T_{\mathrm{b}} = T\left[1 - (c + bT + aT^2)\right], \tag{4}$$

where $a = A_3/\{12q(q+2)(q+4)\}$, $b = \{A_2 - 2A_3\}/\{12q(q+2)\}$, and $c = \{A_1 - A_2 + A_3\}/12q$. For the sake of brevity, the quantities A_1 to A_3 are presented in the Appendix. The expression present in (4) is the major result of this paper.

In the Supplementary Material to this paper, we present two particular null hypotheses: (i) for a subset of $\boldsymbol{\beta}$ and (ii) for a subset of $\boldsymbol{\lambda}$. If there are unknown parameters, they must be replaced by the maximum likelihood estimate under the null hypothesis. Although the derivation of (4) is cumbersome, the result depends on simple matrix operations that can be easily computed using a matrix programming language, such as Ox [25] or R [26].

4. Numerical results

In this section, our objective is to compare, through simulation, the performances of the usual likelihood ratio (LR), score (SR), Wald (W) and gradient (T) tests, and their improved versions LR adjusted via Skovgaard adjustment (LR_a; Equation (5) in [27]), LR with the Bartlett correction factor (LR_b), SR with Bartlett-type correction factor (SR_b), and T with the Bartlett-type correction factor (T_b), considering small and moderate sample sizes in HGLMs. We also include bootstrap tests (LR_{boot}, SR_{boot}, W_{boot} and T_{boot}).

We test the effects on the mean and the dispersion separately. The performances are evaluated according to the proximity of the probability of rejection of the null hypothesis when it is true (probability of type I error) with the respective nominal levels of the tests. We also evaluate the power of the improved and bootstrap tests under study.

The simulations are based on the gamma regression model with systematic components defined by:

$$\mu_\ell = \beta_1 + \beta_2 x_{\ell 2} + \cdots + \beta_p x_{\ell p} \quad \text{and} \quad \log(\phi_\ell) = \lambda_1 + \lambda_2 s_{\ell 2} + \cdots + \lambda_q s_{\ell q}, \tag{5}$$

where the covariables $x_2, \ldots, x_p$ and $s_2, \ldots, s_q$ are obtained from the $\mathcal{U}(1,2)$ distribution independently. These covariables are kept constant during the simulations. All parameters, except those in the null hypothesis, are defined as one. Simulation results for the inverse Gaussian regression model are presented in the Supplementary Material.

All simulations are performed using the matrix programming language Ox [25]. Maximization of the logarithm of the likelihood function for $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ is done by the quasi-Newton method BFGS [24] with first analytic derivatives. The number of Monte Carlo replicates is set at 5,000 and the following nominal levels are considered: $\alpha = 10\%, 5\%, 1\%$. We use $B = 500$ replicates for the bootstrap tests.

For each sample size and nominal level, we calculate the rejection rate of each test, that is, we estimate $\mathbb{P}(LR > z_\alpha)$, $\mathbb{P}(LR_a > z_\alpha)$, $\mathbb{P}(LR_b > z_\alpha)$, $\mathbb{P}(SR > z_\alpha)$, $\mathbb{P}(SR_a > z_\alpha)$,

$\mathbb{P}(W > z_\alpha), \mathbb{P}(T > z_\alpha), \mathbb{P}(T_b > z_\alpha)$, where z_α is the $(1-\alpha)$-quantile of the reference χ^2 distribution. In other words, we calculate for each statistic the percentage of times that the null hypothesis is rejected. For the tests based on the statistics LR_{boot}, SR_{boot}, W_{boot} and T_{boot}, the rejection rates are obtained from the probabilities $\mathbb{P}(LR > \widehat{q}_\alpha)$, $\mathbb{P}(SR > \widehat{q}_\alpha)$, $\mathbb{P}(W > \widehat{q}_\alpha)$ and $\mathbb{P}(T > \widehat{q}_\alpha)$, where $\widehat{q}_\alpha$ is the $(1-\alpha)$ bootstrap quantile estimated as follows.

Let S be any of the uncorrected statistics. Using the bootstrap technique, we can find the S empirical distribution from the observed sample $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$. This method consists of generating B bootstrap samples $(y_1^{*1}, \ldots, y_n^{*B})^\top$ from the assumed model with the parameters replaced by restricted estimates computed using the original sample, under the null hypothesis. Then we calculate the value of the statistic S for each sample, that is, we obtain S^{*b} for $b = 1, 2, \ldots, B$. Finally, these values are placed in ascending order. By setting the significance level α, the percentile $1-\alpha$ of S is estimated by the value $\widehat{q}_\alpha$ such that $\#\{S^{*b} \leq \widehat{q}_\alpha\}/B = 1-\alpha$. Therefore, the test rejects the null hypothesis if $S > \widehat{q}_\alpha$. Further information on bootstrap tests can be found in Efron and Tibshirani [28, Chapter 16].

In Table 2, we present the rejection rates considering hypothesis $\mathcal{H}_0 : \beta_p = 0$, $n = 30$, $q = 2$ and different values of p. We vary p to analyze the effect of the number of covariates present in the model in the various tests. Note that the likelihood ratio (LR) and Wald (W) tests are extremely liberal, rejecting the null hypothesis more often than expected based on the nominal level, for any value of p. The score (SR) and gradient (T) tests are also liberal in almost all cases but have smaller distorted sizes than the likelihood ratio and Wald tests in all cases. For example, when $p = 6$ and $\alpha = 5\%$, the test rejection rates are 11.08% (LR), 7.68% (SR), 15.02% (W) and 9.32% (T). The test T_{b} outperforms the test T, especially when the number of parameters p increases. We also note that the impact of the number of parameters is less accentuated in the corrected tests LR_{a}, LR_{b}, SR_{b} and T_{b}. For the tests LR_{a} and LR_{b}, the Skovgaard adjustment and the Bartlett correction factor, respectively, cause the rejection rates to be closer to the nominal levels considered. For example, for $p = 5$ and $\alpha = 10\%$, the LR test has a rejection rate of 17.18%, while its modified versions have 12.06% (LR_{a}) and 12.72% (LR_{b}). The bootstrap tests, in general, perform better than the other tests but are similar to each other.

In Table 3, we set the null hypothesis as $\mathcal{H}_0 : \beta_p = 0, p = 2, n = 30$ with different values of q. The tests that have their rejection rates most affected by the increase in the number of dispersion parameters are the W, LR and T tests. The other tests show reasonable performance, and the tests LR_{a} and LR_{b} have more unstable empirical sizes relative to their respective nominal levels. The rejection rate of the Wald test for $q = 5$ and $\alpha = 5\%$ is 19.56%, that is, almost four times the nominal level considered, while the rejection rates of the LR, LR_{a}, LR_{b}, SR, SR_{b}, T, T_{b}, LR_{boot}, SR_{boot}, W_{boot} and T_{boot} tests are 11.56%, 7.32%, 6.64%, 5.28%, 4.86%, 9.12%, 4.96%, 5.24%, 5.42%, 5.26%, and 5.72%, respectively. The other previous conclusions obtained from Table 2 are also verified here.

In Table 4, we consider the null hypothesis $\mathcal{H}_0 : \beta_p = 0, p = 3, q = 3$ and $n = 20, 30, 40$. As the sample size increases, the rejection rates of all the tests approach the respective nominal levels. In addition, the SR_{b}, T_{b} and bootstrap tests present rejection rates near the corresponding nominal levels for all values of n. It is worth mentioning here that even for $n = 40$, the tests based on the statistics LR, W, and T are still liberal. For example, for $\alpha = 10\%$, the rejection rates of the respective tests are 14.02% (LR), 16.96% (W) and 12.86% (T). The other previous conclusions show in Tables 2 and 3 can also be seen here.

Table 2. Rejection rates for $\mathcal{H}_0 : \beta_p = 0$; gamma distribution with $n = 30$ and $q = 2$.

p	α(%)	LR	LR_a	LR_b	SR	SR_b	W	T	T_b	LR_{boot}	SR_{boot}	W_{boot}	T_{boot}
	10	12.76	9.72	9.98	10.12	9.34	15.50	11.56	9.56	9.56	9.64	9.54	9.56
2	5	6.68	4.74	4.80	4.50	4.34	9.28	5.48	4.50	4.68	4.60	4.62	4.72
	1	1.56	0.86	0.80	0.70	0.84	3.38	0.78	0.68	0.96	0.90	1.10	0.98
	10	13.40	10.40	10.38	11.14	9.88	15.82	12.54	10.08	9.94	10.08	9.86	9.86
3	5	7.70	4.84	5.04	5.48	4.66	9.94	6.12	4.78	4.66	4.66	4.46	4.84
	1	1.80	1.04	1.12	1.00	0.94	3.36	1.04	0.86	1.02	1.10	1.02	0.98
	10	16.74	11.94	12.66	12.60	10.82	21.58	15.10	11.38	10.98	10.58	10.82	11.12
4	5	10.42	6.10	6.62	6.68	5.74	14.68	8.54	5.62	5.44	5.62	5.64	5.70
	1	2.86	1.44	1.56	1.44	1.38	6.40	1.68	1.26	1.34	1.42	1.28	1.54
	10	17.18	12.06	12.72	13.34	10.58	22.24	15.78	11.56	10.42	10.10	10.80	10.62
5	5	10.74	6.38	6.96	7.10	5.30	14.72	9.02	5.82	5.24	5.22	5.58	5.36
	1	3.28	1.36	1.64	1.36	0.94	6.76	2.00	1.12	1.22	1.08	1.30	1.18
	10	18.14	10.90	12.78	14.22	10.12	22.72	16.72	10.82	9.68	9.50	9.72	9.60
6	5	11.08	5.64	6.72	7.68	5.42	15.02	9.32	5.52	4.98	5.14	5.16	5.06
	1	3.54	1.22	1.64	1.50	0.98	6.66	2.24	1.02	1.14	1.08	1.06	1.26
	10	17.80	11.28	12.20	14.02	10.28	21.08	16.58	10.74	9.46	9.76	9.86	9.42
7	5	10.70	5.36	6.38	7.76	4.66	14.48	9.16	5.32	4.70	4.52	4.62	4.56
	1	3.28	0.96	1.40	1.56	0.90	6.06	1.86	0.86	0.78	0.94	0.76	0.70

Table 3. Rejection rates for $\mathcal{H}_0 : \beta_p = 0$; gamma distribution with $n = 30$ and $p = 2$.

	$q = 2$			$q = 3$		
Statistic	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 1\%$
LR	12.76	6.68	1.56	15.42	9.12	2.54
LR_a	9.72	4.74	0.86	11.52	5.80	1.48
LR_b	9.98	4.80	0.80	12.04	6.02	1.44
SR	10.12	4.50	0.70	11.44	5.68	0.98
SR_b	9.34	4.34	0.84	10.26	5.26	1.04
W	15.50	9.28	3.38	20.28	13.86	5.80
T	11.56	5.48	0.78	14.40	7.54	1.56
T_b	9.56	4.50	0.68	10.88	5.36	1.14
LR_{boot}	9.56	4.68	0.96	10.56	5.44	1.38
SR_{boot}	9.64	4.60	0.90	10.62	5.48	1.28
W_{boot}	9.54	4.62	1.10	10.78	5.36	1.32
T_{boot}	9.56	4.72	0.98	10.82	5.68	1.52
	$q = 4$			$q = 5$		
Statistic	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 1\%$
LR	16.70	9.98	2.84	18.38	11.56	3.56
LR_a	11.80	6.18	1.26	13.74	7.32	1.96
LR_b	11.46	6.00	1.26	12.68	6.64	1.40
SR	10.50	5.26	0.70	11.22	5.28	0.74
SR_b	9.70	5.02	0.84	9.98	4.86	0.84
W	23.90	16.74	8.22	26.98	19.56	10.08
T	14.86	7.88	1.18	16.34	9.12	1.76
T_b	9.98	4.88	0.78	10.38	4.96	0.92
LR_{boot}	9.88	4.94	1.08	10.10	5.24	1.22
SR_{boot}	9.76	5.20	1.10	10.50	5.42	1.24
W_{boot}	9.72	4.80	1.22	10.08	5.26	1.08
T_{boot}	9.96	5.26	1.14	10.66	5.72	1.34

In Table 5, we take the null hypothesis $\mathcal{H}_0 : \lambda_2 = \cdots = \lambda_q = 0$, $q = 3$, $p = 2$ and $n = 20, 30, 40$. The rejection rates of all tests approach their respective nominal levels as the sample size increases. Furthermore, the tests LR, W and T present high distortion of size, with the distortion of W being more accentuated. In contrasts, the tests SR, SR_b and T_b are conservative. We further find that the test LR_a is slightly better than the LR_b test.

Table 4. Rejection rates for $\mathcal{H}_0 : \beta_p = 0$; gamma distribution with $p = 3$ and $q = 3$.

n	α(%)	LR	LR_a	LR_b	SR	SR_b	W	T	T_b	LR_{boot}	SR_{boot}	W_{boot}	T_{boot}
	10	22.94	15.88	14.74	12.38	9.90	33.04	18.58	11.10	10.30	9.82	10.70	10.78
20	5	14.42	9.50	8.36	6.22	5.22	26.30	10.38	5.38	5.36	5.28	5.44	5.52
	1	5.60	2.64	2.24	0.70	0.76	15.46	2.06	1.06	1.02	1.34	1.26	1.58
	10	16.96	12.12	12.32	11.66	10.02	24.16	15.38	10.68	10.12	10.22	10.10	10.42
30	5	10.26	6.44	6.48	5.64	4.88	16.74	8.24	5.28	4.88	5.00	4.86	5.10
	1	2.98	1.40	1.50	1.04	1.02	7.84	1.68	1.16	1.20	1.22	1.18	1.40
	10	14.02	10.50	10.64	10.72	9.52	16.96	12.86	10.14	9.74	9.76	9.78	9.86
40	5	7.64	5.32	5.60	5.54	4.94	10.46	6.94	5.18	4.94	5.06	4.98	5.06
	1	1.84	1.08	1.12	0.98	0.90	3.74	1.40	0.98	1.06	1.24	0.98	1.16

Table 5. Rejection rates for $\mathcal{H}_0 : \lambda_2 = \cdots = \lambda_q = 0$; gamma distribution with $p = 2$ and $q = 3$.

n	α(%)	LR	LR_a	LR_b	SR	SR_b	W	T	T_b	LR_{boot}	SR_{boot}	W_{boot}	T_{boot}
	10	18.98	12.32	13.38	8.22	10.38	34.84	17.70	10.30	10.54	10.68	10.88	10.92
20	5	11.06	6.20	7.10	4.26	4.96	26.52	9.68	4.14	5.36	5.58	5.56	5.46
	1	3.10	1.50	1.68	0.72	0.88	14.92	2.46	0.40	1.40	1.14	1.48	1.30
	10	13.64	10.14	10.98	8.60	9.74	23.30	13.06	9.86	10.24	9.86	9.64	10.14
30	5	7.68	5.14	5.64	4.46	4.82	15.92	7.16	4.32	5.00	4.96	5.42	5.04
	1	2.00	1.20	1.26	0.92	0.80	7.04	1.82	0.60	1.28	1.20	1.06	1.32
	10	12.70	10.42	11.00	9.44	10.18	19.82	12.44	10.34	10.38	9.96	10.72	10.22
40	5	7.10	5.34	5.44	4.68	4.86	13.08	6.82	4.84	5.30	5.02	5.10	5.32
	1	1.78	1.38	1.42	1.24	1.02	4.58	1.68	0.88	1.22	1.38	1.30	1.20

Finally, we find that the corrected gradient test rates (T_b) are closer to the nominal levels than those presented by the gradient test (T). As an example, for $n = 20$, the rejection rates for the levels $\alpha = 10\%$ and $\alpha = 5\%$ are, respectively, 17.70% and 9.68% for T and 10.30% and 4.14% for T_b. The bootstrap tests perform satisfactorily.

In Table 6, we fix the null hypothesis as $\mathcal{H}_0 : \lambda_2 = \cdots = \lambda_q = 0$, $p = 1$, $n = 20, 30, 40$ and $q = 2, 3, 4$. The Wald test (W) is quite liberal, as the number of parameters increases. The likelihood ratio (LR), and gradient (T) tests are also liberal but have a smaller distorted size than the W test. Furthermore, the score (SR) and corrected gradient (T_b) tests are conservative, the latter in almost all cases. We also note that as the sample size increases, the tests' rejection rates approach the respective nominal levels. Among the modified likelihood ratio tests, the LR_a test presents slightly better results. Finally, the bootstrap and SR_b tests lead to a considerable reduction in size distortion.

In Table 7, we present the simulation results to evaluate the non-null rejection rates of the tests (power) LR_a, LR_b, SR_b, T_b, LR_{boot}, SR_{boot}, W_{boot} and T_{boot}. The LR, SR, W and T statistics are not included, because the simulation results showed they are oversized. The powers of the tests were calculated based on estimated critical values that make the size of each test equal to α. We set $n = 30$, $p = 1$, $q = 2$, $\alpha = 10\%$ and consider the hypothesis $\mathcal{H}_1 : \lambda_2 = \epsilon$ for $\epsilon = 0.5, 1.5, 2.5, 3.5$. As expected, the powers of the tests are very similar and increase as ϵ increases.

In Figure 1, we present the graph of the asymptotic quantiles versus the relative discrepancies of the quantiles. The relative discrepancy is defined as the difference between the exact quantile (estimated via simulation) and the asymptotic quantile divided by the latter. The closer the discrepancy is to zero, the better the approximation is of the exact distribution under the null hypothesis of the test statistic by the limit distribution χ^2. We consider the null hypothesis $\mathcal{H}_0 : \beta_3 = 0$, $p = q = 3$ and $n = 40$, that is, asymptotic quantiles are

Table 6. Rejection rates for $\mathcal{H}_0 : \lambda_2 = \cdots = \lambda_q = 0$; gamma distribution with $p = 1$.

q	n	$\alpha(\%)$	LR	LR_a	LR_b	SR	SR_b	W	T	T_b	LR_{boot}	SR_{boot}	W_{boot}	T_{boot}
		10	12.58	10.70	11.16	9.00	10.36	16.92	11.98	10.46	10.42	10.60	10.00	10.48
	20	5	6.80	5.54	5.82	4.16	5.52	10.48	6.30	4.94	5.28	5.76	5.50	5.36
		1	1.90	1.50	1.58	0.84	1.24	4.54	1.62	0.94	1.62	1.42	1.34	1.54
		10	10.54	9.28	9.18	8.40	9.32	13.78	10.34	9.04	9.22	9.56	9.48	9.22
2	30	5	5.44	4.72	4.78	3.90	4.62	7.86	5.30	4.50	4.78	4.96	4.74	4.78
		1	1.22	0.90	0.90	0.66	0.82	2.44	1.02	0.58	1.06	1.14	1.12	1.12
		10	11.52	10.20	10.50	9.20	10.04	13.66	11.24	10.20	10.48	10.20	10.22	10.44
	40	5	5.52	4.92	4.96	4.48	4.98	7.64	5.38	4.74	5.10	5.20	5.22	5.04
		1	1.22	0.94	0.92	0.90	1.04	2.16	1.06	0.72	1.08	1.24	0.94	1.10
		10	14.08	10.92	11.52	8.54	10.10	23.12	12.84	9.90	10.38	10.34	10.60	10.38
	20	5	7.36	5.62	5.96	4.02	4.90	16.14	6.54	4.54	5.34	5.48	5.70	5.22
		1	2.08	1.02	1.16	0.66	0.82	7.50	1.38	0.38	1.18	1.02	1.28	1.18
		10	11.50	9.94	10.30	8.94	10.10	17.98	11.08	9.36	9.82	10.14	9.86	9.80
3	30	5	6.34	5.24	5.44	4.24	4.92	11.30	5.92	4.52	5.20	5.12	4.74	5.20
		1	1.34	1.02	1.06	0.74	0.78	4.20	1.16	0.66	1.24	1.02	1.18	1.22
		10	11.30	9.90	9.98	9.22	10.06	15.82	11.06	9.62	9.88	10.08	9.90	9.86
	40	5	6.20	5.18	5.42	4.72	5.24	9.72	5.72	4.72	5.52	5.38	5.26	5.42
		1	1.34	1.00	1.08	0.94	1.00	3.46	1.16	0.68	1.12	1.24	1.32	1.10
		10	15.54	10.40	12.44	8.22	10.56	31.70	14.10	9.66	10.76	10.86	10.46	10.88
	20	5	8.58	5.44	6.50	4.16	5.00	23.80	7.28	4.14	5.36	5.36	5.30	5.24
		1	2.40	1.16	1.68	1.08	0.16	12.66	1.86	0.20	1.36	1.26	1.16	1.38
		10	12.46	10.30	10.80	8.82	9.82	22.64	11.76	9.46	10.04	10.02	10.00	10.14
4	30	5	6.66	5.12	5.52	4.32	4.86	14.76	6.16	4.12	5.04	5.20	5.06	5.20
		1	1.70	1.32	1.34	1.18	0.94	6.42	1.40	0.60	1.40	1.44	1.12	1.30
		10	12.08	10.68	10.70	9.68	10.70	19.62	11.64	9.92	10.56	10.82	10.14	10.52
	40	5	6.28	5.34	5.40	4.56	4.88	12.56	5.74	4.50	5.34	5.40	5.14	5.36
		1	1.42	1.16	1.26	1.18	0.90	4.44	1.30	0.64	1.40	1.28	1.52	1.34

Table 7. Power of the $\mathcal{H}_1 : \lambda_2 = \epsilon; n = 30, p = 1, q = 2, \alpha = 10\%$; gamma distribution.

ϵ	LR_a	LR_b	SR_b	T_b	LR_{boot}	SR_{boot}	W_{boot}	T_{boot}
0.5	16.06	16.32	15.84	16.04	15.14	15.16	14.82	15.10
1.5	42.76	43.16	41.40	42.84	41.38	40.18	40.08	41.46
2.5	75.78	76.06	72.66	75.84	74.80	71.64	74.54	74.90
3.5	94.06	94.12	91.30	94.06	93.82	90.84	94.40	93.78

obtained from the χ^2_1 distribution. The figure confirms the tendency of the Wald test (W) to be too liberal. The likelihood ratio (LR) and gradient (T) tests are also liberal but less so than W. The other tests reject the null hypothesis less frequently compared to the nominal level. Note that the distribution of the test T_b is closer to the reference distribution than the T distribution. The best agreement between the exact and asymptotic quantiles is obtained by T_b.

5. An application

To illustrate the application of the usual likelihood ratio, score, Wald and gradient statistics and their corrected versions, we consider the dataset analyzed in Simonoff and Tsai [29]. These data represent the monthly return of the market (x) and Acme-Cleveland Corporation (Y) stocks from January 1986 to December 1990.

In the original analysis of these data, the authors suggested a normal heteroscedastic regression model $N(\mu_\ell, \phi_\ell^{-1})$ to predict Y_ℓ in terms of a linear return function of the market x_ℓ, considering the following systematic components for the mean and precision:

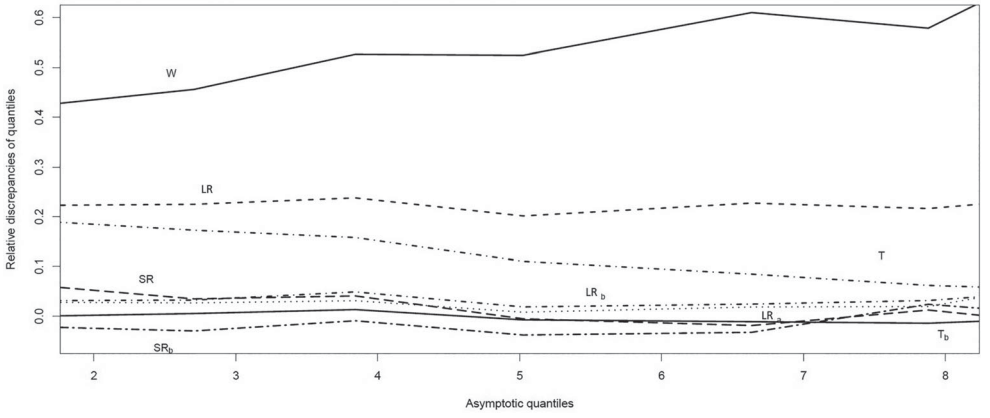

Figure 1. Relative discrepancies of quantiles – gamma model with $n = 40$, $p = 3$ and $q = 3$.

Table 8. Parameter estimates.

Parameter	β_1	β_2	λ_1	λ_2
Estimate	−0.005	1.253	4.410	−8.092
Standard error	0.020	0.249	0.265	4.009

Table 9. Values of test statistics and p-values of test $\mathcal{H}_0 : \lambda_2 = 0$ in the complete database.

Statistic	Value	P-value
LR	3.329	0.068
LR_a	3.095	0.079
LR_b	3.120	0.077
SR	2.698	0.100
SR_b	2.889	0.089
W	4.075	0.044
T	3.316	0.069
T_b	3.086	0.079
LR_{boot}	–	0.076
SR_{boot}	–	0.081
W_{boot}	–	0.077
T_{boot}	–	0.076

$\mu_\ell = \beta_1 + \beta_2 x_\ell$ and $\log(\phi_\ell) = \lambda_1 + \lambda_2 x_\ell$, for $\ell = 1, \ldots, 59$, where observation 22, which indicates that the market lost approximately one-third of its value during two weeks in October 1987 [29, p. 359], was removed from the analysis. Here, we are interested in testing the null hypothesis $\mathcal{H}_0 : \lambda_2 = 0$ (homoscedasticity).

The estimates of the parameters are shown in Table 8. We can observe in Table 9 that if we consider the nominal level of 5%, the null hypothesis is only rejected when the inference is based on the usual Wald test. However, if the nominal level considered is 10%, the null hypothesis is only not rejected by the test score.

Similar to Melo et al. [30], we randomly select a subset of the dataset, with $n = 20$ observations, to illustrate the effect of the different correction factors. Table 10 presents, for this reduced database, the results of the same tests performed for the complete database. Note

Table 10. Values of test statistics and p-values of test $\mathcal{H}_0 : \lambda_2 = 0$ in the reduced database.

Statistic	Value	P-value
LR	5.187	0.023
LR_{a}	3.129	0.077
LR_{b}	4.159	0.041
SR	3.147	0.076
SR_{b}	3.939	0.047
W	8.245	0.004
T	5.094	0.024
T_{b}	3.629	0.057
LR_{boot}	–	0.059
SR_{boot}	–	0.039
W_{boot}	–	0.106
T_{boot}	–	0.050

that the observed values of the test statistics LR, SR, W and T preserve the order shown in the simulation studies (see Section 4: $W > LR > T > SR$. For $\alpha = 0.05$, the conclusions are different between the LR, SR and T tests and their improved versions LR_a, SR_b and T_b, respectively.

6. Concluding remarks

Broadly speaking, the likelihood ratio, Wald and gradient tests may be too liberal when working with small sample sizes, rejecting the null hypothesis more frequently than expected based on the nominal level selected. The score test is the least distorted and can be conservative in some cases. The Bartlett-type correction factors for the score and gradient tests have empirical rejection rates closer to the nominal levels. This result is more pronounced concerning the usual gradient test. The adjusted and corrected likelihood ratio tests present greater size distortion than the corrected score and gradient tests. The bootstrap tests lead to a considerable reduction in observed size distortion for asymptotic tests, especially for the likelihood ratio and Wald tests, whose usual versions are typically liberal. However, they are costly computationally and they add two uncertainties: the number of replications and the size of each replication. On the other hand, the corrected score and gradient tests are less distorted, in most cases analyzed, and are therefore recommended for practical applications. For future work, we suggest analyzing the behaviour of the four tests when the systematic components, in (2) are nonlinear. Since, in a general setting, there is no Bartlett-type correction factor to improve the Wald test, a second suggestion is to derive a modified version of the Wald statistic based on the general result of Magalhães et al. [31].

Acknowledgments

This paper is based on the PhD dissertation of Dr. Fabiana U. Barros submitted to the Institute of Mathematics and Statistics, University of São Paulo, under the supervision of the second and third 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.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Denise A. Botter http://orcid.org/0000-0002-6537-6151
Tiago M. Magalhães http://orcid.org/0000-0003-3814-9532

References

[1] Terrell GR. The gradient statistic. Comput Sci Stat. 2002;34:206–215.
[2] Wilks SS. The large-sample distribution of the likelihood ratio for testing composite hypotheses. Ann Math Stat. 1938;9(1):60–62.
[3] Wald A. Tests of statistical hypotheses concerning several parameters when the number of observations is large. Trans Am Math Soc. 1943;54(3):426–482.
[4] Rao CR. Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. Math Proc Camb Philos Soc. 1948;44(1):50–57.
[5] Rao CR. Score test: historical review and recent developments. In: Balakrishnan N, Kannan N, and Nagaraja HN, editors. Advances in Ranking and Selection, Multiple Comparisons, and Reliability. Boston: Birkhuser; 2005. p. 3–20.
[6] Montoril MH, Souza EA. Estatística gradiente: propriedades e aplicações. Rev Bras Biometr. 2013;31(1):43–60.
[7] Lemonte AJ. The gradient test: another likelihood-based test. 1st ed. London: Academic Press; 2016.
[8] Vargas TM, Ferrari SLP, Lemonte AJ. Gradient statistic: higher-order asymptotics and bartlett-type correction. Electron J Stat. 2013;7:43–61.
[9] Vargas TM, Ferrari SLP, Lemonte AJ. Improved likelihood inference in generalized linear models. Comput Stat Data Anal. 2014;74:110–124.
[10] Medeiros FMC, Ferrari SLP, Lemonte AJ. Improved inference in dispersion models. Appl Math Model. 2017;51:317–328.
[11] Medeiros FMC, Lemonte AJ. Likelihood-based inference in censored exponential regression models. Commun Stat Theory Methods. 2021;50(13):3214–3233.
[12] Magalhães TM, Gallardo DI. Bartlett and bartlett-type corrections for censored data from a weibull distribution. SORT Stat Oper Res Trans. 2020;44(1):127–140.
[13] Botter DA, Cordeiro GM. Improved estimators for generalized linear models with dispersion covariates. J Stat Comput Simul. 1998;62(1–2):91–104.
[14] Barroso LP, Botter DA, Cordeiro GM. Second-order covariance matrix formula for heteroskedastic generalized linear models. Commun Stat Theory Methods. 2013;42(9): 1618–1627.
[15] Botter DA, Cordeiro GM. Bartlett corrections for generalized linear models with dispersion covariates. Commun Stat Theory Methods. 1997;26(2):279–307.
[16] Cordeiro GM, Botter DA, Barroso LP, et al. Three corrected score tests for generalized linear models with dispersion covariates. Commun Stat Theory Methods. 2003;57(4):391–409.
[17] Nelder JA, Wedderburn RWM. Generalized linear models. J R Stat Soc Ser A. 1972;135(3): 370–384.
[18] Brooks SM, Cunha R, Mosley L. Sovereign risk and government change: elections, ideology and experience. Comp Polit Stud. 2022;55(9):1501–1538.
[19] Hussain W, Campbell MT, Jarquin D, et al. Variance heterogeneity genome-wide mapping for cadmium in bread wheat reveals novel genomic loci and epistatic interactions. Plant Genome. 2020;13(1):e20011.
[20] Smyth GK, Huele AF, Verbyla AP. Exact and approximate reml for heteroscedastic regression. Stat Model. 2001;1(3):161–175.

[21] Smyth GK. Generalized linear models with varying dispersion. J R Stat Soc Ser B. 1989;51(1):47–60.
[22] Cox DR, Hinkley DV. Theoretical statistics. London: Chapman and Hall; 1974.
[23] Cox DR, Reid N. Parameter orthogonality and approximate conditional inference (with discussion). J R Stat Soc Ser B. 1987;49(1):1–39.
[24] Nocedal J, Wright S. Numerical optimization. 2nd. New York (NY): Springer; 2006.
[25] Doornik JA. Ox: an object-oriented matrix language. 5th ed. London: Timberlake Consultants Press; 2006.
[26] R Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2017.
[27] Ferrari SLP, A. Cysneiros AHM. Skovgaard's adjustment to likelihood ratio tests in exponential family nonlinear models. Stat Probab Lett. 2008;78:3047–3055.
[28] Efron B, Tibshirani RJ. An introduction to the bootstrap. New York: Chapman & Hall; 1993.
[29] Simonoff JS, Tsai CL. Use of modified profile likelihood for improved tests of constancy of variance in regression. Appl Stat. 1994;43(2):357–370.
[30] Melo TF, Vargas TM, Lemonte AJ, et al. Higher-order asymptotic refinements in the multivariate dirichlet regression model. Commun Stat Simul Comput. 2022;51(1):53–71.
[31] Magalhães TM, Botter DA, Sandoval MC. A general expression for second-order covariance matrices—an application to dispersion models. Braz J Probab Stat. 2021;35(1):37–49.

Appendix

In order to express the corrected gradient statistic, presented in (4), it is helpful to define the following matrices: $\boldsymbol{Z_\beta} = \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{W}\boldsymbol{\Phi}\boldsymbol{X})^{-1}\boldsymbol{X}^\top, \boldsymbol{Z_\lambda} = \boldsymbol{S}(-\boldsymbol{S}^\top \boldsymbol{D_2}\boldsymbol{\Phi}_\mathbf{1}^2\boldsymbol{S})^{-1}\boldsymbol{S}^\top, \boldsymbol{Z_{\beta d}} = \mathrm{diag}(z_{\beta 11}, \ldots, z_{\beta nn})$, $\boldsymbol{Z_{\lambda d}} = \mathrm{diag}(z_{\lambda 11}, \ldots, z_{\lambda nn})$, $\boldsymbol{Z_{\beta_2}} = \boldsymbol{X_2}(\boldsymbol{X_2}^\top \boldsymbol{W}\boldsymbol{\Phi}\boldsymbol{X_2})^{-1}\boldsymbol{X_2}^\top$, $\boldsymbol{Z_{\lambda_2}} = \boldsymbol{S_2}(-\boldsymbol{S_2}^\top \boldsymbol{D_2}\boldsymbol{\Phi}_\mathbf{1}^2\boldsymbol{S_2})^{-1}\boldsymbol{S_2}^\top$, $\boldsymbol{Z_{\beta_2 d}} = \mathrm{diag}(z_{\beta_2 11}, \ldots, z_{\beta_2 nn})$, $\boldsymbol{Z_{\lambda_2 d}} = \mathrm{diag}(z_{\lambda_2 11}, \ldots, z_{\lambda_2 nn})$, $\boldsymbol{F} = \mathrm{diag}\{f_1, \ldots, f_n\}$, $\boldsymbol{G} = \mathrm{diag}\{g_1, \ldots, g_n\}$, $\boldsymbol{B} = \mathrm{diag}\{b_1, \ldots, b_n\}$, $\boldsymbol{C} = \mathrm{diag}\{c_1, \ldots, c_n\}$, $\boldsymbol{E} = \mathrm{diag}\{e_1, \ldots, e_n\}$, $\boldsymbol{\Phi_1} = \mathrm{diag}\{\phi_{11}, \ldots, \phi_{1n}\}$, $\boldsymbol{\Phi_2} = \mathrm{diag}\{\phi_{21}, \ldots, \phi_{2n}\}, \boldsymbol{\Phi_3} = \mathrm{diag}\{\phi_{31}, \ldots, \phi_{3n}\}, \boldsymbol{D_2} = \mathrm{diag}\{d_{21}, \ldots, d_{2n}\}, \boldsymbol{D_3} = \mathrm{diag}\{d_{31}, \ldots, d_{3n}\}$ and $\boldsymbol{D_4} = \mathrm{diag}\{d_{41}, \ldots, d_{4n}\}$, with $f_\ell = \frac{1}{V_\ell}\frac{\mathrm{d}\mu_\ell}{\mathrm{d}\eta_\ell}\frac{\mathrm{d}^2\mu_\ell}{\mathrm{d}\eta_\ell^2}$, $g_\ell = f_\ell - \frac{1}{V_\ell^2}\frac{\mathrm{d}V_\ell}{\mathrm{d}\mu_\ell}(\frac{\mathrm{d}\mu_\ell}{\mathrm{d}\eta_\ell})^3$, $\nu_{1\ell} = \frac{1}{V_\ell^3}(\frac{\mathrm{d}V_\ell}{\mathrm{d}\mu_\ell})^2(\frac{\mathrm{d}\mu_\ell}{\mathrm{d}\eta_\ell})^4$, $\nu_{2\ell} = \frac{1}{V_\ell^2}\frac{\mathrm{d}^2V_\ell}{\mathrm{d}\mu_\ell^2}(\frac{\mathrm{d}\mu_\ell}{\mathrm{d}\eta_\ell})^4$, $\nu_{3\ell} = \frac{1}{V_\ell^2}\frac{\mathrm{d}V_\ell}{\mathrm{d}\mu_\ell}(\frac{\mathrm{d}\mu_\ell}{\mathrm{d}\eta_\ell})^2\frac{\mathrm{d}^2\mu_\ell}{\mathrm{d}\eta_\ell^2}$, $\nu_{4\ell} = \frac{1}{V_\ell}(\frac{\mathrm{d}^2\mu_\ell}{\mathrm{d}\eta_\ell^2})^2$, $\nu_{5\ell} = \frac{1}{V_\ell}\frac{\mathrm{d}\mu_\ell}{\mathrm{d}\eta_\ell}\frac{\mathrm{d}^3\mu_\ell}{\mathrm{d}\eta_\ell^3}$, $b_\ell = -6\nu_{1\ell} + 3\nu_{2\ell} + 12\nu_{3\ell} - 3\nu_{4\ell} - 4\nu_{5\ell}, c_\ell = -4\nu_{1\ell} + 2\nu_{2\ell} + 9\nu_{3\ell} - 3\nu_{4\ell} - 3\nu_{5\ell}, e_\ell = -2\nu_{1\ell} + \nu_{2\ell} + 5\nu_{3\ell} - 2\nu_{4\ell} - 2\nu_{5\ell}$ and $\phi_{k\ell} = \mathrm{d}^k\phi_\ell/\mathrm{d}\tau_\ell^k$ and $d_{k\ell} = \mathrm{d}^k d(\phi_\ell)/\mathrm{d}\phi_\ell^k$, for $k = 1, 2, 3, 4$ and $\ell = 1, \ldots, n$. $\boldsymbol{Z_\beta^{(2)}} = \boldsymbol{Z_\beta} \odot \boldsymbol{Z_\beta}, \boldsymbol{Z_\beta^{(3)}} = \boldsymbol{Z_\beta^{(2)}} \odot \boldsymbol{Z_\beta}$ etc., where $\odot$ represents a direct product.

The quantities A_1 to A_3, where $A_1 = A_{11} + A_{12} + A_{13} + A_{14}$ and $A_2 = A_{21} + A_{22} + A_{23}$, in Equation (4), to define an improved gradient statistic are, respectively:

$$
\begin{aligned}
A_{11} = {} & 3\{\mathbf{1}^\top \boldsymbol{\Phi}(\boldsymbol{F} + 2\boldsymbol{G})[\boldsymbol{Z_{\beta d}}(\boldsymbol{Z_\beta} + \boldsymbol{Z_{\beta_2}})\boldsymbol{Z_{\beta_2 d}} - 2\boldsymbol{Z_{\beta_2 d}}\boldsymbol{Z_{\beta_2}}\boldsymbol{Z_{\beta_2 d}} + 2(\boldsymbol{Z_\beta} - \boldsymbol{Z_{\beta_2}}) \odot \boldsymbol{Z_{\beta_2}^{(2)}}] \\
& \times \boldsymbol{\Phi}(\boldsymbol{F} + 2\boldsymbol{G})\mathbf{1}\} \\
& + 3\{\mathbf{1}^\top \boldsymbol{\Phi_1}\boldsymbol{W}[\boldsymbol{Z_{\beta d}}(\boldsymbol{Z_\lambda} + \boldsymbol{Z_{\lambda_2}})\boldsymbol{Z_{\beta_2 d}} - 2\boldsymbol{Z_{\beta_2 d}}\boldsymbol{Z_{\lambda_2}}\boldsymbol{Z_{\beta_2 d}} + 4(\boldsymbol{Z_\beta} - \boldsymbol{Z_{\beta_2}}) \odot \boldsymbol{Z_{\beta_2}} \odot \boldsymbol{Z_{\lambda_2}} \\
& + 2(\boldsymbol{Z_\lambda} - \boldsymbol{Z_{\lambda_2}}) \odot \boldsymbol{Z_{\beta_2}^{(2)}}]\boldsymbol{\Phi_1}\boldsymbol{W}\mathbf{1}\} \\
& - 3\{\mathbf{1}^\top \boldsymbol{\Phi_1}\boldsymbol{W}[\boldsymbol{Z_{\beta d}}(\boldsymbol{Z_\lambda} + \boldsymbol{Z_{\lambda_2}})\boldsymbol{Z_{\lambda_2 d}} - 2\boldsymbol{Z_{\beta_2 d}}\boldsymbol{Z_{\lambda_2}}\boldsymbol{Z_{\lambda_2 d}}](\boldsymbol{D_3}\boldsymbol{\Phi}_\mathbf{1}^3 + 3\boldsymbol{D_2}\boldsymbol{\Phi_1}\boldsymbol{\Phi_2})\mathbf{1}\} \\
& - 3\{\mathbf{1}^\top (\boldsymbol{D_3}\boldsymbol{\Phi}_\mathbf{1}^3 + 3\boldsymbol{D_2}\boldsymbol{\Phi_1}\boldsymbol{\Phi_2})[\boldsymbol{Z_{\lambda d}}(\boldsymbol{Z_\lambda} + \boldsymbol{Z_{\lambda_2}})\boldsymbol{Z_{\beta_2 d}} - 2\boldsymbol{Z_{\lambda_2 d}}\boldsymbol{Z_{\lambda_2}}\boldsymbol{Z_{\beta_2 d}}]\boldsymbol{\Phi_1}\boldsymbol{W}\mathbf{1}\} \\
& + 3\{\mathbf{1}^\top (\boldsymbol{D_3}\boldsymbol{\Phi}_\mathbf{1}^3 + 3\boldsymbol{D_2}\boldsymbol{\Phi_1}\boldsymbol{\Phi_2})[\boldsymbol{Z_{\lambda d}}(\boldsymbol{Z_\lambda} + \boldsymbol{Z_{\lambda_2}})\boldsymbol{Z_{\lambda_2 d}} - 2\boldsymbol{Z_{\lambda_2 d}}\boldsymbol{Z_{\lambda_2}}\boldsymbol{Z_{\lambda_2 d}} \\
& + 2(\boldsymbol{Z_\lambda} - \boldsymbol{Z_{\lambda_2}}) \odot \boldsymbol{Z_{\lambda_2}^{(2)}}](\boldsymbol{D_3}\boldsymbol{\Phi}_\mathbf{1}^3 + 3\boldsymbol{D_2}\boldsymbol{\Phi_1}\boldsymbol{\Phi_2})\mathbf{1}\}, \\
A_{12} = {} & 12\{\mathbf{1}^\top \boldsymbol{\Phi}(\boldsymbol{F} + \boldsymbol{G})[\boldsymbol{Z_{\beta d}}\boldsymbol{Z_\beta}\boldsymbol{Z_{\beta d}} - \boldsymbol{Z_{\beta_2 d}}\boldsymbol{Z_{\beta_2}}\boldsymbol{Z_{\beta_2 d}} + \boldsymbol{Z_\beta^{(3)}} - \boldsymbol{Z_{\beta_2}^{(3)}}]\boldsymbol{\Phi}(\boldsymbol{F} + \boldsymbol{G})\mathbf{1}\}
\end{aligned}
$$

$$
\begin{aligned}
&+ 12\{\mathbf{1}^\top(\boldsymbol{D}_3\boldsymbol{\Phi}_1^3 + 2\boldsymbol{D}_2\boldsymbol{\Phi}_1\boldsymbol{\Phi}_2)[\boldsymbol{Z}_{\lambda d}\boldsymbol{Z}_\lambda\boldsymbol{Z}_{\lambda d} - \boldsymbol{Z}_{\lambda_2 d}\boldsymbol{Z}_{\lambda_2}\boldsymbol{Z}_{\lambda_2 d} + \boldsymbol{Z}_\lambda^{(3)} - \boldsymbol{Z}_{\lambda_2}^{(3)}]\\
&\times(\boldsymbol{D}_3\boldsymbol{\Phi}_1^3 + 2\boldsymbol{D}_2\boldsymbol{\Phi}_1\boldsymbol{\Phi}_2)\mathbf{1}\},\\
A_{13} = &-6\{\mathbf{1}^\top\boldsymbol{\Phi}(\boldsymbol{F}+2\boldsymbol{G})[(\boldsymbol{Z}_\beta^{(2)} - \boldsymbol{Z}_{\beta_2}^{(2)}) \odot (\boldsymbol{Z}_\beta + \boldsymbol{Z}_{\beta_2}) + (\boldsymbol{Z}_{\beta d} - 3\boldsymbol{Z}_{\beta_2 d})\boldsymbol{Z}_{\beta_2}\boldsymbol{Z}_{\beta_2 d}\\
&+ (\boldsymbol{Z}_{\beta d} + \boldsymbol{Z}_{\beta_2 d})\boldsymbol{Z}_\beta\boldsymbol{Z}_{\beta d} + 2\boldsymbol{Z}_{\beta_2}^{(2)} \odot (\boldsymbol{Z}_\beta - \boldsymbol{Z}_{\beta_2})]\boldsymbol{\Phi}(\boldsymbol{F}+\boldsymbol{G})\mathbf{1}\}\\
&- 6\{\mathbf{1}^\top\boldsymbol{\Phi}_1\boldsymbol{W}[(\boldsymbol{Z}_\beta^{(2)} - \boldsymbol{Z}_{\beta_2}^{(2)}) \odot (\boldsymbol{Z}_\lambda + \boldsymbol{Z}_{\lambda_2}) + 2\boldsymbol{Z}_{\beta_2}^{(2)} \odot (\boldsymbol{Z}_\lambda - \boldsymbol{Z}_{\lambda_2})]\boldsymbol{\Phi}_1\boldsymbol{W}\mathbf{1}\}\\
&+ 6\{\mathbf{1}^\top\boldsymbol{\Phi}_1\boldsymbol{W}[(\boldsymbol{Z}_{\beta d} + \boldsymbol{Z}_{\beta_2 d})\boldsymbol{Z}_\lambda\boldsymbol{Z}_{\lambda d} + (\boldsymbol{Z}_{\beta d} - 3\boldsymbol{Z}_{\beta_2 d})\boldsymbol{Z}_{\lambda_2}\boldsymbol{Z}_{\lambda_2 d}](\boldsymbol{D}_3\boldsymbol{\Phi}_1^3 + 2\boldsymbol{D}_2\boldsymbol{\Phi}_1\boldsymbol{\Phi}_2)\mathbf{1}\}\\
&- 6\{\mathbf{1}^\top(\boldsymbol{D}_3\boldsymbol{\Phi}_1^3 + 3\boldsymbol{D}_2\boldsymbol{\Phi}_1\boldsymbol{\Phi}_2)[(\boldsymbol{Z}_\lambda^{(2)} - \boldsymbol{Z}_{\lambda_2}^{(2)}) \odot (\boldsymbol{Z}_\lambda + \boldsymbol{Z}_{\lambda_2}) + (\boldsymbol{Z}_{\lambda d} - 3\boldsymbol{Z}_{\lambda_2 d})\boldsymbol{Z}_{\lambda_2}\boldsymbol{Z}_{\lambda_2 d}\\
&+ (\boldsymbol{Z}_{\lambda d} + \boldsymbol{Z}_{\lambda_2 d})\boldsymbol{Z}_\lambda\boldsymbol{Z}_{\lambda d} + 2\boldsymbol{Z}_{\lambda_2}^{(2)} \odot (\boldsymbol{Z}_\lambda - \boldsymbol{Z}_{\lambda_2})](\boldsymbol{D}_3\boldsymbol{\Phi}_1^3 + 2\boldsymbol{D}_2\boldsymbol{\Phi}_1\boldsymbol{\Phi}_2)\mathbf{1}\},\\
A_{14} = &\ 6\{\mathbf{1}^\top\boldsymbol{\Phi}\boldsymbol{B}(\boldsymbol{Z}_{\beta d} - \boldsymbol{Z}_{\beta_2 d})\boldsymbol{Z}_{\beta_2 d}\mathbf{1}\} + 6\{\mathbf{1}^\top\boldsymbol{\Phi}_2\boldsymbol{W}[(\boldsymbol{Z}_{\beta d} - \boldsymbol{Z}_{\beta_2 d})\boldsymbol{Z}_{\lambda d} + (\boldsymbol{Z}_{\lambda d} - \boldsymbol{Z}_{\lambda_2 d})\boldsymbol{Z}_{\beta_2 d}]\mathbf{1}\}\\
&+ 6\{\mathbf{1}^\top(\boldsymbol{D}_4\boldsymbol{\Phi}_1^4 + 6\boldsymbol{D}_3\boldsymbol{\Phi}_1^2\boldsymbol{\Phi}_2 + 3\boldsymbol{D}_2\boldsymbol{\Phi}_2^2 + 4\boldsymbol{D}_2\boldsymbol{\Phi}_1\boldsymbol{\Phi}_3)(\boldsymbol{Z}_{\lambda d} - \boldsymbol{Z}_{\lambda_2 d})\boldsymbol{Z}_{\lambda_2 d}\mathbf{1}\}\\
&- 6\{\mathbf{1}^\top\boldsymbol{\Phi}\boldsymbol{C}(\boldsymbol{Z}_{\beta d} - \boldsymbol{Z}_{\beta_2 d})(3\boldsymbol{Z}_{\beta_2 d} + \boldsymbol{Z}_{\beta d})\mathbf{1}\}\\
&- 6\{\mathbf{1}^\top(\boldsymbol{D}_4\boldsymbol{\Phi}_1^4 + 6\boldsymbol{D}_3\boldsymbol{\Phi}_1^2\boldsymbol{\Phi}_2 + 3\boldsymbol{D}_2\boldsymbol{\Phi}_2^2 + 3\boldsymbol{D}_2\boldsymbol{\Phi}_1\boldsymbol{\Phi}_3)(\boldsymbol{Z}_{\lambda d} - \boldsymbol{Z}_{\lambda_2 d})(\boldsymbol{Z}_{\lambda d} + 3\boldsymbol{Z}_{\lambda_2 d})\mathbf{1}\}\\
&+ 12\{\mathbf{1}^\top\boldsymbol{\Phi}\boldsymbol{E}(\boldsymbol{Z}_{\beta d}^{(2)} - \boldsymbol{Z}_{\beta_2 d}^{(2)})\mathbf{1}\}\\
&+ 12\{\mathbf{1}^\top(\boldsymbol{D}_4\boldsymbol{\Phi}_1^4 + 5\boldsymbol{D}_3\boldsymbol{\Phi}_1^2\boldsymbol{\Phi}_2 + 2\boldsymbol{D}_2\boldsymbol{\Phi}_2^2 + 2\boldsymbol{D}_2\boldsymbol{\Phi}_1\boldsymbol{\Phi}_3)(\boldsymbol{Z}_{\lambda d}^{(2)} - \boldsymbol{Z}_{\lambda_2 d}^{(2)})\mathbf{1}\},\\
A_{21} = &-3\left\{\mathbf{1}^\top\boldsymbol{\Phi}(\boldsymbol{F}+2\boldsymbol{G})\left[\frac{1}{4}(\boldsymbol{Z}_{\beta d} - \boldsymbol{Z}_{\beta_2 d})(\boldsymbol{Z}_\beta - \boldsymbol{Z}_{\beta_2})(3\boldsymbol{Z}_{\beta d} + \boldsymbol{Z}_{\beta_2 d})\right.\right.\\
&\left.\left.+(\boldsymbol{Z}_{\beta d} - \boldsymbol{Z}_{\beta_2 d})\boldsymbol{Z}_{\beta_2}(\boldsymbol{Z}_{\beta d} - \boldsymbol{Z}_{\beta_2 d}) + \frac{1}{2}(\boldsymbol{Z}_\beta - \boldsymbol{Z}_{\beta_2})^{(2)} \odot (\boldsymbol{Z}_\beta + 3\boldsymbol{Z}_{\beta_2})\right]\boldsymbol{\Phi}(\boldsymbol{F}+2\boldsymbol{G})\mathbf{1}\right\}\\
&- 3\left\{\mathbf{1}^\top\boldsymbol{\Phi}_1\boldsymbol{W}\left[\frac{1}{4}(\boldsymbol{Z}_{\beta d} - \boldsymbol{Z}_{\beta_2 d})(\boldsymbol{Z}_\lambda - \boldsymbol{Z}_{\lambda_2})(3\boldsymbol{Z}_{\beta d} + \boldsymbol{Z}_{\beta_2 d}) + (\boldsymbol{Z}_{\beta d} - \boldsymbol{Z}_{\beta_2 d})\boldsymbol{Z}_{\lambda_2}(\boldsymbol{Z}_{\beta d} - \boldsymbol{Z}_{\beta_2 d})\right.\right.\\
&\left.\left.+\frac{1}{2}(\boldsymbol{Z}_\beta - \boldsymbol{Z}_{\beta_2})^{(2)} \odot (\boldsymbol{Z}_\lambda + 3\boldsymbol{Z}_{\lambda_2}) + (\boldsymbol{Z}_\beta - \boldsymbol{Z}_{\beta_2}) \odot (\boldsymbol{Z}_\lambda - \boldsymbol{Z}_{\lambda_2}) \odot (\boldsymbol{Z}_\beta + 3\boldsymbol{Z}_{\beta_2})\right]\right.\\
&\left.\times\boldsymbol{\Phi}_1\boldsymbol{W}\mathbf{1}\right\}\\
&+ 3\left\{\mathbf{1}^\top\boldsymbol{\Phi}_1\boldsymbol{W}(\boldsymbol{Z}_{\beta d} - \boldsymbol{Z}_{\beta_2 d})\left[\frac{1}{4}(\boldsymbol{Z}_\lambda - \boldsymbol{Z}_{\lambda_2})(3\boldsymbol{Z}_{\lambda d} + \boldsymbol{Z}_{\lambda_2 d}) + \boldsymbol{Z}_{\lambda_2}(\boldsymbol{Z}_{\lambda d} - \boldsymbol{Z}_{\lambda_2 d})\right]\right.\\
&\left.\times(\boldsymbol{D}_3\boldsymbol{\Phi}_1^3 + 3\boldsymbol{D}_2\boldsymbol{\Phi}_1\boldsymbol{\Phi}_2)\mathbf{1}\right\}\\
&+ 3\left\{\mathbf{1}^\top(\boldsymbol{D}_3\boldsymbol{\Phi}_1^3 + 3\boldsymbol{D}_2\boldsymbol{\Phi}_1\boldsymbol{\Phi}_2)(\boldsymbol{Z}_{\lambda d} - \boldsymbol{Z}_{\lambda_2 d})\left[\frac{1}{4}(\boldsymbol{Z}_\lambda - \boldsymbol{Z}_{\lambda_2})(3\boldsymbol{Z}_{\beta d} + \boldsymbol{Z}_{\beta_2 d})\right.\right.\\
&\left.\left.+\boldsymbol{Z}_{\lambda_2}(\boldsymbol{Z}_{\beta d} - \boldsymbol{Z}_{\beta_2 d})\right]\boldsymbol{\Phi}_1\boldsymbol{W}\mathbf{1}\right\}\\
&- 3\left\{\mathbf{1}^\top(\boldsymbol{D}_3\boldsymbol{\Phi}_1^3 + 3\boldsymbol{D}_2\boldsymbol{\Phi}_1\boldsymbol{\Phi}_2)\left[\frac{1}{4}(\boldsymbol{Z}_{\lambda d} - \boldsymbol{Z}_{\lambda_2 d})(\boldsymbol{Z}_\lambda - \boldsymbol{Z}_{\lambda_2})(3\boldsymbol{Z}_{\lambda d} + \boldsymbol{Z}_{\lambda_2 d})\right.\right.\\
&\left.\left.+(\boldsymbol{Z}_{\lambda d} - \boldsymbol{Z}_{\lambda_2 d})\boldsymbol{Z}_{\lambda_2}(\boldsymbol{Z}_{\lambda d} - \boldsymbol{Z}_{\lambda_2 d}) + \frac{1}{2}(\boldsymbol{Z}_\lambda - \boldsymbol{Z}_{\lambda_2})^{(2)} \odot (\boldsymbol{Z}_\lambda + 3\boldsymbol{Z}_{\lambda_2})\right]\right.
\end{aligned}
$$

$$\times (\boldsymbol{D}_3\boldsymbol{\Phi}_1^3 + 3\boldsymbol{D}_2\boldsymbol{\Phi}_1\boldsymbol{\Phi}_2)1\Big\},$$

$$\begin{aligned}
A_{22} = {} & 6\{\mathbf{1}^\top\boldsymbol{\Phi}(\boldsymbol{F}+2\boldsymbol{G})[(\boldsymbol{Z}_\beta - \boldsymbol{Z}_{\beta_2}) \odot (\boldsymbol{Z}_\beta^{(2)} - \boldsymbol{Z}_{\beta_2}^{(2)}) + (\boldsymbol{Z}_{\beta d} - \boldsymbol{Z}_{\beta_2 d})(\boldsymbol{Z}_\beta\boldsymbol{Z}_{\beta d} - \boldsymbol{Z}_{\beta_2}\boldsymbol{Z}_{\beta_2 d})] \\
& \times \boldsymbol{\Phi}(\boldsymbol{F}+\boldsymbol{G})\mathbf{1}\} \\
& + 6\{\mathbf{1}^\top\boldsymbol{\Phi}_1\boldsymbol{W}[(\boldsymbol{Z}_\lambda - \boldsymbol{Z}_{\lambda_2}) \odot (\boldsymbol{Z}_\beta^{(2)} - \boldsymbol{Z}_{\beta_2}^{(2)})]\boldsymbol{\Phi}_1\boldsymbol{W}\mathbf{1}\} \\
& + 6\{\mathbf{1}^\top(\boldsymbol{D}_3\boldsymbol{\Phi}_1^3 + 3\boldsymbol{D}_2\boldsymbol{\Phi}_1\boldsymbol{\Phi}_2)[(\boldsymbol{Z}_\lambda - \boldsymbol{Z}_{\lambda_2}) \odot (\boldsymbol{Z}_\lambda^{(2)} - \boldsymbol{Z}_{\lambda_2}^{(2)}) \\
& + (\boldsymbol{Z}_{\lambda d} - \boldsymbol{Z}_{\lambda_2 d})(\boldsymbol{Z}_\lambda\boldsymbol{Z}_{\lambda d} - \boldsymbol{Z}_{\lambda_2}\boldsymbol{Z}_{\lambda_2 d})](\boldsymbol{D}_3\boldsymbol{\Phi}_1^3 + 2\boldsymbol{D}_2\boldsymbol{\Phi}_1\boldsymbol{\Phi}_2)\mathbf{1}\} \\
& - 6\{\mathbf{1}^\top\boldsymbol{\Phi}_1\boldsymbol{W}(\boldsymbol{Z}_{\beta d} - \boldsymbol{Z}_{\beta_2 d})(\boldsymbol{Z}_\lambda\boldsymbol{Z}_{\lambda d} - \boldsymbol{Z}_{\lambda_2}\boldsymbol{Z}_{\lambda_2 d})(\boldsymbol{D}_3\boldsymbol{\Phi}_1^3 + 2\boldsymbol{D}_2\boldsymbol{\Phi}_1\boldsymbol{\Phi}_2)\mathbf{1}\}, \\
A_{23} = {} & 3\{\mathbf{1}^\top\boldsymbol{\Phi}(2\boldsymbol{C} - \boldsymbol{B})(\boldsymbol{Z}_{\beta d} - \boldsymbol{Z}_{\beta_2 d})^{(2)}\mathbf{1}\} \\
& + 3\{\mathbf{1}^\top(\boldsymbol{D}_4\boldsymbol{\Phi}_1^4 + 6\boldsymbol{D}_3\boldsymbol{\Phi}_1^2\boldsymbol{\Phi}_2 + 3\boldsymbol{D}_2\boldsymbol{\Phi}_2^2 + 2\boldsymbol{D}_2\boldsymbol{\Phi}_1\boldsymbol{\Phi}_3)(\boldsymbol{Z}_{\lambda d} - \boldsymbol{Z}_{\lambda_2 d})^{(2)}\mathbf{1}\}, \\
A_3 = {} & \mathbf{1}^\top\boldsymbol{\Phi}(\boldsymbol{F}+2\boldsymbol{G})\left[\frac{3}{4}(\boldsymbol{Z}_{\beta d} - \boldsymbol{Z}_{\beta_2 d})(\boldsymbol{Z}_\beta - \boldsymbol{Z}_{\beta_2})(\boldsymbol{Z}_{\beta d} - \boldsymbol{Z}_{\beta_2 d}) + \frac{1}{2}(\boldsymbol{Z}_\beta - \boldsymbol{Z}_{\beta_2})^{(3)}\right]\boldsymbol{\Phi}(\boldsymbol{F}+2\boldsymbol{G})\mathbf{1} \\
& + \mathbf{1}^\top\boldsymbol{\Phi}_1\boldsymbol{W}\left[\frac{3}{4}(\boldsymbol{Z}_{\beta d} - \boldsymbol{Z}_{\beta_2 d})(\boldsymbol{Z}_\lambda - \boldsymbol{Z}_{\lambda_2})(\boldsymbol{Z}_{\beta d} - \boldsymbol{Z}_{\beta_2 d}) + \frac{3}{2}(\boldsymbol{Z}_\beta - \boldsymbol{Z}_{\beta_2})^{(2)} \odot (\boldsymbol{Z}_\lambda - \boldsymbol{Z}_{\lambda_2})\right] \\
& \times \boldsymbol{\Phi}_1\boldsymbol{W}\mathbf{1} \\
& - \frac{3}{2}\mathbf{1}^\top\boldsymbol{\Phi}_1\boldsymbol{W}(\boldsymbol{Z}_{\beta d} - \boldsymbol{Z}_{\beta_2 d})(\boldsymbol{Z}_\lambda - \boldsymbol{Z}_{\lambda_2})(\boldsymbol{Z}_{\lambda d} - \boldsymbol{Z}_{\lambda_2 d})(\boldsymbol{D}_3\boldsymbol{\Phi}_1^3 + 3\boldsymbol{D}_2\boldsymbol{\Phi}_1\boldsymbol{\Phi}_2)\mathbf{1} \\
& + \mathbf{1}^\top(\boldsymbol{D}_3\boldsymbol{\Phi}_1^3 + 3\boldsymbol{D}_2\boldsymbol{\Phi}_1\boldsymbol{\Phi}_2)\left[\frac{3}{4}(\boldsymbol{Z}_{\lambda d} - \boldsymbol{Z}_{\lambda_2 d})(\boldsymbol{Z}_\lambda - \boldsymbol{Z}_{\lambda_2})(\boldsymbol{Z}_{\lambda d} - \boldsymbol{Z}_{\lambda_2 d}) + \frac{1}{2}(\boldsymbol{Z}_\lambda - \boldsymbol{Z}_{\lambda_2})^{(3)}\right] \\
& \times (\boldsymbol{D}_3\boldsymbol{\Phi}_1^3 + 3\boldsymbol{D}_2\boldsymbol{\Phi}_1\boldsymbol{\Phi}_2)\mathbf{1}.
\end{aligned}$$

Once again, although deriving the expressions for A_1 to A_3 entails a great deal of algebra, these expressions only involve a repetition of matrices, $\boldsymbol{F}$, $\boldsymbol{G}$ and $\boldsymbol{W}$, for instance, in simple operations of diagonal matrices, i.e. they are simple expressions to implement.