UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA
DOUTORADO EM ESTATÍSTICA

ANA CRISTINA GUEDES PEREIRA

ESSAYS ON REGRESSION MODELS FOR DOUBLE BOUNDED AND EXTREME-VALUE RANDOM VARIABLES: IMPROVED TESTING INFERENCES AND EMPIRICAL ANALYSES

Recife

2022

ANA CRISTINA GUEDES PEREIRA

ESSAYS ON REGRESSION MODELS FOR DOUBLE BOUNDED AND EXTREME-VALUE RANDOM VARIABLES: IMPROVED TESTING INFERENCES AND EMPIRICAL ANALYSES

Doctoral dissertation submitted to the Graduate Program in Statistics, Department of Statistics, Universidade Federal de Pernambuco as a requirement to obtain a Doctoral degree in Statistics. Concentration area: Mathematical Statistics

Advisor: Francisco Cribari Neto
Co-advisor: Patrícia Leone Espinheira

Recife

2022

Catalogação na fonte
Bibliotecária Nataly Soares Leite Moro, CRB4-1722

P436e Pereira, Ana Cristina Guedes
Essays on regression models for double bounded and extreme-value random variables: improved testing inferences and empirical analyses / Ana Cristina Guedes Pereira. – 2022.
95 f.: il., fig., tab.

Orientador: Francisco Cribari Neto.
Tese (Doutorado) – Universidade Federal de Pernambuco. CCEN, Estatística, Recife, 2022.
Inclui referências e apêndices.

1. Estatística matemática. 2. COVID-19. 3. Correção de Bartlett. 4. Regressão beta. I. Cribari Neto, Francisco (orientador). II. Título.

519.5 CDD (23. ed.) UFPE- CCEN 2022 - 45

ANA CRISTINA GUEDES PEREIRA

ESSAYS ON REGRESSION MODELS FOR DOUBLE BOUNDED AND EXTREME-VALUE RANDOM VARIABLES: IMPROVED TESTING INFERENCES AND EMPIRICAL ANALYSES

Tese apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Doutor em Estatística.

Aprovada em: 21 de fevereiro de 2022.

BANCA EXAMINADORA

Prof. Dr.Francisco Cribari Neto
UFPE

Profª Drª. Fernanda De Bastiani
UFPE

Prof. Dr. Raydonal Ospina Martínez
UFPE

Prof.Dr. Fábio Mariano Bayer
UFSM

Prof. Dr. Gilberto Alvarenga Paula
USP

To God and my mother.

AGRADECIMENTOS

A minha família, em especial a minha mãe, Neide Guedes, que, mesmo possuindo apenas o ensino fundamental incompleto e sendo mãe solo desde minha infância, contribuiu imensamente para minha formação moral e profissional; mulher guerreira e meu maior exemplo de vida e força. Agradeço-lhe pelo amor incondicional, pelo imenso incentivo e por sempre acreditar em mim. Agradeço também as minhas irmãs, Jéssica Guedes e Cecília Guedes, pelo forte apoio, por sempre acreditarem no meu sucesso, por todo cuidado, pelo carinho, pelos momentos de alegria e por me ajudarem a enfrentar dificuldades pessoais, que aconteceram ocasionalmente no decorrer da elaboração da presente tese. Minha mãe e minhas irmãs têm sido fundamentais para que eu consiga alcançar meus objetivos. Assim, dedico a elas todas as minhas conquistas.

A meu orientador, professor Francisco Cribari Neto, pela incrível compreensão, grande apoio, carinho, estímulo, confiança, constante incentivo, amizade, paciência, atenção, excelente orientação e por todos os ensinamentos acadêmicos, em especial pelos grandes esforços e acolhimento nos muitos momentos difíceis. Minha eterna gratidão pela dedicação a mim dedicada na superação de obstáculos. Ter a honra de defender minha tese de doutorado sendo orientada por minha principal referência como pesquisador e professor é algo indescritível.

A minha coorientadora, professora Patrícia Leone Espinheira, pela enorme atenção, afeto, amizade, parceria e cumplicidade, pelos bons momentos de descontração, pelas histórias compartilhadas, pelas emoções vividas e pela grande contribuição e entrega como coorientadora e amiga. Agradeço-lhe imensamente pelo suporte pessoal e profissional, que foram fundamentais. Tenho imensa admiração e inspiração por sua trajetória incrível como pesquisadora e professora.

Aos professores do Programa de Pós-Graduação em Estatística da UFPE, em especial àqueles de quem eu tive a honra de ter sido aluna: professores Alex Ramos, Audrey Cysneiros, Gauss Cordeiro, Francisco Cribari Neto, Klaus Vasconcellos, Raydonal Ospina e Patrícia Espinheira. Tais docentes contribuíram sobremaneira para minha formação teórica em estatística e agregaram fortemente na minha formação profissional e pessoal. A

Valéria Bittencourt, pelo carinho, pela paciência, pela amizade e por sempre haver tratado todos os alunos do Programa com muita atenção e cortesia.

Aos amigos que fiz na pós-graduação, nos momentos de estudos coletivos ou de descontração, pela força e torcida. Em especial, agradeço a Larissa Lima, Wenia Valdevino, Jucelino Matos, Jonas Weverson, Maria Ioneres, Vínicius Scher, José Jairo, Rodolpho Jordan, Fidel Fernandes, Talytha Cavalcante, Yuri Alves, Bruna Palm, Cristine Rauber, Anabeth Radunz, Eduardo Ensslin, Luan Portela, Adenice Ferreira, Alison Santos, Alexssandra Lima, João Eudes, Fernanda Clotilde, Pedro Almeida, Daniel Matos, Saul Souza, Marcelo Santos, José Iraponil, Ranah Duarte e demais colegas com quem convivi no Programa. A meu grande amigo, César Diogo, por seu forte apoio, companheirismo e carinho e também pelos momentos inesquecíveis e divertidos que passamos juntos. A meu amigo, Thiago Oliveira, pelo apoio durante o doutorado.

Aos professores que tive em minha graduação na Universidade Federal do Ceará (UFC), que contribuíram para minha formação e para meu interesse na pós-graduação, em especial ao professor Mauricio Mota e ao professor Juvêncio Nobre.

Ao Hospital das Clínicas (HC) da Universidade Federal de Pernambuco e ao Hospital da Restauração (HR) de Pernambuco, pelo importante tratamento médico e pelo suporte a mim fornecidos; em especial, agradeço às doutoras Angela Pinto, Edna Freitas, Íris Machado, Marcelo Ataíde e às enfermeiras que me acompanharam.

À Capes, pelo apoio financeiro.

Aos participantes da banca examinadora, pela atenção e sugestões.

A todos que contribuíram para a realização deste trabalho.

ABSTRACT

Beta regressions are commonly used with responses that assume values in the standard unit interval, such as rates, proportions and concentration indices. Hypothesis testing inferences on the model parameters are typically performed using the likelihood ratio test. It delivers accurate inferences when the sample size is large, but can otherwise lead to unreliable conclusions. It is thus important to develop alternative tests with superior finite sample behavior. We derive the Bartlett correction to the likelihood ratio test under the more general formulation of the beta regression model, i.e. under varying precision. The model contains two submodels, one for the mean response and a separate one for the precision parameter. Our interest lies in performing testing inferences on the parameters that index both submodels. We use three Bartlett-corrected likelihood ratio test statistics that are expected to yield superior performance when the sample size is small. We present Monte Carlo simulation evidence on the finite sample behavior of the Bartlett-corrected tests relative to the standard likelihood ratio test and to two improved tests that are based on an alternative approach. The numerical evidence shows that one of the Bartlett-corrected typically delivers accurate inferences even when the sample is quite small. An empirical application related to behavioral biometrics is presented and discussed. We also address the issue of performing testing inference in a general extreme value regression model when the sample size is small. The model contains separate submodels for the location and dispersion parameters. It allows practitioners to investigate the impacts of different covariates on extreme events. Testing inferences are frequently based on the likelihood test, including those carried out to determine which independent variables are to be included into the model. The test is based on asymptotic critical values and may be considerably size-distorted when the number of data points is small. In particular, it tends to be liberal, i.e., it yields rates of type I errors that surpass the test's nominal size. We derive the Bartlett correction to the likelihood ratio test and use it to define three Bartlett-corrected test statistics. Even though these tests also use asymptotic critical values, their size distortions vanish faster than that of the unmodified test and thus they yield better control of the type I error frequency. Extensive Monte Carlo evidence and an empirical application that uses COVID-19 related data are presented and discussed.

Keywords: COVID-19; Bartlett correction; beta regression; extreme value; likelihood ratio test.

RESUMO

Regressões beta são comumente usadas com respostas que assumem valores no intervalo de unidade padrão, tais como taxas, proporções e índices de concentração. Inferências via teste de hipóteses sobre os parâmetros do modelo são usualmente realizadas utilizando o teste de razão de verossimilhanças. Tal teste fornece inferências precisas quando o tamanho da amostra é grande, mas pode conduzir a conclusões imprecisas quando o número de observações é pequeno. Portanto, é importante desenvolver testes alternativos com comportamento superior em pequenas amostras. Derivamos o fator de correção de Bartlett para o teste da razão de verossimilhanças sob a formulação mais geral do modelo de regressão beta, ou seja, sob precisão variável. O modelo contém dois submodelos, um para a resposta média e outro para o parâmetro de precisão. Nosso interesse reside na realização de testes sobre os parâmetros que indexam os dois submodelos. Usamos três estatísticas de teste da razão de verossimilhanças corrigidas por Bartlett que devem apresentar desempenho superior quando o tamanho da amostra é pequeno relativamente ao teste usual. Apresentamos resultados de simulações de Monte Carlo sobre os comportamentos em pequenas amostras dos testes corrigidos por Bartlett, do teste da razão de verossimilhanças usual e de dois testes melhorados que se baseiam em uma abordagem alternativa. A evidência numérica apresentada mostra que um dos testes corrigidos por Bartlett tipicamente conduz a inferências precisas, mesmo quando o tamanho da amostra é muito pequeno. Uma aplicação empírica relacionada a biometria comportamental é apresentada e discutida. Também consideramos a realização de teste de hipóteses sobre os parâmetros que indexam um modelo geral de regressão de valor extremo. O modelo contém submodelos separados para os parâmetros de localização e dispersão e permite não linearidades. Com base em tal modelo, é possível avaliar os impactos de diferentes covariáveis sobre a ocorrência de eventos extremos. As inferências de teste são frequentemente baseadas no teste da razão de verossimilhanças, incluindo aquelas realizadas para determinar quais variáveis independentes devem ser incluídas no modelo. Tal teste utiliza valores críticos assintóticos e pode apresentar distorções de tamanho apreciáveis quando o número de observações é pequeno. Em particular, ele tende a ser liberal, ou seja, tipicamente fornece taxas de erro do tipo I que superam o nível de significância selecionado pelo usuário. Derivamos o fator de correção de Bartlett para o teste de razão de verossimilhanças e o utilizamos para definir três estatísticas de teste corrigidas. Embora os testes corrigidos também utilizem valores críticos assintóticos, suas distorções de tamanho convergem para zero mais rapidamente do que as do teste não modificado e, portanto, os novos testes tendem a produzir melhor controle da frequência de erro do tipo I. São apresentados e discutidos resultados de simulações Monte Carlo e também uma aplicação empírica que utiliza dados relacionados à pandemia de COVID-19.

Palavras-chave: COVID-19; correção de Bartlett; regressão beta; valor extremo; teste da razão de verossimilhanças.

LIST OF FIGURES

LIST OF TABLES

CONTENTS

1 BARTLETT-CORRECTED TESTS FOR VARYING PRECISION BETA REGRESSIONS WITH APPLICATION TO ENVIRONMENTAL BIOMETRICS

1.1 INTRODUCTION

Regression models are useful for gaining knowledge on how different variables (known as regressors, covariates or independent variables) impact the mean behavior of a variable of interest (known as dependent variable or response). The beta regression model is the most commonly used model with responses that are double bounded, in particular with responses that assume values in the standard unit interval, $(0,1)$. It was introduced by Ferrari and Cribari-Neto (2004) who used an alternative parameterization for the beta density, which is indexed by mean (μ) and precision (ϕ) parameters. Let Y be a beta-distributed random variable. Its density is

$$b(y;\mu,\phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma(\phi(1-\mu))} y^{\mu\phi-1}(1-y)^{\phi(1-\mu)-1}, \quad 0 < y < 1, \tag{1.1}$$

$0 < \mu < 1, \phi > 0$, where $\Gamma(\cdot)$ is the gamma function. Such a law is quite flexible in the sense that the density in (1.1) can assume different shapes depending on the parameter values. It was used by Ferrari and Cribari-Neto (2004) as the underlying foundation for a regression model in which $y_1,\ldots,y_n$ are independent random variables such that y_i is beta-distributed with mean μ_i (i.e. $\mathbb{E}(y_i) = \mu_i$) and precision parameter ϕ, for $i = 1,\ldots,n$. They showed that the variance of y_i is $\mu_i(1-\mu_i)/(1+\phi)$ which, for a given μ_i, is decreasing in ϕ. The model is thus heteroskedastic since the variance of y_i changes with μ_i. The response means are modeled using a set of covariates and ϕ is assumed constant across observations. This model became known as the fixed precision beta regression model.

A more general beta regression formulation was considered by Kieschnick and McCullough (2003) and formally introduced by Simas, Barreto-Souza and Rocha (2010) who allowed the precision parameter to vary across observations, i.e. y_i is beta-distributed with mean μ_i and precision ϕ_i, $i = 1,\ldots,n$. More flexibility can be achieved in some situations by allowing the precision parameter to be impacted by some covariate values. In such a more general formulation, the variance of y_i is no longer restricted to be a multiple of $\mu_i(1-\mu_i)$. The model includes two separate regression submodels, one for the mean and another for the precision, and became known as the variable precision beta regression model. The fixed precision beta regression model is a particular case of the variable precision counterpart; it is obtained by setting $\phi_1 = \cdots = \phi_n = \phi$.

Fixed and varying precision beta regression modeling have been used in many different fields. A beta regression analysis of the effects of sexual maturity on space use in Atlantic salmon (Salmo salar) parr can be found in Bouchard *et al.* (2020). In Chen, Chiu and Chang (2017) the beta regression model is used to segment and describe the container shipping market by analyzing the relationships between service attributes and likelihood of customer retention for the container shipping industry. Some applications of beta regression modeling in ecology can be found in Douma and Weedon (2019). In Mandal, Srivastav and Simonovic (2016), a statistical downscaling model is developed based on beta regression which allow precipitation state in river basin to be calculated. The beta regression model is used by Mullen, Marshall and McGlynn (2013) to model global solar radiation. For a beta regression analysis of ischemic stroke volume, see Swearingen *et al.* (2011).

In both variants of the beta regression model (fixed and variable precision), parameter estimation is carried out by maximum likelihood. It is common practice to perform likelihood ratio tests and z-tests on the model's parameters. The latter are Wald-type tests and are typically less accurate than the former; see Cribari-Neto and Queiroz (2014). Point estimation and testing inferences are usually accurate when the sample size (n) is large. In some applications, nonetheless, the number of data points is small and it is recommended to make use of inferential tools that are expected to yield reliable inferences in small samples. For instance, Ospina, Cribari-Neto and Vasconcellos (2006) obtained modified parameter estimates that display smaller biases in fixed and variable precision linear beta regression models.

The likelihood ratio test, which is commonly used in beta regression empirical analyses, employs an asymptotic approximation: the critical values used in the test are obtained from the test statistic's asymptotic null distribution, which is known to be χ^2_l, where l is the number of restrictions under evaluation; see the discussion in Chapter 9 of Cox and Hinkley (1979). An asymptotic approximation is used because the test statistic's exact null distribution is unknown. In large samples, the test typically delivers accurate inferences since there is good control of the type I error frequency. In contrast, when the number of data points is small (e.g., less than 50 observations), size distortions can be large. In particular, the test tends to be liberal (oversized): the effective null rejection rates tend to be considerably larger than the selected significance level. When the sample size is quite small, the test's effective null rejection can be much larger than the nominal significance level, as shown by the numerical evidence we report. A Bartlett correction to the likelihood ratio test was derived by Bayer and Cribari-Neto (2013) for beta regressions. A major shortcoming of their result, however, is that it only holds for the fixed precision beta regression model. In this chapter, we overcome such a shortcoming by deriving the Bartlett correction for varying precision beta regressions, which are more commonly used by practitioners. The derivation of the correction becomes more challenging in the more general setting. That happens because the parameters that index the two submodels are not orthogonal in the sense that Fisher's information matrix is not block diagonal, and that renders lengthier and more complex derivations of the quantities involved in the Bartlett correction. We considered three Bartlett-corrected test statistics. It is noteworthy that the size distortions of such tests vanish faster than those of the standard likelihood ratio test as the sample size increases and thus the new tests are expected to outperform the likelihood ratio test in small samples. In particular, the likelihood ratio test's size distortions are $O(n^{-1})$ whereas those of the Bartlett-corrected tests are $O(n^{-2})$. It is noteworthy that, although we do not explicitly consider the test of linear restrictions on the model's parameters, the Bartlett correction we derive can be used when testing such restrictions; see Arellano-Valle, Ferrari and Cribari-Neto (1999).

To motivate our analysis, consider the following important issue in behavioral biometrics: the impact of average intelligence on the prevalence of religious disbelievers. Suppose there is interest in measuring such a net impact using data on n nations. The variable of interest (response) is the proportion of atheists in each country and the covariates include average intelligence and other control variables. Cribari-Neto and Souza (2013) carried out varying precision beta regression analyses and produced estimates of such an impact under different scenarios. Each scenario corresponds to a particular choice of countries. We consider the scenario that uses data on 50 countries. We show that by using corrected likelihood ratio tests we arrive at a varying precision beta regression model which differs from that used by Cribari-Neto and Souza (2013). It is noteworthy that our model yields a better fit than their model. We also note that the maximal estimated impact of intelligence on religious disbelief obtained from our model is considerably larger than that computed from the model in Cribari-Neto and Souza (2013) in low income nations.

Our results also reveal that, as countries become more developed, the maximal impact of intelligence on the prevalence of atheists weakens and the impact becomes, in the plausible range of average intelligence values, more symmetric. To the best of our knowledge, this is the first analysis of how the maximal impact of average intelligence on the prevalence of atheists changes with economic development. This illustrates the importance of using tests with good small sample performance when performing beta regression analyses with samples of small to moderate sizes.

The remainder of the chapter is structured as follows. In first section that follows this introduction, we present the variable precision beta regression model. In the second section, we derive the Bartlett correction to the likelihood ratio test in varying precision beta regressions and use it in three modified test statistics. Our main contribution is that we obtain closed-form expressions for the quantities that allow improved testing inferences to be carried out in varying precision beta regressions. Additionally, we briefly review an alternative small sample correction that is already available in the literature. Unlike the correction we derive, however, it does not yield an improvement in the rate at which size distortions vanish. In particular, the size distortions of our corrected tests vanish at rate $O(n^{-2})$ whereas those of the alternative tests we consider do so at rate $O(n^{-1})$. Monte Carlo simulation evidence is presented in the third section. An empirical application that addresses an important issue in behavioral biometrics is presented and discussed in the fourth section. The fifth section contains some concluding remarks. Technical details related to the derivation of the quantities involved in the Bartlett correction are presented in Appendix A.

1.2 THE BETA REGRESSION MODEL

Let $y = (y_1, \dots, y_n)^\top$ be a vector of independent random variables such that y_i follows the beta distribution with mean μ_i and precision ϕ_i, $i = 1, \dots, n$. Such parameters are modeled as

$$g_1(\mu_i) = \eta_i = \sum_{j=1}^{p} \beta_j x_{ij} \quad \text{and} \quad g_2(\phi_i) = \zeta_i = \sum_{j=1}^{q} \delta_j h_{ij},$$

where $\beta = (\beta_1, \dots, \beta_p)^\top \in \mathbb{R}^p$ and $\delta = (\delta_1, \dots, \delta_q)^\top \in \mathbb{R}^q$ are unknown regression parameters $(p+q < n)$, η_i and ζ_i are linear predictors, $x_{i1} \equiv h_{i1} \equiv 1 \, \forall i$, $x_{i2}, \dots, x_{ip}$ and $h_{i2}, \dots, h_{iq}$ are mean and precision covariates, respectively, and $g_1 : (0,1) \mapsto \mathbb{R}$ and $g_2 : (0,\infty) \mapsto \mathbb{R}$ are strictly monotonic and twice-differentiable link functions. Common choices for g_1 are logit, probit, loglog, cloglog and Cauchy, and common choices for g_2 are log and square root; see Cribari-Neto and Zeileis (2010).

Let $\theta = (\beta^\top, \delta^\top)^\top$ be the vector containing all regression coefficients. The log-likelihood function is

$$\ell(\theta) \equiv \ell(\beta, \delta) = \sum_{i=1}^{n} \ell_i(\mu_i, \phi_i), \tag{1.2}$$

where $\ell_i(\mu_i, \phi_i) = \log\Gamma(\phi_i) - \log\Gamma(\mu_i\phi_i) - \log\Gamma((1-\mu_i)\phi_i) + (\mu_i\phi_i - 1)y_i^* + (\phi_i - 2)y_i^\dagger$, with $y_i^* = \log(y_i/1-y_i)$ and $y_i^\dagger = \log(1-y_i)$. The maximum likelihood estimators of β and δ solve $U = \partial\ell(\beta,\delta)/\partial\theta = \left(U_\beta(\beta,\delta)^\top, U_\delta(\beta,\delta)^\top\right)^\top = 0_{p+q}$, where 0_{p+q} is a $(p+q)$-vector of zeros. They cannot be expressed in closed-form. Maximum likelihood estimates can be obtained by numerically maximizing the model log-likelihood function using a Newton or quasi-Newton optimization algorithm such as the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm; see Nocedal and Wright (2006).

For a recent overview of the beta regression model, see Douma and Weedon (2019). Practitioners can perform beta regression analyses using the `betareg` package developed for the R statistical computing environment; see Cribari-Neto and Zeileis (2010). Finally, it is worth noticing that the beta regression model belongs to the class of Generalized Additive Models for Location Scale and Shape (GAMLSS); see Rigby and Stasinopoulos (2005).

1.3 IMPROVED LIKELIHOOD RATIO TESTS IN BETA REGRESSIONS

At the outset, we consider a general setup. Suppose the interest lies in testing a null hypothesis ($\mathcal{H}_0$) that imposes l restrictions on the k-dimensional parameter vector $\theta = (\beta^\top, \delta^\top)^\top$, where $k = p+q$. To that end, we write $\theta = (\psi^\top, \lambda^\top)^\top$, where $\psi = (\psi_1, \ldots, \psi_l)^\top$ is the vector of parameters of interest and $\lambda = (\lambda_1, \ldots, \lambda_s)^\top$ is the vector of nuisance parameters so that $l + s = p + q$. We wish to test $\mathcal{H}_0 : \psi = \psi^{(0)}$ against $\mathcal{H}_1 : \psi \neq \psi^{(0)}$, where $\psi^{(0)}$ is a given l-vector. The likelihood ratio test statistic is

$$\omega = 2[\ell(\hat{\psi}, \hat{\lambda}) - \ell(\psi^0, \tilde{\lambda})],$$

where $(\hat{\psi}^\top, \hat{\lambda}^\top)$ and $(\psi^{0\top}, \tilde{\lambda}^\top)$ are the unrestricted and restricted maximum likelihood estimators of $(\psi^\top, \lambda^\top)$, respectively. Under the null hypothesis, w is asymptotically distributed as χ^2_l. The test is usually performed using critical values obtained from such an asymptotic null distribution, the approximation error being of order $O(n^{-1})$. That is, under the null hypothesis, $\Pr(\omega > \chi^2_{l;1-\alpha}) = \alpha + O(n^{-1})$, where $\alpha \in (0,1)$ is the test significance level and $\chi^2_{l;1-\alpha}$ is the $(1-\alpha)$th quantile from the χ^2_l distribution. The chi-squared approximation to the null distribution of ω may be poor when the sample size is small (e.g., less than 50 observations) and, as a result, large size distortions may take place.

A correction that became known as 'the Bartlett correction' was developed to improve the likelihood ratio test's small sample behavior. It uses the fact that, under $\mathcal{H}_0$, $\mathbb{E}(\omega) = l + b + O(n^{-2})$, where $b = b(\theta)$ is $O(n^{-1})$. Using such a result, it is possible to define the corrected test statistic

$$\omega_{b1} = \frac{\omega}{1 + b/l}$$

whose expected value equals l when terms of order $O(n^{-2})$ are neglected. The quantity $c = 1 + b/l$ became known as 'the Bartlett correction factor'. A general approach for obtaining the Bartlett correction factor in statistical models was developed by Lawley (1956). His approach requires the derivation of log-likelihood cumulants. The expected value of ω, under the null hypothesis, can be expressed as

$$\mathbb{E}(\omega) = l + \varepsilon_k - \varepsilon_{k-l} + O(n^{-2}),$$

where ε_k and ε_{k-l} are of order $O(n^{-1})$. Here,

$$\varepsilon_k = \sum_{\theta} (\lambda_{rstu} - \lambda_{rstuvw}), \qquad (1.3)$$

where

$$\lambda_{rstu} = \kappa^{rs}\kappa^{tu}\left\{\frac{\kappa_{rstu}}{4} - \kappa^{(u)}_{rst} + \kappa^{(su)}_{rt}\right\} \quad \text{and}$$
$$\lambda_{rstuvw} = \kappa^{rs}\kappa^{tu}\kappa^{vw}\left\{\kappa_{rtv}\left(\frac{\kappa_{suw}}{6} - \kappa^{(u)}_{sw}\right) + \kappa_{rtu}\left(\frac{\kappa_{svw}}{4} - \kappa^{(v)}_{sw}\right) + \kappa^{(v)}_{rt}\kappa^{(u)}_{sw} + \kappa^{(u)}_{rt}\kappa^{(v)}_{sw}\right\}.$$

The above cumulants (κ's) are defined in the Appendix. The indices r, s, t, u, v and w vary over all k parameters in the summation in (1.3). The Bartlett correction factor can then be written as

$$c = 1 + \frac{\varepsilon_k - \varepsilon_{k-l}}{l}.$$

Lawley (1956) also showed that all cumulants of the Bartlett-corrected test statistic agree with those of the reference chi-squared distribution with error of order $O(n^{-3/2})$ which indicates that its null distribution is expected to be well approximated by the limiting chi-squared distribution. Hayakawa (1977) obtained an asymptotic expansion for the null distribution of ω; see also Chesher and Smith (1995), Cordeiro (1987) and Harris (1986). Barndorff-Nielsen and Hall (1988) showed that size distortions of Bartlett-corrected tests are of order $O(n^{-2})$, and not of order $O(n^{-3/2})$, as previously believed.

In what follows, we shall obtain the Bartlett correction factor for the class of varying precision beta regressions. We shall only present the main result. Details on the derivation can be found in Appendix A. It is noteworthy that β and δ are not orthogonal (i.e., Fisher's information matrix is not block diagonal), unlike what happens in the class of generalized linear models. As a consequence, the derivation of the Bartlett correction factor becomes lengthier and more challenging. We shall use the main result in Cordeiro (1993), who wrote the general adjustment factor in matrix form. At the outset, we define some $k \times k$ matrices whose (r,s) elements are

$$A^{(tu)} = \left\{ \frac{\kappa_{rstu}}{4} - \kappa_{rst}^{(u)} + \kappa_{rt}^{(su)} \right\}, \quad P^{(t)} = \{\kappa_{rst}\}, \quad Q^{(u)} = \{\kappa_{su}^{(r)}\},$$

$t, u = 1, \ldots, k$. We derived the log-likelihood cumulants up to fourth order for the class of varying precision beta regression models. These cumulants are presented in Appendix A. Using such results, we obtain matrices $A^{(tu)}$, $P^{(t)}$ and $Q^{(u)}$. It is then possible to write ε_k as

$$\varepsilon_k = \mathrm{tr}\left(K^{-1}(L - M - N)\right), \tag{1.4}$$

where $\mathrm{tr}(\cdot)$ is the trace operator and the (r,s) elements of L, M and N are

$$\begin{aligned}
L_{rs} &= \{\mathrm{tr}(K^{-1}A^{(rs)})\},\\
M_{rs} &= -\frac{1}{6}\{\mathrm{tr}(K^{-1}P^{(r)}K^{-1}P^{(s)})\} + \{\mathrm{tr}(K^{-1}P^{(r)}K^{-1}Q^{(s)^{\top}})\}\\
&\quad - \{\mathrm{tr}(K^{-1}Q^{(r)}K^{-1}Q^{(s)})\},\\
N_{rs} &= -\frac{1}{4}\{\mathrm{tr}(P^{(r)}K^{-1})\,\mathrm{tr}(P^{(s)}K^{-1})\} + \{\mathrm{tr}(P^{(r)}K^{-1})\,\mathrm{tr}(Q^{(s)}K^{-1})\}\\
&\quad - \{\mathrm{tr}(Q^{(r)}K^{-1})\,\mathrm{tr}(Q^{(s)}K^{-1})\},
\end{aligned}$$

$r, s = 1, \ldots, k$. Also, ε_{k-l} is obtained from (2.14) by only considering the nuisance parameters.

The corrected statistic ω_{b1} is the standard Bartlett-corrected likelihood ratio test statistic. In addition to it, we shall also consider two other Bartlett-corrected test statistics that are used in Lemonte, Ferrari and Cribari-Neto (2010). The three test statistics are equivalent up to order $O(n^{-1})$ and are given by

$$\omega_{b1} = \frac{\omega}{c}, \quad \omega_{b2} = \omega \exp\left\{-\frac{(\varepsilon_k - \varepsilon_{k-l})}{l}\right\}, \quad \omega_{b3} = \omega\left\{1 - \frac{(\varepsilon_k - \varepsilon_{k-l})}{l}\right\}.$$

We shall refer to the three corrected test statistics above as 'ratio-like', 'exponentially adjusted' and 'multiplicative-like', respectively. An advantage of ω_{b2} is that it is always positive-valued. In order to use the above test statistics in a given class of models, it is necessary to obtain closed-form expressions for ε_k and ε_{k-l} that are valid for such models. For varying precision beta regressions, these quantities can be computed using Equation (1.4), which is our main result. For details on Bartlett corrections, we refer readers to Cordeiro and Cribari-Neto (2014) and Cribari-Neto and Cordeiro (1996).

An alternative correction to the likelihood ratio test statistic was proposed by Skovgaard (2001) who generalized previous results in Skovgaard (1996). His main result relate to those in Barndorff-Nielsen (1986), Barndorff-Nielsen (1991). The author in Skovgaard (2001) proposed using the following two modified test statistics: $\omega_{a1} = \omega - 2\log\xi$ and $\omega_{a2} = \omega\left(1 - \omega^{-1}\log\xi\right)^2$, the latter having the advantage of always being positive-valued. ξ is a function of several model-based quantities (score function, expected information, observed information, etc.). Closed-form expressions for ξ were derived by several authors considering different underlying models. In particular, for models tailored for double limited responses, they were derived by Guedes, Cribari-Neto and Espinheira (2020) for unit gamma regressions, by Ferrari and Pinheiro (2011) for varying precision beta regression models, and by Rauber, Cribari-Neto and Bayer (2020) for beta regressions with parametric mean link function. The finite sample performances of such corrected tests when used in beta regressions was numerically evaluated by Cribari-Neto and Queiroz (2014). It is noteworthy that the size distortions of the three Bartlett-corrected tests vanish at a faster rate than those of ω, ω_{a1} and ω_{a2} as the sample size increases: $O(n^{-2})$ versus $O(n^{-1})$. As a consequence, it is expected that, for a given sample size, the Bartlett-corrected tests display superior control of the type I error probability.

Finally, we note that there are alternative strategies for achieving accurate hypothesis testing inferences in small samples. For instance, Rocke (1989) proposed a numerical approach for estimating the Bartlett correction factor and Ferrari, Lucambio and Cribari-Neto (2005) obtained the Bartlett correction for generalized linear models using a modified version of the likelihood function that accounts for the impact of nuisance parameters on the inference made on the parameters of interest. We shall not pursue these approaches since, as we shall see, the standard Bartlett corrected test is able to deliver extremely accurate inference in small samples in varying precision beta regressions even when the number of nuisance parameters is large.

1.4 NUMERICAL EVIDENCE

In what follows we shall present Monte Carlo simulation results on the finite sample performances of six tests in varying precision beta regressions, namely: ω, ω_{b1} ('ratio-like'), ω_{b2} ('exponentially adjusted'), ω_{b3} ('multiplicative-like'), ω_{a1} and ω_{a2}. All reported results are based on 10,000 replications and were obtained using the R statistical computing environment; see R Core Team (2021). Log-likelihood maximization was performed using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm with analytical first derivatives. There were no convergence failures. Starting values for β and δ were computed as described in Appendix A of Ferrari, Espinheira and Cribari-Neto (2011) with minor tweaks. The computation of such starting values entails the estimation of two linear regressions. We consider the varying precision beta regression model $\log(\mu_i/(1-\mu_i)) = \beta_1+\beta_2 x_{i2}+\beta_3 x_{i3}+\beta_4 x_{i4}$ and $\log(\phi_i) = \delta_1+\delta_2 h_{i2}+\delta_3 h_{i3}$, $i = 1,\ldots,n$. All covariate values were obtained as $\mathcal{U}(-0.5, 0.5)$ random draws and remained constant for all replications performed for a given sample size. We consider three scenarios. In the first scenario, we test $\mathcal{H}_0 : \beta_4 = 0$, and hence $l = 1$ (one restriction). The true parameter values are $\beta_1 = 1.0$, $\beta_2 = 1.7$, $\beta_3 = 3.5$, $\beta_4 = 0$, $\delta_1 = 3.7$, $\delta_2 = 1.5$ and $\delta_3 = 0.9$. In the second scenario, the interest lies in testing $\mathcal{H}_0$: $\beta_3 = \beta_4 = 0$, thus $l = 2$ (two restrictions). The parameter values in this case are $\beta_1 = 1.0$, $\beta_2 = 1.7$, $\beta_3 = \beta_4 = 0$, $\delta_1 = 3.7$, $\delta_2 = 1.5$ and $\delta_3 = 0.9$. In the third and final scenario, the null hypothesis under evaluation is $\mathcal{H}_0$: $\delta_2 = \delta_3 = 0$, and hence $l = 2$ (two restrictions). The parameter values are $\beta_1 = 1.0$, $\beta_2 = 1.7$, $\beta_3 = 2.5$, $\beta_4 = -3.0$, $\delta_1 = 3.7$ and $\delta_2 = \delta_3 = 0$. We computed the tests' null rejection rates at the $\alpha = 10\%, 5\%, 1\%$ significance levels for different sample sizes ($n \in \{15, 20, 30, 40\}$). They are presented in Tables 1 (first scenario), 2 (second scenario) and 3 (third scenario); all entries are percentages.

The tests' null rejection rates for the first scenario are, as noted, presented in

Table 1. At the outset, we note that the likelihood ratio test ω is considerably liberal, that is, it rejects the null hypothesis too often when it is true. For instance, when $n = 15$ and $\alpha = 10\%$, its null rejection rate exceeds 30%, i.e. it is over three time larger than the nominal significance level. When $n = 20$, it equals 25.2%. The test is considerably oversized even when $n = 40$ (null rejection rate > 15%). The corrected tests display much better control of the type I error frequency, especially the third Bartlett-corrected test (i.e. that based on ω_{b3} – 'multiplicative-like'). For example, when $n = 20$ and $\alpha = 10\%$, its null rejection is 10.4% whereas those of ω_{b1}, ω_{b2}, ω_{a1} and ω_{a2}, are, respectively, 16.7%, 14.6%, 14.7% and 17.2%. All modified tests display small size distortions when $n = 40$; again, ω_{b3} ('mutiplicative-like') is the best performer. Interestingly, ω_{b3} is the only conservative test when the sample size is very small ($n = 15$).

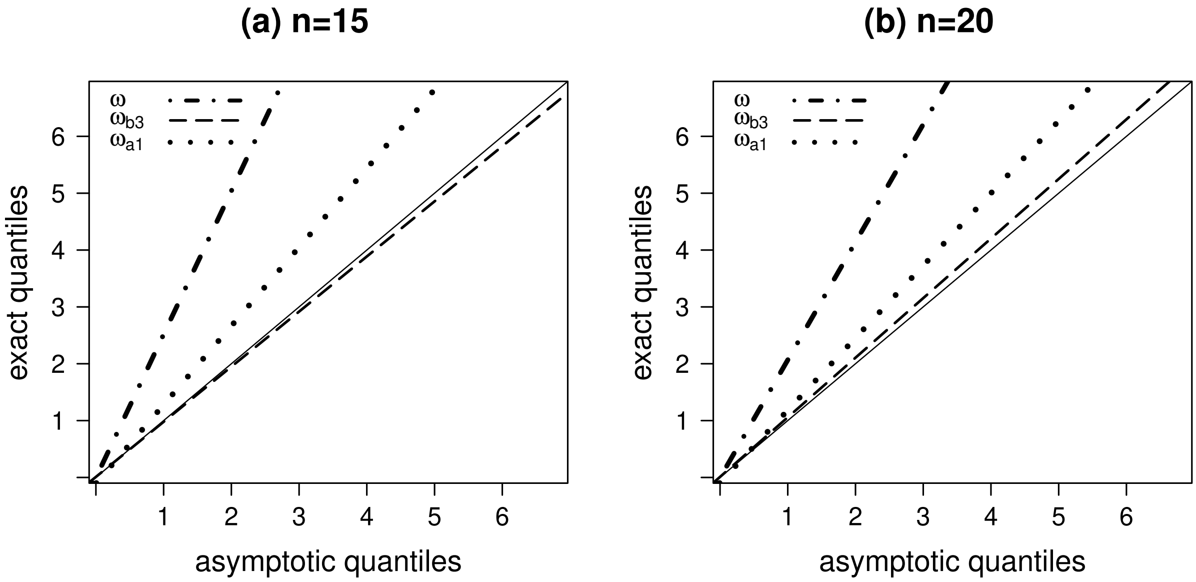
Table 1 – Null rejection rates (%), $\mathcal{H}_0$: $\beta_4 = 0$.

	$\alpha = 10\%$				$\alpha = 5\%$				$\alpha = 1\%$			
	n				n				n			
	15	20	30	40	15	20	30	40	15	20	30	40
ω	30.1	25.2	19.2	15.7	21.8	17.6	12.1	9.0	10.7	7.4	4.0	2.9
ω_{b1}	19.5	16.7	12.8	11.1	11.8	9.8	6.9	5.8	3.8	2.8	1.5	1.2
ω_{b2}	16.6	14.6	11.6	10.6	9.5	8.2	6.0	5.4	2.4	1.8	1.1	1.1
ω_{b3}	8.2	10.4	10.0	9.9	3.5	4.9	4.6	5.0	0.4	0.8	0.7	0.9
ω_{a1}	16.0	14.7	10.6	10.0	10.2	9.3	5.3	5.2	4.0	3.0	1.2	1.1
ω_{a2}	19.4	17.2	11.5	10.4	12.9	11.1	5.9	5.5	5.6	3.9	1.4	1.2

Source: The author (2020).

Figure 1 contains quantile-quantile (QQ) plots of three test statistics, namely: the likelihood ratio test statistic, the best performing Bartlett-corrected test statistic (ω_{b3} – 'mutiplicative-like') and the best performing test statistics obtained from the alternative finite sample correction (ω_{a1}). We plot the exact quantiles of the three test statistics against their asymptotic counterparts (obtained from the χ^2_1 distribution). The included 45^o line indicates perfect agreement between exact and asymptotic null distributions. The left and right panels are for $n = 15$ and $n = 20$, respectively. In both plots, the line that corresponds to ω is considerably above the 45^o line which indicates that the test statistic exact quantiles are much larger than the asymptotic quantiles, and that translates into liberal test behavior, i.e., the test tends to overreject the null hypothesis. The exact quantiles of ω_{a1} also exceed those from the chi-squared distribution, but less dramatically. The null distribution of ω_{b3}, the Bartlett-corrected test statistic ('mutiplicative-like'), is very well approximated by the limiting χ^2_1 distribution since the dashed line is very close to the 45º line.

Figure 1 – Quantile-quantile plots, $\mathcal{H}_0 : \beta_4 = 0$.



Source: The author (2020).

Table 2 contains simulation results for the second scenario, that is, it contains results relative to testing that β_3 and β_4 are jointly equal to zero. Here, $l = 2$. Again, the likelihood ratio test is markedly oversized when the sample size is small, even more so than in the previous scenario. For instance, when $\alpha = 10\%$ and $n = 20$, the estimated size of the test equals 31.1%, i.e. the test's empirical size is over three times larger than the nominal significance level. The corrected tests perform much more reliably. Again, overall, the best performing test is that based on our third Bartlett-corrected test statistic (ω_{b3} – 'mutiplicative-like'). For instance, when $n = 20$ and $\alpha = 10\%$, its null rejection rate is 10.7%; the corresponding figures for ω_{a1} and ω_{a2} (the two alternative corrected tests) are 14.9% and 17.5%, respectively.

Figure 2 contains QQ plots for the second scenario. As in the previous scenario, the null distribution of ω is poorly approximated by the limiting chi-squared distribution and the approximation works better for ω_{b3} (the Bartlett-corrected test statistic, 'mutiplicative-like') than for ω_{a1}.

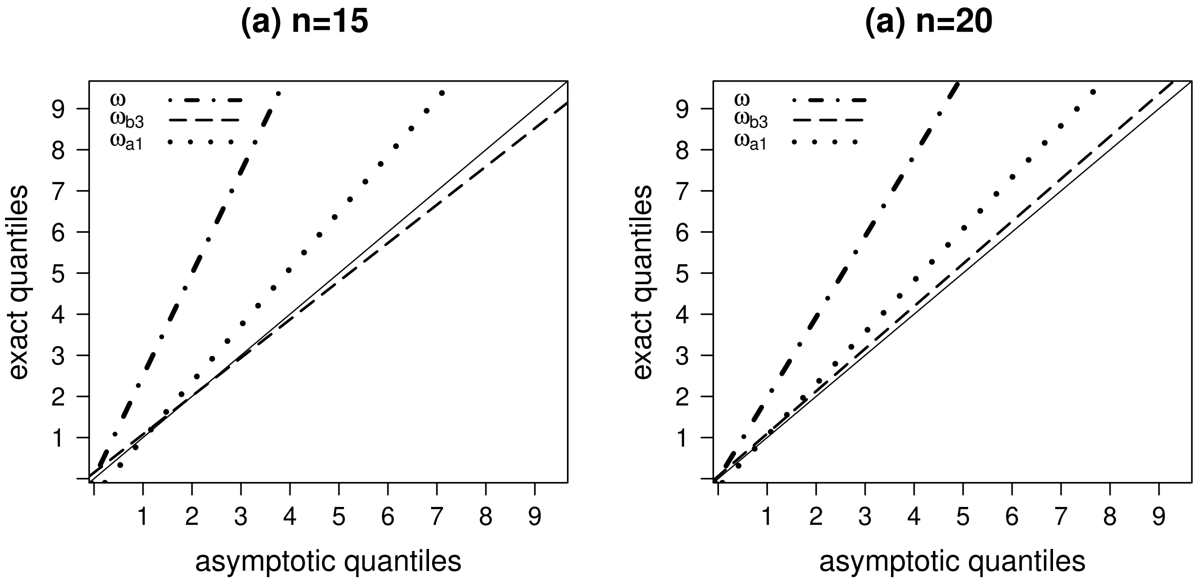
We shall now consider to tests on the coefficients of the precision submodel. The

Table 2 – Null rejection rates (%), $\mathcal{H}_0 : \beta_3 = \beta_4 = 0$.

	$\alpha = 10\%$				$\alpha = 5\%$				$\alpha = 1\%$			
	n				n				n			
	15	20	30	40	15	20	30	40	15	20	30	40
ω	39.6	31.1	21.2	16.6	29.9	22.0	12.9	9.5	16.5	9.6	4.2	2.8
ω_{b1}	23.4	18.2	12.6	10.9	15.2	10.7	6.7	5.6	5.5	2.7	1.5	1.1
ω_{b2}	19.1	15.7	11.5	10.5	11.4	8.6	5.9	5.2	3.3	1.8	1.2	1.0
ω_{b3}	7.4	10.7	9.9	9.6	2.9	4.7	4.8	4.7	0.4	0.7	0.9	0.8
ω_{a1}	16.7	14.9	10.5	9.8	10.8	8.8	5.3	4.8	4.1	2.7	1.1	0.9
ω_{a2}	21.6	17.5	11.3	10.3	14.1	10.7	6.0	5.1	5.8	3.4	1.3	1.0

Source: The author (2020).

Figure 2 – Quantile-quantile plots, $\mathcal{H}_0$: $\beta_3 = \beta_4 = 0$.



Source: The author (2020).

null rejection rates for the third scenario are in Table 3. We test the null hypothesis of fixed precision, i.e. we test $\mathcal{H}_0 : \delta_2 = \delta_3 = 0$ which is equivalent to testing $\mathcal{H}_0 : \phi_1 = \cdots = \phi_n = \phi$ (the precision parameter is constant across observations). The figures in Table 3 indicate, once again, that testing inferences based on ω can be quite unreliable when n is small. Overall, the third Bartlett-corrected ('multiplicative-like') test outperforms all other corrected tests. For instance, when $n = 20$ and $\alpha = 10\%$, its null rejection rate is 9.5% whereas those of ω_{b1} ('ratio-like'), ω_{b2} ('exponentially adjusted'), ω_{a1} and ω_{a2} are 17.1%, 14.4%, 11.7% and 14.1%. We do not present QQ plots for brevity. We note, however, that they show that the null distribution of ω_{b3} ('mutiplicative-like') is well approximated by the limiting χ^2 distribution.

Table 3 – Null rejection rates (%), $\mathcal{H}_0 : \delta_2 = \delta_3 = 0$.

	$\alpha = 10\%$				$\alpha = 5\%$				$\alpha = 1\%$			
	n				n				n			
	15	20	30	40	15	20	30	40	15	20	30	40
ω	39.6	29.8	20.3	16.3	30.2	20.7	12.9	9.3	15.9	8.8	4.4	2.7
ω_{b1}	23.6	17.1	12.9	11.0	14.7	9.7	6.8	5.3	4.8	2.4	1.6	1.1
ω_{b2}	19.1	14.4	12.0	10.3	10.9	7.5	6.2	5.0	2.7	1.8	1.3	1.0
ω_{b3}	7.7	9.5	10.1	9.7	3.3	4.1	5.0	4.7	0.4	0.7	0.8	0.8
ω_{a1}	13.1	11.7	10.3	9.7	7.4	6.1	5.3	4.8	1.9	1.5	1.2	1.0
ω_{a2}	18.0	14.1	11.3	10.2	10.8	7.7	5.9	5.0	3.3	2.0	1.4	1.0

Source: The author (2020).

We also performed simulations using a data generating process that differs from the estimated model, that is, we estimated the tests' non-null rejection rates (powers). We restrict attention to the likelihood ratio test (ω), the best performing Bartlett-corrected test (ω_{b3} – 'mutiplicative-like') and the best performing test obtained using the alternative finite sample correction (ω_{a1}). We consider two sample sizes ($n \in \{20, 40\}$) and two significance levels ($\alpha = 10\%, 5\%$). We test $\mathcal{H}_0 : \beta_4 = 0$ (first scenario), but the data are generated using a value of β_4 that is different from zero; we denote such a value by γ. The null hypothesis is thus false. Since some tests are oversized, all testing inferences are carried out using exact (estimated from the size simulations) critical values. The tests' estimated powers for different values of γ are presented in Table 4. As expected, the tests become more powerful when the sample size is larger and also as the value of γ moves away from zero. Overall, the three tests display similar non-null rejection rates.

We shall now return to the evaluation of the tests' null performances. First, we shall investigate the impact of the number of nuisance parameters on the tests' null behavior. We set the sample size at $n = 40$ and consider the following model:

$$\log\left(\frac{\mu_i}{1-\mu_i}\right) = \beta_1 + \sum_{j=2}^{p} \beta_j x_{ij}$$

$$\log(\phi_i) = \delta_1 + \delta_2 x_{i2} + \delta_3 x_{i3},$$

$i = 1, \ldots, 40$. We test $\mathcal{H}_0$: $\beta_2 = 0$ against $\mathcal{H}_1$: $\beta_2 \neq 0$. The covariate x_2 is a dummy variable that equals 1 for the first twenty observations and 0 otherwise. The values of all other covariates were obtained as random $\mathcal{U}(-0.5, 0.5)$ draws. Table 5 contains the tests' null rejection rates for $p = 3, 4, 5, 6$. The results show that the likelihood ratio test tends to

Table 4 – Nonnull rejection rates (%), $\mathcal{H}_0 : \beta_4 = 0$.

	$n = 20$						$n = 40$					
	$\alpha = 5\%$			$\alpha = 10\%$			$\alpha = 5\%$			$\alpha = 10\%$		
γ	ω	ω_{b3}	ω_{a1}	ω	ω_{b3}	ω_{a1}	ω	ω_{b3}	ω_{a1}	ω	ω_{b3}	ω_{a1}
−1.5	99.1	96.9	99.1	99.7	98.4	99.8	100.0	100.0	100.0	100.0	100.0	100.0
−1.25	95.7	94.2	95.8	98.3	97.3	98.5	99.9	99.9	99.9	100.0	100.0	100.0
−1.0	83.7	83.9	82.8	92.0	91.4	91.8	98.6	98.7	98.9	99.4	99.5	99.5
−0.75	59.3	61.9	57.8	74.1	75.7	73.8	87.8	88.8	89.3	93.6	94.0	94.3
−0.5	31.0	32.2	29.6	45.5	47.2	44.4	54.3	56.1	56.7	68.0	68.8	69.5
0.5	30.5	33.8	30.2	45.1	47.7	45.2	55.6	56.8	56.6	69.6	70.2	70.1
0.75	58.7	62.9	58.1	73.6	76.0	73.2	87.7	88.6	88.5	93.8	94.0	93.9
1.0	83.0	85.0	82.6	91.6	92.6	92.1	98.3	98.4	98.4	99.4	99.4	99.4
1.25	94.9	94.7	94.8	98.0	97.5	98.3	99.8	99.9	99.9	100.0	100.0	100.0
1.5	98.8	97.6	98.8	99.6	98.8	99.7	100.0	100.0	100.0	100.0	100.0	100.0

Source: The author (2020).

become progressively more liberal as the number of nuisance parameters increases. In contrast, the corrected tests are much less sensitive to the number of nuisance parameters, especially ω_{b3} ('mutiplicative-like'), which is the best performing test. Its null rejection rates for the different values of p at the 10% significance level range from 9.9% to 10% whereas those of ω range between 14.1% ($p = 3$) and 17.1% ($p = 6$).

Table 5 – Null rejection rates (%), $\mathcal{H}_0 : \beta_2 = 0$, varying number of nuisance parameters.

	$\alpha = 10\%$				$\alpha = 5\%$				$\alpha = 1\%$			
	p				p				p			
	3	4	5	6	3	4	5	6	3	4	5	6
ω	14.1	14.6	16.1	17.1	7.9	8.4	9.8	10.1	2.2	2.3	2.9	3.1
ω_{b1}	10.6	10.9	11.5	11.9	5.2	5.8	5.9	6.4	1.0	1.2	1.3	1.3
ω_{b2}	10.3	10.5	11.0	11.1	5.0	5.4	5.4	5.8	0.9	1.1	1.1	1.1
ω_{b3}	9.9	10.0	10.0	10.0	4.8	5.1	4.8	4.8	0.8	1.0	0.9	0.8
ω_{a1}	10.1	10.3	10.4	10.6	5.0	5.3	5.3	5.4	0.9	1.1	1.1	1.1
ω_{a2}	10.4	10.6	10.9	11.1	5.2	5.5	5.5	5.8	0.9	1.2	1.2	1.3

Source: The author (2020).

Second, we shall evaluate the tests' finite sample performances when the null hypothesis includes restrictions on the parameters of both submodels simultaneously. The data generating process is

$$\log\left(\frac{\mu_i}{1-\mu_i}\right) = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3}$$
$$\log(\phi_i) = \delta_1 + \delta_2 x_{i2} + \delta_3 x_{i3}.$$

We consider two different null hypotheses, namely: (i) $\mathcal{H}_0$: $\beta_2 = 0, \delta_3 = 0$ ($l = 2$) and (ii) $\mathcal{H}_0$: $\beta_2 = 0, \delta_2 = \delta_3 = 0$ ($l = 3$). The corresponding parameter values are (i) $\beta_1 = 1.0, \beta_2 = 0, \beta_3 = 3.0, \delta_1 = 1.7, \delta_2 = 0.7, \delta_3 = 0$ and (ii) $\beta_1 = 1.5, \beta_2 = 0, \beta_3 = -1.4, \delta_1 = 1.5, \delta_2 = \delta_3 = 0$. The covariate values were obtained as random $\mathcal{U}(-0.5, 0.5)$ draws and $n \in \{15, 20, 30, 40\}$. Table 6 contains the tests' null rejection rates. The test based on the Bartlett-corrected test statistic ω_{b3} ('mutiplicative-like') is the best performer in both cases. For instance, when $l = 2$ and $n = 15$, its null rejection rate at the 10% significance level is 9.5% whereas those of the competing tests range from 11.1% to 25.0%.

Table 6 – Null rejection rates (%), $\mathcal{H}_0 : \beta_2, \delta_3 = 0$ ($l = 2$) and $\mathcal{H}_0 : \beta_2, \delta_2, \delta_3 = 0$ ($l = 3$).

		$\alpha = 10\%$				$\alpha = 5\%$				$\alpha = 1\%$			
		n				n				n			
		15	20	30	40	15	20	30	40	15	20	30	40
$l = 2$	ω	25.0	20.4	15.3	12.4	16.4	12.6	9.0	6.7	6.8	4.5	2.4	1.6
	ω_{b1}	14.6	12.6	11.3	9.8	8.2	6.9	5.8	5.0	1.9	1.2	1.2	1.0
	ω_{b2}	12.8	11.5	10.9	9.6	6.9	6.1	5.6	4.8	1.3	0.9	1.1	0.9
	ω_{b3}	9.5	10.2	10.4	9.5	4.3	5.0	5.2	4.7	0.6	0.6	1.0	0.9
	ω_{a1}	11.1	11.6	10.8	9.5	6.0	6.2	5.5	4.8	1.2	1.1	1.2	1.0
	ω_{a2}	13.1	12.2	11.0	9.6	7.2	6.6	5.7	4.9	1.5	1.3	1.2	1.0
$l = 3$	ω	23.7	18.9	13.3	12.5	15.5	11.6	7.4	6.7	5.8	3.6	1.9	1.6
	ω_{b1}	13.4	11.9	9.8	9.8	7.2	6.2	5.0	4.7	1.9	1.4	1.1	1.1
	ω_{b2}	11.8	11.2	9.6	9.6	6.0	5.7	4.8	4.6	1.4	1.2	1.0	1.0
	ω_{b3}	9.4	10.2	9.3	9.5	4.4	5.0	4.7	4.6	0.8	1.0	0.9	1.0
	ω_{a1}	7.8	8.8	8.8	9.6	3.8	4.4	4.4	4.4	0.9	0.9	0.9	0.9
	ω_{a2}	9.2	9.6	9.1	9.7	4.7	4.8	4.6	4.5	1.3	1.0	0.9	1.0

Source: The author (2020).

Finally, we shall evaluate the impact of different levels of correlations between regressors on the tests' small sample performance. The model is

$$\log\left(\frac{\mu_i}{1-\mu_i}\right) = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3}$$
$$\log(\phi_i) = \delta_1 + \delta_2 x_{i2}.$$

The values of the two regressors are obtained as random draws from the bivariate normal distribution with mean $(0,0)^\top$ and covariance matrix Σ. The diagonal and off-diagonal elements of Σ are, respectively, 1 and ρ. Hence, ρ is the correlation coefficient between x_2 and x_3. We test $\mathcal{H}_0$: $\beta_2 = \beta_3 = 0$ $(l = 2)$. Data generation was carried out using $\beta_1 = 1.0$, $\beta_2 = \beta_3 = 0$, $\delta_1 = 1.7$ and $\delta_2 = 0.1$. Different correlation strengths were considered, ranging from very low to very strong: $\rho \in (0.1, 0.5, 0.75, 0.95)$. The sample sizes are $n \in \{15, 20, 30, 40\}$. Table 7 contains the tests' null rejection rates. Again, the likelihood ratio test ω is quite liberal when n is small, slightly more so under very strong correlation between the two regressors. The Bartlett-corrected tests perform very well for all correlation values, especially ω_{b3} ('multiplicative-like'). Its null rejection rates are once again very close to α. For instance, when $\rho = 0.75$, $n = 15$ and $\alpha = 10\%$ (5%), the test's null rejection rate is 9.7% (4.4%) whereas that of uncorrected test (ω) is 22.3% (14.0%) and those of alternative tests ω_{a1} and ω_{a2} are 18.4% (9.9%) and 20.5% (13.4%), respectively. It is noteworthy that the null rejection rates of the three Bartlett-corrected tests are insensitive to the level of correlation between regressors. For example, when $n = 15$ and $\alpha = 10\%$, the null rejection rates of ω_{b1} ('ratio-like'), ω_{b2} ('exponentially adjusted') and ω_{b3} ('multiplicative-like') for $\rho = (0.1, 0.5, 0.75, 0.95)$ are in $[13.3\%, 13.8\%]$, $[11.9\%, 12.2\%]$ and $[9.4\%, 10.6\%]$, respectively.

Overall, the simulation evidence presented above indicates that the Bartlett-corrected tests achieve excellent control of the type I error frequency, even when the sample size is quite small. In particular, we recommend the use of ω_{b3} which was the best performer.

1.5 BEHAVIORAL BIOMETRICS: INTELLIGENCE AND ATHEISM

We shall now address the behavioral biometrics issue briefly outlined in the Introduction. The interest lies in modeling the impact of average intelligence on the prevalence of religious disbelievers. General intelligence relates to the ability to reason deductively or inductively, think abstractly, use analogies, synthesize information, and apply it to new domains. It is typically measured by the intelligence quotient (IQ) which

Table 7 – Null rejection rates (%), $\mathcal{H}_0: \beta_2 = \beta_3 = 0$ $(l = 2)$; varying correlation between regressors.

		$\alpha = 10\%$				$\alpha = 5\%$				$\alpha = 1\%$			
		n				n				n			
		15	20	30	40	15	20	30	40	15	20	30	40
$\rho = 0.1$	ω	21.5	16.3	15.2	13.1	13.7	9.6	8.5	7.2	4.7	2.9	2.2	1.6
	ω_{b1}	13.5	11.2	10.8	10.5	7.2	5.8	5.2	5.2	1.8	1.4	1.1	1.0
	ω_{b2}	12.2	10.6	10.4	10.3	6.4	5.5	5.1	5.1	1.4	1.2	1.0	1.0
	ω_{b3}	10.6	9.9	9.9	10.2	5.1	5.2	4.8	5.0	0.9	1.1	0.9	1.0
	ω_{a1}	16.9	10.4	10.3	10.5	9.8	5.7	4.9	5.3	2.9	1.2	1.0	1.0
	ω_{a2}	18.1	10.7	10.6	10.6	11.0	5.9	5.1	5.4	3.6	1.4	1.1	1.0
$\rho = 0.5$	ω	21.7	17.1	14.8	13.3	13.4	9.8	8.2	7.2	4.2	2.6	2.1	1.6
	ω_{b1}	13.3	11.4	10.6	10.6	6.8	5.7	5.3	5.5	1.2	1.2	1.2	1.1
	ω_{b2}	12.1	10.8	10.3	10.4	6.1	5.3	5.1	5.4	1.0	1.1	1.1	1.0
	ω_{b3}	10.5	10.1	9.8	10.3	4.8	4.9	4.8	5.3	0.8	0.9	1.0	1.0
	ω_{a1}	19.2	10.5	9.8	10.5	11.8	5.3	4.9	5.5	3.6	1.1	1.1	1.1
	ω_{a2}	21.3	10.9	10.1	10.6	13.6	5.5	5.1	5.5	4.9	1.2	1.1	1.1
$\rho = 0.75$	ω	22.3	17.0	14.7	12.8	14.0	9.8	8.3	6.8	4.4	2.6	2.3	1.8
	ω_{b1}	13.3	11.2	10.7	9.9	6.9	5.8	5.5	5.0	1.3	1.3	1.0	1.2
	ω_{b2}	11.9	10.7	10.3	9.8	5.9	5.4	5.3	4.9	1.0	1.2	1.0	1.1
	ω_{b3}	9.7	10.1	9.9	9.7	4.4	4.9	5.0	4.9	0.6	1.1	0.9	1.1
	ω_{a1}	18.4	10.2	10.0	10.9	9.9	5.3	5.2	5.1	3.3	1.3	1.0	1.1
	ω_{a2}	20.5	10.6	10.2	10.0	13.4	5.6	5.3	5.1	4.8	1.4	1.0	1.1
$\rho = 0.95$	ω	24.0	16.8	14.4	13.1	15.3	9.3	8.2	6.9	5.3	2.6	2.2	1.5
	ω_{b1}	13.8	11.0	10.4	10.5	7.1	5.4	5.3	5.1	1.7	1.2	1.2	0.9
	ω_{b2}	12.0	10.5	10.1	10.3	6.1	5.1	5.1	5.1	1.2	1.1	1.2	0.9
	ω_{b3}	9.4	9.8	9.8	10.1	4.3	4.7	4.8	5.0	0.8	0.9	1.0	0.9
	ω_{a1}	15.5	10.3	9.8	10.5	9.3	5.0	4.9	5.2	2.6	1.1	1.2	0.9
	ω_{a2}	17.9	10.7	10.0	10.9	10.6	5.2	5.1	5.3	3.7	1.1	1.3	0.9

Source: The author (2020).

is a score obtained from standardized tests. Average IQ scores have been computed for a large number of countries; see e.g. Lynn and Meisenberg (2010) and Lynn and Vanhanen (2006). There is evidence that intelligence negatively correlates with religious belief at the individual level; see e.g. Ganzach and Gotlibovski (2013). The negative correlation holds even when religiosity and performance on analytic thinking are measured in separate sessions; see Pennycook *et al.* (2016). It also holds when computed from a cross section of nations and from the U.S. states; see Lynn, Harvey and Nyborg (2009) and Reeve and Basalik (2011). There are evolutionary reasons for the inverse relationship between intelligence and religious belief. More specifically, the Savanna-IQ Interaction Hypothesis predicts that more intelligent individuals are more likely to acquire and hold evolutionarily novel values and preferences (e.g., liberalism and atheism and, for men, sexual exclusivity) relative to less intelligent individuals. It also predicts that intelligence does not impact the acquisition and espousal of evolutionarily familiar values (e.g., children, marriage, family, and friends). We refer readers to Kanazawa (2010).

Several regression analysis were performed to measure the net impact of changes in intelligence levels on the prevalence of atheists; see Zuckerman, Silberman and Hall (2013) for details. A beta regression analysis was carried out by Cribari-Neto and Souza (2013). They used data on 124 nations and showed that the net impact of average intelligence on the prevalence of religious disbelievers is always positive, gains strength up to a certain level of average intelligence and then weakens. The same data set ($n = 124$) was analyzed by Rauber, Cribari-Neto and Bayer (2020) using a beta regression model that includes a parametric mean link function and by Guedes, Cribari-Neto and Espinheira (2020) using the unit gamma regression model. In what follows, we shall consider a different data set. On page 487 of their paper, Cribari-Neto and Souza (2013) briefly mention a beta regression analysis that was performed using data on the fifty countries with the largest prevalence of atheists ($n = 50$) which they call 'scenario 3'. Since our interest lies in small sample inferences, we shall pursue that modeling. A novel feature of such data is that they do not include countries for which the prevalence of atheists is very small (close to zero).

The response variable (y) is the proportion of atheists in each country and the covariates are: average intelligence quotient (x_2), average intelligence quotient squared (x_3), life expectancy in 2007 in years (x_4), the logarithm of the ratio between trade volume (the sum of imports and exports) and gross national product (x_5), and per capita income adjusted for purchasing power parity (x_6). Additionally, the following interactions are used: $x_7 = x_5 \times x_6$ and $x_8 = x_4 \times x_5$; the latter was not considered by the original authors. Except for x_8, these are the same variables used by Cribari-Neto and Souza (2013). Average intelligence is the independent variable of main interest and the remaining regressors are control variables. Also, $n = 50$ (fifty countries with the largest prevalence of religious disbelievers). The data and computer code used in the empirical analysis that follows can be obtained at <https://github.com/acguedes/beta-Bartlett>.

Cribari-Neto and Souza (2013) fitted the following beta regression model to the data (Model $\mathcal{M}_1$):

$$\log\left(\frac{\mu_i}{1-\mu_i}\right) = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6} + \beta_7 x_{i7}$$
$$\sqrt{\phi_i} = \delta_1 + \delta_2 x_{i2} + \delta_3 x_{i4} + \delta_4 x_{i6}.$$

We noticed that an improved fit according to standard model selection criteria and pseudo-R^2 (see below) can be achieved by adding x_8 to the mean submodel and by only using x_2 in the precision submodel, since the x_4 and x_5 lose statistical significance when the mean submodel includes the interaction between these two variables. Our model (Model $\mathcal{M}_2$) is then

$$\log\left(\frac{\mu_i}{1-\mu_i}\right) = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6} + \beta_7 x_{i7} + \beta_8 x_{i8}$$
$$\sqrt{\phi_i} = \delta_1 + \delta_2 x_{i2}.$$

All parameter estimates of the above model are statistically significant at the 5% significance level according to the z test, and its pseudo-R^2, as defined by Ferrari and Cribari-Neto (2004), is superior to that of the model fitted by Cribari-Neto and Souza (2013): 0.3719 vs 0.3216. Model $\mathcal{M}_2$ is also favored by the three most commonly used model selection criteria when compared to Model $\mathcal{M}_1$, AIC (-61.0555 vs -55.3188), AICC (-55.4144 vs -48.3714) and BIC (-41.9352 vs -34.2865). We shall investigate whether x_4 and x_5 should be excluded from our model by testing whether β_4 and β_5 equal zero (individually and jointly). We shall use three tests, namely: the likelihood ratio test (ω), the best performing Bartlett-corrected test (ω_{b3} – 'mutiplicative-like') and the best performing test based on the alternative small sample correction (ω_{a1}).

At the outset, we test the exclusion of x_4 from Model $\mathcal{M}_2$, that is, we test $\mathcal{H}_0 : \beta_4 = 0$. The p-values of the ω, ω_{b3} and ω_{a1} tests are 0.0258, 0.0443 and 0.0295, respectively. The first and third tests clearly reject the null hypothesis at $\alpha = 5\%$ whereas the p-value of the Bartlett-corrected test is very close to 0.05 which renders uncertainty about the exclusion of x_4 from the model. Next, we test $\mathcal{H}_0 : \beta_5 = 0$. We obtain the following p-values for ω, ω_{b3} and ω_{a1}: 0.0303, 0.0505 and 0.0332, respectively. The first and third tests clearly reject the removal of x_5 from the model at the 5% significance level; the null hypothesis is not rejected by the Bartlett-corrected test. Finally, we test the joint exclusion of both covariates, i.e. we test $\mathcal{H}_0 : \beta_4 = \beta_5 = 0$, and obtain the following p-values for ω, ω_{b3} and ω_{a1}: 0.0726, 0.1121 and 0.0840, respectively. The null hypothesis is not rejected by the three tests at the 5% nominal level, but only the Bartlett-corrected test maintains that inference at the 10% nominal level. That is, such a test provides more evidence in favor of the removal of x_4 and x_5 from the mean submodel.

Based on the above testing inference, we arrive at the following reduced model (Model $\mathcal{M}_{2R}$), which is our final model:

$$\log\left(\frac{\mu_i}{1-\mu_i}\right) = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i6} + \beta_5 x_{i7} + \beta_6 x_{i8}$$

$$\sqrt{\phi_i} = \delta_1 + \delta_2 x_{i2}.$$

The estimates of $\beta_1,\ldots,\beta_6$ (standard errors in parenthesis) are, respectively, 22.9423 (7.6472), -0.7583 (0.1942), 0.0044 (0.0011), 0.1866 (0.0545), -0.0483 (0.0136), 0.0265 (0.0055). For the precision submodel, we obtain $\hat{\delta}_1 = 22.0333$ (5.0312) and $\hat{\delta}_2 = -0.1934$ (0.0498). The model pseudo-R^2 is 0.3455; it is higher than that of the model estimated by Cribari-Neto and Souza (2013). Additionally, AIC $= -59.8094$, AICC $= -56.2972$ and BIC $= -44.5133$. It is noteworthy that these criteria clearly favor our reduced model relative to the model presented in Cribari-Neto and Souza (2013); recall that for that model, AIC $= -55.3188$, AICC $= -48.3714$ and BIC $= -34.2865$. The difference in AIC (AICC) [BIC] in favor of Model $\mathcal{M}_{2R}$ is of nearly 5 points (nearly 8 points) [over 10 points]. When the difference in AIC values exceeds 4, one can conclude that there is considerably less support for the model with larger AIC; see Burnham and Anderson (2004). The evidence in favor of our reduced model is thus strong.

Asymptotic confidence intervals with nominal coverage $(1-\alpha)\times 100\%$ for the parameters of Model $\mathcal{M}_{2R}$ can be obtained using the asymptotic normality of the corresponding maximum likelihood estimators. In particular, for $j = 1,\ldots,6$ and $k = 1,2$, $\hat{\beta}_j + z_{1-\alpha/2}\,\mathrm{se}(\hat{\beta}_j)$ and $\hat{\delta}_k + z_{1-\alpha/2}\,\mathrm{se}(\hat{\delta}_k)$ are asymptotic confidence intervals for β_j and δ_k with nominal coverage $(1-\alpha)\times 100\%$, respectively, the asymptotic standard errors, se, being obtained from Fisher's information matrix inverse evaluated at the maximum likelihood estimates. Here, $z_{1-\alpha/2}$ denotes the $1-\alpha/2$ standard normal quantile. Table 8 contains the lower and upper limits (LLCI_a and ULCI_a) of such intervals for the parameters that index Model $\mathcal{M}_{2R}$ for $1-\alpha = 0.95$. Following Das, Dhar and Pradhan (2018), we also computed approximate confidence intervals based on the test statistics ω, ω_{b3} and

ω_{a1} which was done by finding the set of parameter values such that the test statistic is smaller than $\chi^2_{1;0.95}$ for each parameter and each test statistic. Such intervals are also presented in Table 8. For instance, the confidence intervals for β_5 constructed using ω_{b3} and ω_{a1} are $[-0.0796, -0.0169]$ and $[-0.0781, -0.0145]$, respectively; the corresponding asymptotic interval estimate is $[-0.0749, -0.0216]$. It is noteworthy that none of the reported confidence intervals contains the value zero.

Table 8 – Lower (LLCI) and upper (ULCI) asymptotic confidence intervals limits for the parameters of Model $\mathcal{M}_{2R}$; standard asymptotic confidence interval and confidence intervals constructed using the test statistics ω, ω_{b3} and ω_{a1}.

	LLCI_a	ULCI_a	LLCI_ω	ULCI_ω	$\text{LLCI}_{\omega_{b3}}$	$\text{ULCI}_{\omega_{b3}}$	$\text{LLCI}_{\omega_{a1}}$	$\text{ULCI}_{\omega_{a1}}$
β_1	7.9538	37.9302	3.7532	39.1532	3.7532	40.3532	3.7532	41.5532
β_2	−1.1389	−0.3776	−1.1985	−0.4794	−1.1985	−0.4182	−1.1985	−0.4794
β_3	0.0022	0.0065	0.0011	0.0068	0.0011	0.0069	0.0014	0.0072
β_4	0.0798	0.2935	0.0674	0.3066	0.0622	0.3144	0.0505	0.3066
β_5	−0.0749	−0.0216	−0.0778	−0.0184	−0.0796	−0.0169	−0.0781	−0.0145
β_6	0.0159	0.0371	0.0137	0.0393	0.0127	0.0403	0.0137	0.0396
δ_1	12.1722	31.8941	8.7085	32.7085	8.7085	33.5085	8.7085	32.3085
δ_2	−0.2910	−0.0957	−0.2926	−0.0566	−0.2926	−0.0566	−0.2926	−0.0566

Source: The author (2020).

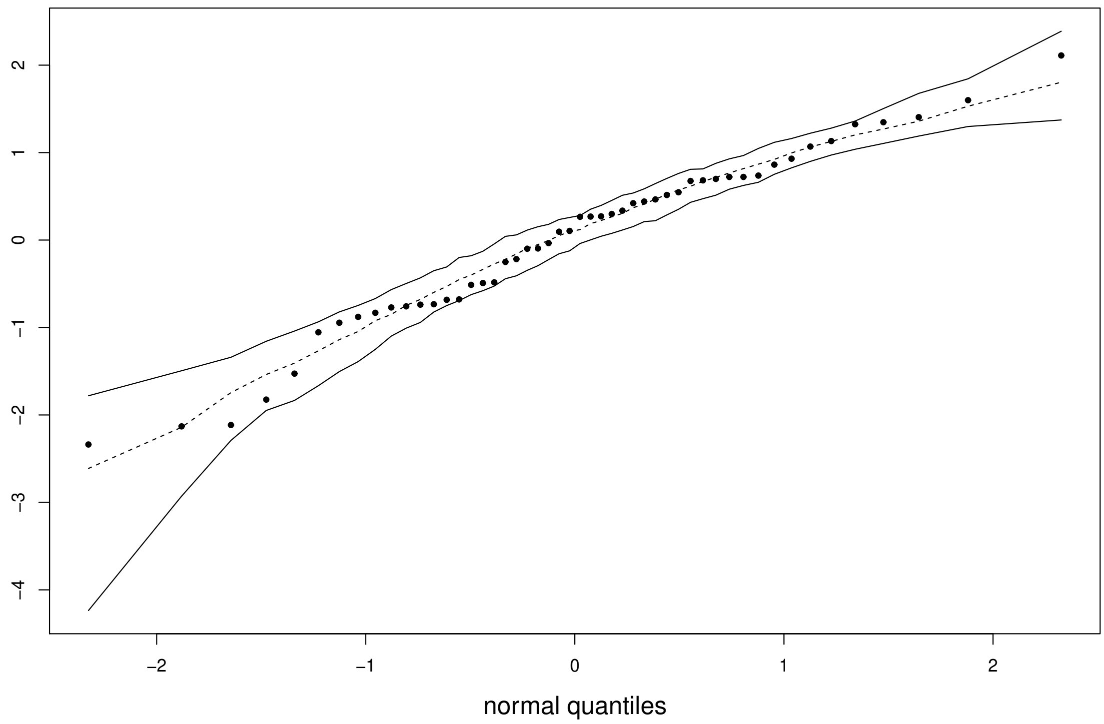
The model used by Cribari-Neto and Souza (2013) (Model $\mathcal{M}_1$) and our reduced model (Model $\mathcal{M}_{2R}$) are non-nested. In order to distinguish between them using a hypothesis test, we performed the J test as outlined by Cribari-Neto and Lucena (2015). When the test is applied to two non-nested models, say Models m_1 and m_2, each model is sequentially tested against the other, i.e. we test Model m_1 against Model m_2, and then we test Model m_2 against Model m_1. It is thus possible to accept one model as the true model and reject the alternative model, to accept both models (i.e. to conclude that the two models are empirically indistinguishable) or to reject both models. Since the J testing inference is reached using the likelihood ratio test, we have also performed the test using the two corrected tests. We first test Model $\mathcal{M}_1$, i.e. the model fitted by Cribari-Neto and Souza (2013), against our reduced model (Model $\mathcal{M}_{2R}$). The p-values of the tests based on ω, ω_{b3} and ω_{a1} are, respectively, 0.0036, 0.0825, 0.0233. All tests reject Model $\mathcal{M}_1$ (i.e. the model used by the authors) at the 10% significance level; the test based on ω_{a1} (ω) yields rejection at $\alpha = 5\%$ (1%). Next, Model $\mathcal{M}_{2R}$ is tested against Model $\mathcal{M}_1$. The

p-values of the tests that use ω, ω_{b3} and ω_{a1} are 0.0364, 0.1100 and 0.2001, respectively. Interestingly, our model is rejected at the 5% significance level by the likelihood ratio test whereas that inference is reversed when the small sample corrections are applied: the two corrected tests do not yield rejection of the model, nor even at $\alpha = 10\%$. That is, our model is not rejected by the two corrected tests.

It is noteworthy that if we consider the three sets of tests, i.e. the tests of $\mathcal{H}_0 : \beta_4 = 0$, $\mathcal{H}_0 : \beta_5 = 0$ and $\mathcal{H}_0 : \beta_4 = \beta_5 = 0$, it is clear that the Bartlett-corrected test was the test that most emphatically suggested the removal of both x_4 and x_5 from the mean submodel of Model $\mathcal{M}_2$.

We constructed a residual normal probability with simulated envelopes using the combined residuals of Espinheira, Santos and Cribari-Neto (2017) from our fitted model (Model $\mathcal{M}_{2R}$); see Figure 3. The envelope bands were constructed using 100 replications The plot shows that there is no evidence against the correct specification of our model since all points lie inside the two envelope bands.

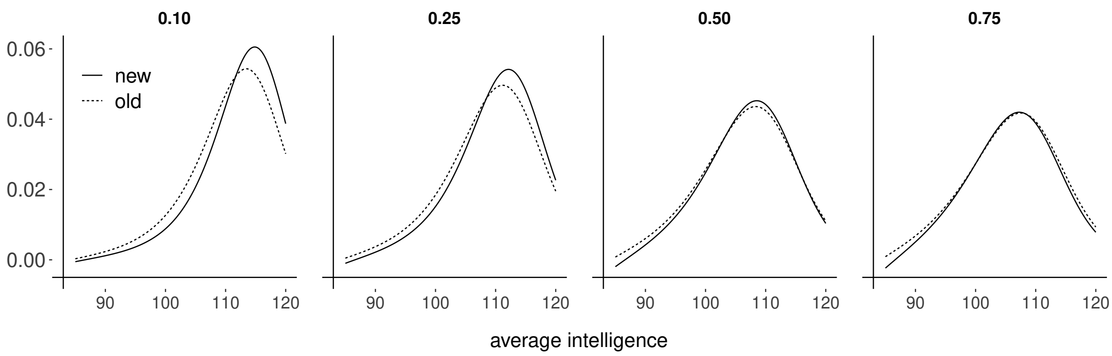
Figure 3 – Residual normal probability plot.



Source: The author (2020).

In Cribari-Neto and Souza (2013), Figure 4, the authors plot an estimate of $\partial\mu_i/\partial x_{i2}$ against a sequence of values of average intelligence by setting all other covariates at their median values. In Figure 4 we present a panel of similar plots each containing two impact curves, namely: (i) that obtained from our reduced model ('new') and (ii) that obtained using the model fitted by Cribari-Neto and Souza (2013) ('old'). That is, 'new' and 'old' in Figure 4 refer to Models $\mathcal{M}_{2R}$ and $\mathcal{M}_1$, respectively. Instead of only fixing the covariates other than average intelligence at their median values, we do so at four different quantiles: 0.10, 0.25, 0.50 (median) and 0.75. The number above each panel indicates at which quantile the covariates other than those related to average intelligence were fixed. We note that the two impact curves become more similar (more dissimilar) as the quantile at which the regressors values are set increases (decreases). Such covariates tend to assume larger values for more developed nations since they relate to per capita income, life expectancy and integration to international trade. In particular, the former two variables are highly correlated with economic development. It then follows that one gets a somewhat different functional form of the impact of average intelligence on the prevalence of religious disbelievers in lower income countries when our reduced model is used relative to the model used by the original authors. At the lowest quantile (0.10), the maximal impact computed from our model (Model $\mathcal{M}_{2R}$) is over 11% larger than that obtained using the alternative model (Model $\mathcal{M}_1$). When the covariates values are set at their medians, the figure drops to nearly 4%. To the best of our knowledge, our analysis provides the first measure of the decline in the maximal impact of average intelligence on the prevalence of religious disbelievers and also of the changes in the functional form of such an impact as nations become more developed.

Figure 4 – Impact curves of average intelligence on the prevalence os atheists.



Source: The author (2020).

1.6 CONCLUDING REMARKS

The beta regression model is widely used to model responses that assume values in $(0, 1)$. In the initial formulation of the model, the precision parameter was assumed constant for all observations, i.e. all responses in the sample share the same precision. This model became known as the fixed precision beta regression model. A more general and more flexible formulation of the model was later proposed. It allows both distribution parameters to vary across observations. Most empirical applications employ this version of the model, which is known as the varying precision beta regression model. It contains two submodels, one for mean and another for the precision.

In both variants of the regression model, testing inferences are usually performed using the likelihood ratio test. Such a test employs an asymptotic approximation, and as a consequence it can be quite size distorted when the sample size is small. In particular, it tends to be liberal (oversized), i.e. it overrejects the null hypothesis when such a hypothesis is true. Since many applications of the beta regression model are based on samples of small to moderate sizes, it is important to develop alternative tests with superior finite sample behavior, i.e. tests that yield better control of the type I error frequency. Bayer and Cribari-Neto (2013) derived a Bartlett correction to the likelihood ratio test that can be used to achieve more accurate testing inferences. Their result, nonetheless, only applies to the more restrictive model formulation, namely: the fixed precision model. Since many applications employ the varying precision formulation of the beta regression model, their small sample correction cannot be used. In this chapter we derived the Bartlett correction to the likelihood ratio test in full generality. Our correction can thus be used to construct modified likelihood ratio tests to be used in varying precision beta regression analyses. We considered three Bartlett-corrected tests. Monte Carlo simulation evidence revealed that one of such tests typically delivers very accurate inferences even when the sample size is quite small. Its small sample performance was numerically compared to those of two tests that are based on an alternative correction. Overall, the results favor the test that employs the Bartlett correction. A novel feature of our Bartlett-correction tests is that their size distortions are guaranteed to vanish at a faster rate than that of the likelihood ratio test: $O(n^{-2})$ vs $O(n^{-1})$.

We presented and discussed an empirical application that involved an important issue in evolutionary biometrics, namely: the relationship between average intelligence and the prevalence of religious disbelievers. Using data on 50 countries, we showed that by using our Bartlett-corrected testing inferences we arrive at a beta regression model slightly different from that previously used in the literature. It is noteworthy that our model displays superior fit and yields a noticeably different functional form of the impact of intelligence on religious disbelief in low income countries. This empirical application illustrates the usefulness of the Bartlett correction derived in this chapter.

A direction for future research is the extension of our analytical results for testing inferences in inflated beta regression models introduced by Ospina and Ferrari (2012) which include both continuous and discrete components and thus allow for response values that are exactly equal to 0 or 1.

2 BARTLETT-CORRECTED TESTS FOR GENERAL EXTREME-VALUE REGRESSIONS WITH APPLICATION TO MAXIMAL COVID-19 MORTALITY IN BRAZIL

2.1 INTRODUCTION

Extreme phenomena are of interest in many fields such as meteorology, finance, hydrology, civil engineering, astronomy, metallurgy, insurance, oceanography, biomedicine, among others; see, e.g., Kotz and Nadarajah (2000). The desire to gain knowledge on the occurrence of extreme events has led to the development of statistical methods for modeling phenomena such as natural disasters, earthquakes, landslides, financial crises, and pandemics. Extreme events have specific attributes, and hence they should not be modeled like events in which the interest typically lies in the central region of the population distribution. In the statistical theory of extremes, parametric and semi-parametric statistical models are developed to deal with rare events, i.e., with events that occur with very small probabilities. It has been shown that the normalized (that is, standardized using sequences of real numbers) maximal realization of an event necessarily follows one of following three laws as the sample size tends to infinity: Weibull, Fréchet or Gumbel; see Gumbel (1958). This fact is routinely explored when developing new modeling strategies for extreme events.

It is typically of interest to model the dependence of extreme events on conditioning (explanatory, independent) variables. A general extreme-value regression model was introduced by Barreto-Souza and Vasconcellos (2011) under the assumption that the extreme phenomenon of interest follows the Gumbel law. The model comprises of two submodels for location and dispersion parameters which are modeled by separate nonlinear predictors. Testing inferences on the parameters that index the two submodels are typically carried out using the likelihood ratio test which uses asymptotic critical values, i.e., critical values obtained from the test statistic's asymptotic null distribution. This approximation may be poor when the sample size is small since the test statistic's exact null distribution may not be close to its asymptotic counterpart. As a result, the likelihood ratio test may be considerably size-distorted. In particular, as revealed by our numerical results, it tends to be liberal, i.e., its exact type I error probability typically exceeds the test's significance level. This liberal behavior may, for instance, lead practitioners to include in their models regressors that have no explanatory power on the phenomenon under study. It is thus important to obtain alternative tests with superior control of the type I error frequency.

We derive the Bartlett correction (BARTLETT, 1954) to the likelihood ratio test and use it to define three Bartlett-corrected test statistics. The three modified tests are also based on asymptotic critical values. They are improved in the sense that their size distortions vanish faster than those of the unmodified test as the sample size increases. The derivation of the Bartlett correction factor is lengthy and cumbersome because the regression parameters in the two submodels are not orthogonal. We provide a closed-form expression for such a correction factor whose numerical evaluation only requires simple matrix operations. It can thus be easily computed in empirical applications that are based on a small number of data points. We present the results of Monte Carlo simulations that show that the proposed tests typically outperform the likelihood ratio test in small samples. Indeed, our numerical evidence shows that the new tests behave remarkably well even when the sample size is very small, the model is nonlinear and there is varying precision.

The motivation for our work is a public health problem, specifically the COVID-19 pandemic in Brazil. Our objective is to determine how the maximal COVID-19 mortality (per 100,000 inhabitants, weekly moving average) in each federative unit is impacted by factors (covariates) such as number of new COVID-19 cases, political positioning of the state governor in relation to the federal government, number of health care facilities, percentage change in the number of frontline physicians facing the disease, amount of chloroquine and hydroxycholoroquine received by the local government, percentage change in the number of ventilators, proportion of local hospitals that are private, and relative availability of intensive care unit beds. We obtained and organized data from several different sources on these variables for the 27 Brazilian federative units. We fitted the generalized extreme-value model to the data and, through hypothesis testing, arrived at reduced models. We arrived at a particular model using a Bartlett corrected test and at two different models when inferences were based on the likelihood ratio test and on a test that adopts an alternative finite sample correction. Measures of goodness-of-fit favored the model obtained using the Bartlett test. Interestingly, this model, unlike the alternative models, did not include among its regressors the number of boxes of chloroquine and hydroxychloroquine pills received by the local government from the federal government. As is well known, the medical literature has already established the inefficacy of these drugs.

The chapter is structured as follows. In Section 2.2, we present the general extreme-value regression model. In Section 2.3, we derive the Bartlett correction factor to the likelihood ratio test in such a class of models and use it to define three corrected test statistics. They can be used to test restrictions on the parameters that index the location and dispersion submodels of the general extreme-value regression model. Monte Carlo simulation evidence on the corrected tests' finite sample performance is presented in Section 2.4. In Section 2.5, we model the maximal COVID-19 mortality in Brazilian states. Some concluding remarks are offered in Section 2.6. Technical details on the derivation of the Bartlett correction factor for the likelihood ratio test statistic in the general extreme value regression model can be found in Appendix C.

2.2 THE GENERAL EXTREME-VALUE REGRESSION MODEL

There are two approaches in extreme-value theory, namely the block maximums (BM) method and the peaks-over-threshold (POT) method. In the BM method, the sample period is divided into periods of equal amplitude and the maximal data value in each period is selected; see Gumbel (1958). For further details on the BM method, see Ferreira and Haan (2015). In the POT method, by contrast, we select the observations that exceed a certain high threshold.

The Fisher-Tippett-Gnedenko Theorem (Fisher and Tippett (1928), Gnedenko (1943)) is a general result related to the asymptotic distribution of extreme order statistics. Let $V_1, \ldots, V_n$ be a sequence of independent and identically distributed random variables. Extreme value theory is interested in the behavior of the extremes of samples that are their maximum or minimum value, given by $M_n = \max\{V_1, \ldots, V_n\}$ and $\tilde{M}_n = \min\{V_1, \ldots, V_n\}$, respectively. The exact behavior of M_n is not known. Under certain regularity conditions, it is possible to gain knowledge on the approximate behavior of M_n when $n \to \infty$. In particular, asymptotic theory reveals the existence of a family of models that can be fitted to observed data on M_n; see Coles *et al.* (2001) and Smith (2003). The rule for extrapolating models based on mathematical bounds as finite level approximations was called by Coles *et al.* (2001) *the extreme value paradigm*. It is of interest to determine the possible limiting distributions of M_n. We have that $\Pr(M_n \leq v) = F(v)^n$, where $F(\cdot)$ is the distribution function of V, for $v < v_+$, where v_+ is the smallest value such that $F(v) = 1$, and $\lim_{n\to\infty} F(v)^n = 0$. The distribution of M_n is thus degenerate at v_+. The extreme value distribution arises when it is possible to obtain real-valued sequences $a_n \in \mathbb{R}_+$ and $b_n \in \mathbb{R}$ such that

$$\Pr\left(\frac{M_n - b_n}{a_n} \leq v\right) = F\Big(a_n v + b_n\Big)^n \to G(v) \quad \forall v \in \mathbb{R}$$

as $n \to \infty$, where G is a nondegenerate cumulative distribution function. Such a distribution is known as the general extreme values (GEV) family:

$$G(v) = \exp\left(-\left(1+\xi\left(\frac{v-\mu}{\phi}\right)\right)^{-1/\xi}\right), \quad \{v \in \mathbb{R} : 1+\xi(v-\mu)/\phi > 0\}, \mu, \xi \in \mathbb{R}, \phi > 0.$$

Here, μ, ϕ and ξ are location, scale and shape parameters, respectively.

According to Gnedenko (1943), Theorem 1, there are only three possible limiting distributions for $M_n^* = (M_n - b_n)/a_n$. By making $\xi \to 0$, $\xi > 0$ and $\xi < 0$ one obtains the classes of distributions, named as extreme-value distributions of types I, II and III; they are also called Gumbel, Fréchet and Weibull distributions, respectively. The most frequently used is Gumbel's law. The limiting distribution is Gumbel when $a_n = 1$, $b_n = \log(n)$ and $\xi \to 0$. For more details on modeling extremes, see Beirlant *et al.* (2006) and Castillo *et al.* (2005). Details on the extreme-value distributions can be found in Chapter 22 of Johnson, Kotz and Balakrishnan (1995). Kusumoto and Takeuchi (2020) contains a study on the uniform rate of convergence of the normalized sample maximum value (M_n^*) with respect to the Kolmogorov distance based on Stein equations.

The Gumbel distribution is commonly used for modeling events that occur with low probability. In particular, it is quite useful in risk modeling, and is used in finance, insurance, economics, risk management, hydrology, etc. Let Y be a random variable that follows the Gumbel distribution, denoted $Y \sim \text{Gumbel}(\mu, \phi)$. Its cumulative distribution and density functions are, respectively,

$$F(y; \mu, \phi) = \exp\left(-\exp\left(-\frac{y-\mu}{\phi}\right)\right)$$

and

$$f(y; \mu, \phi) = \frac{1}{\phi}\exp\left(\left(-\frac{y-\mu}{\phi}\right) - \exp\left(-\frac{y-\mu}{\phi}\right)\right),$$

$y \in \mathbb{R}$, where $\mu \in \mathbb{R}$ and $\phi > 0$ are location and dispersion parameters, respectively. The distribution mean, variance, and quantile function are

$$\mathbb{E}(Y) = \mu + \phi\gamma, \quad \text{Var}(Y) = \frac{\pi^2\phi^2}{6} \quad \text{and} \quad F^{-1}(p) = \mu - \phi\log(-\log(p)),$$

$0 < p < 1$, where $\gamma = \lim_{n\to\infty}\left(\sum_{k=1}^{n} k^{-1} - \log n\right) \approx 0.5772156649$ is the Euler-Mascheori constant and $F(\cdot)$ is the distribution function of Y.

In some cases the interest lies in modeling minimum values, $\tilde{M}_n = \min\{V_1,\ldots,V_n\}$, as, e.g., in Smith and Naylor (1987). That can de done by taking $-V_i$ which implies that $\tilde{M}_n = -M_n$.

Let $Y = (Y_1,\ldots,Y_n)^\top$ be a vector of independent random variables such that each y_i is Gumbel-distributed with location parameter μ_i and dispersion parameter ϕ_i, $i = 1,\ldots,n$. The general extreme value regression model proposed by Barreto-Souza and Vasconcellos (2011) is defined by

$$g_1(\mu) = \eta_1 = f_1(X;\beta) \tag{2.1}$$

and

$$g_2(\phi) = \eta_2 = f_2(Z;\theta), \tag{2.2}$$

where $\beta = (\beta_1,\ldots,\beta_p)^\top \in \mathbb{R}^p$ and $\theta = (\theta_1,\ldots,\theta_q)^\top \in \mathbb{R}^q$ are vectors of unknown regression parameters $(p+q<n)$, $\eta_1 = (\eta_{11},\ldots,\eta_{1n})^\top$, $\eta_2 = (\eta_{21},\ldots,\eta_{2n})^\top$, $\mu = (\mu_1,\ldots,\mu_n)^\top$ and $\phi = (\phi_1,\ldots,\phi_n)^\top$. It is worth noting that under nonlinearity the dimensions of the β and θ parameter vectors may differ from the number of covariates in the respective submodels. Also, X is the $n\times p_1$ matrix of location covariates whose ith line is $x_i = (x_{i1},\ldots,x_{ip_1})^\top$ and Z is the $n\times q_1$ matrix of dispersion covariates with ith line $z_i = (z_{i1},\ldots,z_{iq_1})^\top$. Such matrices can coincide partially, totally or not at all. Also, $f_1(\cdot,\cdot)$ and $f_2(\cdot,\cdot)$ are possibly nonlinear functions which are continuous and twice continuously differentiable in the second argument such that the ranks of $J_1 = \partial\eta_1/\partial\beta$ and $J_2 = \partial\eta_2/\partial\theta$ are p and q, respectively, and $g_1(\cdot)$ and $g_2(\cdot)$ are link functions which are strictly monotonic and twice continuously differentiable with domains $\mathbb{R}$ and $\mathbb{R}^+$, respectively.

Let $\upsilon = (\beta^\top,\theta^\top)^\top$ be the vector of regression coefficients of the model given in Equations (2.1) and (2.2). The log-likelihood function is

$$\ell(\upsilon) \equiv \ell(\beta,\theta) = \sum_{i=1}^{n} \ell_i(\mu_i,\phi_i), \tag{2.3}$$

where

$$\ell_i(\mu_i,\phi_i) = \log(f(y_i;\mu_i,\phi_i)) = -\log(\phi_i) - \frac{y_i-\mu_i}{\phi_i} - \exp\left(-\frac{y_i-\mu_i}{\phi_i}\right).$$

It is possible to express $\ell(\upsilon)$ in matrix form as

$$\ell(\upsilon) = -\iota^{\top}\{\Phi^* + T + T^*\}\,\iota,$$

where $\Phi^* = \mathrm{diag}(\log(\phi_1),\ldots,\log(\phi_n))$, $T = \mathrm{diag}(t_1,\ldots,t_n)$, $T^* = \mathrm{diag}(t_1^*,\ldots,t_n^*)$ with $t_i = (y_i-\mu_i)/\phi_i$ and $t_i^* = \exp(-t_i)$, ι being an n-vector of ones. It is possible to show that $\mathbb{E}(T) = \gamma$, $\mathbb{E}(T^*) = 1$, $\mathbb{E}(T^*T) = \gamma - 1$, $\mathbb{E}(T^*T^2) = \Gamma^{(2)}(2)$, where $\Gamma^{(2)}(\cdot)$ is the second derivative of the gamma function. For further details on these results, see Appendix B.

The score function is obtained by differentiating the log-likelihood function with respect to the regression parameters. We have that

$$U \equiv U(\upsilon) = \left(U_\beta(\beta,\theta)^\top, U_\theta(\beta,\theta)^\top\right)^\top,$$

where

$$U_{\beta_j}(\beta,\theta) = \frac{\partial \ell(\beta,\theta)}{\partial \beta_j} = \sum_{i=1}^{n} \frac{1}{\phi_i}(1-t_i^*)\left(\frac{\partial \mu_i}{\partial \eta_{1i}}\right)\left(\frac{\partial \eta_{1i}}{\partial \beta_j}\right), \quad j = 1,\ldots,p, \tag{2.4}$$

and

$$U_{\theta_J}(\beta,\theta) = \frac{\partial \ell(\beta,\theta)}{\partial \theta_J} = \sum_{i=1}^{n} \frac{1}{\phi_i}[-1 + t_i(1-t_i^*)]\left(\frac{\partial \phi_i}{\partial \eta_{2i}}\right)\left(\frac{\partial \eta_{2i}}{\partial \theta_J}\right), \quad J = 1,\ldots,q. \tag{2.5}$$

Let $\Phi = \mathrm{diag}(\phi_1,\ldots,\phi_n)$, $H_1 = \mathrm{diag}(\partial\mu_1/\partial\eta_{11},\ldots,\partial\mu_n/\partial\eta_{1n})$, $H_2 = \mathrm{diag}(\partial\phi_1/\partial\eta_{21},\ldots,$ $\partial\phi_n/\partial\eta_{2n})$ and let $\mathcal{I}$ denote the n-dimensional identity matrix. It is possible to write Equations (2.4) and (2.5) in matrix form as

$$\frac{\partial \ell(\beta,\theta)}{\partial \beta} = J_1^\top \Phi^{-1} H_1 (\mathcal{I} - T^*)\,\iota \quad \text{and} \quad \frac{\partial \ell(\beta,\theta)}{\partial \theta} = J_2^\top \Phi^{-1} H_2 [-\mathcal{I} + T(\mathcal{I} - T^*)]\,\iota,$$

respectively.

The maximum likelihood estimator of υ, say $\hat{\upsilon}$, solves $U(\upsilon) = 0_{p+q}$, where 0_{p+q} is the $(p+q)$-dimensional vector of zeros. It is not possible to express $\hat{\upsilon}$ in closed-form. Parameter estimates are typically obtained by numerically maximizing $\ell(\upsilon)$ using a Newton or quasi-Newton nonlinear optimization algorithm, such as the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm; see Nocedal and Wright (2006).

The Hessian matrix J is

$$\mathrm{J} = \begin{bmatrix} \mathrm{J}_{\beta\theta} & \mathrm{J}_{\beta\theta} \\ \mathrm{J}_{\theta\beta} & \mathrm{J}_{\theta\theta} \end{bmatrix} = \begin{bmatrix} \frac{\partial^2 \ell(\beta,\theta)}{\partial\beta\partial\beta^\top} & \frac{\partial^2 \ell(\beta,\theta)}{\partial\beta\partial\theta^\top} \\ \frac{\partial^2 \ell(\beta,\theta)}{\partial\theta\partial\beta^\top} & \frac{\partial^2 \ell(\beta,\theta)}{\partial\theta\partial\theta^\top} \end{bmatrix}.$$

Let $(jl)_i = \partial^2\eta_{1i}/\partial\beta_j\partial\beta_l$, $(j,l)_i = (\partial\eta_{1i}/\partial\beta_j)\times(\partial\eta_{1i}/\partial\beta_l)$, $(JL)_i = \partial^2\eta_{2i}/\partial\theta_J\partial\theta_L$, $(J,L)_i = (\partial\eta_{2i}/\partial\theta_J)\times(\partial\eta_{2i}/\partial\theta_L)$, etc. It follows that

$$\frac{\partial^2\ell(\beta,\theta)}{\partial\beta_j\partial\beta_l} = \sum_{i=1}^{n}\left\{-\frac{t_i^*}{\phi_i^2}\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)^2(j,l)_i + \left(\frac{1-t_i^*}{\phi_i}\right)\left(\frac{\partial^2\mu_i}{\partial\eta_{1i}^2}(j,l)_i + \frac{\partial\mu_i}{\partial\eta_{1i}}(jl)_i\right)\right\},$$

$$\frac{\partial^2\ell(\beta,\theta)}{\partial\beta_j\partial\theta_L} = \frac{\partial^2\ell(\beta,\theta)}{\partial\theta_L\partial\beta_j} = -\sum_{i=1}^{n}\left\{\frac{1}{\phi_i^2}[(t_i-1)t_i^*+1]\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)(j,L)_i\right\} \quad \text{and}$$

$$\frac{\partial^2\ell(\beta,\theta)}{\partial\theta_J\partial\theta_L} = \sum_{i=1}^{n}\left\{\left(\frac{1-2t_i(1-t_i^*)-t_i^2t_i^*}{\phi_i^2}\right)\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)^2(J,L)_i + \left(\frac{-1+t_i-t_it_i^*}{\phi_i}\right)\right.$$
$$\left.\times\left(\frac{\partial^2\phi_i}{\partial\eta_{2i}^2}(J,L)_i + \frac{\partial\phi_i}{\partial\eta_{2i}}(JL)_i\right)\right\}.$$

Fisher's information matrix is

$$\mathrm{I} \equiv \mathrm{I}(\beta,\theta) = \begin{bmatrix} \mathrm{I}_{\beta\beta} & \mathrm{I}_{\beta\theta} \\ \mathrm{I}_{\theta\beta} & \mathrm{I}_{\theta\theta} \end{bmatrix},$$

where

$$\mathrm{I}_{\beta\beta} = \mathbb{E}\left[-\frac{\partial^2\ell(\beta,\theta)}{\partial\beta_j\partial\beta_l}\right] = J_1^\top\Phi^{-1}H_1^2\Phi^{-1}J_1,$$

$$\mathrm{I}_{\beta\theta} = I_{\theta\beta}^\top = \mathbb{E}\left[-\frac{\partial^2\ell(\beta,\theta)}{\partial\beta_j\partial\theta_L}\right] = (\gamma-1)J_1^\top\Phi^{-1}H_1\Phi^{-1}H_2J_2 \quad \text{and}$$

$$\mathrm{I}_{\theta\theta} = \mathbb{E}\left[-\frac{\partial^2\ell(\beta,\theta)}{\partial\theta_J\partial\theta_L}\right] = \left(1+\Gamma^{(2)}(2)\right)J_2^\top\Phi^{-1}H_2^2\Phi^{-1}J_2.$$

It is noteworthy that the parameters β and θ are not orthogonal since Fisher's information matrix is not block-diagonal.

Several authors have modeled the occurrence of extreme values. Chan *et al.* (2008) developed point and interval estimation for the parameters that index the minimum extreme-value linear regression model with type II censoring. Bali (2003) introduced a more general extreme-value distribution using the Box-Cox transformation. Residuals, generalized leverage measures, an expression for Cook's distance, and local influence analysis for the general extreme-value regression model were developed by Oliveira Jr., Cribari-Neto and Nobre (2020). Generalized maximum likelihood estimators for the nonstationary generalized extreme value model were defined by Adlouni *et al.* (2007). Aryal and Tsokos (2009) used the quadratic rank transmutation map to generate a flexible family of probability distributions taking the extreme-value distribution as the base value distribution. Miladinovic and Tsokos (2009) used a modified Gumbel distribution to characterize the failure times of a given system. Calabrese and Osmetti (2013) introduced a generalized extreme-value regression model for binary dependent variables. For empirical analyses based on extreme-value regression models, see Calabrese and Giudici (2015).

2.3 IMPROVED LIKELIHOOD RATIO TESTS IN THE GENERAL EXTREME VALUE REGRESSION MODEL

As noted earlier, the chi-squared distribution may not yield a good approximation to the exact null distribution of the likelihood ratio test statistic when the sample size is not large. It is thus important to explore alternative testing strategies with superior performance in small samples. A promising strategy is that which involves modifying the likelihood ratio test statistic using what is known as 'the Bartlett correction factor'. This approach has been successfully used in different classes of models; see, e.g., Bayer and Cribari-Neto (2013), Botter and Cordeiro (1997), Chan, Chen and Yau (2014), Guedes, Cribari-Neto and Espinheira (2021), Lemonte, Ferrari and Cribari-Neto (2010) and Magalhães and Gallardo (2020). This is the approach we shall pursue.

We will adopt a general framework in which it is possible to perform inference on any set of parameters. To that end, we partition the parameter vector of the general extreme-value regression model, $\upsilon = (\beta^\top, \theta^\top)^\top$, as $\upsilon = (\rho^\top, \delta^\top)^\top$, where ρ is an l-dimensional vector of parameters of interest and δ is an r-dimensional vector of nuisance parameters. Our interest lies in making inferences on ρ; more specifically, we wish to test the null hypothesis $\mathcal{H}_0 : \rho = \rho^{(0)}$ against the composite alternative hypothesis $\mathcal{H}_1 : \rho \neq \rho^{(0)}$, where $\rho^{(0)}$ is a given l-dimensional vector. For instance, we may test $\mathcal{H}_0 : \rho^{(0)} = 0_l$, where 0_l is an l-dimensional vector of zeros, to determine whether the corresponding regressors should be in the model. The likelihood ratio test statistic for the test of $\mathcal{H}_0$: $\rho = \rho^{(0)}$ is

$$\omega = 2\left[\ell(\hat{\rho}, \hat{\delta}) - \ell(\rho^0, \tilde{\delta})\right], \tag{2.6}$$

where $(\hat{\rho}\top, \hat{\delta}^\top)^\top$ and $(\rho^{0\top}, \tilde{\delta}^\top)^\top$ are, respectively, the unrestricted and restricted maximum likelihood estimators of $(\rho^\top, \delta^\top)^\top$. Under the null hypothesis, ω is asymptotically distributed as χ^2_l; *veja* te discussion in Chapter 9 of Cox and Hinkley (1979). The test is then carried out using critical values from such an asymptotic distribution so that $\mathcal{H}_0$ is rejected at significance level $\alpha \in (0,1)$ if $\omega > \chi^2_{l,1-\alpha}$, i.e., if the test statistic value exceeds the $1-\alpha$ quantile of the limiting null distribution. It worth noticing that the test may yield unreliable inferences since the exact null distribution of ω may be poorly approximated by χ^2_l when n is small. A more reliable test can be obtained by resorting to second order asymptotic theory.

Bartlett (1937) proposed applying an adjustment factor to the likelihood ratio statistic. He obtained a scalar correction factor c that made the expected value of the new statistic equal to or very close to the expected value of the reference χ^2 distribution. By doing so, he achieved a better approximation to the null distribution of the modified test statistic by the chi-squared distribution. Under the null hypothesis, he computed the expected value of ω up to order n^{-1} and showed that $\mathbb{E}(\omega) = l + b + O(n^{-2})$, where $b = b(\upsilon)$ is of order $O(n^{-1})$. He then defined the quantity $c = 1 + b/l$. Thus, the first Bartlett-corrected test statistic is expressed as $\omega_{b1} = \omega/c$ and its expected value equals l if we neglect terms of order $O(n^{-2})$. He further established the agreement between the first three cumulants of ω_{b1} with those of the chi-squared distribution with error of order n^{-1}. Bartlett (1938) obtained similar results in the context of multivariate analysis; see also Bartlett (1947) and Bartlett (1954).

A more general approach came much later when Lawley (1956) obtained an expansion of $\ell(\hat{\rho}, \hat{\delta})$, under null hypothesis, up to order n^{-1} which involves derivatives of the log-likelihood function up to the fourth order. He then showed that the expected value of ω up to order $O(n^{-1})$ can be expressed as

$$2\mathbb{E}\left[\ell(\hat{\rho}, \hat{\delta}) - \ell(\rho, \delta)\right] = k + \epsilon_k + O(n^{-2}). \tag{2.7}$$

Here, ϵ_k is of order n^{-1} and can expressed as

$$\epsilon_k = \sum_{\upsilon_1, \ldots, \upsilon_k} (\lambda_{rstu} - \lambda_{rstuvw}). \tag{2.8}$$

The summation in (2.8) denoting the sum over all combinations of $p+q$ parameters of the vector $\upsilon = (\beta^\top, \theta^\top)^\top$, i.e., the indices r, s, t, u, v and w range in the vectors β and θ,

$$\lambda_{rstu} = \kappa^{rs}\kappa^{tu}\left\{\frac{\kappa_{rstu}}{4} - \kappa_{rst}^{(u)} + \kappa_{rt}^{(su)}\right\},$$
$$\lambda_{rstuvw} = \kappa^{rs}\kappa^{tu}\kappa^{vw}\left\{\kappa_{rtv}\left(\frac{\kappa_{suw}}{6} - \kappa_{sw}^{(u)}\right) + \kappa_{rtu}\left(\frac{\kappa_{svw}}{4} - \kappa_{sw}^{(v)}\right) + \kappa_{rt}^{(v)}\kappa_{sw}^{(u)} + \kappa_{rt}^{(u)}\kappa_{sw}^{(v)}\right\},$$

where $-\kappa^{rs}$ is the (r,s) element of the inverse of Fisher's information matrix, I^{-1}. He also showed that

$$2\mathbb{E}\left[\ell(\rho^0, \tilde{\delta}) - \ell(\rho, \delta)\right] = k - l + \epsilon_{k-l} + O(n^{-2}), \tag{2.9}$$

where ϵ_{k-l} is of order n^{-1} and can be written as

$$\epsilon_{k-l} = \sum_{\upsilon_{l+1},\ldots,\upsilon_k} (\lambda_{rstu} - \lambda_{rstuvw})\,. \tag{2.10}$$

Notice that the summation in (2.10) only operates on the $k-l$ nuisance parameters. It is thus possible to write the likelihood ratio test statistic ω given in (2.6) as

$$\omega = 2\left\{\left[\ell(\hat{\rho},\hat{\delta}) - \ell(\rho,\delta)\right] - \left[\ell(\rho^0,\tilde{\delta}) - \ell(\rho,\delta)\right]\right\}.$$

It follows from (2.7) and (2.9) that, under the null hypothesis, the expected value of ω is

$$\begin{aligned}\mathbb{E}(\omega) &= 2\mathbb{E}\left\{\left[\ell(\hat{\rho},\hat{\delta}) - \ell(\rho,\delta)\right] - \left[\ell(\rho^0,\tilde{\delta}) - \ell(\rho,\delta)\right]\right\} = l + \epsilon_k - \epsilon_{k-l} + O(n^{-2}) \\ &= l\left(1 + \frac{\epsilon_k - \epsilon_{k-l}}{l}\right) + O(n^{-2}).\end{aligned}$$

We thus arrive at the following Bartlett-corrected test statistic:

$$\omega_{b1} = \frac{\omega}{1 + \frac{\epsilon_k - \epsilon_{k-l}}{l}}.$$

It is easy to see that $\mathbb{E}(\omega_{b1}) = l + O(n^{-2})$. More generally, all of its cumulants coincide with those of the chi-squared distribution with error of order $O(n^{-3/2})$; see Lawley (1956). The quantity $c = 1 + (\epsilon_k - \epsilon_{k-l})/l$ is known as 'the Bartlett correction factor'.

The following two test statistics, denoted ω_{b2} and ω_{b3}, are equivalent to ω_{b1} up to order $O(n^{-1})$:

$$\omega_{b2} = \omega \exp\left(-\frac{(\varepsilon_k - \varepsilon_{k-l})}{l}\right) \quad \text{and} \quad \omega_{b3} = \omega\left[1 - \frac{(\varepsilon_k - \varepsilon_{k-l})}{l}\right];$$

see Lemonte, Ferrari and Cribari-Neto (2010). We note that ω_{b2} has the advantage of only assuming positive values.

According to Cordeiro (1987), Chesher and Smith (1995) and Hayakawa (1977), under null hypothesis, the corrected test statistic ω_{b1} is distributed as χ^2_l up to order $O(n^{-1})$. It is worth noting that the correction factor c may depend on unknown parameters. When that happens, they can be replaced by their restricted maximum likelihood estimates with no impact on the order of the approximation. Barndorff-Nielsen and Hall (1988) showed that the size distortions of the Bartlett-corrected tests are of order $O(n^{-2})$, and not of order $O(n^{-3/2})$, as previously believed. It is thus noteworthy that the size distortions of the Bartlett-corrected tests vanish much faster than that of the likelihood ratio test as

$n \to \infty$: $O(n^{-2})$ vs $O(n^{-1})$. For further details on Bartlett corrections, we refer readers to Cordeiro and Cribari-Neto (2014) and Cribari-Neto and Cordeiro (1996).

We aim to obtain Bartlett-corrected tests to be used in the general extreme-value regression model with varying dispersion. The Bartlett correction factor depends on the quantity ϵ_k, defined in (2.8), which in turn depends on joint cumulants of log-likelihood derivatives. It is noteworthy that the model parameters β and θ are not orthogonal, that is, Fisher's information matrix is not block diagonal, unlike what happens in the class of generalized linear models. As a consequence, the derivation of a closed-form expression for ϵ_k becomes quite cumbersome. The Bartlett correction factor for use in general extreme-value regressions can be obtained using the result developed by Cordeiro (1993). The author presents a general matrix formula for the the adjustment factor. In order to use his result, we need to obtain closed-form expressions for the log-likelihood cumulants up to the fourth order for the general extreme-value regression model in full generality, i.e., considering nonlinearity and varying dispersion. These cumulants are presented in Appendix C. From them, we obtain the matrices of order k which are used in the calculation of the correction factor whose (r,s) elements are

$$A^{(tu)} = \left\{ \frac{\kappa_{rstu}}{4} - \kappa_{rst}^{(u)} + \kappa_{rt}^{(su)} \right\}, \quad P^{(t)} = \left\{ \kappa_{rst} \right\}, \quad Q^{(u)} = \left\{ \kappa_{su}^{(r)} \right\}, \tag{2.11}$$

$t,u = 1,\ldots,k$. We then define the matrices L, $M^{(1)}$, $M^{(2)}$, $M^{(3)}$, $N^{(1)}$, $N^{(2)}$ and $N^{(3)}$ whose (r,s) elements are

$$\begin{aligned}
L_{rs} &= \left\{ \mathrm{tr}(\mathrm{I}^{-1} A^{(rs)}) \right\}, \\
M_{rs}^{(1)} &= \left\{ \mathrm{tr}(\mathrm{I}^{-1} P^{(r)} \mathrm{I}^{-1} P^{(s)}) \right\}, \\
M_{rs}^{(2)} &= \left\{ \mathrm{tr}(\mathrm{I}^{-1} P^{(r)} \mathrm{I}^{-1} Q^{(s)^\top}) \right\}, \\
M_{rs}^{(3)} &= \left\{ \mathrm{tr}(\mathrm{I}^{-1} Q^{(r)} \mathrm{I}^{-1} Q^{(s)}) \right\}, \\
N_{rs}^{(1)} &= \left\{ \mathrm{tr}(P^{(r)} \mathrm{I}^{-1}) \, \mathrm{tr}(P^{(s)} \mathrm{I}^{-1}) \right\}, \\
N_{rs}^{(2)} &= \left\{ \mathrm{tr}(P^{(r)} \mathrm{I}^{-1}) \, \mathrm{tr}(Q^{(s)} \mathrm{I}^{-1}) \right\}, \\
N_{rs}^{(3)} &= \left\{ \mathrm{tr}(Q^{(r)} \mathrm{I}^{-1}) \, \mathrm{tr}(Q^{(s)} \mathrm{I}^{-1}) \right\},
\end{aligned} \tag{2.12}$$

$r,s = 1,\ldots,k$, where $\mathrm{tr}(\cdot)$ is the trace operator. We use the matrices given in (2.11) and

(2.12) to express the following quantities in matrix form:

$$\begin{aligned}
&\sum_{\upsilon_1,\ldots,\upsilon_k} \lambda_{rstu} = \mathrm{tr}(\mathrm{I}^{-1}L),\\
&\sum_{\upsilon_1,\ldots,\upsilon_k} \kappa^{rs}\kappa^{tu}\kappa^{vw}\left\{\frac{\kappa_{rtv}\kappa_{suw}}{6} - \kappa_{rtv}\kappa_{sw}^{(u)} + \kappa_{rt}^{(v)}\kappa_{sw}^{(u)}\right\} = -\frac{\mathrm{tr}(\mathrm{I}^{-1}M^{(1)})}{6}\\
&+\mathrm{tr}(\mathrm{I}^{-1}M^{(2)}) - \mathrm{tr}(\mathrm{I}^{-1}M^{(3)}),\\
&\kappa^{rs}\kappa^{tu}\kappa^{vw}\left\{\frac{\kappa_{rtu}\kappa_{svw}}{4} - \kappa_{rtu}\kappa_{sw}^{(v)} + \kappa_{rt}^{(u)}\kappa_{sw}^{(v)}\right\} = -\frac{\mathrm{tr}(\mathrm{I}^{-1}N^{(1)})}{4}\\
&+\mathrm{tr}(\mathrm{I}^{-1}N^{(2)}) - \mathrm{tr}(\mathrm{I}^{-1}N^{(3)}).
\end{aligned} \tag{2.13}$$

Since the trace operator is linear, it follows from (2.8) and from the expressions in (2.13) that ε_k can be expressed in matrix notation as

$$\begin{aligned}
\varepsilon_k = \mathrm{tr}\left(\mathrm{I}^{-1}L\right) - \Bigg[-\frac{\mathrm{tr}\left(\mathrm{I}^{-1}M^{(1)}\right)}{6} + \mathrm{tr}\left(\mathrm{I}^{-1}M^{(2)}\right) - \mathrm{tr}\left(\mathrm{I}^{-1}M^{(3)}\right) - \frac{\mathrm{tr}\left(\mathrm{I}^{-1}N^{(1)}\right)}{4}\\
+\mathrm{tr}\left(\mathrm{I}^{-1}N^{(2)}\right) - \mathrm{tr}\left(\mathrm{I}^{-1}N^{(3)}\right)\Bigg] = \mathrm{tr}\left(\mathrm{I}^{-1}\left[L - \left(-\frac{M^{(1)}}{6} + M^{(2)} - M^{(3)} - \frac{N^{(1)}}{4}\right.\right.\right.\\
\left.\left.\left. +\ N^{(2)} - N^{(3)}\right)\right]\right) = \mathrm{tr}\left(\mathrm{I}^{-1}(L - M - N)\right).
\end{aligned} \tag{2.14}$$

The quantity ε_{k-l} given in (2.10) can be obtained analogously to the term (2.14) by only considering the $k-l$ nuisance parameters.

It is important to note that as n increases the size distortions of the three Bartlett corrected tests converge to zero considerably faster than those of the standard likelihood ratio test: the former are $O(n^{-2})$ whereas the latter is $O(n^{-1})$.

2.4 NUMERICAL EVIDENCE

In what follows we will present results from Monte Carlo simulations that were carried out to evaluate the performance of the likelihood ratio test (ω) and its Bartlett-corrected versions (ω_{b1}, ω_{b2} and ω_{b3}) in general extreme value regressions. Our focus is on small sample inference. Maximum likelihood estimates are obtained by numerically maximizing the logarithm of the likelihood function using the BFGS quasi-Newton nonlinear optimization algorithm with analytic first derivatives. There were very few BFGS convergence failures. Gumbel random number generation is performed using the inversion method. Uniform random number generation, which is needed for use with the inversion method, was carried out using the Mersenne-Twister algorithm. The number of Monte Carlo replications is 10,000. All computations were carried out using the R statistical computing environment (R Core Team, 2021). Since the likelihood ratio test is liberal, power simulations were performed using exact critical values obtained from the size simulations. We thus compare the powers of size-adjusted tests.

The optimization algorithm used for maximum likelihood estimation requires the specification of starting values for the parameters β and θ. In the linear model with fixed dispersion, we use as starting value for β, denoted by $\beta^{(0)}$, the ordinary least squares estimator of the linear regression of $g_1(y)$ on X, that is, $\beta^{(0)} = (X^\top X)^{-1} X^\top g_1(y)$, where $y = (y_1, \ldots, y_n)^\top$. Additionally, the starting value used for ϕ, denoted by $\phi^{(0)}$, is $\phi^{(0)} = (1/n)\left(6\hat{\sigma}_i^2/\pi^2\right)^{0.5}$, where $\hat{\sigma}_i^2 = \breve{\sigma}^2/[g_1'(\breve{\mu}_i)]^2$, with $\breve{\sigma}^2 = \breve{e}^\top \breve{e}/(n-p)$, $\breve{e} = g_1(y) - \breve{\mu}$, $\breve{\mu} = (\breve{\mu}_1, \ldots, \breve{\mu}_n)^\top$ and $\breve{\mu}_i = g_1^{-1}\left(x_i^\top \beta^{(0)}\right)$. Here, x_i is the ith row of X written as a column vector. To define $\breve{\sigma}^2$ we use (i) $\text{Var}(y_i) \approx \text{Var}[g_1(y_i)]/[g_1'(\mu_i)]^2$ (ESPINHEIRA; SANTOS; CRIBARI-NETO, 2017) and (ii) $\widehat{\text{Var}}(y_i) = \breve{\sigma}^2$. In the linear model with varying dispersion, $\beta^{(0)}$ is as before. The starting value for θ, say $\theta^{(0)}$, follows from starting values for $\phi_1, \ldots, \phi_n$: $\phi_i^{(0)} = \left(6\hat{\sigma}_i^2/\pi^2\right)^{0.5}$. We use $\theta^{(0)} = (Z^\top Z)^{-1} Z^\top g_2(\phi^{(0)})$. For the nonlinear model, starting values for the parameters were computed following the suggestion made in Section 4 of Espinheira, Santos and Cribari-Neto (2017) for nonlinear beta regressions with the appropriate changes; for details, see Appendix D.

At the outset, we consider the following fixed dispersion linear general extreme value regression model:

$$\mu_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4},$$

which we call 'Model I'. We test three null hypotheses, namely: $\mathcal{H}_0 : \beta_2 = 0$ $(l = 1)$, $\mathcal{H}_0 : \beta_2 = \beta_3 = 0$ $(l = 2)$ and $\mathcal{H}_0 : \beta_2 = \beta_3 = \beta_4 = 0$ $(l = 3)$. In the first case, the true parameter values are $\beta_1 = 1.0$, $\beta_2 = 0$, $\beta_3 = 1.0$ and $\beta_4 = -4.5$; in the second case, $\beta_1 = 1.0$, $\beta_2 = 0$, $\beta_3 = 0$ and $\beta_4 = -4.5$; in the third and final case, $\beta_1 = 1.0$, $\beta_2 = 0$, $\beta_3 = 0$ and $\beta_4 = 0$. In all three cases, $\phi = 5$. The covariate values were obtained as random $\mathcal{N}(0,1)$ draws. The significance levels are $\alpha = 10\%, 5\%, 1\%$ and the sample sizes are $n = 15, 20, 30, 40$. The tests null rejection rates are presented in Table 9 (all entries are percentages).

The figures in Table 9 lead to interesting conclusions. First, the likelihood ratio test is considerably liberal when the sample size is small. For example, when $n = 15$ and $\alpha = 10\%$, the test's null rejection rates range from 16.7% to 19.8%. Second, the

Table 9 – Null rejection rates (%), Model I.

	n	$\alpha = 10\%$ ω	ω_{b1}	ω_{b2}	ω_{b3}	$\alpha = 5\%$ ω	ω_{b1}	ω_{b2}	ω_{b3}	$\alpha = 1\%$ ω	ω_{b1}	ω_{b2}	ω_{b3}
$l = 1$	15	16.7	11.0	9.9	8.2	10.5	5.6	4.8	3.7	3.0	1.1	0.9	0.7
	20	16.0	10.8	10.2	9.1	9.2	5.3	4.9	4.3	2.7	1.1	1.0	0.8
	30	13.1	10.1	9.8	9.5	7.7	5.3	5.1	4.8	1.8	1.0	0.9	0.9
	40	12.0	10.1	10.0	9.9	7.0	5.3	5.3	5.2	1.6	1.0	1.0	1.0
$l = 2$	15	19.1	11.0	9.7	8.2	11.7	5.5	4.7	3.7	3.5	1.2	0.9	0.7
	20	17.2	10.9	10.1	9.1	9.9	5.4	5.0	4.4	2.9	1.0	0.9	0.7
	30	14.2	10.3	10.0	9.8	8.1	5.2	5.1	4.8	1.9	1.0	0.9	0.9
	40	13.1	10.2	10.0	9.9	7.2	5.2	5.1	5.0	1.8	1.2	1.1	1.1
$l = 3$	15	19.8	11.0	10.0	8.6	12.2	5.7	5.1	4.2	3.9	1.2	1.0	0.8
	20	17.4	10.8	10.1	9.4	10.1	5.5	5.2	4.8	2.8	1.1	1.0	0.8
	30	14.4	10.2	9.9	9.7	7.8	5.4	5.2	4.9	2.0	1.0	1.0	0.9
	40	13.3	10.3	10.2	10.0	7.2	5.3	5.2	5.0	1.8	1.1	1.0	1.0

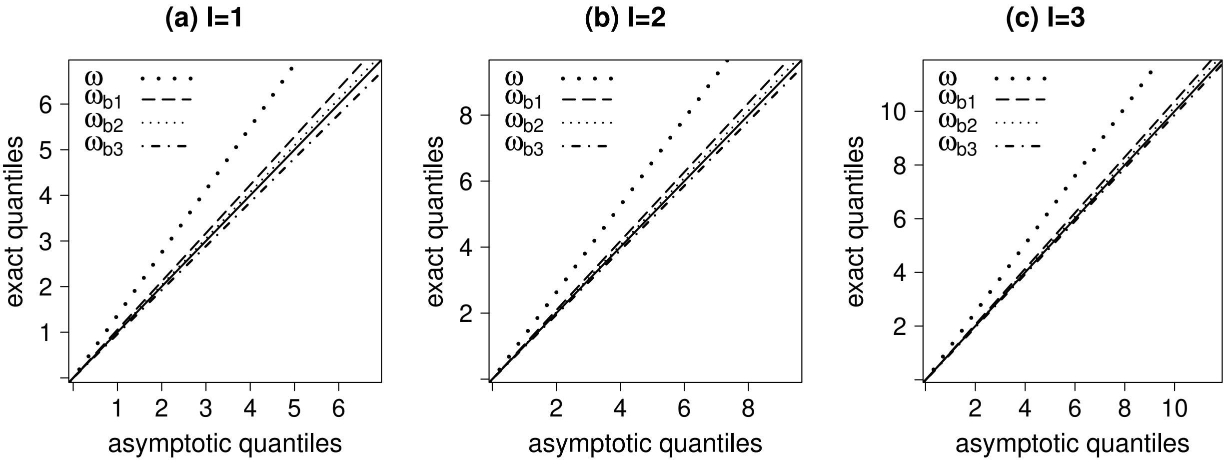
Source: The author (2021).

Bartlett-corrected tests (ω_{b1}, ω_{b2} and ω_{b3}) display much better control of the type I error frequency, i.e., they are much less size-distorted. For instance, when $n = 15$, $\alpha = 5\%$ and $l = 3$, their null rejection rates are 5.7% (ω_{b1}), 5.1% (ω_{b2}) and 4.2% (ω_{b3}) whereas that of the likelihood ratio test is 12.2%. Also, the sizes of the corrected tests are not impacted by the number of restrictions under test.

Figure 5 contains quantile-quantile (QQ) plots in which the exact quantiles of ω, ω_{b1}, ω_{b2} and ω_{b3} are plotted against the corresponding asymptotic quantiles. The 45º line indicates perfect agreement between exact and asymptotic null distributions. The sample size is $n = 20$ and the three panels correspond to $l = 1$ (left), $l = 2$ (center), and $l = 3$ (right). It is noteworthy that the exact quantiles of ω considerably exceed the respective chi-squared quantiles. The asymptotic approximation used in the test is thus quite poor. By contrast, the lines that correspond to the three Bartlett-corrected test statistics are quite close to the 45º line. There is thus very good agreement between their exact quantiles and the asymptotic quantiles.

The second set of simulations uses the following varying dispersion general

Figure 5 – Quantile-quantile plots, (a)$\mathcal{H}_0$: $\beta_2 = 0$ $(l = 1)$, (b)$\mathcal{H}_0$: $\beta_2 = \beta_3 = 0$ $(l = 2)$ and (c)$\mathcal{H}_0$: $\beta_2 = \beta_3 = \beta_4 = 0$ $(l = 3)$, n =20, Model I.



Source: The author (2021).

extreme-value regression model as data-generating process:

$$\mu_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3}$$
$$\log(\phi_i) = \theta_1 + \theta_2 z_{i2},$$

$i = 1, \ldots, n$, which we call 'Model II'. We test the null hypothesis $\mathcal{H}_0 : \beta_3 = 0$ against a two-sided alternative hypothesis. The true parameter values are $\beta_1 = 1.0$, $\beta_2 = 4.5$, $\beta_3 = 0$, $\theta_1 = -4.7$ and $\theta_2 = 0.4$. The covariate values were obtained as random draws from the Fisher-Snedecor distribution with 1 and 6 degrees-of-freedom, $\mathcal{F}(1,6)$. The significance levels are $\alpha = 10\%, 5\%, 1\%$ and the sample sizes are $n = 15, 20, 30, 40, 50, 60$. Table 10 contains the tests' null rejection rates (all entries are percentages).

Table 10 – Null rejection rates (%), $\mathcal{H}_0 : \beta_3 = 0$, Model II.

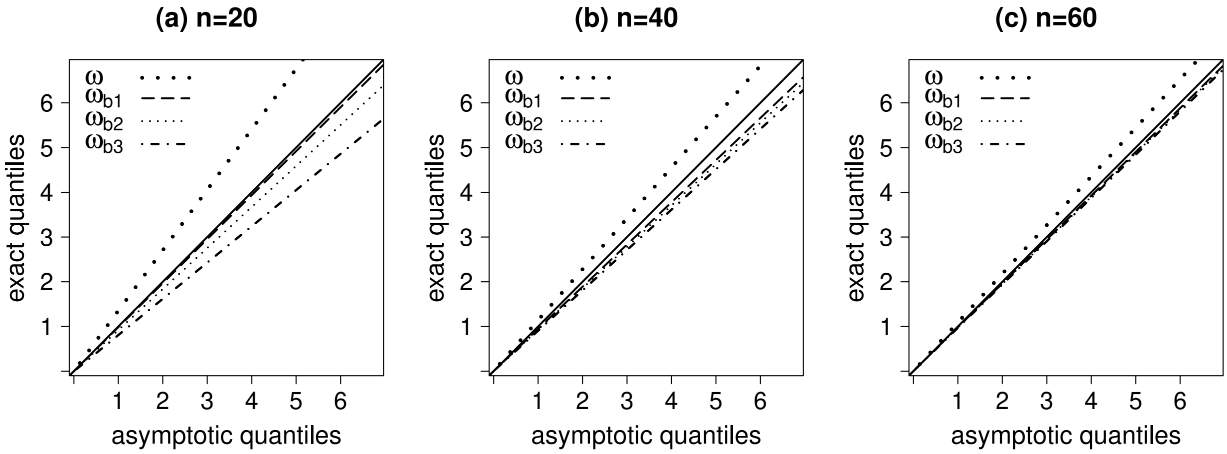
	$\alpha = 10\%$				$\alpha = 5\%$				$\alpha = 1\%$			
n	ω	ω_{b1}	ω_{b2}	ω_{b3}	ω	ω_{b1}	ω_{b2}	ω_{b3}	ω	ω_{b1}	ω_{b2}	ω_{b3}
15	16.1	9.6	8.4	6.7	9.7	5.0	4.1	3.3	2.9	1.0	0.8	0.6
20	15.4	9.7	8.6	7.2	9.0	4.7	4.2	3.5	2.6	0.9	0.8	0.7
30	13.8	9.2	8.6	7.5	7.9	4.0	4.5	3.4	2.0	0.7	0.6	0.5
40	12.7	9.5	9.2	8.8	6.9	4.6	4.4	4.1	1.6	0.8	0.7	0.7
50	12.1	9.7	9.5	9.4	6.5	5.0	5.0	4.9	1.5	0.9	0.9	0.9
60	11.6	9.7	9.6	9.4	5.8	4.5	4.4	4.4	1.4	1.1	1.1	1.0

Source: The author (2021).

The figures in Table 10 show that the likelihood ratio test is oversized, especially when the sample size is small. Consider, e.g., $n = 15$ and $\alpha = 5\%$. The test's null rejection is 9.7%, i.e., the test's exact size is almost twice the nominal size. The Bartlett-corrected tests perform much better. For the same sample size and significance level, their null rejection rates are 5.0% (ω_{b1}), 4.1% (ω_{b2}) and 3.3% (ω_{b3}). Such tests yield much better control of the type I error frequency. As expected, the size distortions of all tests diminish as the sample size increases.

Figure 6 contains QQ plots for three sample sizes: 20 (left), 40 (center) and 60 (right) observations. In all three panels, there is considerable disagreement between the exact and asymptotic quantiles of ω, especially for the smallest sample size. Two of the modified tests display conservative behavior (lines below the 45$^{\text{o}}$ line) when $n = 20$: ω_{b2} and ω_{b3}. There is very good agreement between the exact and asymptotic quantiles for the three Bartlett-corrected tests when $n = 40, 60$; when $n = 20$, the exact null distribution of ω_{b1} is very well approximated by its asymptotic counterpart.

Figure 6 – Quantile-quantile plots, $\mathcal{H}_0 : \beta_3 = 0$, Model II.



Source: The author (2021).

We now consider a nonlinear data-generating mechanism. In particular, data generation is performed using the following fixed dispersion general extreme-value regression model:

$$\mu_i = \beta_1 + \mathrm{e}^{(\beta_2 x_{i2})} + \beta_3 x_{i3} + \beta_4 x_{i4},$$

which we call 'Model III'. The interest lies in testing $\mathcal{H}_0 : \beta_3 = 0$ $(l = 1)$ and the true parameter values are $\beta_1 = 1.0$, $\beta_2 = 1.0$, $\beta_3 = 0$, $\beta_4 = -5.0$ and $\phi = 5$. The values of x_2, x_3 and x_4 were selected, respectively, as random draws from the following uniform distributions: $\mathcal{U}(0,5)$, $\mathcal{U}(0,1)$ and $\mathcal{U}(1,2)$. The significance levels are as before and the sample sizes are $n = 15, 20, 30, 40, 50, 60$. Here, X is an $n \times 4$ matrix with ith row given by $\left(1, x_{i2}\mathrm{e}^{\beta_2 x_{i2}}, x_{i3}, x_{i4}\right)$. The second and third order derivatives of the nonlinear predictor with respect to the $\beta's$ are $\partial^2\eta_{1i}/\partial\beta_j\partial\beta_l = x_{i2}^2\mathrm{e}^{\beta_2 x_{i2}}$ and $\partial^3\eta_{1i}/\partial\beta_j\partial\beta_l\partial\beta_m = x_{i2}^3\mathrm{e}^{\beta_2 x_{i2}}$, respectively.

Table 11 – Null rejection rates (%), $\mathcal{H}_0 : \beta_3 = 0$, Model III.

	$\alpha = 10\%$				$\alpha = 5\%$				$\alpha = 1\%$			
n	ω	ω_{b1}	ω_{b2}	ω_{b3}	ω	ω_{b1}	ω_{b2}	ω_{b3}	ω	ω_{b1}	ω_{b2}	ω_{b3}
15	13.8	8.0	7.5	6.7	7.2	4.1	4.1	3.8	2.1	1.4	1.5	1.4
20	13.7	9.3	8.9	8.2	7.3	4.5	4.2	3.9	1.3	0.8	0.8	0.7
30	13.0	9.5	9.2	8.9	6.7	4.3	4.2	4.0	1.2	0.7	0.7	0.7
40	12.8	10.3	10.1	9.9	7.1	5.2	5.0	4.9	1.6	1.1	1.1	1.1
50	11.8	9.3	9.1	9.0	5.9	4.5	4.4	4.4	1.3	0.9	0.9	0.9
60	11.5	9.6	9.5	9.4	6.0	4.8	4.8	4.7	1.2	0.9	0.9	0.9

Source: The author (2021).

Table 11 contains the tests' null rejection rates for testing $\mathcal{H}_0 : \beta_3 = 0$ in Model III. As in the previous simulations, the likelihood ratio test is liberal, i.e., its null rejection rates exceed the chosen significance levels. By contrast, the Bartlett-corrected tests display good control of the type I error frequency, being slightly conservative when n is quite small. For instance, when $n = 20$ and $\alpha = 10\%$, the null rejection rates of the modified tests are 9.3% (ω_{b1}), 8.9%(ω_{b2}) and 8.2% (ω_{b3}) whereas that of ω is 13.7%. The first corrected test, ω_{b1}, displays very good performance even when $n = 20$.

Next, we consider a general extreme value regression model with both nonlinearity and varying dispersion. The model is

$$\mu_i = \beta_1 + \beta_2 x_{i2} - \mathrm{e}^{(\beta_3 x_{i3})}$$

$$\log(\phi_i) = \theta_1 + \theta_2 z_{i2},$$

$i = 1, \ldots, n$. We shall refer to it as 'Model IV'. The null hypothesis under test is $\mathcal{H}_0 : \beta_2 = 0$ (hence, $l = 1$) and the following parameter values were used for data generation: $\beta_1 = 1.5$, $\beta_2 = 0$, $\beta_3 = 4.5$, $\theta_1 = 0.3$ and $\theta_2 = -2.0$. The values of x_{i2} were obtained as random $\mathcal{U}(0, 2)$ draws, and the values of x_{i3} and z_{i2} were obtained by random sampling from $\mathcal{U}(0, 1)$. Here, X is an $n \times 3$ matrix with ith line given by $\left(1, x_{i2}, -x_{i3}\mathrm{e}^{\beta_3 x_{i3}}\right)$ and Z is an $n \times 2$ matrix whose ith line is $(1, z_{i2})$. The second and third order derivatives of η_{1i} with respect to the

β's are, respectively,

$$\begin{cases} \frac{\partial^2 \eta_{1i}}{\partial\beta_j\partial\beta_l} = -x_{i3}^2 \mathrm{e}^{\beta_3 x_{i3}}, & \text{if } (j,l) = (3,3), \\ \frac{\partial^3 \eta_{1i}}{\partial\beta_j\partial\beta_l\partial\beta_m} = -x_{i3}^3 \mathrm{e}^{\beta_3 x_{i3}}, & \text{if } (j,l,m) = (3,3,3), \\ 0, & \text{otherwise.} \end{cases}$$

Table 12 contains the null rejection rates of the likelihood ratio test and its three Bartlett-corrected variants for $\alpha = 10\%, 5\%, 1\%$ and $n = 15, 20, 30, 40, 50, 60$.

Table 12 – Null rejection rates (%), $\mathcal{H}_0 : \beta_2 = 0$, Model IV.

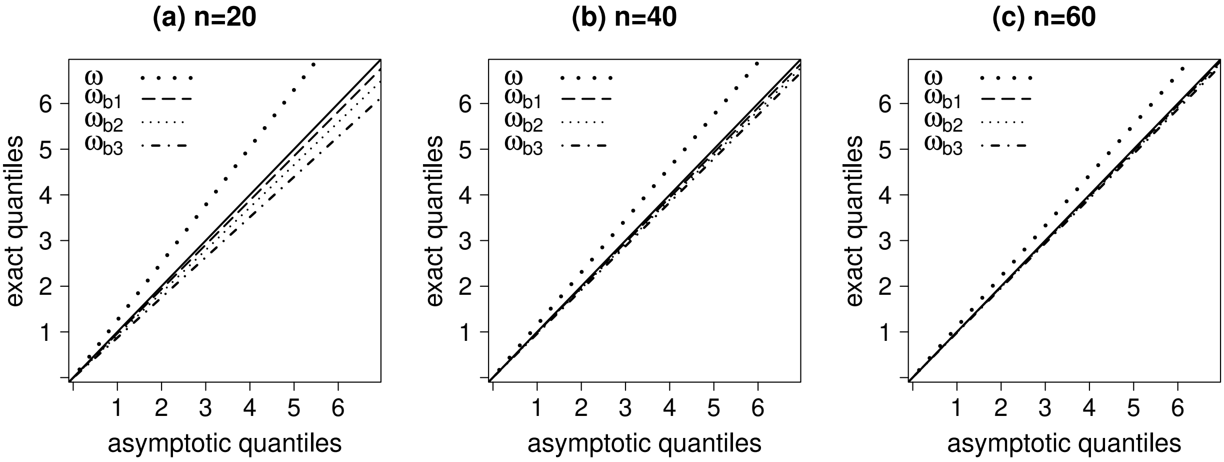
	$\alpha = 10\%$				$\alpha = 5\%$				$\alpha = 1\%$			
n	ω	ω_{b1}	ω_{b2}	ω_{b3}	ω	ω_{b1}	ω_{b2}	ω_{b3}	ω	ω_{b1}	ω_{b2}	ω_{b3}
15	18.2	10.8	9.9	7.9	10.9	5.9	5.4	4.0	3.5	1.6	1.5	1.1
20	15.7	9.8	9.1	8.0	8.9	5.1	4.6	3.8	2.6	1.1	0.9	0.8
30	13.8	9.9	9.3	8.7	7.3	4.6	4.3	4.1	1.9	0.9	0.8	0.8
40	12.3	9.5	9.2	9.0	6.7	4.6	4.5	4.3	1.5	1.0	0.9	0.8
50	12.3	9.9	9.7	9.5	6.6	4.9	4.7	4.6	1.5	0.9	0.9	0.9
60	11.8	9.6	9.5	9.4	6.3	4.9	4.8	4.7	1.2	0.8	0.8	0.8

Source: The author (2021).

The numerical results reported in Table 12 show that the likelihood ratio is quite oversized in small samples; e.g.: empirical size of 18.2% for $n = 15$ and $\alpha = 10\%$. Interestingly, the Bartlett-corrected tests perform remarkably well in this most challenging scenario (nonlinearity coupled with varying dispersion). For instance, for $\alpha = 10\%$ and $n = 15$, the null rejection rates of ω_{b1} and ω_{b2} are, respectively, 10.8% and 9.9%; the third modified test is slightly conservative (7.9%). It is also worth noticing that the Bartlett-corrected tests yield very good control of the type I error frequency for $n \geq 20$ even at $\alpha = 1\%$ (extreme right tail of the test statistics' asymptotic null distribution).

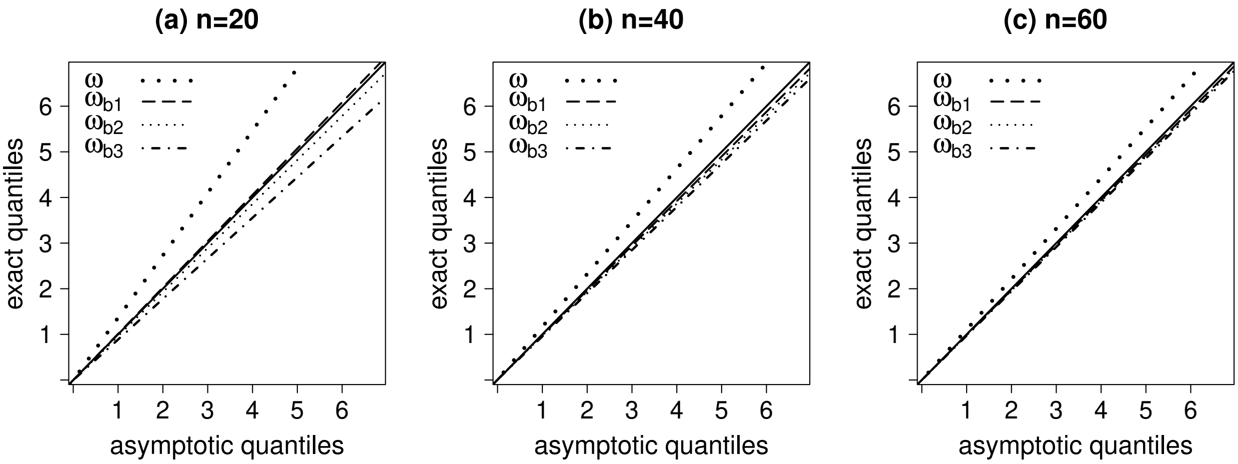
Figures 7 and 8 contain QQ plots for three sample sizes: 20 (left), 40 (center) and 60 (right) data points. They correspond to Models III and IV, respectively. Again, the exact quantiles of w greatly exceed the corresponding chi-squared quantiles, especially when $n < 60$. By contrast, there is very good agreement between the quantiles of the Bartlett-corrected test statistics and their asymptotic counterparts, especially when $n \geq 40$. When $n = 20$, such an agreement is quite good for two modified test statistics, namely: ω_{b1} and ω_{b2}. In this extreme situation (very small sample size, nonlinearity, varying dispersion), the test based on ω_{b3} is slightly conservative. Conservativeness is arguably less damaging than liberalness since the effective type I frequency does not exceed the test's significance level, which is commonly viewed as the maximum acceptable probability of committing a type I error.

Figure 7 – Quantile-quantile plots, $\mathcal{H}_0 : \beta_3 = 0$, Model III.



Source: The author (2021).

Figure 8 – Quantile-quantile plots, $\mathcal{H}_0$: $\beta_2 = 0$, Model IV.



Source: The author (2021).

Next, we shall evaluate the tests' ability to correctly detect that the null hypothesis under evaluation is false, i.e., we shall evaluate the tests' powers. We test $\mathcal{H}_0 : \beta_2 = 0$ in Model IV. Data generation is performed, nonetheless, using a value of β_2 that differs from zero; we denote such a value by γ. Table 13 contains the tests' non-null rejection rates for values of γ ranging from -2 to 2, for $n = 15, 30$ and for two significance levels (10% and 5%). As explained earlier, since the likelihood ratio is quite liberal, we compare the powers of size-adjusted tests. The results show that the tests become more powerful as γ moves away from zero (in either direction). Additionally, the powers of the four tests are very similar for all values of γ; when $n = 15$, the power of ω is very slightly superior to those of the three modified tests.

2.5 MODELING MAXIMAL COVID-19 MORTALITY IN BRAZIL

Next, we will model data on COVID-19, a disease caused by the SARS-CoV-2 virus, from Brazil. The modeling will involve performing hypothesis testing inferences to select the best fitting general extreme-value model. In our empirical analysis, we will use the method of maximal values per blocks of observations.

Several authors have modeled COVID-19 data; e.g., the 2020 COVID-19 outbreak in India was analyzed by Pandey *et al.* (2020) and predictions were produced for the number of cases in the coming weeks; Wu *et al.* (2020) evaluated whether long-term exposure to air pollution increases the severity of COVID-19 health outcomes, including death; Ribeiro *et al.* (2021) introduced a new regression model and used it to identify the covariates that impacted COVID-19 mortality rates in the first COVID-19 wave in the United States. Pro *et al.* (2021) modeled associations between county-level COVID-19 mortality, pre-pandemic county-level excessive drinking, and county rurality in the United States. Cordeiro *et al.* (2021) modeled coronavirus death rates in the first wave of twenty European countries using varying precision beta regressions.

Our interest lies in modeling the maximum value of the weekly moving average of daily deaths per 100,000 inhabitants; this will be our response variable (y). Data on such a maximal mortality was obtained for each of the 27 federative units/states of Brazil. Therefore, the sample size is $n = 27$. The data were obtained from the 'CoronaCities platform' (<https://coronacidades.org>) for the period ranging from April 26, 2020 to June 5, 2021.

COVID-19 is a respiratory infection caused by the SARS-CoV-2 virus that has high transmissibility. The SARS-CoV-2 virus was discovered in bronchoalveolar lavage samples from patients with pneumonia of unknown cause in Wuhan, China in December 2019. In Brazil, the first case of COVID-19 was detected on February 26, 2020 in the state of São Paulo and on March 20, 2020, SARS-CoV-2 transmission became community-wide. By June 2021 the country counted more than 18 million cases and more than 516,000 deaths from COVID-19. Increased health resources in the public sector was important in

Table 13 – Non-null rejection rates (%), $\mathcal{H}_0 : \beta_2 = 0$, Model IV.

	$n = 15$							
	$\alpha = 5\%$				$\alpha = 10\%$			
γ	ω	ω_{b1}	ω_{b2}	ω_{b3}	ω	ω_{b1}	ω_{b2}	ω_{b3}
−2.0	74.4	72.1	70.8	69.4	86.4	85.6	85.6	85.2
−1.75	63.3	61.1	59.7	58.1	77.9	77.2	76.8	76.3
−1.5	50.3	47.9	46.4	44.6	66.9	66.0	65.4	65.0
−1.25	37.3	35.1	34.0	32.8	53.2	52.6	52.0	51.5
−1.0	25.9	24.6	23.8	22.8	39.3	38.7	38.3	37.7
−0.75	16.7	15.8	15.1	14.7	27.7	27.2	27.0	26.5
−0.5	10.2	9.6	9.3	8.9	18.2	18.1	18.0	17.7
0.5	11.3	10.5	10.2	10.0	19.7	19.4	19.2	19.1
0.75	19.4	18.3	17.7	17.2	30.7	30.1	29.7	29.4
1.0	30.8	28.9	28.2	27.3	45.0	44.1	43.7	43.2
1.25	44.9	42.5	41.4	40.0	59.5	58.5	58.0	57.2
1.5	59.2	56.6	55.3	54.0	72.2	71.0	70.7	70.0
1.75	71.4	69.1	67.9	66.5	82.3	81.5	81.2	80.6
2.0	81.3	79.3	78.5	77.2	89.4	88.7	88.8	88.6
	$n = 30$							
−2.0	98.8	98.8	98.8	98.8	99.6	99.6	99.6	99.6
−1.75	95.9	95.9	95.9	95.9	98.3	98.3	98.3	98.3
−1.5	88.9	88.8	88.8	88.8	94.2	94.2	94.2	94.2
−1.25	75.2	75.1	75.0	75.1	84.6	84.6	84.6	84.5
−1.0	56.4	56.1	56.1	56.2	68.6	68.6	68.6	68.5
−0.75	36.0	35.6	35.6	35.6	47.8	47.8	47.7	47.7
−0.5	19.3	19.0	19.0	19.0	28.2	28.2	28.2	28.1
0.5	21.3	21.1	21.0	21.0	30.6	30.6	30.6	30.6
0.75	40.8	40.5	40.5	40.5	52.2	52.2	52.2	52.1
1.0	62.3	62.1	62.0	62.0	72.2	72.2	72.2	72.2
1.25	79.6	79.3	79.3	79.3	86.9	86.9	86.9	86.9
1.5	91.1	91.0	91.0	91.0	94.7	94.6	94.6	94.6
1.75	96.7	96.7	96.6	96.7	98.2	98.2	98.2	98.2
2.0	98.9	98.9	98.9	98.9	99.4	99.4	99.4	99.4

Source: The author (2021).

the fight against COVID-19, since the National Health Survey (PNS) conducted in 2019 by the Brazilian Institute of Geography and Statistics (IBGE) showed that 71.5% of the Brazilian population uses the Unified Public Health System (SUS) exclusively; 28.5% of the population has some type of private health insurance plan. Thus, it is important to understand how the maximum weekly moving average of deaths is impacted by variables related to available health resources in each state, such as total the number of ventilators; the number of physicians; the number of intensive care unit (ICU) beds; the number of emergency care units (UPA), health posts, and basic health units (UBS); and the proportion of hospitals that are private (not public). Additionally, it is important to understand the association between variables related to partisan politics with the response variable, since in many instances there were conflicting views involving the state governors on the one hand and the president of the country, Jair Bolsonaro, on the other hand. Thus, it is important to analyze how the political standing of governors towards the federal government relates to the peaks in death tolls.

For each state, we obtained and organized data on the following explanatory variables:

- x_2: weekly moving average of daily cases per 100,000 inhabitants in the week in which such a moving average was the highest (source: CoronaCidades platform);
- x_3: dummy variable that equals 1 if the state governor opposes the federal government and 0 otherwise (source: Veja magazine, March 22, 2021 (<https://is.gd/OpKtUG>));
- x_4: total number of UPA, UBS, health clinics and public hospitals per 100,000 inhabitants in the week in which the weekly moving average of daily deaths per 100,000 inhabitants was the highest (source: National Registry of Health Establishments (CNES) and DATASUS);
- x_5: ratio between the total number of physicians in June 2020 and December 2019; we consider (a) the number of physicians on the COVID-19 frontline per 100,000 inhabitants in June 2020 (source: CNN Brasil/CFM) and (b) the number of physicians per 100,000 inhabitants in December 2019 (source: CNES);
- x_6: logarithm of the mean monthly number of boxes of chloroquine and hydroxychloroquine pills received from the federal government up to the week with the maximal mortality (source: Brazilian Open Data Portal);
- x_7: percentage change in the number of available ventilators between December 2019

and the week in which mortality peaked, composed of (a) number of ventilators per 100,000 inhabitants delivered by the date on which the maximum weekly moving average of deaths per state was reached (source: Brazilian Open Data Portal) and (b) number of existing ventilators per 100,000 inhabitants in December 2019 (source: CNES);

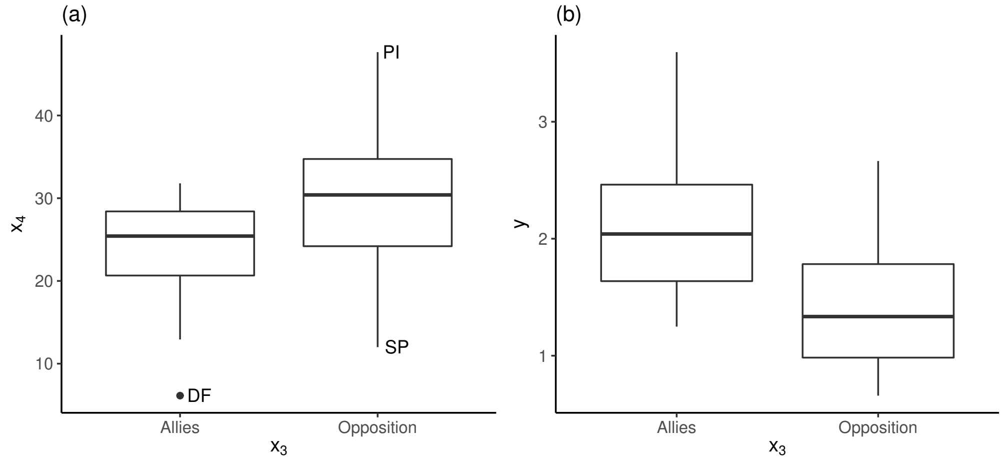
- x_8: proportion of hospitals that are private, considering general hospitals, specialized hospitals and day hospitals (hospitals where the maximum stay is 12 hours) in January 2020 (source: Brazilian Hospital Foundation (FBH));
- x_9: logarithm of the ratio between (i) the total number of ICU beds per 100,000 inhabitants in the week of maximal mortality (source: CNES) and (ii) the total number of ICU beds per 100,000 inhabitants in December 2019 (source: DATASUS);
- z_2: total number of existing ICU beds per 100,000 inhabitants in the week in which the weekly moving average of daily deaths per 100,000 inhabitants was the highest (source: National Registry of Health Establishments (CNES));
- $z_3 = x_3 \times [\log(z_2)]^2$.

We will comment on the choice of the above regressors later.

Figure 9 contains two boxplots. The left panel (panel (a)) relates to the total number of UPA, UBS, health clinics and public hospitals per 100,000 inhabitants (x_3) split in two groups which correspond to whether the state governor supports or opposes the federal government. The right panel (panel (b)) is for the response variable with the same two-group split. 'DF', 'PI' and 'SP' stand for Federal District, the state of Piauí and the state of São Paulo, respectively. It is clear from the two boxplots that there are typically fewer hospitals in the states whose governors support the president of Brazil and that the maximal COVID-19 mortality levels are generally higher in such states.

In Figure 10 we present scatter plots of the continuous covariates that will be considered for the first submodel and the maximum value of the weekly moving average of deaths per 100,000 inhabitants (y). Such graphs are indicative of the direction and intensity of the relationship between each regressor and the response variable. For example, y tends to decrease as the total number of health care facilities (x_4) increases. Additionally, the higher the proportion of private hospitals (x_8) the higher the peak mortality from COVID-19; this positive relationship is due to the fact that only 28.5% of Brazilians have some type of health insurance. We note that the maximal mortality (maximal value of

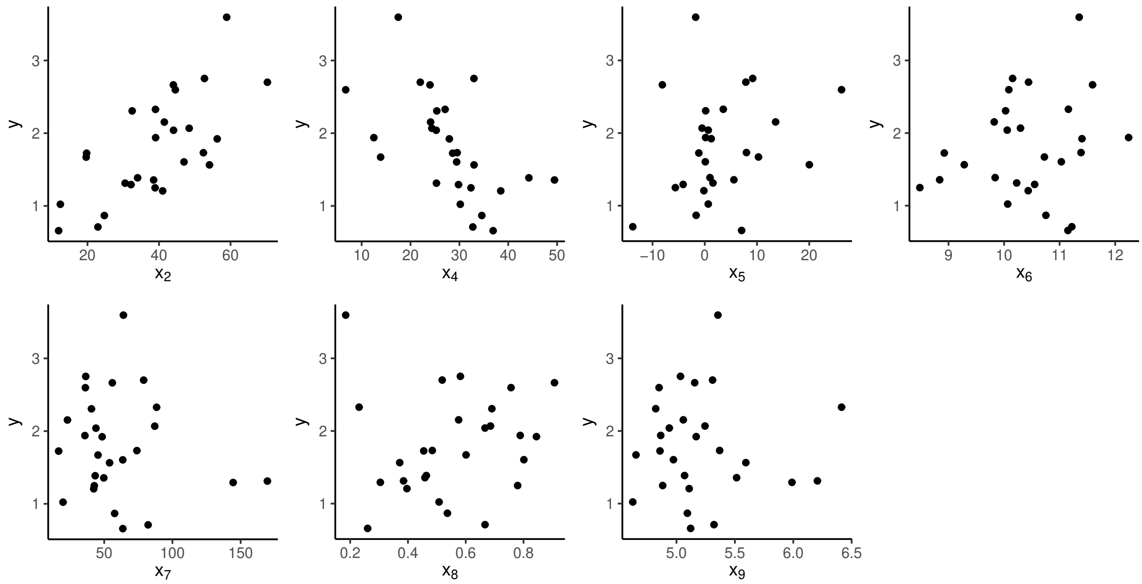
Figure 9 – (a) Boxplot of the number of hospitals per 100,000 inhabitants for states whose governors support and oppose the federal government, (b) boxplot of maximal mortality levels for the same two groups.



Source: The author (2021).

y_i) in the scatter plots in Figure 10 corresponds to the state of Amazonas, located in the North of Brazil; the maximum daily moving average of deaths occurred on February 4, 2021.

Figure 10 – Dispersion plots for the some continuous covariates.



Source: The author (2021).

In the following, we will fit general extreme-value regression models to the Brazilian COVID-19 data described above. Testing inferences will be based on the likelihood ratio test (ω) and on our first Bartlett-corrected test statistic (ω_{b1}). We will also consider another modified likelihood ratio test statistic that is available in the literature, namely: $\tilde{\omega}^* = \omega - 2\log(\zeta)$, a corrected test statistic obtained by Ferrari and Pinheiro

(2012) for use with the general extreme-value regression model; see their paper for a closed-form expression for ζ. It tends to deliver inferences that are more accurate than those achieved using ω, but the size distortions of the test vanish at the same rate as those of the likelihood ratio test as $n \to \infty$, namely $O(n^{-1})$. Recall that the size distortions of our Bartlett-corrected tests decay to zero at a much faster rate: $O(n^{-2})$. We will include testing inferences based on $\tilde{\omega}^*$ for comparison with those achieved using ω_{b1}.

We initially fitted the following variable dispersion extreme-value regression model (Model M1):

$$\text{M1:} \quad \begin{array}{l} \mu_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6} + \beta_7 x_{i7} + \beta_8 x_{i8} + \beta_9 x_{i9} \\ \log(\phi_i) = \theta_1 + \theta_2 z_{i2}^{3/2} + \theta_3 z_{i3} \end{array},$$

$i = 1, \ldots, 27$. The regressors used in the first submodel (submodel for μ) were selected based on the absolute values of their correlations with the response (y). We then fitted the fixed dispersion counterpart of Model M1 and obtained the standardized residuals from such a model fit. The ith standardized residual for the variable dispersion model is

$$r_i^s = \frac{y_i - \hat{\mu}_i - \gamma\hat{\phi}_i}{\sqrt{\pi^2 \hat{\phi}_i^2 / 6}} = \frac{\sqrt{6}}{\pi}\left(\frac{y_i - \hat{\mu}_i}{\hat{\phi}_i} - \gamma\right);$$

see Oliveira Jr., Cribari-Neto and Nobre (2020). For the fixed dispersion model, it suffices to replace $\hat{\phi}_i$ with $\hat{\phi}$ in the above expression. We then used such residuals to select the regressors to be included in the second submodel (submodel for ϕ). We did so by (i) plotting the residuals against the regressors and some functions of the regressors, (ii) computing the Pearson correlations between them and the regressors and also with some functions of them, and (iii) computing the Pearson correlations between the squared regressors and the covariates and also with some functions of them. In the residual plots, we searched for visible patterns. We selected two explanatory variables for the second submodel based on such an analysis.

All parameters in Model M1 are statistically non-null at the 10% significance level according to individual z tests. We will now further investigate the statistical significance of the covariates x_{i3}, x_{i6} and x_{i9}. After that, we will investigate the covariates in the second submodel. At first, we will individually test the exclusion of epidemiological covariates, which are (i) the logarithm of the average monthly number of boxes of chloroquine and hydroxychloroquine pills received from the federal government until mortality peaked (x_6) and (ii) the logarithm of the ratio between the total number of existing ICU beds per

100,000 inhabitants until reaching the maximum weekly moving average of deaths and the total number of existing ICU beds per 100,000 inhabitants in December 2019 (x_9). That is, we test $\mathcal{H}_0 : \beta_6 = 0$ and $\mathcal{H}_0 : \beta_9 = 0$. The p-values of the ω, ω_{b1} and $\tilde{\omega}^*$ tests for the exclusion of x_9 from the model are 0.1291, 0.3027 and 0.0976, respectively. Hence, unlike the test based on $\tilde{\omega}^*$, those based on ω and ω_{b1} do not reject the null hypothesis at the 10% significance level; in particular, the p-value of the Bartlett-corrected test is quite large. It follows that x_9 should be removed from Model M1 according to the likelihood ratio test and also according to our Bartlett-corrected test.

Next, we test $\mathcal{H}_0 : \beta_6 = 0$. The p-values are 0.0595 (ω), 0.1618 (ω_{b1}) and 0.0051 ($\tilde{\omega}^*$). The tests based on ω and $\tilde{\omega}^*$ reject $\mathcal{H}_0$ at 10%; by contrast, the Bartlett-correct test does not reject the null hypothesis at 10%. Our test thus indicates that x_6 should be removed from Model M1. The conclusion reached on the basis of the Bartlett-corrected test that the volume of chloroquine/hydroxychloroquine received by the state has no impact on the maximal mortality is in agreement with findings from medical studies that point to the lack of effectiveness of chloroquine/hydroxychloroquine in treating patients with COVID-19; see, e.g., Reis *et al.* (2021) and Rosenberg *et al.* (2020). We removed x_6 and x_9 from Model M1 and then tested $\mathcal{H}_0 : \theta_2 = 0$, the tests' p-values being 0.2766, 0.4071 and 0.0250 (ω, ω_{b1} and $\tilde{\omega}^*$, respectively). The likelihood ratio test and the Bartlett-corrected test agree that z_2 should be dropped from the second submodel; a different conclusion is reached when the test with the alternative finite sample correction is used since it rejects the null hypothesis at the 5% significance level. Finally, we test $\mathcal{H}_0 : \beta_3 = 0$. The p-values are 0.2546 (ω), 0.3456 (ω_{b1}) and 0.2558 ($\tilde{\omega}^*$), and the null hypothesis is not rejected; as a consequence, x_3 should be dropped from the model. The three tests thus agree that the independent variable related to partisan politics does not impact μ.

By following the testing inferences based on the Bartlett-corrected test statistic (ω_{b1}) we arrive at the following model (Model M2):

$$\text{M2:} \quad \begin{array}{l} \mu_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i4} + \beta_4 x_{i5} + \beta_5 x_{i7} + \beta_6 x_{i8} \\ \log(\phi_i) = \theta_1 + \theta_2 z_{i3} \end{array},$$

$i = 1, \ldots, 27$. All parameters are significantly non-null at the 5% significance level according to individual z tests. The maximum likelihood parameter estimates (standard errors in parentheses) are $\hat{\beta}_1 = 5.6962$ (0.7950), $\hat{\beta}_2 = 0.0303$ (0.0033), $\hat{\beta}_3 = -0.0582$ (0.0061), $\hat{\beta}_4 = -2.2602$ (0.6197), $\hat{\beta}_5 = -0.0070$ (0.0017), $\hat{\beta}_6 = -1.4950$ (0.2595), $\hat{\theta}_1 = -1.6541$ (0.1980) and $\hat{\theta}_2 = 0.0541$ (0.0237).

Following the testing inferences based on $\tilde{\omega}^*$, the test statistic that employs an alternative small sample adjustment, we arrive at the following model (Model M3):

$$\text{M3:} \quad \begin{array}{l} \mu_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i4} + \beta_4 x_{i5} + \beta_5 x_{i6} + \beta_6 x_{i7} + \beta_7 x_{i8} + \beta_8 x_{i9} \\ \log(\phi_i) = \theta_1 + \theta_2 z_{i2}^{3/2} + \theta_2 z_{i3} \end{array},$$

$i = 1, \ldots, 27$. The point estimates are $\hat{\beta}_1 = 5.6728$ (0.3032), $\hat{\beta}_2 = 0.0217$ (0.0013), $\hat{\beta}_3 = -0.0789$ (0.0059), $\hat{\beta}_4 = -2.2993$ (0.2988), $\hat{\beta}_5 = 0.0598$ (0.0225), $\hat{\beta}_6 = -0.0101$ (0.0006), $\hat{\beta}_7 = -1.2454$ (0.1132), $\hat{\beta}_8 = 0.4277$ (0.0734), $\hat{\theta}_1 = -5.0565$ (0.3465), $\hat{\theta}_2 = 0.0121$ (0.0013) $\hat{\theta}_2 = 0.1867$ (0.0239). Interestingly, unlike Model M2, this model includes as an explanatory variable the logarithm of the mean monthly number of boxes of chloroquine and hydroxychloroquine pills received up to the moment when deaths peaked (x_6). As noted earlier, it has been already established in the medical literature that these drugs have no efficacy against COVID-19; again, we refer readers to Reis *et al.* (2021) and Rosenberg *et al.* (2020). It is thus expected that x_6 should not impact maximal mortality levels. Such a covariate is not in the model selected on the basis of the Bartlett-corrected test (Model M2). The correlation between y and x_6 is slightly positive, but weak: 0.1753. It becomes even weaker when it is computed without the state with the highest mortality (Amazonas) in the data: 0.0826. Manaus's (the capital of Amazonas) health-care system experienced a shortage of oxygen and partially collapsed in early 2021. It turns out that Amazonas is the state with the fifth largest volume of chloroquine and hydroxychloroquine pills received from the federal government, only behind São Paulo, Rio Grande do Sul, Santa Catarina and Ceará; the values of x_6 for the latter three states and for Amazonas are nearly equal (11.5929, 11.4023, 11.3850 and 11.3523, respectively). The model we arrived at using Bartlett-corrected testing inferences does not include the volume of drugs that are known to be ineffective against Covid-19 as a relevant independent variable for explaining fluctuations in maximal mortality.

Using the conclusions reached based on ω (i.e., based on the likelihood ratio test) we arrive at a model (say, Model M4) that is similar to M2, but includes x_6 as a regressor in the submodel for μ (first submodel). The remaining covariates in both submodels coincide.

Neither of the three models (M2, M3 and M4) includes x_3, the partisan political dummy variable. Thus, we conclude that the political stand of the state governor (support or opposition to the federal government) does not impact the maximal mortality level. This variable, nonetheless, plays a role in determining the dispersions, since z_3 (which is a function of x_3) is a covariate in the submodels for ϕ in the three models. Dispersion is larger for states in which the governor opposes the federal government.

How do Models M1 through M4 compare? Which model yields the best fit? We will compare the three models using two sets of goodness-of-fit measures, namely: (i) model selection (information) criteria and (ii) pseudo-R^2 measures. A pseudo-R^2 is an overall measure of the quality of the model fit that assumes values between zero (no fit at all) and one (perfect fit). We propose a pseudo-R^2 for use with the generalized extreme-value regression model that is similar to that introduced by Ferrari and Cribari-Neto (2004) for beta regressions. Their measure needs to be modified to account for the fact that here μ_i differs from the mean of y_i. Here, the goodness-of-fit measure is given by the square of the correlation coefficient between y_i and $\hat{\mu}_i + \gamma\hat{\phi}_i$, hats denoting evaluation at the maximum likelihood estimates. We will refer to this measure as pseudo-R_1^2. Following Bayer and Cribari-Neto (2017), we also consider the following adjusted measure which includes a penalization term linked to the model dimension: pseudo-$R_2^2 = 1 - (1 - \text{pseudo-}R_1^2) \times [(n-1)/(n-p-q)]$. It is noteworthy that, unlike pseudo-R_1^2, pseudo-R_2^2 can assume negative negative values when pseudo-R_1^2 is very small, n is small and the number of parameters is large. Models M2 and M4 share the same pseudo-R_1^2: approximately 0.71. Such a value is much lower for Models M1 and M3 (approx. 0.46 and 0.09, respectively). The value of pseudo-R_2^2 favors Model M2 over M4: approx. 0.61 vs 0.59, respectively; the corresponding values for M1 and M4 are quite small (approx. 0.06 and negative, respectively).

As noted above, we followed Ferrari and Cribari-Neto (2004) when proposing a pseudo-R^2 measure for the general extreme-value regression model. Their measure was introduced in the context of beta regressions where μ_i is the mean of y_i, and is only impacted by the first submodel, that postulated for μ. It is given by the square of the correlation coefficient between the response transformed by the mean link function and the estimated linear predictor of the mean submodel. In the generalized extreme-value regression model, by contrast, μ_i does not equal the mean of y_i. Our pseudo-R_1^2 is thus given by the square of the correlation coefficient between y_i and $\hat{\mu}_i + \gamma\hat{\phi}_i$. Unlike what happens in beta regressions, it follows that here misspecification of the second submodel (that for ϕ) greatly impacts pseudo-R_1^2. This explains why the value of such a measure is so low for Model M3: this model includes a covariate in the submodel for ϕ which was found to be irrelevant by the two alternative testing criteria.

We will now compare the four models using model selection criteria. We consider the AIC (Akaike Information Criterion), AICc (Corrected Akaike Information Criterion) and BIC (Bayesian Information Criterion); for details on these crteria, we refer readers to Burnham and Anderson (2004). We note that the AICc is particularly useful here since it incorporates a small sample correction and our sample size is small ($n = 27$). Table 14 contains the values of three model selection criteria computed for the four models. The AIC values are similar for Models M1, M2 and M3, being slightly larger for Model M4. Model M2 is clearly favored by the AICc and BIC over all competing models. Indeed, the difference in AICc and BIC values in favor of Model M2 relative to M3 are 9.5383 and 3.8258, respectively. Thus, overall, the information criteria favor the model selected on the basis of Bartlett-corrected tests.

Table 14 – Model selection criteria.

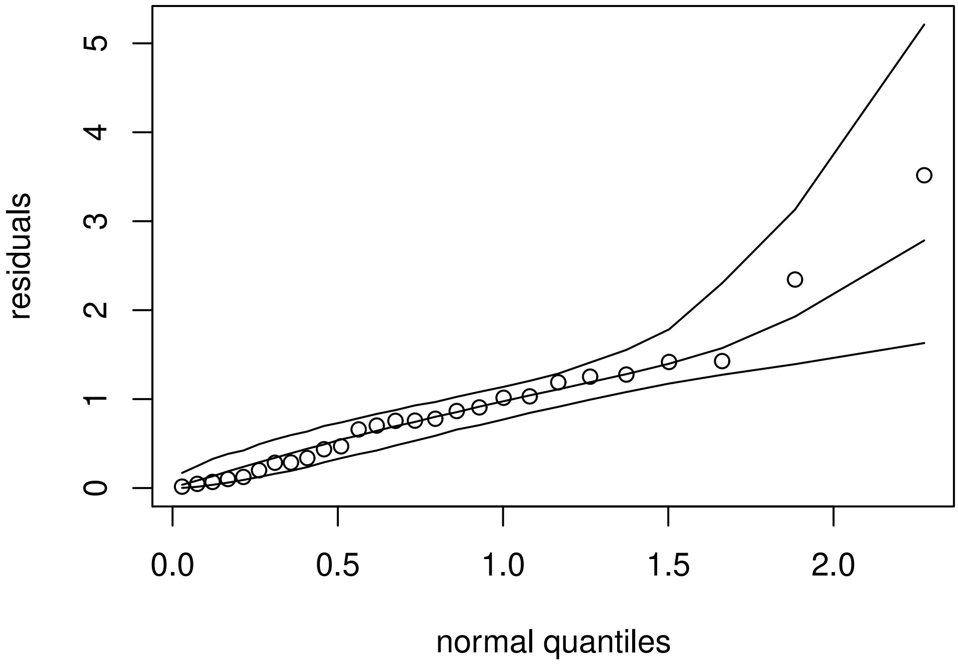
	AIC	AICc	BIC
Model M1	28.6968	50.9825	44.2469
Model M2	29.0338	37.0338	39.4005
Model M3	28.9721	46.5721	43.2263
Model M4	31.0309	41.6191	42.6934

Source: The author (2021).

Using the residuals from Model M2, we produced residual a half-normal probability plot with simulated envelopes for such a model which is given in Figure 11. The envelopes were constructed using 1000 replications. All residuals fall within the envelopes, and hence there is evidence that the model is correctly specified. We do not present similar plots for the other models for brevity, but we note that there were residuals outside the envelopes for Models M1 and M3.

Model M2 provides us with some interesting conclusions on the behavior of the maximum weekly moving average of deaths per 100,000 inhabitants. The coefficient estimate associated with the total number of health units is negative, which indicates that maximum mortality tends to be higher in states with fewer health units. Additionally,

Figure 11 – Residual half-normal probability plot for Model M2.



Source: The author (2021).

the response variable tends to decrease when there is a relative increase in the number of physicians. We also note that the maximum moving average of deaths decreases as the state receives additional ventilators. It is worth noting that on the basis of our first Bartlett-corrected test we conclude that the availability of chloroquine and hydroxychloroquine pills in the state does not positively or negatively impact the maximal mortality. The empirical application at hand showcases the importance of performing reliable testing inferences when the sample size is small. In particular, it is important to use tests that have good control of the type I error frequency even when the number of data points is small.

Neither of the three models (M2, M3 and M4) includes x_3, the partisan political dummy variable. Thus, we conclude that the political stand of the state governor (support or opposition to the federal government) does not impact the Gumbel location. This variable, nonetheless, plays a role in determining the dispersions, since z_3 (which is a function of x_3) is a covariate in the submodels for ϕ in the three models. Dispersion is larger for states in which the governor opposes the federal government. Since x_3 indirectly impacts ϕ_i, it also impacts the mean maximal mortality levels. The result implies that it is larger for states that are political opposition to the federal government.

2.6 CONCLUDING REMARKS

Extreme value modeling is used in many fields to uncover information on rare events. Oftentimes the occurrence of such events is impacted by changes in certain conditioning variables, and when that happens it is useful to resort to regression analysis. In this chapter, we focused on testing inferences in a general extreme-value regression model that accommodates both nonlinearity and varying dispersion. The model comprises of two submodels, one for the location parameter and a separate one for the dispersion parameter. The independent variables used in these submodel may not coincide, may coincide partially or may be the same. Testing inferences are often based on the likelihood ratio test. Such a test uses asymptotic (approximate) critical values and may be considerably size-distorted when the sample size is not large. Specifically, it tends to be quite liberal (oversized). This may lead, for instance, a practitioner to conclude that a given regressor must be in one or in both (linear or nonlinear) predictors used in the model when in truth such a regressor has no impact on the occurrence of the extreme event of interest. We addressed this shortcoming of the test by modifying its test statistic in a way that the limiting null distribution of the transformed test statistic better approximates the exact, unknown test statistic null distribution. As a result, the asymptotic critical values used for deciding whether the null hypothesis should be rejected are close to the true/ideal (unknown) critical values which implies that better control of the type I error frequency is achieved. More specifically, we derived the Bartlett correction factor to the likelihood ratio test statistic. The derivation was rather lengthy and cumbersome due to the non-orthogonality between the parameters that index the two submodels. The formula we provided for the correction factor is, nonetheless, rather simple and can be easily implemented into statistical software and computing environments. We used the correction factor to define three Bartlett-corrected test statistics. A noteworthy feature of our tests is that their size distortions vanish at a much faster rate than that of the likelihood ratio test: $O(n^{-2})$ vs $O(n^{-1})$. Additionally, all cumulants of our test statistics agree with those of the asymptotic null reference distribution to order $O(n^{-3/2})$.

Results from Monte Carlo simulations showed that the modified tests display much better control of the type I error frequency than the likelihood ratio test. In other words, they are much less prone to size distortion. Indeed, their null performance is remarkable even when the sample size is very small, there is nonlinearity and the model incorporates varying dispersion. For instance, when the sample contains only 15 data points, the significance level is 5%, the model is nonlinear and subject to nonconstant dispersion, the empirical size of the likelihood ratio test is 10.9% (over twice the nominal size!) whereas the exact sizes of the three modified tests developed in this chapter range from 4.0% to 5.9% (Table 12 in Section 2.4). In particular, we recommend the use of ω_{b1} which was the best performing test.

We modeled the maximum COVID-19 mortality in Brazil. The response variable used was the maximum value per federation unit of the weekly moving average of deaths per 100,000 inhabitants. Several explanatory variables were considered, including one of a political partisan nature that did not prove relevant. The sample size is small (27 observations). We performed hypothesis tests to arrive at the model that provides the best fit. In particular, we used the likelihood ratio test, one of the Bartlett-corrected tests that we derived and a test that incorporates an alternative correction for small samples. Interestingly, the likelihood ratio test and the alternative test led to two models (which we called Models M4 and M3, respectively) and the Bartlett test led to another, more parsimonious model (which we called Model M2). We note that Models M3 and M4 included in the first submodel a covariate related to the volume of chloroquine and hypoxycholoroquine received by the local government from the federal government. This result is surprising and was not expected, given the ineffectiveness of such medications in treating COVID-19. This regressor was not part of the model selected via the Bartlett-corrected test.

REFERENCES

ADLOUNI, S. E.; OUARDA, T. B. M. J.; ZHANG, X.; ROY, R.; BOBÉE, B. Generalized maximum likelihood estimators for the nonstationary generalized extreme value model. **Water Resources Research**, v. 43, n. 3, 2007.

ARELLANO-VALLE, R.; FERRARI, S. L. P.; CRIBARI-NETO, F. Bartlett and Bartlett-type corrections for testing linear restrictions. **Applied Economics Letters**, v. 6, n. 9, p. 547–549, 1999.

ARYAL, G. R.; TSOKOS, C. P. On the transmuted extreme value distribution with application. **Nonlinear Analysis: Theory, Methods & Applications**, Elsevier, v. 71, n. 12, p. e1401–e1407, 2009.

BALI, T. G. The generalized extreme value distribution. **Economics letters**, Elsevier, v. 79, n. 3, p. 423–427, 2003.

BARNDORFF-NIELSEN, O. E. Inference on full or partial parameters based on the standardized signed log likelihood ratio. **Biometrika**, v. 73, n. 2, p. 307–322, 1986.

BARNDORFF-NIELSEN, O. E. Modified signed log likelihood ratio. **Biometrika**, v. 78, n. 3, p. 557–563, 1991.

BARNDORFF-NIELSEN, O. E.; HALL, P. On the level-error after Bartlett adjustment of the likelihood ratio statistic. **Biometrika**, v. 75, n. 2, p. 374–378, 1988.

BARRETO-SOUZA, W.; VASCONCELLOS, K. L. P. Bias and skewness in a general extreme-value regression model. **Computational Statistics & Data Analysis**, Elsevier, v. 55, n. 3, p. 1379–1393, 2011.

BARTLETT, M. S. Properties of sufficiency and statistical tests. **Proceedings of the Royal Society of London. Series A – Mathematical and Physical Sciences**, v. 160, n. 901, p. 268–282, 1937.

BARTLETT, M. S. Further aspects of the theory of multiple regression. In: CAMBRIDGE UNIVERSITY PRESS. **Mathematical Proceedings of the Cambridge Philosophical Society**. [S.l.], 1938. v. 34, n. 1, p. 33–40.

BARTLETT, M. S. Multivariate analysis. **Supplement to the Journal of the Royal Statistical Society**, JSTOR, v. 9, n. 2, p. 176–197, 1947.

BARTLETT, M. S. A note on the multiplying factors for various χ^2 approximations. **Journal of the Royal Statistical Society B**, JSTOR, v. 16, p. 296–298, 1954.

BAYER, F. M.; CRIBARI-NETO, F. Bartlett corrections in beta regression models. **Journal of Statistical Planning and Inference**, v. 143, n. 3, p. 531–547, 2013.

BAYER, F. M.; CRIBARI-NETO, F. Model selection criteria in beta regression with varying dispersion. **Communications in Statistics - Simulation and Computation**, v. 46, n. 4, p. 729–746, 2017.

BEIRLANT, J.; GOEGEBEUR, Y.; SEGERS, J.; TEUGELS, J. L. **Statistics of Extremes: Theory and Applications**. New York: John Wiley & Sons, 2006.

BOTTER, D. A.; CORDEIRO, G. M. Bartlett corrections for generalized linear models with dispersion covariates. **Communications in statistics - Theory and methods**, v. 26, n. 2, p. 279–307, 1997.

BOUCHARD, C.; LANGE, F.; GUéRAUD, F.; RIVES, J.; TENTELIER, C. Sexual maturity increases mobility and heterogeneity in individual space use in Atlantic salmon (Salmo salar) parr. **Fish Biology**, v. 96, n. 4, p. 925–938, 2020.

BURNHAM, K. P.; ANDERSON, D. R. Multimodel inference: Understanding AIC and BIC in model selection. **Sociological Methods Research**, v. 33, n. 2, p. 261–304, 2004.

CALABRESE, R.; GIUDICI, P. Estimating bank default with generalised extreme value regression models. **Journal of the Operational Research Society**, v. 66, n. 11, p. 1783–1792, 2015.

CALABRESE, R.; OSMETTI, S. A. Modelling small and medium enterprise loan defaults as rare events: The generalized extreme value regression model. **Journal of Applied Statistics**, v. 40, n. 6, p. 1172–1188, 2013.

CASTILLO, E.; HADI, A. S.; BALAKRISHNAN, N.; SARABIA, J. M. Extreme value and related models with applications in engineering and science. John Wiley & Sons, Wiley Hoboken, 2005.

CHAN, N. H.; CHEN, K.; YAU, C. Y. On the Bartlett correction of empirical likelihood for Gaussian long-memory time series. **Electronic Journal of Statistics**, v. 8, n. 1, p. 1460–1490, 2014.

CHAN, P. S.; NG, H.; BALAKRISHNAN, N.; ZHOU, Q. Point and interval estimation for extreme-value regression model under type-ii censoring. **Computational Statistics & Data Analysis**, v. 52, n. 8, p. 4040–4058, 2008.

CHEN, K. K.; CHIU, R.-H.; CHANG, C.-T. Using beta regression to explore the relationship between service attributes and likelihood of customer retention for the container shipping industry. **Transportation Research Part E: Logistics and Transportation Review**, v. 104, p. 1–16, 2017.

CHESHER, A.; SMITH, R. J. Bartlett corrections to likelihood ratio tests. **Biometrika**, Oxford University Press, v. 82, n. 2, p. 433–436, 1995.

COLES, S.; BAWA, J.; TRENNER, L.; DORAZIO, P. **An Introduction to Statistical Modeling of Extreme Values**. New York: Springer, 2001.

CORDEIRO, G. M. On the corrections to the likelihood ratio statistics. **Biometrika**, Oxford University Press, v. 74, n. 2, p. 265–274, 1987.

CORDEIRO, G. M. General matrix formulae for computing Bartlett corrections. **Statistics & Probability Letters**, v. 16, n. 1, p. 11–18, 1993.

CORDEIRO, G. M.; CRIBARI-NETO, F. **An Introduction to Bartlett Correction and Bias Reduction**. New York: Springer, 2014. ISBN 9783642552540.

CORDEIRO, G. M.; FIGUEIREDO, D.; SILVA, L.; ORTEGA, E. M. M.; PRATAVIERA, F. Explaining COVID-19 mortality rates in the first wave in Europe. **Model Assisted Statistics and Applications**, v. 16, p. 211–221, 2021.

COX, D. R.; HINKLEY, D. V. **Theoretical Statistics**. New York: Chapman & Hall/CRC, 1979.

CRIBARI-NETO, F.; CORDEIRO, G. M. On Bartlett and Bartlett-type corrections. **Econometric Reviews**, v. 15, n. 4, p. 339–367, 1996.

CRIBARI-NETO, F.; LUCENA, S. E. F. Non-nested hypothesis testing in the class of varying dispersion beta regressions. **Journal of Applied Statistics**, v. 42, n. 5, p. 967–985, 2015.

CRIBARI-NETO, F.; QUEIROZ, M. P. F. On testing inference in beta regressions. **Journal of Statistical Computation and Simulation**, v. 84, n. 1, p. 186–203, 2014.

CRIBARI-NETO, F.; SOUZA, T. C. Religious belief and intelligence: Worldwide evidence. **Intelligence**, v. 41, p. 482–489, 2013.

CRIBARI-NETO, F.; ZEILEIS, A. Beta regression in R. **Journal of Statistical Software**, v. 34, n. 2, p. 1–24, 2010.

DAS, U.; DHAR, S. S.; PRADHAN, V. Corrected likelihood-ratio tests in logistic regression using small-sample data. **Communications in Statistics - Theory and Methods**, v. 47, n. 17, p. 4272–4285, 2018.

DOUMA, J. C.; WEEDON, J. T. Analysing continuous proportions in ecology and evolution: A practical introduction to beta and dirichlet regression. **Methods in Ecology and Evolution**, v. 10, n. 9, p. 1412–1430, 2019.

ESPINHEIRA, P. L.; SANTOS, E. G.; CRIBARI-NETO, F. On nonlinear beta regression residuals. **Biometrical Journal**, v. 59, n. 3, p. 445–461, 2017.

FERRARI, S. L. P.; CRIBARI-NETO, F. Beta regression for modelling rates and proportions. **Journal of Applied Statistics**, v. 31, n. 7, p. 799–815, 2004.

FERRARI, S. L. P.; ESPINHEIRA, P. L.; CRIBARI-NETO, F. Diagnostic tools in beta regression with varying dispersion. **Statistica Neerlandica**, v. 65, n. 3, p. 337–351, 2011.

FERRARI, S. L. P.; LUCAMBIO, F.; CRIBARI-NETO, F. Improved profile likelihood inference. **Journal of Statistical Planning and Inference**, v. 134, n. 2, p. 373–391, 2005.

FERRARI, S. L. P.; PINHEIRO, E. C. Improved likelihood inference in beta regression. **Journal of Statistical Computation and Simulation**, v. 81, n. 4, p. 431–443, 2011.

FERRARI, S. L. P.; PINHEIRO, E. C. Small-sample likelihood inference in extreme-value regression models. **Journal of Statistical Computation and Simulation**, v. 84, n. 3, p. 582–595, 2012.

FERREIRA, A.; HAAN, L. D. On the block maxima method in extreme value theory: Pwm estimators. **The Annals of statistics**, Institute of Mathematical Statistics, v. 43, n. 1, p. 276–298, 2015.

FISHER, R. A.; TIPPETT, L. H. C. Limiting forms of the frequency distribution of the largest or smallest member of a sample. **Mathematical Proceedings of the Cambridge Philosophical Society**, v. 24, n. 2, p. 180–190, 1928.

GANZACH, Y.; GOTLIBOVSKI, C. Intelligence and religiosity: Within families and over time. **Intelligence**, v. 41, n. 5, p. 546–552, 2013.

GNEDENKO, B. Sur la distribution limite du terme maximum d'une serie aleatoire. **Annals of Mathematics**, v. 44, n. 3, p. 423–453, 1943.

GRADSHTEYN, I. S.; RYZHIK, I. M. **Table of Integrals, Series, and Products**. 7th. ed. [S.l.]: Elsevier, 2007.

GUEDES, A. C.; CRIBARI-NETO, F.; ESPINHEIRA, P. L. Modified likelihood ratio tests for unit gamma regressions. **Journal of Applied Statistics**, v. 47, n. 9, p. 1562–1586, 2020.

GUEDES, A. C.; CRIBARI-NETO, F.; ESPINHEIRA, P. L. Bartlett-corrected tests for varying precision beta regressions with application to environmental biometrics. **PLOS ONE**, v. 16, n. 6, p. e0253349, 2021.

GUMBEL, E. J. **Statistics of Extremes**. New York: Columbia University Press, 1958.

HARRIS, P. A note on Bartlett adjustments to likelihood ratio tests. **Biometrika**, Oxford University Press, v. 73, n. 3, p. 735–737, 1986.

HAYAKAWA, T. The likelihood ratio criterion and the asymptotic expansion of its distribution. **Annals of the Institute of Statistical Mathematics**, Springer, v. 29, n. 1, p. 359–378, 1977.

JOHNSON, N. L.; KOTZ, S.; BALAKRISHNAN, N. **Continuous Univariate Distributions - Volume 2**. New York: John Wiley & Sons, 1995.

KANAZAWA, S. Why liberals and atheists are more intelligent. **Social Psychology Quarterly**, v. 73, n. 1, p. 33–57, 2010.

KIESCHNICK, R.; MCCULLOUGH, B. D. Regression analysis of variates observed on $(0, 1)$: Percentages, proportions and fractions. **Statistical Modelling**, v. 3, n. 3, p. 193–213, 2003.

KOTZ, S.; NADARAJAH, S. **Extreme Value Distributions: Theory and Applications**. River Edge: World Scientific, 2000.

KUSUMOTO, H.; TAKEUCHI, A. Remark on rates of convergence to extreme value distributions via the Stein equations. **Extremes**, Springer, v. 23, n. 3, p. 411–423, 2020.

LAWLEY, D. N. A general method for approximating to the distribution of likelihood ratio criteria. **Biometrika**, v. 43, n. 3-4, p. 295–303, 1956.

LEMONTE, A. J.; FERRARI, S. L. P.; CRIBARI-NETO, F. Improved likelihood inference in Birnbaum-Saunders regressions. **Computational Statistics & Data Analysis**, v. 54, n. 5, p. 1307–1316, 2010.

LYNN, R.; HARVEY, J.; NYBORG, H. Average intelligence predicts atheism rates across 137 nations. **Intelligence**, v. 37, n. 1, p. 11–15, 2009.

LYNN, R.; MEISENBERG, G. National IQs calculated and validated for 108 nations. **Intelligence**, v. 38, n. 4, p. 353–360, 2010.

LYNN, R.; VANHANEN, T. **IQ and the Wealth of Nations**. Augusta: Washington Summit Publishers, 2006.

MAGALHãES, T. M.; GALLARDO, D. I. Bartlett and Bartlett-type corrections for censored data from a Weibull distribution. **SORT**, v. 44, n. 1, p. 127–140, 2020.

MANDAL, S.; SRIVASTAV, R. K.; SIMONOVIC, S. P. Use of beta regression for statistical downscaling of precipitation in the Campbell River Basin, British Columbia, Canada. **Journal of Hydrology**, v. 538, p. 49–62, 2016.

MILADINOVIC, B.; TSOKOS, C. P. Ordinary, Bayes, empirical Bayes, and non-parametric reliability analysis for the modified Gumbel failure model. **Nonlinear Analysis: Theory, Methods & Applications**, Elsevier, v. 71, n. 12, p. e1426–e1436, 2009.

MULLEN, L.; MARSHALL, L.; MCGLYNN, B. A beta regression model for improved solar radiation predictions. **Journal of Applied Meteorology and Climatology**, v. 52, n. 8, p. 1923–1938, 2013.

NOCEDAL, J.; WRIGHT, S. J. **Numerical Optimization**. 2nd. ed. New York: Springer, 2006.

OLIVEIRA JR., J. V.; CRIBARI-NETO, F.; NOBRE, J. S. Influence diagnostics and model validation for the generalized extreme value nonlinear regression model. **Journal of Statistical Computation and Simulation**, v. 90, n. 3, p. 515–549, 2020.

OSPINA, R.; CRIBARI-NETO, F.; VASCONCELLOS, K. L. P. Improved point and interval estimation for a beta regression model. **Computational Statistics & Data Analysis**, v. 51, n. 2, p. 960–981, 2006. Erratum: 55, 2011, 2445.

OSPINA, R.; FERRARI, S. L. P. A general zero-or-one inflated beta regression models. **Computational Statistics & Data Analysis**, v. 56, n. 6, p. 1609–1623, 2012.

PANDEY, G.; CHAUDHARY, P.; GUPTA, R.; PAL, S. SEIR and regression model based COVID-19 outbreak predictions in India. **arXiv preprint arXiv:2004.00958v1**, 2020.

PENNYCOOK, G.; M., R. R.; KOEHLER, D. J.; FUGELSANG, J. A. Atheists and agnostics are more reflective than religious believers: Four empirical studies and a meta-analysis. **PLOS ONE**, v. 11, n. 4, p. e0153039, 2016. Erratum: 12(4):e0176586.

PRO, G.; GILBERT, P. A.; BALDWIN, J. A.; BROWN, C. C.; YOUNG, S.; ZALLER, N. Multilevel modeling of county-level excessive alcohol use, rurality, and COVID-19 case fatality rates in the US. **PLOS ONE**, v. 16, n. 6, p. e0253466, 2021.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2021. Disponível em: <http://www.R-project.org/>.

RAUBER, C.; CRIBARI-NETO, F.; BAYER, F. M. Improved testing inferences for beta regressions with parametric mean link function. **AStA Advances in Statistical Analysis**, v. 104, n. 4, p. 687–717, 2020.

REEVE, C. L.; BASALIK, D. A state level investigation of the associations among intellectual capital, religiosity and reproductive health. **Intelligence**, v. 39, n. 1, p. 64–73, 2011.

REIS, G.; SILVA, E. A. S. M.; SILVA, D. C. M.; THABANE, L.; SINGH G.; PARK, J. J. H.; FORREST, J. I.; HARARI, O.; SANTOS, C. V. Q.; ALMEIDA, A. P. F. G.; NETO, A. D. F.; SAVASSI, L. C. M.; MILAGRES, A. C.; TEIXEIRA, M. M.; SIMPLICIO, M. I. C.; RIBEIRO, L. B.; OLIVEIRA, R.; MILLS, E. J.; INVESTIGATORS, T. Effect of early treatment with hydroxychloroquine or lopinavir and ritonavir on risk of hospitalization among patients with COVID-19: The TOGETHER randomized clinical trial. **JAMA Network Open**, v. 4, n. 4, p. e216468, 2021.

RIBEIRO, T. F.; CORDEIRO, G. M.; PEÑA-RAMÍREZ, F. A.; GUERRA, R. R. A new quantile regression for the covid-19 mortality rates in the united states. **Computational and Applied Mathematics**, v. 40, n. 7, p. 1–16, 2021.

RIGBY, R. A.; STASINOPOULOS, M. D. Generalized additive models for location, scale and shape. **Applied Statistics**, v. 54, n. 3, p. 507–554, 2005.

ROCKE, D. M. Bootstrap Bartlett adjustment in seemingly unrelated regression. **Journal of the American Statistical Association**, v. 84, n. 406, p. 598–601, 1989.

ROSENBERG, E. S.; DUFORT, E. M.; UDO, T.; WILBERSCHIED, L. A.; KUMAR, J.; TESORIERO, J.; WEINBERG, P.; KIRKWOOD, J.; MUSE, A.; DEHOVITZ, J.; BLOG, D. S.; BRAD, H.; HOLTGRAVE, D. R.; ZUCKER, H. A. Association of treatment with hydroxychloroquine or azithromycin with in-hospital mortality in patients with COVID-19 in New York State. **JAMA**, American Medical Association, v. 323, n. 24, p. 2493–2502, 2020.

SIMAS, A. B.; BARRETO-SOUZA, W.; ROCHA, A. V. Improved estimators for a general class of beta regression models. **Computational Statistics & Data Analysis**, v. 54, n. 2, p. 348–366, 2010.

SKOVGAARD, I. M. An explicit large-deviation approximation to one-parameter tests. **Bernoulli**, v. 2, n. 2, p. 145–165, 1996.

SKOVGAARD, I. M. Likelihood asymptotics. **Scandinavian Journal of Statistics**, v. 28, n. 1, p. 3–32, 2001.

SMITH, R. L. Statistics of extremes, with applications in environment, insurance and finance. **Extreme values in Finance, Telecommunications and the Environment**, Chapman and Hall/CRC Press London, v. 1, p. 78, 2003.

SMITH, R. L.; NAYLOR, J. C. A comparison of maximum likelihood and Bayesian estimators for the three-parameter Weibull distribution. **Journal of the Royal Statistical Society C**, v. 36, n. 3, p. 358–369, 1987.

SWEARINGEN, C. J.; TILLEY, B. C.; ADAMS, R. J.; RUMBOLDT, Z.; NICHOLAS, J. S.; BANDYOPADHYAY, D.; WOOLSONE, R. F. Application of beta regression to analyze ischemic stroke volume in NINDS rt-PA clinical trials. **Neuroepidemiology**, v. 37, n. 2, p. 73–82, 2011.

WU, X.; NETHERY, R. C.; SABATH, M. B.; BRAUN, D.; DOMINICI, F. Air pollution and COVID-19 mortality in the United States: Strengths and limitations of an ecological regression analysis. **Science Advances**, v. 6, n. 45, p. eabd4049, 2020.

ZUCKERMAN, M.; SILBERMAN, J.; HALL, J. A. The relation between intelligence and religiosity: A meta-analysis and some proposed explanations. **Personality and Social Psychology Review**, v. 17, n. 4, p. 325–354, 2013.

APPENDIX A – VARYING PRECISION BETA REGRESSION LOG-LIKELIHOOD CUMULANTS

We shall now present the varying precision beta regression model log-likelihood cumulants up to fourth order. We shall use lower and upper case case letters to index derivatives of (2) with respect to the components of β and δ, respectively. We use tensor notation: $\kappa_{rs} = \mathbb{E}\left(\partial^2 \ell(\theta)/\partial\beta_r\partial\beta_s\right)$, $\kappa_{rst} = \mathbb{E}\left(\partial^3 \ell(\theta)/\partial\beta_r\partial\beta_s\partial\beta_t\right)$, $\kappa_{rstu} = \mathbb{E}\left(\partial^4 \ell(\theta)/\partial\beta_r\partial\beta_s\partial\beta_t\partial\beta_u\right)$, etc., $r,s,t,u = 1,\ldots,k$. Additionally, we use the following notation for derivatives of the above cumulants: $\kappa_{rs}^{(t)} = \partial\kappa_{rs}/\partial\beta_t$, $\kappa_{rs}^{(tu)} = \partial\kappa_{rs}/\partial\beta_t\partial\beta_u$, $\kappa_{rst}^{(u)} = \partial\kappa_{rst}/\partial\beta_u$, etc.

It can be shown that

$$\frac{\partial\mu_i}{\partial\eta_i} = \frac{1}{g_1'(\mu_i)}, \quad \frac{\partial}{\partial\mu_i}\frac{\partial\mu_i}{\partial\eta_i} = \frac{-g_1''(\mu_i)}{(g_1'(\mu_i))^2}, \quad \frac{\partial}{\partial\mu_i}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2 = \frac{-2g_1''(\mu_i)}{(g_1'(\mu_i))^3},$$

$$\frac{\partial}{\partial\mu_i}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^3 = \frac{-3g_1''(\mu_i)}{(g_1'(\mu_i))^4}, \quad \frac{\partial^2}{\partial\mu_i^2}\left(\frac{\partial\mu_i}{\partial\eta_i}\right) = \frac{-g_1'''(\mu_i)g_1'(\mu_i) + 2(g_1''(\mu_i))^2}{(g_1'(\mu_i))^3},$$

$$\frac{\partial\phi_i}{\partial\zeta_i} = \frac{1}{g_2'(\phi_i)}, \quad \frac{\partial}{\partial\phi_i}\frac{\partial\phi_i}{\partial\zeta_i} = \frac{-g_2''(\phi_i)}{(g_2'(\phi_i))^2}, \quad \frac{\partial}{\partial\phi_i}\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)^2 = \frac{-2g_2''(\phi_i)}{(g_2'(\phi_i))^3},$$

$$\frac{\partial}{\partial\phi_i}\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)^3 = \frac{-3g_2''(\phi_i)}{(g_2'(\phi_i))^4}, \quad \frac{\partial^2}{\partial\phi_i^2}\left(\frac{\partial\phi_i}{\partial\zeta_i}\right) = \frac{-g_2'''(\phi_i)g_2'(\phi_i) + 2(g_2''(\phi_i))^2}{(g_2'(\phi_i))^3}.$$

Let $w_i = \psi'(\mu_i\phi_i) + \psi'((1-\mu_i)\phi_i)$, $m_i = \psi''(\mu_i\phi_i) - \psi''((1-\mu_i)\phi_i)$ and also

$$a_i = 3\left(\frac{\partial}{\partial\mu_i}\frac{\partial\mu_i}{\partial\eta_i}\right)\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2, \quad t_i = 3\left(\frac{\partial}{\partial\phi_i}\frac{\partial\phi_i}{\partial\zeta_i}\right)\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)^2,$$

$$c_i = \phi_i[\mu_i w_i - \psi'((1-\mu_i)\phi_i)],\ b_i = \frac{\partial\mu_i}{\partial\eta_i}\left[\left(\frac{\partial^2}{\partial\mu_i^2}\frac{\partial\mu_i}{\partial\eta_i}\right)\frac{\partial\mu_i}{\partial\eta_i} + \left(\frac{\partial}{\partial\mu_i}\frac{\partial\mu_i}{\partial\eta_i}\right)^2\right],$$

$$v_i = \frac{\partial\phi_i}{\partial\zeta_i}\left[\left(\frac{\partial^2}{\partial\phi_i^2}\frac{\partial\phi_i}{\partial\zeta_i}\right)\frac{\partial\phi_i}{\partial\zeta_i} + \left(\frac{\partial}{\partial\phi_i}\frac{\partial\phi_i}{\partial\zeta_i}\right)^2\right],$$

$$d_i = (1-\mu_i)^2\psi'((1-\mu_i)\phi_i) + \mu^2\psi'(\mu_i\phi_i) - \psi'(\phi_i),\ s_i = (1-\mu_i)^3\psi''((1-\mu_i)\phi_i)$$
$$+ \mu_i^3\psi''(\mu_i\phi_i) - \psi''(\phi_i),\ u_i = -\phi_i\left[2w_i + \phi\frac{\partial w_i}{\partial\phi_i}\right],\ r_i = \left[2\frac{\partial\mu_i^*}{\partial\phi_i} + \phi\frac{\partial^2\mu_i^*}{\partial\phi_i^2}\right]\frac{\partial\mu_i}{\partial\eta_i},$$

where $\psi'(\cdot)$ and $\psi''(\cdot)$ is the trigamma and tetragamma functions, respectively. The following derivatives are needed for obtaining the log-likelihood cumulants:

$$
\begin{aligned}
\frac{\partial \mu_i^*}{\partial \mu_i} &= \phi_i \psi'(\mu_i\phi_i) + \phi_i \psi'((1-\mu_i)\phi_i) = \phi_i w_i,\ \frac{\partial \mu_i^*}{\partial \phi_i} = \mu_i \psi'(\mu_i\phi_i) - (1-\mu_i)\psi'((1-\mu_i)\phi_i) \\
&= \frac{c_i}{\phi_i}, \quad \frac{\partial \mu_i^\dagger}{\partial \mu_i} = -\phi_i \psi'((1-\mu_i)\phi_i), \quad \frac{\partial \mu_i^\dagger}{\partial \phi_i} = (1-\mu_i)\psi'((1-\mu_i)\phi_i) - \psi'(\phi_i), \\
\frac{\partial w_i}{\partial \mu_i} &= \phi_i \psi''(\mu_i\phi_i) - \phi_i \psi''((1-\mu_i)\phi_i) = \phi_i m_i, \\
\frac{\partial w_i}{\partial \phi_i} &= \mu_i \psi''(\mu_i\phi_i) + (1-\mu_i)\psi''((1-\mu_i)\phi_i), \\
\frac{\partial m_i}{\partial \mu_i} &= \phi_i \psi'''(\mu_i\phi_i) + \phi_i \psi'''((1-\mu_i)\phi_i), \quad \frac{\partial m_i}{\partial \phi_i} = \mu_i \psi'''(\mu_i\phi_i) - (1-\mu_i)\psi'''((1-\mu_i)\phi_i), \\
\frac{\partial a_i}{\partial \mu_i} &= 3\left(\frac{\partial \mu_i}{\partial \eta_i}\right)\left[\left(\frac{\partial^2}{\partial \mu_i^2}\frac{\partial \mu_i}{\partial \eta_i}\right)\left(\frac{\partial \mu_i}{\partial \eta_i}\right) + 2\left(\frac{\partial}{\partial \mu_i}\frac{\partial \mu_i}{\partial \eta_i}\right)^2\right], \quad \frac{\partial u_i}{\partial \mu_i} = -\phi_i^2\left(3m_i + \phi_i \frac{\partial m_i}{\partial \phi_i}\right), \\
\frac{\partial t_i}{\partial \phi_i} &= 3\left(\frac{\partial \phi_i}{\partial \zeta_i}\right)\left[\left(\frac{\partial^2}{\partial \phi_i^2}\frac{\partial \phi_i}{\partial \zeta_i}\right)\left(\frac{\partial \phi_i}{\partial \zeta_i}\right) + 2\left(\frac{\partial}{\partial \phi_i}\frac{\partial \phi_i}{\partial \zeta_i}\right)^2\right], \quad \frac{\partial}{\partial \mu_i}\frac{\partial w_i}{\partial \phi_i} = m_i + \phi_i \frac{\partial m_i}{\partial \phi_i}, \\
\frac{\partial b_i}{\partial \mu_i} &= \left(\frac{\partial}{\partial \mu_i}\frac{\partial \mu_i}{\partial \eta_i}\right)^3 + \left(\frac{\partial \mu_i}{\partial \eta_i}\right)\left[\left(\frac{\partial^3}{\partial \mu_i^3}\frac{\partial \mu_i}{\partial \eta_i}\right)\left(\frac{\partial \mu_i}{\partial \eta_i}\right) + 4\left(\frac{\partial^2}{\partial \mu_i^2}\frac{\partial \mu_i}{\partial \eta_i}\right)\left(\frac{\partial}{\partial \mu_i}\frac{\partial \mu_i}{\partial \eta_i}\right)\right], \\
\frac{\partial v_i}{\partial \phi_i} &= \left(\frac{\partial}{\partial \phi_i}\frac{\partial \phi_i}{\partial \zeta_i}\right)^3 + \left(\frac{\partial \phi_i}{\partial \zeta_i}\right)\left[\left(\frac{\partial^3}{\partial \phi_i^3}\frac{\partial \phi_i}{\partial \zeta_i}\right)\left(\frac{\partial \phi_i}{\partial \zeta_i}\right) + 4\left(\frac{\partial^2}{\partial \phi_i^2}\frac{\partial \phi_i}{\partial \zeta_i}\right)\left(\frac{\partial}{\partial \phi_i}\frac{\partial \phi_i}{\partial \zeta_i}\right)\right], \\
\frac{\partial^2 \mu_i^*}{\partial \phi_i^2} &= \mu_i^2 \psi''(\mu_i\phi_i) - (1-\mu_i)^2\psi''((1-\mu_i)\phi_i), \quad \frac{\partial c_i}{\partial \mu_i} = \phi_i\left(w_i + \phi_i \frac{\partial w_i}{\partial \phi_i}\right), \\
\frac{\partial^2 w_i}{\partial \phi_i^2} &= \mu_i^2 \psi'''(\mu_i\phi_i) + (1-\mu_i)^2\psi'''((1-\mu_i)\phi_i), \quad \frac{\partial c_i}{\partial \phi_i} = \frac{\partial \mu_i^*}{\partial \phi_i} + \phi_i \frac{\partial^2 \mu_i^*}{\partial \phi_i^2}, \\
\frac{\partial^3 \mu_i^*}{\partial \phi_i^3} &= \mu_i^3 \psi'''(\mu_i\phi_i) - (1-\mu_i)^3\psi'''((1-\mu_i)\phi_i), \quad \frac{\partial s_i}{\partial \mu_i} = 3\frac{\partial^2 \mu_i^*}{\partial \phi_i^2} + \phi_i \frac{\partial^3 \mu_i^*}{\partial \phi_i^3}, \\
\frac{\partial s_i}{\partial \phi_i} &= \mu_i^4 \psi'''(\mu_i\phi_i) + (1-\mu_i)^4\psi'''((1-\mu_i)\phi_i) - \psi'''(\phi_i), \\
\frac{\partial r_i}{\partial \mu_i} &= \left(2\frac{\partial \mu_i^*}{\partial \phi_i} + \phi_i \frac{\partial^2 \mu_i^*}{\partial \phi_i^2}\right)\left(\frac{\partial}{\partial \mu_i}\frac{\partial \mu_i}{\partial \eta_i}\right) + \left(2w_i + 4\frac{\partial w_i}{\partial \phi_i} + \phi_i^2 \frac{\partial^2 w_i}{\partial \phi_i^2}\right)\left(\frac{\partial \mu_i}{\partial \eta_i}\right), \\
\frac{\partial r_i}{\partial \phi_i} &= \left(3\frac{\partial^2 \mu_i^*}{\partial \phi_i} + \phi_i \frac{\partial^3 \mu_i^*}{\partial \phi_i^3}\right)\left(\frac{\partial \mu_i}{\partial \eta_i}\right), \quad \frac{\partial u_i}{\partial \phi_i} = -2w_i - \phi_i\left(4\frac{\partial w_i}{\partial \phi_i} + \phi_i \frac{\partial^2 w_i}{\partial \phi_i^2}\right).
\end{aligned}
$$

The log-likelihood derivatives with respect to the components of $\theta = (\beta^\top, \delta^\top)^\top$ are given by

$$U_{rs} = \sum_{i=1}^{n} \left\{ \left[-\phi_i^2 w_i \left(\frac{\partial \mu_i}{\partial \eta_i} \right) - \phi_i (y_i^* - \mu_i^*) \left(\frac{\partial}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \right) \right] \left(\frac{\partial \mu_i}{\partial \eta_i} \right) \right\} x_{ir} x_{is},$$

$$U_{rR} = \sum_{i=1}^{n} [-c_i + (y_i^* - \mu_i^*)] \left(\frac{\partial \phi_i}{\partial \zeta_i} \right) \left(\frac{\partial \mu_i}{\partial \eta_i} \right) x_{ir} z_{iR},$$

$$U_{RS} = \sum_{i=1}^{n} \left\{ -d_i \left(\frac{\partial \phi_i}{\partial \zeta_i} \right)^2 + [(y_i^\dagger - \mu_i^\dagger) + \mu_i (y_i^* - \mu_i^*)] \left(\frac{\partial}{\partial \phi_i} \frac{\partial \phi_i}{\partial \zeta_i} \right) \left(\frac{\partial \phi_i}{\partial \zeta_i} \right) \right\} z_{iR} z_{iS},$$

$$U_{rst} = \sum_{i=1}^{n} \left\{ -\phi_i \left[\phi_i w_i a_i + \phi_i^2 m_i \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^3 + (y_i^* - \mu_i^*) b_i \right] \right\} x_{ir} x_{is} x_{it},$$

$$U_{rsR} = \sum_{i=1}^{n} \left\{ \left[-c_i \left(\frac{\partial}{\partial \eta_i} \frac{\partial \mu_i}{\partial \eta_i} \right) + (y_i^* - \mu_i^*) \left(\frac{\partial}{\partial \eta_i} \frac{\partial \mu_i}{\partial \eta_i} \right) + u_i \left(\frac{\partial \mu_i}{\partial \eta_i} \right) \right] \left(\frac{\partial \mu_i}{\partial \eta_i} \right) \left(\frac{\partial \phi_i}{\partial \zeta_i} \right) \right\} x_{ir}$$

$$\times x_{is} z_{iR},$$

$$U_{rRS} = \sum_{i=1}^{n} \left\{ \left(\frac{\partial}{\partial \zeta_i} \frac{\partial \phi_i}{\partial \zeta_i} \right) \left(\frac{\partial \phi_i}{\partial \zeta_i} \right) \left(\frac{\partial \mu_i}{\partial \eta_i} \right) [(y_i^* - \mu_i^*) - c_i] - \left(\frac{\partial \phi_i}{\partial \zeta_i} \right)^2 r_i \right\} x_{ir} z_{iR}$$

$$\times z_{iS},$$

$$U_{RST} = \sum_{i=1}^{n} \left\{ -s_i \left(\frac{\partial \phi_i}{\partial \zeta_i} \right)^3 - d_i \left(\frac{\partial}{\partial \phi_i} \left(\frac{\partial \phi_i}{\partial \zeta_i} \right)^2 \right) \left(\frac{\partial \phi_i}{\partial \zeta_i} \right) - d_i \left(\frac{\partial}{\partial \phi_i} \frac{\partial \phi_i}{\partial \zeta_i} \right) \left(\frac{\partial \phi_i}{\partial \zeta_i} \right)^2 \right.$$

$$\left. + [(y_i^\dagger - \mu_i^\dagger) + \mu_i (y_i^* - \mu_i^*)] v_i \right\}$$

$$\times z_{iR} z_{iS} z_{iT},$$

$$U_{rstu} = \sum_{i=1}^{n} \left\{ -\phi_i \left[\phi_i^2 \left(m_i \frac{\partial}{\partial \mu_i} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^3 + \frac{\partial m_i}{\partial \mu_i} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^3 \right) + \phi_i \left[\left(\frac{\partial a_i}{\partial \mu_i} + b_i \right) w_i \right. \right. \right.$$

$$\left. \left. \left. + \frac{\partial w_i}{\partial \mu_i} a_i \right] - (y_i^* - \mu_i^*) \frac{\partial b_i}{\partial \mu_i} \right] \frac{\partial \mu_i}{\partial \eta_i} \right\} x_{ir} x_{is} x_{it} x_{iu},$$

$$U_{rstR} = \sum_{i=1}^{n} \left\{ \left[-\phi_i \left[\phi_i \left(3m_i + \phi_i \left(\frac{\partial m_i}{\partial \phi_i} \right) \right) \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^3 + a_i \left(2w_i + \phi_i \left(\frac{\partial w_i}{\partial \phi_i} \right) \right) \right. \right. \right.$$

$$\left. \left. \left. + b_i \left(\frac{\partial \mu_i^*}{\partial \phi_i} \right) \right] + (y_i^* - \mu_i^*) b_i \right] \left(\frac{\partial \phi_i}{\partial \zeta_i} \right) \right\} x_{ir} x_{is} x_{it} z_{iR},$$

$$U_{rsRS} = \sum_{i=1}^{n} \left\{ \left[[(y_i^* - \mu_i^*) - c_i] \left(\frac{\partial}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \right) \left(\frac{\partial}{\partial \phi_i} \frac{\partial \phi_i}{\partial \zeta_i} \right) \left(\frac{\partial \phi_i}{\partial \zeta_i} \right) - \left(\frac{\partial r_i}{\partial \mu_i} \right) \left(\frac{\partial \phi_i}{\partial \zeta_i} \right)^2 \right. \right.$$

$$\left. \left. - \left(\frac{\partial \mu_i^*}{\partial \phi_i} \right) \left(\frac{\partial \mu_i}{\partial \eta_i} \right) \left(\frac{\partial \phi_i}{\partial \zeta_i} \right)^2 + [u_i + (y_i^* - \mu_i^*)] \left(\frac{\partial \mu_i}{\partial \eta_i} \right) \left(\frac{\partial \phi_i}{\partial \zeta_i} \right) \left(\frac{\partial}{\partial \phi_i} \frac{\partial \phi_i}{\partial \zeta_i} \right) \right] \left(\frac{\partial \mu_i}{\partial \eta_i} \right) \right\}$$

$$\times x_{ir} x_{is} z_{iR} z_{iS},$$

$$U_{rRST} = \sum_{i=1}^{n} \left\{ [(y_i^* - \mu_i^*) - c_i] \left(\frac{\partial \mu_i}{\partial \eta_i} \right) v_i - r_i \left(\frac{\partial}{\partial \phi_i} \left(\frac{\partial \phi_i}{\partial \zeta_i} \right)^3 \right) - \left(\frac{\partial s_i}{\partial \mu_i} \right) \left(\frac{\partial \mu_i}{\partial \eta_i} \right) \left(\frac{\partial \phi_i}{\partial \zeta_i} \right)^3 \right\}$$

$$\times x_{ir} z_{iR} z_{iS} z_{iT}, \quad U_{RSTU} = \sum_{i=1}^{n} \left\{ \left[- \left(\frac{\partial s_i}{\partial \phi_i} \right) \left(\frac{\partial \phi_i}{\partial \zeta_i} \right)^3 - 2 s_i \left(\frac{\partial}{\partial \phi_i} \left(\frac{\partial \phi_i}{\partial \zeta_i} \right)^3 \right) \right. \right.$$

$$\left. \left. - d_i \left(\left(\frac{\partial t_i}{\partial \phi_i} \right) + v_i \right) + [(y_i^\dagger - \mu_i^\dagger) + \mu_i (y_i^* - \mu_i^*)] \left(\frac{\partial v_i}{\partial \phi_i} \right) \right] \left(\frac{\partial \phi_i}{\partial \zeta_i} \right) \right\} z_{iR} z_{iS} z_{iT} z_{iU}.$$

Using the above results, we arrive, after long derivations, at the following expressions for the relevant varying precision beta regression model cumulants:

$$
\begin{aligned}
\kappa_{rs} &= -\sum_{i=1}^{n}\left\{\phi_i^2 w_i \left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2\right\} x_{ir}x_{is}, \quad \kappa_{rR} = -\sum_{i=1}^{n}\left\{c_i\left(\frac{\partial\mu_i}{\partial\eta_i}\right)\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)\right\} x_{ir}z_{iR},\\
\kappa_{RS} &= -\sum_{i=1}^{n}\left\{d_i\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)^2\right\} z_{iR}z_{iS}, \quad \kappa_{rst} = -\sum_{i=1}^{n}\left\{\phi_i^3 m_i\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^3 + \phi_i^2 w_i a_i\right\} x_{ir}x_{is}x_{it},\\
\kappa_{rsR} &= \sum_{i=1}^{n}\left\{\left[u_i\left(\frac{\partial\mu_i}{\partial\eta_i}\right) - c_i\left(\frac{\partial}{\partial\mu_i}\frac{\partial\mu_i}{\partial\eta_i}\right)\right]\left(\frac{\partial\mu_i}{\partial\eta_i}\right)\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)\right\} x_{ir}x_{is}z_{iR},\\
\kappa_{rRS} &= -\sum_{i=1}^{n}\left\{r_i\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)^2 + c_i\left(\frac{\partial}{\partial\phi_i}\frac{\partial\phi_i}{\partial\zeta_i}\right)\left(\frac{\partial\mu_i}{\partial\eta_i}\right)\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)\right\} x_{ir}z_{iR}z_{iS},\\
\kappa_{RST} &= -\sum_{i=1}^{n}\left\{s_i\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)^3 + d_i\frac{\partial}{\partial\phi_i}\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)^3\right\} z_{iR}z_{iS}z_{iT}, \; \kappa_{rstu} = \sum_{i=1}^{n}\left\{-\phi_i\left[\phi_i^2\left(\frac{\partial m_i}{\partial\mu_i}\right.\right.\right.\\
&\times \left.\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^3 + m_i\frac{\partial}{\partial\mu_i}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^3\right) + \phi_i\left(\left(\frac{\partial a_i}{\partial\mu_i} + b_i\right)w_i + \frac{\partial w_i}{\partial\mu_i}a_i\right)\Bigg]\frac{\partial\mu_i}{\partial\eta_i}\Bigg\} x_{ir}x_{is}x_{it}x_{iu},\\
\kappa_{rstR} &= \sum_{i=1}^{n}\left\{-\phi_i\left[\phi_i\left(3m_i + \phi_i\left(\frac{\partial m_i}{\partial\phi_i}\right)\right)\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^3 + a_i\left(2w_i + \phi_i\left(\frac{\partial w_i}{\partial\phi_i}\right)\right) + b_i\left(\frac{\partial\mu_i^*}{\partial\phi_i}\right)\right]\right.\\
&\times \left.\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)\right\} x_{ir}x_{is}x_{it}z_{iR}, \quad \kappa_{rsRS} = \sum_{i=1}^{n}\left\{-c_i\left(\frac{\partial}{\partial\mu_i}\frac{\partial\mu_i}{\partial\eta_i}\right)\left(\frac{\partial}{\partial\phi_i}\frac{\partial\phi_i}{\partial\zeta_i}\right)\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)\right.\\
&- \left.\left(\frac{\partial r_i}{\partial\mu_i}\right)\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)^2 - \left(\frac{\partial\mu_i^*}{\partial\phi_i}\right)\left(\frac{\partial\mu_i}{\partial\eta_i}\right)\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)^2 + u_i\left(\frac{\partial\mu_i}{\partial\eta_i}\right)\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)\left(\frac{\partial}{\partial\phi_i}\frac{\partial\phi_i}{\partial\zeta_i}\right)\right\}\frac{\partial\mu_i}{\partial\eta_i}\\
&\times x_{ir}x_{is}z_{iR}z_{iS}, \; \kappa_{rRST} = -\sum_{i=1}^{n}\left\{c_i\left(\frac{\partial\mu_i}{\partial\eta_i}\right)v_i + r_i\left(\frac{\partial}{\partial\phi_i}\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)^3\right) + \left(\frac{\partial s_i}{\partial\mu_i}\right)\left(\frac{\partial\mu_i}{\partial\eta_i}\right)\right.\\
&\times \left.\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)^3\right\} x_{ir}z_{iR}z_{iS}z_{iT}, \; \kappa_{RSTU} = -\sum_{i=1}^{n}\left\{\left[\left(\frac{\partial s_i}{\partial\phi_i}\right)\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)^3 + 2s_i\left(\frac{\partial}{\partial\phi_i}\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)^3\right)\right.\right.\\
&+ \left.\left.d_i\left(\frac{\partial t_i}{\partial\phi_i} + v_i\right)\right]\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)\right\} z_{iR}z_{iS}z_{iT}z_{iU}.
\end{aligned}
$$

Also, we obtained the following expressions for the first order derivatives of the log-likelihood cumulants:

$$\kappa_{rs}^{(u)} = \sum_{i=1}^{n}\left\{-\phi_i^2\left[\phi_i m_i\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^3 + \frac{2}{3}w_i a_i\right]\right\} x_{ir}x_{is}x_{iu},\ \kappa_{rs}^{(R)} = \sum_{i=1}^{n}\left\{u_i\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)\right\}$$
$$\times\, x_{ir}x_{is}z_{iR},$$
$$\kappa_{rR}^{(u)} = -\sum_{i=1}^{n}\left\{\left[\left(\frac{\partial c_i}{\partial\mu_i}\right)\left(\frac{\partial\mu_i}{\partial\eta_i}\right) + c_i\left(\frac{\partial}{\partial\mu_i}\frac{\partial\mu_i}{\partial\eta_i}\right)\right]\left(\frac{\partial\mu_i}{\partial\eta_i}\right)\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)\right\} x_{ir}x_{iu}z_{iR},$$
$$\kappa_{rR}^{(S)} = -\sum_{i=1}^{n}\left\{\left[\left(\frac{\partial c_i}{\partial\phi_i}\right)\left(\frac{\partial\phi_i}{\partial\zeta_i}\right) + c_i\left(\frac{\partial}{\partial\phi_i}\frac{\partial\phi_i}{\partial\zeta_i}\right)\right]\left(\frac{\partial\mu_i}{\partial\eta_i}\right)\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)\right\} x_{ir}z_{iR}z_{iS},$$
$$\kappa_{RS}^{(u)} = -\sum_{i=1}^{n}\left\{r_i\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)^2\right\} x_{iu}z_{iR}z_{iS},\quad \kappa_{RS}^{(T)} = -\sum_{i=1}^{n}\left\{s_i\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)^3 + \frac{2}{3}d_i t_i\right\} z_{iR}z_{iS}z_{iT},$$
$$\kappa_{rst}^{(u)} = \sum_{i=1}^{n}\left\{-\phi_i^2\left[\phi_i\left(m_i\left(\frac{\partial}{\partial\mu_i}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^3 + a_i\right) + \frac{\partial m_i}{\partial\mu_i}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^3\right)\right.\right.$$
$$\left.\left. +\ w_i\frac{\partial a_i}{\partial\mu_i}\right]\frac{\partial\mu_i}{\partial\eta_i}\right\} x_{ir}x_{is}x_{it}x_{iu},$$
$$\kappa_{rst}^{(R)} = -\sum_{i=1}^{n}\left\{\phi_i\left[\left(3\phi_i m_i + \phi_i^2\left(\frac{\partial m_i}{\partial\phi_i}\right)\right)\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^3 + \left(2w_i + \phi_i\left(\frac{\partial w_i}{\partial\phi_i}\right)\right)a_i\right]\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)\right\}$$
$$\times\, x_{ir}x_{is}x_{it}z_{iR},$$
$$\kappa_{rsR}^{(t)} = \sum_{i=1}^{n}\left\{\left[2u_i\left(\frac{\partial}{\partial\mu_i}\frac{\partial\mu_i}{\partial\eta_i}\right)\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2 - c_i b_i + \left(\frac{\partial u_i}{\partial\mu_i}\right)\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^3 - \left(\frac{\partial c_i}{\partial\mu_i}\right)\left(\frac{\partial}{\partial\mu_i}\frac{\partial\mu_i}{\partial\eta_i}\right)\right.\right.$$
$$\left.\left.\times\ \left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2\right]\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)\right\} x_{ir}x_{is}x_{it}z_{iR},\ \kappa_{rsR}^{(S)} = \sum_{i=1}^{n}\left\{\left\{\left[\left(\frac{\partial u_i}{\partial\phi_i}\right)\left(\frac{\partial\phi_i}{\partial\zeta_i}\right) + u_i\left(\frac{\partial}{\partial\phi_i}\frac{\partial\phi_i}{\partial\zeta_i}\right)\right]\right.\right.$$
$$\left.\left.\times\ \left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2 - \left[z_i\left(\frac{\partial\phi_i}{\partial\zeta_i}\right) + c_i\left(\frac{\partial}{\partial\phi_i}\frac{\partial\phi_i}{\partial\zeta_i}\right)\right]\left(\frac{\partial}{\partial\mu_i}\frac{\partial\mu_i}{\partial\eta_i}\right)\left(\frac{\partial\mu_i}{\partial\eta_i}\right)\right\}\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)\right\}$$
$$\times\, x_{ir}x_{is}z_{iR}z_{iS},$$
$$\kappa_{rRS}^{(s)} = \sum_{i=1}^{n}\left\{\left[-\left(\frac{\partial r_i}{\partial\mu_i}\right)\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)^2 - \left[\left(\frac{\partial c_i}{\partial\mu_i}\right)\left(\frac{\partial\mu_i}{\partial\eta_i}\right) + c_i\left(\frac{\partial}{\partial\mu_i}\frac{\partial\mu_i}{\partial\eta_i}\right)\right]\left(\frac{\partial}{\partial\phi_i}\frac{\partial\phi_i}{\partial\zeta_i}\right)\right.\right.$$
$$\left.\left.\times\ \left(\frac{\partial\phi_i}{\partial\zeta_i}\right)\right]\left(\frac{\partial\mu_i}{\partial\eta_i}\right)\right\} x_{ir}x_{is}z_{iR}z_{iS},\quad \kappa_{rRS}^{(T)} = -\sum_{i=1}^{n}\left\{\left(\frac{\partial r_i}{\partial\phi_i}\right)\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)^3 + z_i\left(\frac{\partial\mu_i}{\partial\eta_i}\right)\right.$$
$$\left.\times\ \left(\frac{\partial\phi_i}{\partial\zeta_i}\right)^2\left(\frac{\partial}{\partial\phi_i}\frac{\partial\phi_i}{\partial\zeta_i}\right) + r_i\left(\frac{\partial}{\partial\phi_i}\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)^2\right)\left(\frac{\partial\phi_i}{\partial\zeta_i}\right) + c_i v_i\left(\frac{\partial\mu_i}{\partial\eta_i}\right)\right\} x_{ir}z_{iR}z_{iS}z_{iT},$$
$$\kappa_{RST}^{(r)} = -\sum_{i=1}^{n}\left\{\left[\left(\frac{\partial s_i}{\partial\mu_i}\right)\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)^3 + \left(\frac{\partial d_i}{\partial\mu_i}\right)\left(\frac{\partial}{\partial\phi_i}\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)^3\right)\right]\left(\frac{\partial\mu_i}{\partial\eta_i}\right)\right\} x_{ir}z_{iR}z_{iS}z_{iT},$$
$$\kappa_{RST}^{(U)} = -\sum_{i=1}^{n}\left\{\left[\left(\frac{\partial s_i}{\partial\phi_i}\right)\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)^3 + 2s_i\left(\frac{\partial}{\partial\phi_i}\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)^3\right) + d_i\left(\frac{\partial t_i}{\partial\phi_i}\right)\right]\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)\right\}$$
$$\times\, z_{iR}z_{iS}z_{iT}z_{iU}.$$

The second order derivatives of the log-likelihood cumulants can expressed as follows:

$$
\begin{aligned}
\kappa_{rs}^{(tu)} &= -\sum_{i=1}^{n}\left\{\phi_i^2\left[\phi_i\left(m_i\left(\frac{\partial}{\partial\mu_i}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^3+\frac{2}{3}a_i\right)+\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^3\frac{\partial m_i}{\partial\mu_i}\right)+\frac{2}{3}w_i\frac{\partial a_i}{\partial\mu_i}\right]\frac{\partial\mu_i}{\partial\eta_i}\right\} \\
&\quad\times x_{ir}x_{is}x_{it}x_{iu}, \\
\kappa_{rs}^{(Rt)} &= \sum_{i=1}^{n}\left\{\left[\left(\frac{\partial u_i}{\partial\mu_i}\right)\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2+u_i\left(\frac{\partial}{\partial\mu_i}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2\right)\right]\left(\frac{\partial\mu_i}{\partial\eta_i}\right)\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)\right\}x_{ir}x_{is}x_{it}z_{iR}, \\
\kappa_{rs}^{(RS)} &= \sum_{i=1}^{n}\left\{\left[\left(\frac{\partial u_i}{\partial\phi_i}\right)\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)+u_i\left(\frac{\partial}{\partial\phi_i}\frac{\partial\phi_i}{\partial\zeta_i}\right)\right]\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)\right\}x_{ir}x_{is}z_{iR}z_{iS}, \\
\kappa_{rR}^{(st)} &= \sum_{i=1}^{n}\left\{\left[-\phi_i^2\left(m_i+\left(\frac{\partial}{\partial\mu_i}\frac{\partial w_i}{\partial\phi_i}\right)\right)\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^3-\left(\frac{\partial c_i}{\partial\mu_i}\right)\left(\frac{\partial}{\partial\mu_i}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^3\right)-c_ib_i\right]\right. \\
&\quad\times\left.\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)\right\}x_{ir}x_{is}x_{it}z_{iR}, \\
\kappa_{rR}^{(Ss)} &= \sum_{i=1}^{n}\left\{\left[-z_i\left(\frac{\partial}{\partial\mu_i}\frac{\partial\mu_i}{\partial\eta_i}\right)\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)^2-c_i\left(\frac{\partial}{\partial\mu_i}\frac{\partial\mu_i}{\partial\eta_i}\right)\left(\frac{\partial}{\partial\phi_i}\frac{\partial\phi_i}{\partial\zeta_i}\right)\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)-\left(\frac{\partial z_i}{\partial\mu_i}\right)\right.\right. \\
&\quad\times\left.\left.\left(\frac{\partial\mu_i}{\partial\eta_i}\right)\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)^2-\left(\frac{\partial c_i}{\partial\mu_i}\right)\left(\frac{\partial}{\partial\phi_i}\frac{\partial\phi_i}{\partial\zeta_i}\right)\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)\left(\frac{\partial\mu_i}{\partial\eta_i}\right)\right]\left(\frac{\partial\mu_i}{\partial\eta_i}\right)\right\}x_{ir}x_{is}z_{iR}z_{iS}, \\
\kappa_{rR}^{(ST)} &= -\sum_{i=1}^{n}\left\{z_i\left(\frac{\partial}{\partial\phi_i}\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)^3\right)\left(\frac{\partial\mu_i}{\partial\eta_i}\right)+c_iv_i\left(\frac{\partial\mu_i}{\partial\eta_i}\right)+\left(\frac{\partial z_i}{\partial\phi_i}\right)\left(\frac{\partial\mu_i}{\partial\eta_i}\right)\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)^3\right\} \\
&\quad\times x_{ir}z_{iR}z_{iS}z_{iT}, \quad \kappa_{RS}^{(rs)} = -\sum_{i=1}^{n}\left\{\left(\frac{\partial r_i}{\partial\mu_i}\right)\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)^2\left(\frac{\partial\mu_i}{\partial\eta_i}\right)\right\}x_{ir}x_{is}z_{iR}z_{iS}, \\
\kappa_{RS}^{(rT)} &= -\sum_{i=1}^{n}\left\{\left[\left(\frac{\partial r_i}{\partial\phi_i}\right)\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)^2+r_i\left(\frac{\partial}{\partial\phi_i}\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)^2\right)\right]\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)\right\}x_{ir}z_{iR}z_{iS}z_{iT}, \\
\kappa_{RS}^{(TU)} &= -\sum_{i=1}^{n}\left\{\left[\left(\frac{\partial s_i}{\partial\phi_i}\right)\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)^3+s_i\left(\frac{\partial}{\partial\phi_i}\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)^3\right)+s_i\left(\frac{2}{3}t_i\right)+d_i\left(\frac{2}{3}\frac{\partial t_i}{\partial\phi_i}\right)\right]\right. \\
&\quad\times\left.\left(\frac{\partial\phi_i}{\partial\zeta_i}\right)\right\}z_{iR}z_{iS}z_{iT}z_{iU}.
\end{aligned}
$$

APPENDIX B – RESULTS RELATED TO THE GUMBEL LAW

We will now present some useful results related to the Gumbel law. Let $Y \sim \text{Gumbel}(\mu, \phi)$. We obtain the cumulative distribution function and the probability density function of the random variable $T = (Y - \mu)/\phi$ which are, respectively,

$$F_T(t) = \Pr(T \leq t) = \Pr\left(\frac{Y-\mu}{\phi} \leq t\right) = \Pr(Y \leq \mu + \phi t) = \exp(-\exp(-t)),$$
$$f_T(t) = \exp(-t)\exp(-\exp(-t)),$$

$t \in \mathbb{R}$. Hence, $T \sim \text{Gumbel}(0,1)$, with $\mathbb{E}(T) = \gamma$. We also establish that the cumulative distribution function and the probability density function of $T^* = \exp(-T)$ are given, respectively, by

$$F_{T^*}(t^*) = \Pr(T^* \leq t^*) = \Pr(\exp(-T) \leq t^*) = 1 - \Pr(T \leq -\ln(t^*)) = 1 - \exp(-t^*),$$
$$f_{T^*}(t^*) = \exp(-t^*),$$

$t^* \in \mathbb{R}_+$. It thus follows that $T^* \sim \text{exponencial}(1)$, with $\mathbb{E}(T^*) = 1$. Next, we will obtain the expected value of $T^n \exp(-dt)$, for $d \in \mathbb{R}$. Notice that

$$\mathbb{E}(T^n \exp(-dt)) = \int_{-\infty}^{\infty} t^n \exp(-dt)\exp(-t - \exp(-t))dt. \tag{B.1}$$

Using the transformation $W = \exp(-T)$ or, alternatively, $T = -\ln(W)$ together with Equation (4.358.5) in Gradshteyn and Ryzhik (2007), it follows that

$$\int_0^{\infty} z^{v-1}\exp(-\mu z)\ln(z)^n dz = \frac{d^n}{dv^n}\left[\mu^{-v}\Gamma(v)\right],$$

$n = 1, 2, 3, 4, \ldots$, with $\psi(z) = \Gamma^{(1)}(z)/\Gamma(z)$. It is thus possible to write Equation (B.1) as

$$\mathbb{E}(T^n \exp(-ct)) = (-1)^n \Gamma^{(n)}(1 + c). \tag{B.2}$$

Using Equation (B.2), we can easily compute expected values of functions of T and T^*. It is also useful to recall that the score function has zero mean, i.e., $\mathbb{E}(U) = 0$. Therefore, we obtain

$$\mathbb{E}(T^*) = (-1)^0\Gamma^{(0)}(2) = \Gamma(2) = 1, \quad \mathbb{E}(TT^*) = (-1)^1\Gamma^{(1)}(2) = \gamma - 1,$$
$$\mathbb{E}(T^2T^*) = (-1)^2\Gamma^{(2)}(2) = \Gamma^{(2)}(2), \quad \mathbb{E}(T^3T^*) = (-1)^3\Gamma^{(3)}(2) = -\Gamma^{(3)}(2),$$
$$\mathbb{E}(T^4T^*) = (-1)^4\Gamma^{(4)}(2) = \Gamma^{(4)}(2).$$

APPENDIX C – GENERAL EXTREME-VALUE REGRESSION MODEL LOG-LIKELIHOOD CUMULANTS

In this appendix, we will derive the general extreme-value regression model log-likelihood cumulants up to fourth order. These cumulants are needed for obtaining the Bartlett correction factor to the likelihood ratio test statistic. The derivatives of the log-likelihood function given in (2.3) with respect to β are denoted by the lower case subscripts j,l,m,u and for those with respect to θ we use the upper case subscripts J,L,M,U. Additionally, we use the following notation for the derivatives with respect to the components of β:

$$U_j = \frac{\partial \ell(\beta,\theta)}{\partial \beta_j}, \quad U_{jl} = \frac{\partial^2 \ell(\beta,\theta)}{\partial \beta_j \partial \beta_l}, \quad U_{jlm} = \frac{\partial^3 \ell(\beta,\theta)}{\partial \beta_j \partial \beta_l \partial \beta_m}, \quad U_{jlmu} = \frac{\partial^4 \ell(\beta,\theta)}{\partial \beta_j \partial \beta_l \partial \beta_m \partial \beta_u}.$$

The cumulants are $\kappa_{jl} = \mathbb{E}(U_{jl})$, $\kappa_{jlm} = \mathbb{E}(U_{jlm})$, $\kappa_{jlmu} = \mathbb{E}(U_{jlmu})$, etc. Their derivatives with respect to the elements of β are denoted by

$$\kappa_{jl}^{(m)} = \frac{\partial \kappa_{jl}}{\partial \beta_m}, \quad \kappa_{jl}^{(mu)} = \frac{\partial \kappa_{jl}}{\partial \beta_m \partial \beta_u}, \quad \kappa_{jlm}^{(u)} = \frac{\partial \kappa_{jlm}}{\partial \beta_u}, \quad \text{etc.}$$

The log-likelihood derivatives with respect to the components of θ, the corresponding cumulants and their derivatives are expressed analogously. The mixed cumulants with respect to the components of β and θ are written, e.g., as $\kappa_{jL} = \mathbb{E}(U_{jL}) = \mathbb{E}(\partial^2 \ell(\beta,\theta)/\partial \beta_j \partial \theta_L)$.

Let $t_i = (y_i - \mu_i)/\phi_i$, $t_i^* = \exp(-(y_i - \mu_i)/\phi_i)$ and $\nu_i = -1 + t_i - t_i t_i^*$. We obtain the following derivatives:

$$\begin{aligned}
\frac{\partial \nu_i}{\partial \mu_i} &= -\frac{(1 + t_i t_i^* - t_i^*)}{\phi_i}, \\
\frac{\partial^2 \nu_i}{\partial \mu_i^2} &= \frac{t_i^*(2 - t_i)}{\phi_i^2}, \\
\frac{\partial \nu_i}{\partial \phi_i} &= \frac{t_i t_i^* - t_i - t_i^2 t_i^*}{\phi_i}, \\
\frac{\partial^2 \nu_i}{\partial \phi_i^2} &= \frac{2t_i - 2t_i t_i^* + 4t_i^2 t_i^* - t_i^3 t_i^*}{\phi_i^2}, \\
\frac{\partial^3 \nu_i}{\partial \phi_i^3} &= -\frac{(6t_i - 6t_i t_i^* + 18t_i^2 t_i^* - 9t_i^3 t_i^* + t_i^4 t_i^*)}{\phi_i^3} \\
\frac{\partial^2 \nu_i}{\partial \mu_i \partial \phi_i} &= \frac{(1 - t_i^* + 3t_i t_i^* - t_i^2 t_i^*)}{\phi_i^2}, \\
\frac{\partial^3 \nu_i}{\partial \mu_i \partial \phi_i^2} &= -\frac{(2 - 2t_i^* + 10t_i t_i^* - 7t_i^2 t_i^* + t_i^3 t_i^*)}{\phi_i^3} \\
\frac{\partial^3 \nu_i}{\partial \phi_i \partial \mu_i^2} &= -\frac{(4t_i^* - 5t_i t_i^* + t_i^2 t_i^*)}{\phi_i^3}.
\end{aligned}$$

After long and tedious algebra, it is possible to write the log-likelihood derivatives with respect to the parameters in $\upsilon = (\beta^\top, \theta^\top)^\top$ as

$$U_{jl} = \sum_{i=1}^{n} \left\{ -\frac{t_i^*}{\phi_i} \left(\frac{\partial \mu_i}{\partial \eta_{1i}} \right)^2 (j,l)_i + \frac{(1-t_i^*)}{\phi_i} \left[\left(\frac{\partial^2 \mu_i}{\partial \eta_{1i}^2} \right) (j,l)_i + \left(\frac{\partial \mu_i}{\partial \eta_{1i}} \right) (jl)_i \right] \right\},$$

$$U_{jL} = -\sum_{i=1}^{n} \left\{ \frac{[(t_i-1)t_i^*+1]}{\phi_i} \left(\frac{\partial \mu_i}{\partial \eta_{1i}} \right) \left(\frac{\partial \phi_i}{\partial \eta_{2i}} \right) (j,L)_i \right\},$$

$$U_{JL} = \sum_{i=1}^{n} \left\{ -\frac{t_i[1+t_i^*(t_i-1)]}{\phi_i^2} \left(\frac{\partial \phi_i}{\partial \eta_{2i}} \right)^2 (J,L)_i + \nu_i \left[\frac{1}{\phi_i} \left(\left(\frac{\partial^2 \phi_i}{\partial \eta_{2i}^2} \right) (J,L)_i \right.\right.\right.$$
$$\left.\left.\left. + \left(\frac{\partial \phi_i}{\partial \eta_{2i}} \right) (JL)_i \right) - \frac{1}{\phi_i^2} \left(\frac{\partial \phi_i}{\partial \eta_{2i}} \right)^2 (J,L)_i \right] \right\},$$

$$U_{jlm} = \sum_{i=1}^{n} \left\{ \left[-\frac{t_i^*}{\phi_i^3} \left(\frac{\partial \mu_i}{\partial \eta_{1i}} \right)^3 - \frac{3t_i^*}{\phi_i^2} \left(\frac{\partial^2 \mu_i}{\partial \eta_{1i}^2} \right) \left(\frac{\partial \mu_i}{\partial \eta_{1i}} \right) \right] (j,l,m)_i - \frac{t_i^*}{\phi_i^2} \left(\frac{\partial \mu_i}{\partial \eta_{1i}} \right)^2 \right.$$
$$\times [(jm,l)_i + (lm,j)_i + (jl,m)_i] + \frac{(1-t_i^*)}{\phi_i} \left[\left(\frac{\partial^3 \mu_i}{\partial \eta_{1i}^3} \right) (j,l,m)_i + \left(\frac{\partial^2 \mu_i}{\partial \eta_{1i}^2} \right) \right.$$
$$\left.\left. \times [(jm,l)_i + (lm,j)_i + (jl,m)_i] + \left(\frac{\partial \mu_i}{\partial \eta_{1i}} \right) (jlm)_i \right] \right\},$$

$$U_{jlM} = \sum_{i=1}^{n} \left\{ \frac{1}{\phi_i} \left(\frac{\partial^2 \nu_i}{\partial \mu_i^2} \right) \left(\frac{\partial \mu_i}{\partial \eta_{1i}} \right)^2 \left(\frac{\partial \phi_i}{\partial \eta_{2i}} \right) (j,l,M)_i + \frac{1}{\phi_i} \left(\frac{\partial \nu_i}{\partial \mu_i} \right) \left(\frac{\partial \phi_i}{\partial \eta_{2i}} \right) \left[\left(\frac{\partial^2 \mu_i}{\partial \eta_{1i}^2} \right) \right.\right.$$
$$\left.\left. \times (j,l,M)_i + \left(\frac{\partial \mu_i}{\partial \eta_{1i}} \right) (jl,M)_i \right] \right\},$$

$$U_{jLM} = \sum_{i=1}^{n} \left\{ \frac{1}{\phi_i} \left(\frac{\partial^2 \nu_i}{\partial \mu_i \partial \phi_i} \right) \left(\frac{\partial \phi_i}{\partial \eta_{2i}} \right)^2 \left(\frac{\partial \mu_i}{\partial \eta_{1i}} \right) (j,L,M)_i + \frac{1}{\phi_i} \left(\frac{\partial \nu_i}{\partial \mu_i} \right) \left(\frac{\partial \mu_i}{\partial \eta_{1i}} \right) \right.$$
$$\left. \times \left[\left(\left(\frac{\partial^2 \phi_i}{\partial \eta_{2i}^2} \right) - \frac{1}{\phi_i} \left(\frac{\partial \phi_i}{\partial \eta_{2i}} \right)^2 \right) (j,L,M)_i + \left(\frac{\partial \phi_i}{\partial \eta_{2i}} \right) (j,LM)_i \right] \right\},$$

$$U_{JLM} = \sum_{i=1}^{n} \left\{ \left[\left(-\frac{2}{\phi_i^2} \left(\frac{\partial \nu_i}{\partial \phi_i} \right) + \frac{1}{\phi_i} \left(\frac{\partial^2 \nu_i}{\partial \phi_i^2} \right) + \frac{2\nu_i}{\phi_i} \right) \left(\frac{\partial \phi_i}{\partial \eta_{2i}} \right)^3 + 3 \left(\frac{1}{\phi_i} \left(\frac{\partial \nu_i}{\partial \phi_i} \right) - \frac{\nu_i}{\phi^2} \right) \right.\right.$$
$$\left. \times \left(\frac{\partial^2 \phi_i}{\partial \eta_{2i}^2} \right) \left(\frac{\partial \phi_i}{\partial \eta_{2i}} \right) \right] (J,L,M)_i + \frac{1}{\phi_i} \left(\left(\frac{\partial \nu_i}{\partial \phi_i} \right) - \frac{\nu_i}{\phi_i} \right) \left(\frac{\partial \phi_i}{\partial \eta_{2i}} \right)^2 [(JM,L)_i$$
$$+ (LM,J)_i + (JL,M)_i] + \frac{\nu_i}{\phi_i} \left[\left(\frac{\partial^3 \phi_i}{\partial \eta_{2i}^3} \right) (J,L,M)_i + \left(\frac{\partial^2 \phi_i}{\partial \eta_{2i}^2} \right) [(JM,L)_i \right.$$
$$\left.\left. + (LM,J)_i + (JL,M)_i] + \left(\frac{\partial \phi_i}{\partial \eta_{2i}} \right) (JLM)_i \right] \right\},$$

$$
\begin{aligned}
U_{jlmu} = \sum_{i=1}^{n} \Bigg\{ & -\frac{1}{\phi_i^2}\left[\frac{t_i^*}{\phi_i^2}\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)^4 + \frac{6t_i^*}{\phi_i}\left(\frac{\partial^2\mu_i}{\partial\eta_{1i}^2}\right)\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)^2 + 3t_i^*\left(\frac{\partial^2\mu_i}{\partial\eta_{1i}^2}\right)^2\right.\\
& \left. + \; 4t_i^*\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)\left(\frac{\partial^3\mu_i}{\partial\eta_{1i}^3}\right)\right](j,l,m,u)_i + \left[-\frac{t_i^*}{\phi_i^3}\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)^3 - \frac{3t_i^*}{\phi_i^2}\left(\frac{\partial^2\mu_i}{\partial\eta_{1i}^2}\right)\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)\right]\\
& \times [(ju,l,m)_i + (lu,j,m)_i + (mu,j,l)_i + (jm,l,u)_i + (lm,j,u)_i\\
& + \; (jl,m,u)_i] - \frac{t_i^*}{\phi_i^2}\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)^2[(jmu,l)_i + (jm,lu)_i + (lmu,j)_i + (lm,ju)_i\\
& + \; (jlu,m)_i + (jl,mu)_i + (jlm,u)_i] + \frac{(1-t_i^*)}{\phi_i}\left[\left(\frac{\partial^4\mu_i}{\partial\eta_{1i}^4}\right)(j,l,m,u)_i\right.\\
& + \; \left(\frac{\partial^3\mu_i}{\partial\eta_{1i}^3}\right)[(ju,l,m)_i + (lu,j,m)_i + (mu,j,l)_i + (jm,l,u)_i + (lm,j,u)_i\\
& + \; (jl,m,u)_i] + \left(\frac{\partial^2\mu_i}{\partial\eta_{1i}^2}\right)[(jmu,l)_i + (jm,lu)_i + (lmu,j)_i + (lm,ju)_i\\
& \left. + \; (jlu,m)_i + (jl,mu)_i + (jlm,u)_i] + \left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)(jlmu)_i\right]\Bigg\},
\end{aligned}
$$

$$
\begin{aligned}
U_{jlMU} = \sum_{i=1}^{n} \Bigg\{ & \left[\frac{1}{\phi_i}\left(-\frac{1}{\phi_i}\left(\frac{\partial^2\nu_i}{\partial\mu_i^2}\right)\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)^2 + \left(\frac{\partial^3\nu_i}{\partial\phi_i\partial\mu_i^2}\right)\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)^2 + \left(\frac{\partial^2\nu_i}{\partial\mu_i^2}\right)\left(\frac{\partial^2\phi_i}{\partial\eta_{2i}^2}\right)\right)\right.\\
& \times \left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)^2 + \frac{1}{\phi_i}\left(-\frac{1}{\phi_i}\left(\frac{\partial\nu_i}{\partial\mu_i}\right)\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)^2 + \left(\frac{\partial^2\nu_i}{\partial\phi_i\partial\mu_i}\right)\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)^2 + \left(\frac{\partial\nu_i}{\partial\mu_i}\right)\left(\frac{\partial^2\phi_i}{\partial\eta_{2i}^2}\right)\right)\\
& \left.\times \left(\frac{\partial^2\mu_i}{\partial\eta_{1i}^2}\right)\right](j,l,M,U)_i + \frac{1}{\phi_i}\left[\left(\frac{\partial^2\nu_i}{\partial\mu_i^2}\right)\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)^2 + \left(\frac{\partial\nu_i}{\partial\mu_i}\right)\left(\frac{\partial^2\mu_i}{\partial\eta_{1i}^2}\right)\right]\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)\\
& \times (j,l,M,U)_i + \frac{1}{\phi_i}\left[-\frac{1}{\phi_i}\left(\frac{\partial\nu_i}{\partial\mu_i}\right)\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)^2 + \left(\frac{\partial^2\nu_i}{\partial\phi_i\partial\mu_i}\right)\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)^2 + \left(\frac{\partial\nu_i}{\partial\mu_i}\right)\right.\\
& \left.\times \left(\frac{\partial^2\phi_i}{\partial\eta_{2i}^2}\right)\right]\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)(jl,M,U)_i + \frac{1}{\phi_i}\left(\frac{\partial\nu_i}{\partial\mu_i}\right)\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)(jl,MU)_i\Bigg\},
\end{aligned}
$$

$$
\begin{aligned}
U_{jLMU} = \sum_{i=1}^{n} \Bigg\{ \frac{1}{\phi_i} \Bigg[-\frac{2}{\phi_i} \left(\frac{\partial^2 \nu_i}{\partial \mu_i \partial \phi_i} \right) + \left(\frac{\partial^3 \nu_i}{\partial \mu_i \partial \phi_i^2} \right) + \frac{2}{\phi_i^2} \left(\frac{\partial \nu_i}{\partial \mu_i} \right) \Bigg] \left(\frac{\partial \phi_i}{\partial \eta_{2i}} \right)^3 \left(\frac{\partial \mu_i}{\partial \eta_{1i}} \right) \\
\times (j, L, M, U)_i + \frac{3}{\phi_i} \left[\left(\frac{\partial^2 \nu_i}{\partial \phi_i \partial \mu_i} \right) - \frac{1}{\phi_i} \left(\frac{\partial \nu_i}{\partial \mu_i} \right) \right] \left(\frac{\partial^2 \phi_i}{\partial \eta_{2i}^2} \right) \left(\frac{\partial \phi_i}{\partial \eta_{2i}} \right) \left(\frac{\partial \mu_i}{\partial \eta_{1i}} \right) \\
\times (j, L, M, U)_i + \frac{1}{\phi_i} \left[\left(\frac{\partial^2 \nu_i}{\partial \phi_i \partial \mu_i} \right) - \frac{1}{\phi_i} \left(\frac{\partial \nu_i}{\partial \mu_i} \right) \right] \left(\frac{\partial \phi_i}{\partial \eta_{2i}} \right)^2 \left(\frac{\partial \mu_i}{\partial \eta_{1i}} \right) [(j, LU, M)_i \\
+ (j, MU, L)_i + (j, LM, U)_i] + \frac{1}{\phi_i} \left(\frac{\partial \nu_i}{\partial \mu_i} \right) \left(\frac{\partial \mu_i}{\partial \eta_{1i}} \right) \Bigg[\left(\frac{\partial^3 \phi_i}{\partial \eta_{2i}^3} \right) (j, L, M, U)_i \\
+ \left(\frac{\partial^2 \phi_i}{\partial \eta_{2i}^2} \right) [(j, LU, M)_i + (j, MU, L)_i + (j, LM, U)_i] + \left(\frac{\partial \phi_i}{\partial \eta_{2i}} \right) (j, LM, U)_i \Bigg] \Bigg\},
\end{aligned}
$$

$$
\begin{aligned}
U_{JLMU} = \sum_{i=1}^{n} \Bigg\{ \frac{1}{\phi_i} \Bigg[\left(\frac{6}{\phi^2} \left(\frac{\partial \nu_i}{\partial \phi_i} \right) - \frac{3}{\phi_i} \left(\frac{\partial^2 \nu_i}{\partial \phi_i^2} \right) + \left(\frac{\partial^3 \nu_i}{\partial \phi_i^3} \right) - \frac{6 \nu_i}{\phi_i^3} \right) \left(\frac{\partial \phi_i}{\partial \eta_{2i}} \right)^4 + \frac{1}{\phi_i} \\
\times \left(-\frac{12}{\phi_i} \left(\frac{\partial \nu_i}{\partial \phi_i} \right) + 6 \left(\frac{\partial^2 \nu_i}{\partial \phi_i^2} \right) + \frac{12 \nu_i}{\phi_i^2} \right) \left(\frac{\partial \phi_i}{\partial \eta_{2i}} \right)^2 \left(\frac{\partial^2 \phi_i}{\partial \eta_{2i}^2} \right) + \frac{4}{\phi_i} \left(\left(\frac{\partial \nu_i}{\partial \phi_i} \right) - \frac{\nu_i}{\phi_i} \right) \\
\times \left(\frac{\partial^3 \phi_i}{\partial \eta_{2i}^3} \right) \left(\frac{\partial \phi_i}{\partial \eta_{2i}} \right) + \frac{3}{\phi_i} \left(\left(\frac{\partial \nu_i}{\partial \phi_i} \right) - \frac{\nu_i}{\phi_i} \right) \left(\frac{\partial^2 \phi_i}{\partial \eta_{2i}^2} \right)^2 \Bigg] (J, L, M, U)_i + \frac{1}{\phi_i} \Bigg[\Bigg(-\frac{2}{\phi_i} \left(\frac{\partial \nu_i}{\partial \phi_i} \right) \\
+ \left(\frac{\partial^2 \nu_i}{\partial \phi_i^2} + \frac{2 \nu_i}{\phi_i^2} \right) \Bigg) \left(\frac{\partial \phi_i}{\partial \eta_{2i}} \right)^3 + 3 \left(\left(\frac{\partial \nu_i}{\partial \phi_i} \right) - \frac{\nu_i}{\phi_i} \right) \left(\frac{\partial^2 \phi_i}{\partial \eta_{2i}^2} \right) \left(\frac{\partial \phi_i}{\partial \eta_{2i}} \right) \Bigg] [(JU, L, M)_i \\
+ (LU, J, M)_i + (MU, J, L)_i + (LM, J, U)_i + (JM, L, U)_i (JL, M, U)_i] + \frac{1}{\phi_i} \Bigg[\left(\frac{\partial \nu_i}{\partial \phi_i} \right) \\
- \frac{\nu_i}{\phi_i} \Bigg] \left(\frac{\partial \phi_i}{\partial \eta_{2i}} \right)^2 [(JMU, L)_i + (JM, LU)_i + (LMU, J)_i + (LM, JU)_i + (JLU, M)_i \\
+ (JL, MU)_i + (JLM, U)_i] + \frac{\nu_i}{\phi_i} \Bigg[\left(\frac{\partial^4 \phi_i}{\partial \eta_{2i}^4} \right) (J, L, M, U)_i + \left(\frac{\partial^3 \phi_i}{\partial \eta_{2i}^3} \right) [(JU, L, M)_i \\
+ (LU, J, M)_i + (MU, J, L)_i + (JM, L, U)_i + (LM, J, U)_i + (JL, M, U)_i] + \left(\frac{\partial^2 \phi_i}{\partial \eta_{2i}^2} \right) \\
\times [(JMU, L)_i + (JM, LU)_i + (LMU, J)_i + (LM, JU)_i + (JLU, M)_i + (JL, MU)_i \\
+ (JLM, U)_i] + \left(\frac{\partial \phi_i}{\partial \eta_{2i}} \right) (JLMU)_i \Bigg] \Bigg\}.
\end{aligned}
$$

By computing the expected values of the above derivatives we arrive at the following cumulants:

$$
\begin{aligned}
\kappa_{jl} &= -\sum_{i=1}^{n}\left\{\frac{1}{\phi_i^2}\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)^2 (j,l)_i\right\},\\
\kappa_{jL} &= -\sum_{i=1}^{n}\left\{\frac{(\gamma-1)}{\phi_i^2}\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)(j,L)_i\right\},\\
\kappa_{JL} &= -\sum_{i=1}^{n}\left\{\frac{(\Gamma^{(2)}(2)+1)}{\phi_i^2}\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)(J,L)_i\right\},\\
\kappa_{jlm} &= -\sum_{i=1}^{n}\left\{\frac{1}{\phi_i^2}\left[\frac{1}{\phi_i}\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)^3 + 3\left(\frac{\partial^2\mu_i}{\partial\eta_{1i}^2}\right)\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)\right](j,l,m)_i\right.\\
&\quad\left. + \frac{1}{\phi_i^2}\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)^2 [(jm,l)_i + (lm,j)_i + (jl,m)_i]\right\},\\
\kappa_{jlM} &= \sum_{i=1}^{n}\left\{\frac{1}{\phi_i^2}\left[\frac{(3-\gamma)}{\phi_i}\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)^2\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right) + (1-\gamma)\left(\frac{\partial^2\mu_i}{\partial\eta_{1i}^2}\right)\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)\right](j,l,M)_i\right.\\
&\quad\left. - \frac{(\gamma-1)}{\phi_i^2}\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)(jl,M)_i\right\},\\
\kappa_{jLM} &= \sum_{i=1}^{n}\left\{\frac{[4(\gamma-1)-\Gamma^{(2)}(2)]}{\phi_i^3}\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)^2\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)(j,L,M)_i + \frac{(1-\gamma)}{\phi_i^2}\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)\right.\\
&\quad\left. \times\left[\left(\frac{\partial^2\phi_i}{\partial\eta_{2i}^2}\right)(j,L,M)_i + \left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)(j,LM)_i\right]\right\},\\
\kappa_{JLM} &= \sum_{i=1}^{n}\left\{\left[\frac{(\Gamma^{(3)}(2)+6\Gamma^{(2)}(2)+4)}{\phi_i^3}\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)^3 - \frac{3(\Gamma^{(2)}(2)+1)}{\phi_i^2}\left(\frac{\partial^2\phi_i}{\partial\eta_{2i}^2}\right)\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)\right]\right.\\
&\quad\left. \times (J,L,M)_i - \frac{(\Gamma^{(2)}(2)+1)}{\phi_i^2}\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)^2 [(JM,L)_i + (LM,J)_i + (JL,M)_i]\right\},\\
\kappa_{jlmu} &= \sum_{i=1}^{n}\left\{-\frac{1}{\phi_i}\left[\frac{1}{\phi_i^3}\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)^4 + \frac{6}{\phi_i^2}\left(\frac{\partial^2\mu_i}{\partial\eta_{1i}^2}\right)\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)^2 + \frac{3}{\phi_i}\left(\frac{\partial^2\mu_i}{\partial\eta_{1i}^2}\right)^2\right.\right.\\
&\quad\left. + \frac{4}{\phi_i}\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)\left(\frac{\partial^3\mu_i}{\partial\eta_{1i}^3}\right)\right](j,l,m,u)_i - \frac{1}{\phi_i^2}\left[\frac{1}{\phi_i}\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)^3 + 3\left(\frac{\partial^2\mu_i}{\partial\eta_{1i}^2}\right)\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)\right]\\
&\quad \times [(ju,l,m)_i + (lu,j,m)_i + (mu,j,l)_i + (jm,l,u)_i + (lm,j,u)_i\\
&\quad + (jl,m,u)_i] - \frac{1}{\phi_i^2}\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)^2 [(jmu,l)_i + (jm,lu)_i + (lmu,j)_i + (lm,ju)_i\\
&\quad \left. + (jlu,m)_i + (jl,mu)_i + (jlm,u)_i]\right\},\\
\kappa_{jlmU} &= \sum_{i=1}^{n}\left\{\left[\frac{(4-\gamma)}{\phi_i^4}\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)^3 + \frac{3(3-\gamma)}{\phi_i^3}\left(\frac{\partial^2\mu_i}{\partial\eta_{1i}^2}\right)\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right) - \frac{(\gamma-1)}{\phi_i^2}\left(\frac{\partial^3\mu_i}{\partial\eta_{1i}^3}\right)\right]\right.\\
&\quad \times\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)(j,l,m,U)_i + \left[\frac{(3-\gamma)}{\phi_i^3}\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)^2 - \frac{(\gamma-1)}{\phi_i^2}\left(\frac{\partial^2\mu_i}{\partial\eta_{1i}^2}\right)\right]\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)\\
&\quad \times [(jm,l,U)_i + (lm,j,U)_i + (jl,m,U)_i] - \frac{(\gamma-1)}{\phi_i^2}\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)\\
&\quad \left. \times (jlm,U)_i\right\},
\end{aligned}
$$

$$\kappa_{jLMU} = \sum_{i=1}^{n} \Bigg\{ \Bigg[-\frac{[18(\gamma-1)-9\Gamma^{(2)}(2)-\Gamma^{(3)}(2)]}{\phi_i^4} \left(\frac{\partial \phi_i}{\partial \eta_{2i}}\right)^3 + \frac{[12(\gamma-1)-3\Gamma^{(2)}(2)]}{\phi_i^3}$$

$$\times \left(\frac{\partial^2 \phi_i}{\partial \eta_{2i}^2}\right)\left(\frac{\partial \phi_i}{\partial \eta_{2i}}\right) - \frac{(\gamma-1)}{\phi_i^2}\left(\frac{\partial^3 \phi_i}{\partial \eta_{2i}^3}\right)\Bigg] \left(\frac{\partial \mu_i}{\partial \eta_{1i}}\right)(j,L,M,U)_i + \Bigg[\frac{(4(\gamma-1)-\Gamma^{(2)}(2))}{\phi_i^3}$$

$$\times \left(\frac{\partial \phi_i}{\partial \eta_{2i}}\right)^2 - \frac{(\gamma-1)}{\phi_i^2}\left(\frac{\partial^2 \phi_i}{\partial \eta_{2i}^2}\right)\Bigg]\left(\frac{\partial \mu_i}{\partial \eta_{1i}}\right)[(j,LU,M)_i+(j,MU,L)_i+(j,LM,U)_i]$$

$$- \frac{(\gamma-1)}{\phi_i^2}\left(\frac{\partial \mu_i}{\partial \eta_{1i}}\right)\left(\frac{\partial \phi_i}{\partial \eta_{2i}}\right)(j,LM,U)_i \Bigg\},$$

$$\kappa_{jlMU} = \sum_{i=1}^{n} \Bigg\{ \Bigg[-\frac{[4-5(\gamma-1)+\Gamma^{(2)}(2)]}{\phi_i^4}\left(\frac{\partial \mu_i}{\partial \eta_{1i}}\right)^2\left(\frac{\partial \phi_i}{\partial \eta_{2i}}\right)^2 + \frac{[3(\gamma-1)-\Gamma^{(2)}(2)]}{\phi_i^3}$$

$$\times \left(\frac{\partial^2 \mu_i}{\partial \eta_{1i}^2}\right)\left(\frac{\partial \phi_i}{\partial \eta_{2i}}\right)^2 + \left(\frac{(3-\gamma)}{\phi_i^3}\left(\frac{\partial \mu_i}{\partial \eta_{1i}}\right)^2 - \frac{(\gamma-1)}{\phi_i^2}\left(\frac{\partial^2 \mu_i}{\partial \eta_{1i}^2}\right)\right)\left(\left(\frac{\partial^2 \phi_i}{\partial \eta_{2i}^2}\right)\right.$$

$$\left.\left. - \frac{1}{\phi_i}\left(\frac{\partial \phi_i}{\partial \eta_{2i}}\right)^2\right)\right](j,l,M,U)_i + \Bigg[\frac{(3(\gamma-1)-\Gamma^{(2)}(2))}{\phi_i^3}\left(\frac{\partial \phi_i}{\partial \eta_{2i}}\right)^2\left(\frac{\partial \mu_i}{\partial \eta_{1i}}\right)$$

$$- \frac{(\gamma-1)}{\phi_i^2}\left(\frac{\partial \mu_i}{\partial \eta_{1i}}\right)\left(\left(\frac{\partial^2 \phi_i}{\partial \eta_{2i}^2}\right) - \frac{1}{\phi_i}\left(\frac{\partial \phi_i}{\partial \eta_{2i}}\right)^2\right)\Bigg](jl,M,U)_i + \Bigg[\frac{(3-\gamma)}{\phi_i^3}\left(\frac{\partial \mu_i}{\partial \eta_{1i}}\right)^2$$

$$- \frac{(\gamma-1)}{\phi_i^2}\left(\frac{\partial^2 \mu_i}{\partial \eta_{1i}^2}\right)\Bigg]\left(\frac{\partial \phi_i}{\partial \eta_{2i}}\right)(j,l,MU)_i - \frac{(\gamma-1)}{\phi_i^2}\left(\frac{\partial \mu_i}{\partial \eta_{1i}}\right)\left(\frac{\partial \phi_i}{\partial \eta_{2i}}\right)(jl,MU)_i \Bigg\},$$

$$\kappa_{JLMU} = \sum_{i=1}^{n} \Bigg\{ \Bigg[-\frac{(36\Gamma^{(2)}(2)+18+12\Gamma^{(3)}(2)+\Gamma^{(4)}(2))}{\phi_i^4}\left(\frac{\partial \phi_i}{\partial \eta_{2i}}\right)^4 + \frac{(36\Gamma^{(2)}(2)+24+6\Gamma^{(3)}(2))}{\phi_i^3}$$

$$\times \left(\frac{\partial \phi_i}{\partial \eta_{2i}}\right)^2\left(\frac{\partial^2 \phi_i}{\partial \eta_{2i}^2}\right) - \frac{4(\Gamma^{(2)}(2)+1)}{\phi_i^2}\left(\frac{\partial^3 \phi_i}{\partial \eta_{2i}^3}\right)\left(\frac{\partial \phi_i}{\partial \eta_{2i}}\right) - \frac{3(\Gamma^{(2)}(2)+1)}{\phi_i^2}\left(\frac{\partial^2 \phi_i}{\partial \eta_{2i}^2}\right)^2\Bigg]$$

$$\times (J,L,M,U)_i + \Bigg[\frac{(6\Gamma^{(2)}(2)+4+\Gamma^{(3)}(2))}{\phi_i^3}\left(\frac{\partial \phi_i}{\partial \eta_{2i}}\right)^3 - \frac{3(\Gamma^{(2)}(2)+1)}{\phi_i^2}\left(\frac{\partial^2 \phi_i}{\partial \eta_{2i}^2}\right)$$

$$\times \left(\frac{\partial \phi_i}{\partial \eta_{2i}}\right)\Bigg][(JU,L,M)_i+(LU,J,M)_i+(MU,J,L)_i+(JM,L,U)_i+(LM,J,U)_i$$

$$+ (JL,M,U)_i] - \frac{(\Gamma^{(2)}(2)+1)}{\phi_i^2}\left(\frac{\partial \phi_i}{\partial \eta_{2i}}\right)^2[(JMU,L)_i+(JM,LU)_i+(LMU,J)_i$$

$$+ (LM,JU)_i+(JLU,M)_i+(JL,MU)_i+(JLM,U)_i]\Bigg\}.$$

Next, we obtain the following closed-form expressions for the first order derivatives of the log-likelihood cumulants:

$$\kappa_{jl}^{(m)} = -\sum_{i=1}^{n} \left\{ \frac{1}{\phi_i^2} \left[2\left(\frac{\partial \mu_i}{\partial \eta_{1i}}\right)\left(\frac{\partial^2 \mu_i}{\partial \eta_{1i}^2}\right)(j,l,m)_i + \left(\frac{\partial \mu_i}{\partial \eta_{1i}}\right)^2 [(jm,l)_i + (j,lm)_i] \right] \right\},$$

$$\kappa_{jl}^{(M)} = \sum_{i=1}^{n} \left\{ \frac{2}{\phi_i^3} \left(\frac{\partial \mu_i}{\partial \eta_{1i}}\right)^2 \left(\frac{\partial \phi_i}{\partial \eta_{2i}}\right)(j,l,M)_i \right\},$$

$$\kappa_{jL}^{(m)} = -\sum_{i=1}^{n} \left\{ \frac{(\gamma-1)}{\phi_i^2} \left[\left(\frac{\partial^2 \mu_i}{\partial \eta_{1i}^2}\right)(j,L,m)_i + \left(\frac{\partial \mu_i}{\partial \eta_{1i}}\right)(jm,L)_i \right] \left(\frac{\partial \phi_i}{\partial \eta_{2i}}\right) \right\},$$

$$\kappa_{jL}^{(M)} = \sum_{i=1}^{n} \left\{ \frac{(\gamma-1)}{\phi_i^2} \left[\frac{2}{\phi_i}\left(\frac{\partial \mu_i}{\partial \eta_{1i}}\right)\left(\frac{\partial \phi_i}{\partial \eta_{2i}}\right)^2 (j,L,M)_i - \left(\frac{\partial \mu_i}{\partial \eta_{1i}}\right)\left(\frac{\partial^2 \phi_i}{\partial \eta_{2i}^2}\right)(j,L,M)_i \right.\right.$$
$$\left.\left. - \left(\frac{\partial \mu_i}{\partial \eta_{1i}}\right)\left(\frac{\partial \phi_i}{\partial \eta_{2i}}\right)(j,LM)_i \right] \right\},$$

$$\kappa_{JL}^{(m)} = 0,$$

$$\kappa_{JL}^{(M)} = \sum_{i=1}^{n} \left\{ \frac{(\Gamma^{(2)}(2)+1)}{\phi_i^2} \left[2\left(\frac{1}{\phi_i}\left(\frac{\partial \phi_i}{\partial \eta_{2i}}\right)^3 - \left(\frac{\partial^2 \phi_i}{\partial \eta_{2i}^2}\right)\left(\frac{\partial \phi_i}{\partial \eta_{2i}}\right) \right)(J,L,M)_i \right.\right.$$
$$\left.\left. - \left(\frac{\partial \phi_i}{\partial \eta_{2i}}\right)^2 [(JM,L)_i + (J,LM)_i] \right] \right\},$$

$$\kappa_{jlm}^{(u)} = -\sum_{i=1}^{n} \left\{ \frac{3}{\phi_i^2} \left[\frac{1}{\phi_i}\left(\frac{\partial \mu_i}{\partial \eta_{1i}}\right)^2 \left(\frac{\partial^2 \mu_i}{\partial \eta_{1i}^2}\right) + \left(\frac{\partial^2 \mu_i}{\partial \eta_{1i}^2}\right)^2 + \left(\frac{\partial \mu_i}{\partial \eta_{1i}}\right)\left(\frac{\partial^3 \mu_i}{\partial \eta_{1i}^3}\right) \right] (j,l,m,u)_i \right.$$
$$+ \left[\frac{1}{\phi_i^3}\left(\frac{\partial \mu_i}{\partial \eta_{1i}}\right)^3 + \frac{3}{\phi_i^2}\left(\frac{\partial \mu_i}{\partial \eta_{1i}}\right)\left(\frac{\partial^2 \mu_i}{\partial \eta_{1i}^2}\right) \right] [(ju,l,m)_i + (lu,j,m)_i + (mu,j,l)_i]$$
$$+ \frac{2}{\phi_i^2}\left(\frac{\partial \mu_i}{\partial \eta_{1i}}\right)\left(\frac{\partial^2 \mu_i}{\partial \eta_{1i}^2}\right)[(lm,j,u)_i + (jm,l,u)_i + (jl,m,u)_i] + \frac{1}{\phi_i^2}\left(\frac{\partial \mu_i}{\partial \eta_{1i}}\right)^2$$
$$\left. \times [(jmu,l)_i + (jm,lu)_i + (lmu,j)_i + (lm,ju)_i + (jlu,m)_i + (jl,mu)_i] \right\},$$

$$\kappa_{jlm}^{(M)} = \sum_{i=1}^{n} \left\{ \frac{3}{\phi_i^3} \left[\frac{1}{\phi_i}\left(\frac{\partial \mu_i}{\partial \eta_{1i}}\right)^3 + 2\left(\frac{\partial^2 \mu_i}{\partial \eta_{1i}^2}\right)\left(\frac{\partial \mu_i}{\partial \eta_{1i}}\right) \right] \left(\frac{\partial \phi_i}{\partial \eta_{2i}}\right)(j,l,m,U)_i + \frac{2}{\phi_i^3} \right.$$
$$\left. \times \left(\frac{\partial \mu_i}{\partial \eta_{1i}}\right)^2 \left(\frac{\partial \phi_i}{\partial \eta_{2i}}\right)[(jm,l,U)_i + (lm,j,U)_i + (jl,m,U)_i] \right\},$$

$$\kappa_{jlM}^{(u)} = \sum_{i=1}^{n} \left\{ \frac{1}{\phi^2} \left[\frac{2(3-\gamma)}{\phi}\left(\frac{\partial \phi_i}{\partial \eta_{2i}}\right)\left(\frac{\partial \mu_i}{\partial \eta_{1i}}\right)\left(\frac{\partial^2 \mu_i}{\partial \eta_{1i}^2}\right) + (1-\gamma)\left(\frac{\partial^3 \mu_i}{\partial \eta_{1i}^3}\right)\left(\frac{\partial \phi_i}{\partial \eta_{2i}}\right) \right] \right.$$
$$\times (j,l,M,u)_i + \frac{(3-\gamma)}{\phi_i^3}\left(\frac{\partial \mu_i}{\partial \eta_{1i}}\right)^2 \left(\frac{\partial \phi_i}{\partial \eta_{2i}}\right)[(ju,l,M)_i + (j,lu,M)_i]$$
$$+ \frac{(1-\gamma)}{\phi_i^2}\left(\frac{\partial^2 \mu_i}{\partial \eta_{1i}^2}\right)\left(\frac{\partial \phi_i}{\partial \eta_{2i}}\right)[(ju,l,M)_i + (j,lu,M)_i + (jl,M,u)_i]$$
$$\left. + \frac{(1-\gamma)}{\phi_i^2}\left(\frac{\partial \phi_i}{\partial \eta_{2i}}\right)\left(\frac{\partial \mu_i}{\partial \eta_{1i}}\right)(jlu,M)_i \right\},$$

$$\kappa_{jlM}^{(U)} = \sum_{i=1}^{n}\left\{\frac{(3-\gamma)}{\phi_i^3}\left[\left(\left(\frac{\partial^2\phi_i}{\partial\eta_{2i}^2}\right) - \frac{3}{\phi_i}\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)^2\right)(j,l,M,U)_i + \left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)(j,l,MU)_i\right]\right.$$
$$\times\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)^2 + \frac{(1-\gamma)}{\phi_i^2}\left[\left(\left(\frac{\partial^2\phi_i}{\partial\eta_{2i}^2}\right) - \frac{2}{\phi_i}\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)^2\right)\left(\left(\frac{\partial^2\mu_i}{\partial\eta_{1i}^2}\right)(j,l,M,U)_i\right.\right.$$
$$\left.\left.\left. + \left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)(jl,M,U)_i\right) + \left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)\left(\left(\frac{\partial^2\mu_i}{\partial\eta_{1i}^2}\right)(j,l,MU)_i + \left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)(jl,MU)_i\right)\right]\right\},$$

$$\kappa_{jLM}^{(u)} = \sum_{i=1}^{n}\left\{\left[\frac{(4(\gamma-1)-\Gamma^{(2)}(2))}{\phi_i^3}\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)^2 + \frac{(1-\gamma)}{\phi_i^2}\left(\frac{\partial^2\phi_i}{\partial\eta_{2i}^2}\right)\right]\left[\left(\frac{\partial^2\mu_i}{\partial\eta_{1i}^2}\right)(j,L,M,u)_i\right.\right.$$
$$\left. + \left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)(ju,L,M)_i\right] + \frac{(1-\gamma)}{\phi_i^2}\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)\left[\left(\frac{\partial^2\mu_i}{\partial\eta_{1i}^2}\right)(j,LM,u)_i\right.$$
$$\left.\left. + \left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)(ju,LM)_i\right]\right\},$$

$$\kappa_{jLM}^{(U)} = \sum_{i=1}^{n}\left\{\left[\frac{(4(\gamma-1)-\Gamma^{(2)}(2))}{\phi_i^3}\left(2\left(\frac{\partial^2\phi_i}{\partial\eta_{2i}^2}\right)\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right) - \frac{1}{\phi_i}\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)^3\right) + \frac{2(\gamma-1)}{\phi_i^3}\right.\right.$$
$$\left.\times\left(\frac{\partial^2\phi_i}{\partial\eta_{2i}^2}\right)\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right) - \frac{(\gamma-1)}{\phi_i^2}\left(\frac{\partial^3\phi_i}{\partial\eta_{2i}^3}\right)\right]\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)(j,L,M,U)_i + \frac{1}{\phi_i^2}$$
$$\times\left[\frac{(4(\gamma-1)-\Gamma^{(2)}(2))}{\phi_i}\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)^2 - (\gamma-1)\left(\frac{\partial^2\phi_i}{\partial\eta_{2i}^2}\right)\right]\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)[(j,LU,M)_i$$
$$+ (j,L,MU)_i] + \frac{(1-\gamma)}{\phi_i^2}\left[\left(\left(\frac{\partial^2\phi_i}{\partial\eta_{2i}^2}\right) - \frac{2}{\phi_i}\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)^2\right)(j,LM,U)_i\right.$$
$$\left.\left. + \left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)(j,LMU)_i\right]\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)\right\},$$

$$\kappa_{JLM}^{(u)} = 0,$$

$$\kappa_{JLM}^{(U)} = \sum_{i=1}^{n}\left\{\frac{3}{\phi_i^2}\left[\frac{(4+6\Gamma^{(2)}(2)+\Gamma^{(3)}(2))}{\phi_i}\left(\left(\frac{\partial^2\phi_i}{\partial\eta_{2i}^2}\right)\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)^2 - \frac{1}{\phi_i}\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)^4\right)\right.\right.$$
$$\left. - (\Gamma^{(2)}(2)+1)\left(\left(\frac{\partial^3\phi_i}{\partial\eta_{2i}^3}\right)\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right) + \left(\frac{\partial^2\phi_i}{\partial\eta_{2i}^2}\right)^2 - \frac{2}{\phi_i}\left(\frac{\partial^2\phi_i}{\partial\eta_{2i}^2}\right)\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)^2\right)\right]$$
$$\times (J,L,M,U)_i + \frac{1}{\phi_i^2}\left[\frac{(4+6\Gamma^{(2)}(2)+\Gamma^{(3)}(2))}{\phi_i}\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)^3 - 3(\Gamma^{(2)}(2)+1)\right.$$
$$\left.\times\left(\frac{\partial^2\phi_i}{\partial\eta_{2i}^2}\right)\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)\right][(JU,L,M)_i + (LU,J,M)_i + (MU,J,L)_i]$$
$$- \frac{2(\Gamma^{(2)}(2)+1)}{\phi_i^2}\left(\left(\frac{\partial^2\phi_i}{\partial\eta_{2i}^2}\right)\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right) - \frac{1}{\phi_i}\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)^3\right)[(JM,L,U)_i$$
$$+ (LM,J,U)_i + (JL,M,U)_i] - \frac{(\Gamma^{(2)}(2)+1)}{\phi_i^2}\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)^2[(JMU,L)_i$$
$$\left. + (JM,LU)_i + (LMU,J)_i + (LM,JU)_i + (JLU,M)_i + (JL,MU)_i]\right\}.$$

Finally, we compute the second order derivatives of the log-likelihood cumulants which can expressed as follows:

$$
\begin{aligned}
\kappa_{jl}^{(mu)} &= \sum_{i=1}^{n}\left\{-\frac{2}{\phi_i^2}\left[\left(\left(\frac{\partial^2\mu_i}{\partial\eta_{1i}^2}\right)^2+\left(\frac{\partial^3\mu_i}{\partial\eta_{1i}^3}\right)\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)\right)(j,l,m,u)_i+\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)\left(\frac{\partial^2\mu_i}{\partial\eta_{1i}^2}\right)\right.\right.\\
&\times\ [(ju,l,m)_i+(lu,j,m)_i+(mu,j,l)_i+(jm,l,u)_i+(lm,j,u)_i]]\\
&\left.-\frac{1}{\phi_i^2}\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)^2[(jmu,l)_i+(jm,lu)_i+(lmu,j)_i+(lm,ju)_i]\right\},\\
\kappa_{jl}^{(Mu)} &= \sum_{i=1}^{n}\left\{\frac{2}{\phi_i^3}\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)\left[2\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)\left(\frac{\partial^2\mu_i}{\partial\eta_{1i}^2}\right)(j,l,M,u)_i+\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)^2\right.\right.\\
&\times\ [(ju,l,M)_i+(j,lu,M)_i]]\},\\
\kappa_{jl}^{(MU)} &= \sum_{i=1}^{n}\left\{\frac{2}{\phi_i^3}\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)^2\left[\left(\left(\frac{\partial^2\phi_i}{\partial\eta_{2i}^2}\right)-\frac{3}{\phi_i}\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)^2\right)(j,l,M,U)_i+\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)\right.\right.\\
&\times\ (j,l,MU)_i]\},\\
\kappa_{jL}^{(mu)} &= \sum_{i=1}^{n}\left\{\frac{(1-\gamma)}{\phi_i^2}\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)\left[\left(\frac{\partial^3\mu_i}{\partial\eta_{1i}^3}\right)(j,L,m,u)_i+\left(\frac{\partial^2\mu_i}{\partial\eta_{1i}^2}\right)[(ju,L,m)_i\right.\right.\\
&\left.\left.+\ (j,L,mu)_i+(jm,L,u)_i]+\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)(jmu,L)_i\right]\right\},\\
\kappa_{jL}^{(Mu)} &= \sum_{i=1}^{n}\left\{(\gamma-1)\left[\left(\frac{2}{\phi_i^3}\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)^2-\frac{1}{\phi_i^2}\left(\frac{\partial^2\phi_i}{\partial\eta_{2i}^2}\right)\right)\left(\left(\frac{\partial^2\mu_i}{\partial\eta_{1i}^2}\right)(j,L,M,u)_i\right.\right.\right.\\
&\left.+\ \left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)(ju,L,M)_i\right)-\frac{1}{\phi_i^2}\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)\left(\left(\frac{\partial^2\mu_i}{\partial\eta_{1i}^2}\right)(LM,j,u)_i+\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)\right.\\
&\times\ (LM,ju)_i)]\},\\
\kappa_{jL}^{(MU)} &= (\gamma-1)\sum_{i=1}^{n}\left\{\left[-\frac{6}{\phi_i^4}\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)^3+\frac{6}{\phi_i^3}\left(\frac{\partial^2\phi_i}{\partial\eta_{2i}^2}\right)\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)-\frac{1}{\phi_i^2}\left(\frac{\partial^3\phi_i}{\partial\eta_{2i}^3}\right)\right]\right.\\
&\times\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)(j,L,M,U)_i+\left[\frac{2}{\phi_i^3}\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)^2-\frac{1}{\phi_i^2}\left(\frac{\partial^2\phi_i}{\partial\eta_{2i}^2}\right)\right]\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)[(j,LU,M)_i\\
&\left.+(j,L,MU)_i+(LM,j,U)_i]-\frac{1}{\phi_i^2}\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)(LMU,j)_i\right\},\\
\kappa_{JL}^{(mu)} &= 0,\quad \kappa_{JL}^{(mU)}=0,\\
\kappa_{JL}^{(MU)} &= (\Gamma^{(2)}(2)+1)\sum_{i=1}^{n}\left\{-\frac{2}{\phi_i^2}\left[\frac{3}{\phi_i^2}\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)^4-\frac{5}{\phi_i}\left(\frac{\partial^2\phi_i}{\partial\eta_{2i}^2}\right)\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)^2+\left(\frac{\partial^3\phi_i}{\partial\eta_{2i}^3}\right)\right.\right.\\
&\times\left.\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)+\left(\frac{\partial^2\phi_i}{\partial\eta_{2i}^2}\right)^2\right](J,L,M,U)_i+\frac{2}{\phi_i^2}\left[\frac{1}{\phi_i}\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)^3-\left(\frac{\partial^2\phi_i}{\partial\eta_{2i}^2}\right)\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)\right]\\
&\times[(JM,L,U)_i+(LM,J,U)_i+(JU,L,M)_i+(LU,J,M)_i+(MU,J,L)_i]\\
&\left.-\frac{1}{\phi_i^2}\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)^2[(JMU,L)_i+(JM,LU)_i+(LMU,J)_i+(LM,JU)_i]\right\}.
\end{aligned}
$$

APPENDIX D – DETAILS ON HOW TO SELECT STARTING VALUES FOR THE ITERATIVE SCHEME USED FOR MAXIMUM LIKELIHOOD PARAMETER ESTIMATION UNDER NONLINEARITY

We shall now provide details on how to select starting values for the iterative scheme used for maximum likelihood parameter estimation under nonlinearity. At the outset, we consider $p_1 = p$, where p_1 is the number of columns of X and $f_1(x_i,\beta)$, a nonlinear function. Expanding the latter in Taylor series around $\beta^{(0)}$ up to first order, we obtain

$$f_1(x_i,\beta) \approx f_1(x_i,\beta^{(0)}) + \sum_{j=1}^{p} \left[\frac{\partial f_1(x_i,\beta)}{\partial \beta_j}\right]_{\beta=\beta^{(0)}} (\beta_j - \beta_j^{(0)}), \tag{D.1}$$

where $\beta^{(0)} = (\beta_1^{(0)},\ldots,\beta_p^{(0)})^\top$ is the vector of starting values for the estimation of β. Let $[\partial f_1(x_i,\beta)/\partial\beta_j]_{\beta=\beta^{(0)}} = c_{ij}^{(0)}$ and $\psi_j^{(0)} = (\beta_j - \beta_j^{(0)})$. It is possible to write (D.1) as

$$f_1(x_i,\beta) - f_1(x_i,\beta^{(0)}) \approx \sum_{j=1}^{p} c_{ij}^{(0)} \psi_j^{(0)} \tag{D.2}$$

In (D.2), we consider $g_1(y_i) \approx g_1(\mu_i) = f_1(x_i,\beta)$, então $g_1(y_i) \approx f_1(x_i,\beta)$. Thus, $g_1(y_i) - f_1(x_i,\beta^{(0)}) \approx f_1(x_i,\beta) - f_1(x_i,\beta^{(0)})$. Next, we consider the least squares estimator of $\psi^{(0)}$, which is given by

$$\psi^{(0)} \approx (J_1^{(0)\top} J_1^{(0)})^{-1} J_1^{(0)\top} \left[g_1(y) - f_1(X,\beta^{(0)})\right],$$

where $J_1^{(0)} = [\partial\eta_1/\partial\beta]_{\beta=\beta^{(0)}}$. We have that $\psi_j^{(0)} = (\hat{\beta}_j - \beta_j^{(0)})$; thus, $\beta_j^{(1)} = \psi_j^{(0)} + \beta_j^{(0)}$. It follows that the iterative scheme is $\beta_j^{(m)} = \psi_j^{(m-1)} + \beta_j^{(m-1)}$. We arrive at

$$\beta^{(m)} \approx \left(J_1^{(m-1)\top} J_1^{(m-1)}\right)^{-1} J_1^{(m-1)\top} \left[g_1(y) - f_1(X,\beta)^{(m-1)}\right] + \beta^{(m-1)}.$$

Setting $m = 1$, we obtain the starting value for β:

$$\beta_{NL}^{(0)} = \left(J_1^{(0)\top} J_1^{(0)}\right)^{-1} J_1^{(0)\top} \left[g_1(y) - f_1(X,\beta_L^{(0)})\right]$$

with $\beta_L^{(0)} = (X^\top X)^{-1} X^\top g(y)$, X denoting the $n \times p_1$ matrix of mean regressors of the linear model.

To determine the starting value for θ, we consider $q_1 = q$, where q_1 is the number of columns of Z and $g_2(\phi_i) = f_2(z_i^\top,\theta)$, a non-linear function. Here, z_i is the ith row of Z written as a column vector. Analogously to what was done previously, we have that

$$\theta_{NL}^{(0)} = \left(J_2^{(0)\top} J_2^{(0)}\right)^{-1} J_2^{(0)\top} \left[g_2(\phi_{NL}^{(0)}) - f_2(Z,\theta_L^{(0)})\right]$$

with

$$\theta_L^{(0)} = (Z^\top Z)^{-1} Z^\top g_2(\phi_L^{(0)}), \quad J_2^{(0)} = \left[\frac{\partial\eta_2}{\partial\theta}\right]_{\theta=\theta^{(0)}}, \quad \phi_{Li}^{(0)} = \sqrt{\frac{6\check{\sigma}_{Li}^2}{\pi^2}} \quad \text{and} \quad \phi_{NLi}^{(0)} = \sqrt{\frac{6\check{\sigma}_{NLi}^2}{\pi^2}},$$

where $\phi_L^{(0)} = (\phi_{L1}^{(0)}, \ldots, \phi_{Ln}^{(0)})^\top$, $\check{\sigma}^2_{Li} = \check{e}_L^\top \check{e}_L / \{(n-p)[g_1'(\check{\mu}_{Li})]^2\}$, $\check{\mu}_{Li} = g_1^{-1}(x_i^\top \beta_L^{(0)})$, $\check{e}_L = g_1(y) - \check{\mu}_L$, $\check{\sigma}^2_{NLi} = \check{e}_{NL}^\top \check{e}_{NL} / \{(n-p)[g_1'(\check{\mu}_{NLi})]^2\}$, $\check{\mu}_{NLi} = g_1^{-1}(f_1(x_i, \beta_{NL}^{(0)}))$, $\check{e}_{NL} = g(y) - \check{\mu}_{NL}$.

It is important to note that in one or both submodels there may be fewer regressors than parameters, i.e., $p_1 < p$ and/or $q_1 < q$. In that case, we proceed as follows. At the outset, we attribute values to some parameters in order to obtain a predictor formed by covariates that do not involve extra unknown parameters. For example, in $g(\mu_i) = \beta_1 + \beta_2[x_{i2}/(x_{i2} + \beta_3)]$, we provide an initial value for β_3. We note that the choice of values attributed to some of the parameters must consider: (i) the characteristics of the regressors and (ii) the mathematical relationship of the regressors to the functions in the nonlinear predictors. Next, we compute $(X^\top X)^{-1} X^\top g_1(y)$, X being based on a linear predictor. Thus, we obtain $\beta_L^{(0)}$ and, subsequently, $\beta_{NL}^{(0)}$.