Universidade Federal de Pernambuco
Centro de Ciências Exatas e da Natureza
Programa de Pós-Graduação em Estatística

JOSÉ VALDENIR DE OLIVEIRA JUNIOR

DIAGNOSTIC ANALYSIS IN GENERALIZED EXTREME VALUE NONLINEAR REGRESSION MODELS

Recife

2019

José Valdenir de Oliveira Junior

DIAGNOSTIC ANALYSIS IN GENERALIZED EXTREME VALUE NONLINEAR REGRESSION MODELS

Master's thesis submitted to the Graduate Program in Statistics, Department of Statistics, Universidade Federal de Pernambuco as a requirement to obtain a Master's degree in Statistics.
Main Area: Exact Sciences and Earth

Advisor: Prof. Dr. Francisco Cribari-Neto
Coadvisor: Prof. Dr. Juvêncio Santos Nobre

Recife

2019

Catalogação na fonte
Bibliotecária Monick Raquel Silvestre da S. Portes, CRB4-1217

O48d Oliveira Júnior, José Valdenir de
Diagnostic analysis in generalized extreme value nonlinear regression models / José Valdenir de Oliveira Júnior. – 2019.
66 f.: il., fig., tab.

Orientador: Francisco Cribari Neto.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CCEN, Estatística, Recife, 2019.
Inclui referências e apêndice.

1. Estatística. 2. Análise de influência. 3. Regressão não-linear. I. Cribari Neto, Francisco (orientador). II. Título.

310 CDD (23. ed.) UFPE- MEI 2019-039

JOSÉ VALDENIR DE OLIVEIRA JUNIOR

DIAGNOSTIC ANALYSIS IN GENERALIZED EXTREME VALUE NONLINEAR REGRESSION MODELS

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

Aprovada em: 25 de fevereiro de 2019.

BANCA EXAMINADORA

Prof.(º) Francisco Cribari Neto
UFPE

Prof.(º) Klaus Leite Pinto Vasconcellos
UFPE

Prof.(º) Aluísio de Souza Pinheiro
UNICAMP

Aos persistentes e doces sorrisos de minha mãe,
à fé inabalável e incomensurável de meu pai
e à amizade verdadeira de meu irmão mais velho.

ACKNOWLEDGEMENTS

They say that two years is enough time to make relatively shocking changes. In concluding this work, I come across a different self than I was. A little more callous and knowing of maybe something new, since I knew this area, I already wanted to be a profound student of the same, perhaps because it is the exact science that is inexact or because it is one of the few sciences that with its own alphabet allows the conversation with any other area. Two years ago I was a newly formed, full of dreams and fears about the academic world. A master's degree in a different area of mine would require a lot of focus from me, this focus being able to take my own spirit of orbit while ruining my body and mind. However here I arrived, here I came. I certainly did not do it in the best way, in the mildest way, in the most knowledgeable way or even in the most recommended way. However, I carry the peace of mind that I came here in my own way. Leaving my house, my family, my friends, my land ... The price I paid was greater than I thought, precisely because I was not ready to face the adversities that happened on the side of here. However, even if so many windstorms passes by and tear their hearts, there are people which made this work possible, without them, this work would not make sense or simply would not be mine.

The great woman who gave me life, persistence, and a strong temper. Maria Lucia Vidal de Oliveira, my mother, thank you so much for the many sacrifices you have already made for me. Thank you for every call, every day I've been here in Recife, thank you for always trying to pull my laughter, even on days that my only desire was to plunge into deep silence. Thank you for pampering me even now, as an adult. Such coldness in the world can only be neutralized by the sweetness of your smile and the simplicity of you gaze. Our personality has given me problems, but they have always guided me to defend myself from unpleasant or outrageous situations. Thank you for advising me uninterruptedly because the blindness of my youth sometimes puts me in abysses and evils that only the wisdom of your maturity and attention allow the deviation. I love you.

José Valdenir de Oliveira, my father, thank you for pulling me in the balance, for always making me aware that from the guard to the janitor, from the receptionist to the teacher, everyone is human and the same. Not needing fleshy words, vague words or inattentive moralism. Thank you for always reminding me that I must be in my own way, as transparent as possible, but always bearing in mind that not everyone is showing what one really is. Thank you for your friend and attentive shoulder every day of my stay here, thank you for always trying to dissolve my pain with you and always try to encourage me to get up once more. Thank you for giving me faith, for

sustaining me and for always being there with your friendly embrace and your frank counsels. Thank you for all the times you donated your blood and sweat just to see me performing or approaching to realize a dream. I love you.

I would also like to thank and dedicate this work to my sister, Dayana, recovering the bond with you was the best gift I could receive during my master's degree.

It is not possible to walk without the help of friends, in my second family I emphasize here the bond of brotherhood that was created in the deepest way. Nathan Carneiro Parente, Aniki, thank you for every belief in my potential, for the incisive counsels and holders of care, and zeal, but always being true with you and me. Thank you for supporting my crises and eating salt with me all these years. I have immeasurable pride in calling you and saying: My brother, I love you. Giselle, nee-chan, even if our dates are always intermittent, when you are here, you are always attentive, affectionate and extremely truthful. My advice is always too careful, but it is not often that you advise me, and therefore we are a flow. My sister, I love you. Daniel, Dan, the one who came last and made to be an extremely philosophical and fluid genius, A friend with whom I know that in his way he will always have a way of communicating with me, through laughter, music, reflection or even Even for silence, I love you, little brother.

There is no way to strengthen the basis of the mind without strengthening the basis of body and spirit. To this I thank Moy Fah, my beloved and esteemed school of Kung Fu Wushu Shaolin of the North and my master Shifu Giovani Nunes. I also thank all my brothers and sisters at the school, Rafael, Camila, Aspasia, Hyggor, Clara, Marcus, Thai, Glenda, Átila, Wesley, Emerson, Yane, Cassia, Lucas, Victor, San, Elton and more. Fah!

I would also like to thank the friends of UFC times, the random pack: Monalisa, Isabelle, Yolanda, Joana, Pedro, Kennedy, Raimundo, Anderson, Venícius, Diaulas, and Laresca. As well as thanking Claudia, Erione, Valdi, Taty, and Edleusa.

I would like to thank the few, but many, friends I made here during my stay: Vinicius Scher, Kalline Costa, Nayara Luíza, Henrique Santos, Yuri Martí, Eduarda Sousa, Érica Nogueira, Lucas de Oliveira, Elmon Noronha, Abel Borges, Anabeth Radünz, Eduardo Ensslin, Rommy Camasca, Fernando Maia and César Diogo.

I would like to thank the teachings of my undergraduate professors, especially the affectionate memory of them: Professor Diego, Professor Pablo, Professor Ana Maria, Professor Jaqueline, Professor Ricardo, professor Ronald. Master's professors, Professor Alex, Professor Audrey, Professor Getúlio, Professor Gauss, Professor Fernanda.

In particular, I would like to thank my illustrious advisor Professor Francisco Cribari-Neto, who

proved to be not only an intelligent teacher but also a wise teacher. Thanks for the patience and teacher teachings! And to my teacher Professor Juvêncio Santos Nobre, for whom I share a quasi-paternal relationship and that I consider a friend. Thanks for all the help and for accompanying me since my graduation.

I would like to thank Valéria, a strong-minded woman who always has her best smile and helps us all students. To Stefano who always looked after all the technical support of our department and to the teachers Professor Cysneiros and Professor Raydonal, who coordinated the program during the years I spent here.

I thank Recife, Pernambuco and every citizen of this land who welcomed me.

I would like to thank CAPES for financial support.

ABSTRACT

In this dissertation, we consider an important class of regression models, namely: the class of generalized extreme value nonlinear regression models. Such models are commonly used in many fields to model extremal events. The main model foundations involve extreme value theory, which provides underlying laws for scenarios in which the data may contain atypical observations which results from the phenomenon of interest and not the result of measurement or recording error. In particular, we develop residual based diagnostic analysis, local influence analysis, generalized Cook's distance and generalized leverage for the generalized extreme value nonlinear regression model. Since the expected value of the dependent variable is determined by the two parameters that index the distribution, we model each parameter separately and also both parameters jointly, thus considering three possible scenarios. Additionally, we present a model misspecification test that can be used to determine whether the fitted model is incorrectly specified. We provide Monte Carlo simulation results on the finite sample behavior of the test. The results show that the test performs well both in terms of size and power. The size simulations were performed by generating the data from the postulated model whereas in the power simulations the fitted model is different from that used for data generation. The local influence analysis is carried out using three different perturbation schemes. We show that the diagnostic procedures that focus on the scale parameter are typically less stable and more computationally challenging than that on the other model parameter. We also propose two residuals for use with the model: the standardized and deviance residuals. Empirical applications based on simulated and observed data are presented and discussed. All numerical results were obtained using the Julia programming language.

Keywords: Extreme value theory. Influence analysis. Misspecification test. Nonlinear regression. Outlier.

RESUMO

A presente dissertação considera uma importante classe de modelos de regressão, a saber: a classe de modelos de regressão generalizados de valores extremos não-linear. Esses modelos são comumente utilizados em diversos campos do conhecimento para modelar eventos extremos. A fundamentação principal do modelo envolve a teoria de valores extremos, que propõe técnicas de modelagem a serem usadas em cenários em que os dados podem conter observações atípicas, resultantes do fenômeno de interesse e não de erro de medição. Em particular, na presente dissertação, nós desenvolvemos análise de diagnóstico baseada em resíduos, análise de influência local, distância de Cook generalizada e alavancagem generalizada para o modelo de regressão generalizado de valores extremos não-linear. Uma vez que o valor esperado da variável dependente é determinado pelos dois parâmetros que compõem a distribuição, modelamos cada um dos parâmetros separadamente e também conjuntamente, considerando, assim, três possíveis cenários. Também apresentamos um teste de especificação correta. A hipótese nula é a de que o modelo está corretamente especificado e a hipótese alternativa é a de que a especificação do modelo está incorreta. Apresentamos resultados de simulação de Monte Carlo que mostram que o teste proposto funciona bem em amostras finitas, apresentando baixas distorções de tamanho e poder elevado. As simulações de tamanho foram realizadas gerando-se os dados do modelo postulado, enquanto que nas simulações de poder o modelo ajustado difere do modelo do qual os dados foram gerados. A análise de influência local é desenvolvida a partir de três esquemas distintos de perturbação dos dados. Mostramos que as técnicas de diagnóstico que focam no parâmetro de escala são tipicamente menos estáveis e mais árduas computacionalmente que as que focam no outro parâmetro. Dois novos resíduos são também propostos, a saber: o resíduo padronizado e o resíduo desvio. Aplicações empíricas baseadas em dados simulados e reais são apresentadas e discutidas. Todos os resultados numéricos foram obtidos utilizando a linguagem de programação Julia.

Palavras-chave: Análise de influência. Outlier. Regressão Não-Linear. Teoria de valores extremos. Teste de especificação incorreta.

LIST OF FIGURES

LIST OF TABLES

CONTENTS

1 ESTIMATION AND INFERENCE IN GENERALIZED EXTREME VALUE MODELS

1.1 INTRODUCTION

Extreme value theory is an important tool for modeling certain phenomena and has been gaining attention in recent years. (COLES; PERICCHI; SISSON, 2003) discuss the growing dissatisfaction with the standard methodology for modeling extreme values. The classic approach consists of adopting an asymptotic model to describe stochastic variation at extreme occurrences of a process. It ignores possible errors at interpretation and may lead to inaccurate inferences. Also, the standard asymptotic theory is not valid, for example, for studying extreme quantiles. In several applications, the interest lies in the maximum (or minimum) sample value, which renders the distribution of such a random variable important. Additionally, extreme value theory is useful when the sample contains an atypical observation, which is considered to be a faithful realization of the phenomenon under study, and not the result of a data recording error. The main idea of (GUMBEL, 1935) was to propose three distributions for such random variables that are valid as the number of data points increases.

Most of the works that make use of extreme values distribution aim at modeling natural phenomena: flood frequency (ROSSI; FIORENTINO; VERSACE, 1984; WAYLEN; WOO, 1982), wind load (SIMIU *et al.*, 2001), industrial management (WANG; DEY, 2010; CALABRESE; OSMETTI, 2013). There are also applications in medicine (PARK; SOHN, 2006). (COLES; PERICCHI; SISSON, 2003) showed the improvements when the extreme value theory is adopted instead of the classical approach. There are also applications for mixture of extreme values distribution (WAYLEN; WOO, 1982; YUE *et al.*, 1999).

Let $Y_1, Y_2, \ldots, Y_n$ be a sequence of independent and identically distributed random variables with distribution $F_Y(y)$ and let $M_n = \max(Y_{(1)}, Y_{(2)}, Y_{(3)}, \ldots, Y_{(n)})$. The distribution for M_n is given by $F_{M_n}(y) = F_Y(y)^n$. The extreme value distribution G arises when it is possible to obtain real-valued sequences a_n and b_n such that

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq y\right) = F_Y(a_n y + b_n)^n \to G(y). \tag{1.1}$$

This property is called 'max stability'. The limiting distribution may assume three forms, namely

- Gumbel (Extreme Value type I):

$$F(y) = e^{-e^{-\frac{y-\mu}{\sigma}}}, \qquad y \in \mathbb{R};$$

- Fréchet (Extreme Value type II):

$$F(y) = e^{-\left(\frac{y-\mu}{\sigma}\right)^{-\varepsilon}}, \qquad \varepsilon > 0 \text{ and } y \in (\mu - \sigma/\varepsilon, \infty);$$

- Reversed Weibull (Extreme Value type III):

$$F(y) = e^{-\left(\frac{\mu-y}{\sigma}\right)^{\varepsilon}}, \qquad \varepsilon < 0 \text{ and } y \in (-\infty, \mu - \sigma/\varepsilon).$$

The parameters μ,σ,ε are, respectively, location, scale and shape parameters. Here, $\mu \in \mathbb{R}$, $\sigma > 0$. Figure 1 shows density curves for some parameters values for GEV distribution, illustrating the extreme values distributions presented in this section. (LITTELL, 2006) shows that the corresponding sequences in (1.1) for each member of extreme value distribution family are:

$$\begin{aligned} \text{I} &: a_n = 1 \text{ and } b_n = -\log n, \\ \text{II} &: a_n = n^{-1/\varepsilon} \text{ and } b_n = 0, \\ \text{III} &: a_n = n^{1/\varepsilon} \;\; b_n = 0. \end{aligned}$$

It is noteworthy that extreme value theory can be applied to minimal values after a change of sign $(-Y)$. In that case, it is possible to obtain the distributions of minimal values (KOTZ; NADARAJAH, 2000).

The Gumbel distribution is more often used than the Fréchet distribution or the reversed Weibull distribution. A decisive factor in the choice between such distributions is the difficulty in estimating the parameter ε. In Fréchet or reversed Weibull cases, the estimation of ε usually occurs by choosing values in the interval $[-0.5, 0) \cup (0, 0.5]$.

Extreme value modeling has seen many improvements since Gumbel's original proposal. The generalized extreme value (GEV) distribution proposed by (JENKINS, 1955) covers the three distributions previously mentioned, by considering restrictions on the parameter space. Other noteworthy developments on the GEV distribution are some criteria that can be

used to select one of the three extreme value distributions (HOSKING, 1984) and a multivariate extension proposed by (ESCALANTE-SANDOUAL, 1998).

This chapter unfolds as follows. Section 1.2 presents the generalized extreme value distribution. We show different parameters estimation approaches in Section 1.3. Section 1.4 presents a useful hypothesis test for extreme value distribution selection. Some empirical applications based on observed and simulated data are presented in Section 1.5. Finally, some concluding remarks are presented in Section 1.6.

1.2 THE GENERALIZED EXTREME VALUE (GEV) DISTRIBUTION

Initially introduced by (JENKINS, 1955), the GEV distribution covers the three main extreme value distributions. The cumulative probability function (cdf) and the probability density function (pdf) of a random variable Y that follows the $GEV(\mu,\sigma,\varepsilon)$ law are given, respectively, by

$$F(y)=\begin{cases} e^{-1+\varepsilon((y-\mu)/\sigma))^{-1/\varepsilon}}, & I_{(-\infty,\mu-\sigma/\varepsilon]}(y) \quad \text{for} \quad \varepsilon<0, \\ & I_{[\mu-\sigma/\varepsilon,\infty)}(y) \quad \text{for} \quad \varepsilon>0, \\ e^{-e^{-(y-\mu)/\sigma}}, & I_{(-\infty,\infty)}(y) \quad \text{for} \quad \varepsilon=0, \end{cases}$$

and

$$f(y)=\begin{cases} e^{-(1+\varepsilon((y-\mu)/\sigma))^{-1/\varepsilon}}\frac{1}{\sigma}\left\{1+\varepsilon\left(\frac{y-\mu}{\sigma}\right)\right\}^{-\frac{1}{\varepsilon}-1}\times & I_{(-\infty,\mu-\sigma/\varepsilon]}(y) \quad \text{for} \quad \varepsilon<0, \\ & \times \quad I_{[\mu-\sigma/\varepsilon,\infty)}(y) \quad \text{for} \quad \varepsilon>0, \\ e^{-e^{-(y-\mu)/\sigma}}\frac{1}{\sigma}e^{-(y-\mu)/\sigma}\times & I_{(-\infty,\infty)}(y) \quad \text{for} \quad \varepsilon=0. \end{cases}$$

When $\varepsilon=0$ the GEV distribution reduces to the Gumbel distribution. The other two distributions are the Fréchet distribution, which corresponds to $\varepsilon>0$, and the reversed Weibull distribution, which corresponds to $\varepsilon<0$. (KOTZ; NADARAJAH, 2000) showed that the standard GEV distribution can be obtained by taking $z=(y-\mu)/\sigma$ and making the proper changes to the parameter space.

The value of the parameter ε determines most of common statistical measures such as, for example, the mean and the variance:

$$\mathbb{E}(Y)=\begin{cases} \mu+\sigma(\Gamma(1-\varepsilon)-1)/\varepsilon & \text{if} \quad \varepsilon<1, \\ \mu+\sigma\gamma & \text{if} \quad \varepsilon=0, \\ \infty & \text{if} \quad \varepsilon\geq 1, \end{cases}$$

and

$$\mathrm{Var}(Y) = \begin{cases} \sigma^2(\Gamma(1-2\varepsilon) - (\Gamma(1-\varepsilon))^2)/\varepsilon^2 & \text{if} \quad \varepsilon < 1/2, \\ \frac{\pi^2\sigma^2}{6} & \text{if} \quad \varepsilon = 0, \\ \infty & \text{if} \quad \varepsilon \geq 1/2. \end{cases}$$

Here, γ is the Euler constant, which is approximately equal to 0.5772. Several restrictions on GEV exists for restricted values of ε. When $\varepsilon > 1/2$ the second and higher moments do not exist.

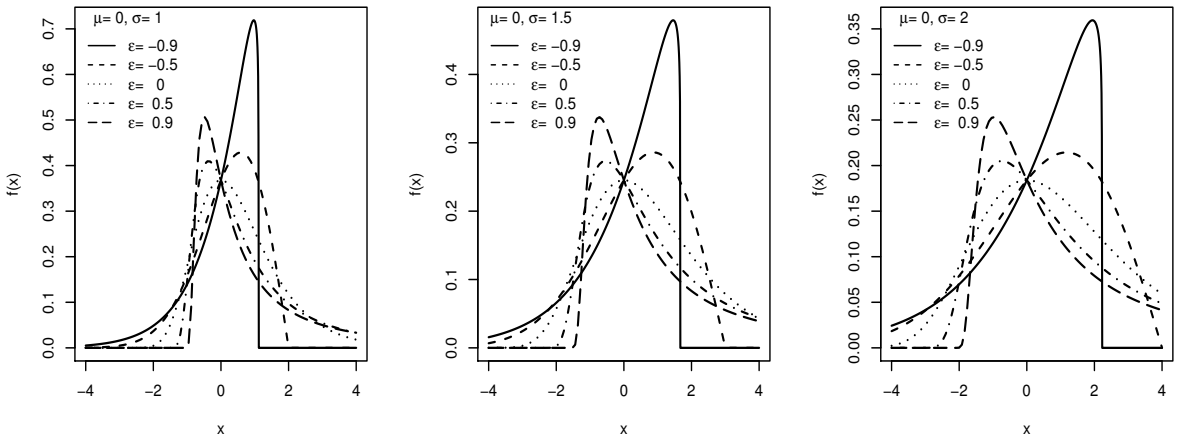

Figure 1 – GEV $(\mu, \sigma, \varepsilon)$ densities for some parameter values.

When the shape parameter equals zero, the GEV distribution is closed by basic constant operations (sum and product), which means that if $Y \sim GEV(\mu, \sigma, 0)$, then $Z = aY + b$ is distributed as $GEV(a\mu + b, a\sigma, 0)$.

The GEV distribution (in the Gumbel case) is closely related to the logistic distribution. Let $Y_1, Y_2 \sim GEV(\mu, \sigma, 0)$. Using, for example, the moment generating function we obtain that

$$Y_1 - Y_2 \sim Logistic(0, \sigma).$$

(III, 1975) showed that the generalized Pareto distribution can be used as an alternative to the GEV law. This distribution covers the Pareto, exponential and uniform distributions. (BALI, 2003) generalized both generalized distributions by using the Box-Cox transformation formula, thus arriving at the Box-Cox-generalized extreme value distribution.

The maximum likelihood estimators of the parameters that index the GEV distribution cannot be expressed in closed-form. Estimates can be obtained by using numerical methods. The estimation presents some problems for ε, due to parametric restrictions. (LITTELL, 2006) shows that if $\varepsilon < -1$ the maximum likelihood estimator does not exist; when $-1 < \varepsilon < 0$ and $0 < \varepsilon < 1$, the estimation is difficult since the contour curves show flat regions as in Figure 2. Finally, when $\varepsilon = 0$ the curves are more defined, which renders the directions produced by quasi-Newton methods more accurate, thus making estimation process easier.

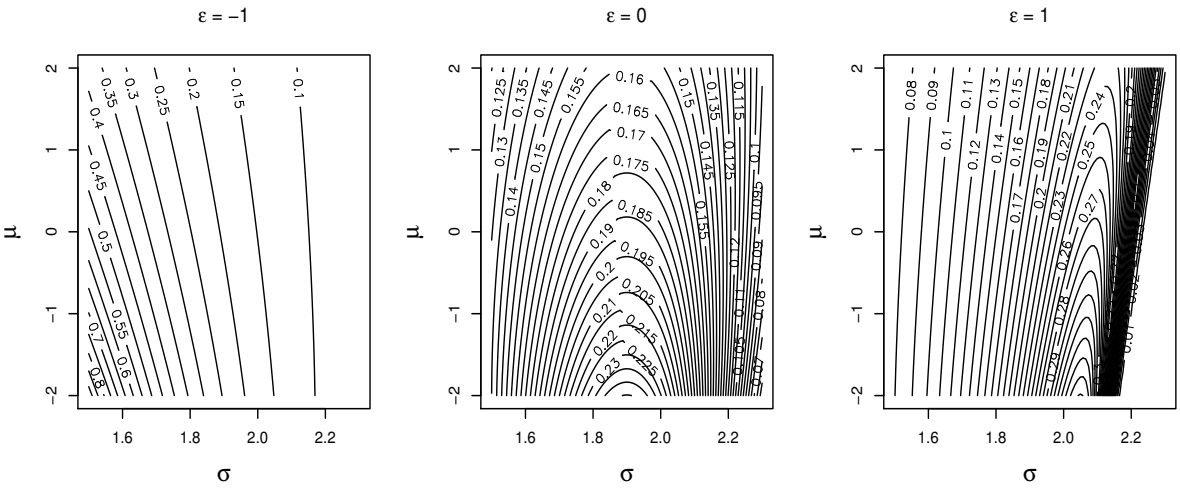

Figure 2 – Contour curves of the GEV profile likelihood for μ and σ.

The contour curves for the Gumbel distribution ($\varepsilon = 0$) show a regular scenario with well defined curves. However, the profile likelihood function shows a region apparently flat for critical values of ε, which may make the optimization process fail to converge.

(LITTELL, 2006) notes that most papers use values between $-1/2$ and $1/2$, thus avoiding numerical optimization instabilities. However, this difficulty does not make the GEV distribution useless,since the generalize extreme value distribution can explain certain phenomena better than the Gumbel distribution as shown by (COLES; PERICCHI; SISSON, 2003).

Probability weighted moments estimation for the GEV distribution (HOSKING; WALLIS; WOOD, 1985) is a good alternative to maximum likelihood estimation for small sample sizes and when information on ε is not available. Other good alternatives are the method of moments, the ranked set estimation method (KOTZ; NADARAJAH, 2000) and generalized least squares (PARK; SOHN, 2006).

1.3 GEV PARAMETER ESTIMATION

1.3.1 The maximum likelihood method

The maximum likelihood estimator enjoys, under regularity conditions (SEN; SINGER; LIMA, 2009), desirable asymptotic properties: consistency, asymptotic efficiency and asymptotic normality. Let $\mathbf{Y} = (y_1, y_2, \ldots, y_n)^\top$ be a vector of independent and identically distributed random variables, each following the generalized extreme value distribution. Let $d_i = \varepsilon(y_i - \mu)/\sigma$. The log-likelihood function is given by

$$\ell(\mu, \sigma, \varepsilon | \mathbf{y}) = -\sum_{i=1}^{n} (1+d_i)^{-\frac{1}{\varepsilon}} - n\log(\sigma) + \left(-\frac{1}{\varepsilon} - 1\right) \sum_{i=1}^{n} \log(1+d_i).$$

Differentiating the log-likelihood function with respect to each parameter, we obtain the score function $U_\nu = (U_\mu, U_\sigma, U_\varepsilon)^\top$, where

$$\begin{aligned}
U_\mu &= \frac{\partial \ell}{\partial \mu} = -\frac{1}{\sigma} \sum_{i=1}^{n} (1+d_i)^{-\frac{1}{\varepsilon}-1} + \left(\frac{1+\varepsilon}{\sigma}\right) \sum_{i=1}^{n} \frac{1}{1+d_i}, \\
U_\sigma &= \frac{\partial \ell}{\partial \sigma} = \sum_{i=1}^{n} \frac{d_i}{\varepsilon\sigma} (1+d_i)^{-\frac{1}{\varepsilon}-1} - \frac{n}{\sigma} + \left(-\frac{1}{\varepsilon} - 1\right) \sum_{i=1}^{n} \frac{-d_i}{\sigma(1+d_i)}, \\
U_\varepsilon &= \frac{\partial \ell}{\partial \varepsilon} = \sum_{i=1}^{n} -(1+d_i)^{\frac{1}{\varepsilon}-1} \left[(1+d_i)\frac{1}{\varepsilon^2}\log(1+d_i) + \frac{1}{\varepsilon}\left(\frac{d_i}{\varepsilon}\right)\right] + \\
&\quad + \frac{1}{\varepsilon^2} \sum_{i=1}^{n} \log(1+d_i) + \left(-\frac{1}{\varepsilon} - 1\right) \sum_{i=1}^{n} \frac{d_i}{\varepsilon(1+d_i)}.
\end{aligned}$$

The above system of equations cannot be solved analytically. Parameter estimates can be obtained by numerically maximizing the log-likelihood function using a Newton or a quasi-Newton method. Some numerical methods require the Hessian matrix or Fisher's information matrix, such as, for example the Newton-Raphson method and Fisher's scoring method. Both matrices can be found, for example, in (SHI, 1995). Some references on maximum likelihood estimation on GEV distribution are (PRESCOTT; WALDEN, 1980) and (HOSKING; WALLIS; WOOD, 1985), who propose initial values for the parameters when using numerical optimization methods.

A GEV Bayesian estimation approach was outlined by (MARTINS; STEDINGER, 2000). The authors use the beta distribution as the a priori distribution for the shape parameter, thus avoiding some computational problems, such as convergence by local maximum or non-convergence. (PRESCOTT; WALDEN, 1983) consider GEV parameter estimation under censoring.

1.3.2 The method of moments

The method of moments estimators can be obtained as the solution to a system of equations obtained by equating population and sample moments. Oftentimes, this solution is explicit, in contrast to the maximum likelihood method. However, the method of moments estimators does not enjoy the same proprieties of the maximum likelihood estimator. In particular, it shows a lack of efficiency.

(MARTINS; STEDINGER, 2000) showed that the system of equations that is used in generalized extreme value estimation is

$$\begin{aligned}
\tilde{\mu} &= \overline{Y} - \frac{\tilde{\sigma}}{\tilde{\varepsilon}}\{1-\Gamma(1+\tilde{\varepsilon})\},\\
\tilde{\sigma} &= \frac{S|\tilde{\varepsilon}|}{\{\Gamma(1+2\tilde{\varepsilon})-[\Gamma(1-\tilde{\varepsilon})]^2\}^{1/2}},\\
\kappa &= \text{sign}(\tilde{\varepsilon})\frac{-\Gamma(1+3\tilde{\varepsilon})+3\Gamma(1+\tilde{\varepsilon})\Gamma(1+2\tilde{\varepsilon})-2[\Gamma(1+\tilde{\varepsilon})]^3}{\{\Gamma(1+2\tilde{\varepsilon})-[\Gamma(1-\tilde{\varepsilon})]^2\}^{3/2}},
\end{aligned}$$

where $\overline{Y}$, S and κ are, respectively, the sample mean, the standard deviation and the sample skewness and $\tilde{\mu}$, $\tilde{\sigma}$ and $\tilde{\varepsilon}$ are the method of moments estimators of μ, σ and ε, respectively. The last equation does not admit a explicit solution, requiring an iterative method. Recall that for certain values on ε the moments are not finite. Consider the GEV type I (Gumbel) distribution. When the shape parameter equals zero, (KOTZ; NADARAJAH, 2000) showed that

$$\tilde{\sigma} = \frac{\sqrt{6}}{\pi}S \quad \text{and} \quad \tilde{\mu} = \overline{Y} - \gamma\tilde{\sigma}.$$

By comparing the variance formulas proposed by (OLIVEIRA, 1963) with the Cramér-Rao lower bounds (CORSINI *et al.*, 1995), (KOTZ; NADARAJAH, 2000) concluded that the estimator for μ has about 95% efficiency whereas the estimator for σ has efficiency of only 55%. For the Fréchet distribution, the system of equations used with the method of moments is presented by (SINGH, 1998). An alternative was given by (HILL, 1975), which presented a estimator for shape parameter when it is known that $\varepsilon > 0$. The Hill estimator is given by

$$M_n^{(1)} = \frac{1}{k}\sum_{i=0}^{k-1} \log y_{n-i} - \log y_{n-k} \qquad (k < n),$$

where $k = k(n)$ is a sequence of integer numbers, such as $y_{n-k} \xrightarrow{\mathbb{P}} n/k$. Based on Hill estimator, (DEKKERS; EINMAHL; HAAN, 1989) proposed an alternative estimation strategy. The estimator is given by

$$\tilde{\varepsilon} = M_n^{(1)} + \frac{1}{2}\left\{1 - \frac{(M_n^{(1)})^2}{M_n^{(2)}}\right\}^{-1},$$

where

$$M_n^{(j)} = \frac{1}{k}\sum_{i=0}^{k-1} (\log Y_{(n-i)} - \log Y_{(n-k)})^j$$

The authors list the sufficient conditions for consistency. Assume the max stability property. Then:

- If $y^*(F) > 0$, $k(n)/n \to 0$, and $\lim_{n\to\infty} k(n) \to \infty$, then $\lim_{n\to\infty} \tilde{\varepsilon} \xrightarrow{\mathbb{P}} \varepsilon$.
- If $y^*(F) > 0$, $k(n)/n \to 0$, and $\lim_{n\to\infty} k(n)/(\log n)^\delta \to \infty$ for some $\delta > 0$, then $\lim_{n\to\infty} \tilde{\varepsilon} \xrightarrow{\text{a. s}} \varepsilon$.

Here, $y^*(F) = \sup\{y|F(y) < 1\}$ and $y^*(F) > 0$. If such conditions are satisfied, then $\tilde{\varepsilon}$ can be used free of parametric restrictions.

1.3.3 The probability-weighted moments method

The probability-weighted moment (PWM) of a random variable Y with cdf $F(\cdot)$ is given by

$$M_{p,r,s} = \mathbb{E}[Y^p\{F(Y)\}^r\{1-F(Y)\}^s],$$

where p, r and s are real numbers. PWM's are particularly useful when the quantile function $y(F)$ can be expressed in closed-form. It has some particular cases when r or s can be equal to zero. When both r and s equal zero $M_{p,r,s}$ reduces to $\mathbb{E}[Y^p]$. Based on the probability-weighted moments, (GREENWOOD *et al.*, 1979) proposed the probability-weighted moments method.

This method is similar to the moments method but the usual sample moments are replaced by probability weighted moments.

In order to use the method proposed, we consider the moment $\tau_r = M_{1,r,0}$ which is estimated by

$$\hat{\tau}_r[p_{i,n}] = \frac{\sum_{i=1}^{n} p_{i,n}^r y_{(i)}}{n}.$$

$p_{i,n}$ is a plotting position, a distribution-free estimate of $F(y_i)$. Reasonable candidates for $p_{i,n}$ are $(i-a)/n$, $0 < a < 1$, $(i-a)/(n+1-2a)$, $-0.5 < a < 0.5$, and

$$p_{i,n} = \frac{i-1}{n-1}.$$

The method consists of evaluating the first p expressions for τ (for a distribution with p parameters), constructing a system of equations, and then solving it to obtain parameters expressions that are functions of $\tau_0, \ldots, \tau_{p-1}$.

For the generalized extreme value distribution, the inverse distribution function (quantile function) is

$$y(F) = \mu + \sigma\{1-(-\log F)^{\varepsilon}\}/\varepsilon \quad \text{when} \quad \varepsilon \neq 0.$$

If $\varepsilon = 0$, we obtain

$$y(F) = \mu - \sigma(-\log F).$$

The GEV rth probability-weighted moment for $\varepsilon \neq 0$ is given by

$$\tau_r = \frac{\mu + \sigma\{1-(r+1)^{-\varepsilon}\Gamma(1+\varepsilon)\}/\varepsilon}{r+1}, \quad \varepsilon > -1.$$

The restriction $\varepsilon > -1$ is adopted to avoid the nonexistence of moments. It follows that

$$\begin{aligned} \tau_0 &= \mu + (\sigma\{1-\Gamma(1+\varepsilon)\}/\varepsilon) \\ 2\tau_1 - \tau_0 &= \sigma\Gamma(1+\varepsilon)(1-2^{-\varepsilon})/\varepsilon \\ \frac{3\tau_2 - \tau_0}{2\tau_1 - \tau_0} &= \frac{1-3^{-\varepsilon}}{1-2^{-\varepsilon}}. \end{aligned}$$

The last equation requires the use of an iterative method to be solved. However, the term on the right hand side is almost linear and can be approximated by a low-order polynomial function. By doing so, we obtain (HOSKING; WALLIS; WOOD, 1985)

$$\ddot{\mu} = \hat{\tau}_0 + \ddot{\sigma}\{\Gamma(1+\ddot{\varepsilon}) - 1\}/\ddot{\varepsilon},$$
$$\ddot{\sigma} = \frac{(2\hat{\tau}_1 - \hat{\tau}_0)\ddot{\varepsilon}}{\Gamma(1+\ddot{\varepsilon})(1-2^{-\ddot{\varepsilon}})},$$
$$\ddot{\varepsilon} = 7.8590c + 2.9554c^2,$$

where

$$c = \frac{2\hat{\tau}_1 - \hat{\tau}_0}{3\hat{\tau}_2 - \hat{\tau}_0} - \frac{\log 2}{\log 3}$$

and $\ddot{\mu}$, $\ddot{\sigma}$ and $\ddot{\varepsilon}$ are the probability-weighted moments estimators of μ, σ and ε, respectively. Using the quantile function for the case where $\varepsilon = 0$, i.e., the Gumbel distribution quantile function, the estimators are

$$\ddot{\mu} = \hat{\tau}_0 - \gamma\ddot{\sigma},$$
$$\ddot{\sigma} = \frac{2\hat{\tau}_1 - \hat{\tau}_0}{\log(2)}.$$

The results in (CHERNOFF; GASTWIRTH; JOHNS, 1967) may be used to prove that the vector of estimators $(\ddot{\mu}, \ddot{\sigma}, \ddot{\varepsilon})$ is asymptotically normally distributed. The probability-weighted moments method is recommended when the sample size is small. (HOSKING; WALLIS; WOOD, 1985) presented an application which makes use of MLE inaccurate in empirical applications, due to nonregularity of the log-likelihood function, causing ocasional nonconvergence of Newton-Raphson method . Occasionally, the probability-weighted moments estimator offers smaller standard deviations for estimators than maximum likelihood estimators.

1.4 GEV HYPOTHESIS TESTING INFERENCES

It is sometimes useful to check whether the assumed distribution is valid. Most of studies find that the generalized extreme value distribution is more adequate than another arbitrary continuous distribution (F). For this test the hypothesis are

$H_0 : F$ is a member of the GEV class of distributions

$H_1 : F$ is not a member of the GEV class of distributions.

When the null hypothesis is not rejected, the researcher may proceed to the next test, which is used to identify which distribution (Gumbel, Fréchet or reversed Weibull) is most appropriate.

Let $\mathbf{Y} = (y_1, y_2, \ldots, y_n)^*$ be a random sample. The test statistic proposed by (KOTZ; NADARAJAH, 2000) is

$$\phi_n(\mathbf{Y}) = \sqrt{n} \min_{\text{a,b}} \max_{\text{y}} |F_n(y) - F_n^2(ay+b)|,$$

where $F_n(y)$ is the empirical distribution function. The motivation behind this statistic lies in the max stability property; see Equation (1.1). (KOTZ; NADARAJAH, 2000) showed that

$$\mathbb{P}(\phi_n(y) > r) \leq \mathbb{P}\left(\sqrt{n} \min_{n} \max_{y \in \{y_1,\ldots,y_n\}} |U_n(y) - U_n^2(y^a)| > r\right),$$

where U_n denotes the empirical distribution function of a random sample taken from the standard uniform distribution. The authors used the fact that $U(y) = U^2(\sqrt{y})$ and constructed a table of critical values for the test, which can be found in (KOTZ; NADARAJAH, 2000). The test is typically conservative in finite samples thus rejecting the null hypothesis when is true less often than expected for $\varepsilon \geq -1$ (the regular case). The test becomes less size-distorted as σ increases.

When the null hypothesis of the test presented above is not rejected it is useful to perform a separate test in order to identify which particular GEV distribution should be used. (LITTELL, 2006) showed that

Type I (Gumbel): has been applied to meteorological extremes to model extremely high temperatures and predict high return levels of wind speed.

Type II (Fréchet): has been used to estimate probabilities of extreme occurrences on Germany's stock index and to predict behavior of solar proton peak fluxes.

Type III (R.Weibull): has been used to partitioning and floorplanning problems, window glasses and for assessing the magnitude of future earthquakes.

The sign of the shape parameter is the indicator of the distribution type, since $\varepsilon = 0$, $\varepsilon > 0$ and $\varepsilon < 0$ imply types I, II and III respectively. (HOSKING, 1984) introduced six test statistics that

can be used to test $H_0:\ \varepsilon = 0$ (Gumbel) against $H_1:\ \varepsilon \neq 0$ (not Gumbel), namely:

$$
\begin{aligned}
T_{LR} &= 2\{L(y;\hat{\nu}) - L(y;\tilde{\nu})\},\\
T^*_{LR} &= (1 - 2.8/n)T_{LR},\\
T_W &= n^{-1}\hat{\varepsilon}^2/\hat{\mathfrak{F}}^{33},\\
T^*_W &= n^{-1}\hat{\varepsilon}^2/\hat{\mathfrak{J}}^{33},\\
T_{LM} &= n^{-1}\tilde{\mathfrak{F}}^{33}\tilde{d}^2,\\
T^*_{LM} &= n^{-1}\tilde{\mathfrak{J}}^{33}\tilde{d}^2,\\
\tilde{d} &= \sum_{i=1}^{n}\{\hat{z}_i + \hat{z}_i^2/2(\exp(-\hat{z}_i) - 1),
\end{aligned}
$$

where $L(y;\nu)$ is the likelihood function, $\mathfrak{F}^{33}$ and $\mathfrak{J}^{33}$ are the $(3,3)$ elements of Fisher's information matrix and from the observed information matrix, respectively, which are given by (PRESCOTT; WALDEN, 1980), and $z_i = (y_i - \mu)/\sigma$. Hats indicate evaluation under the alternative hypothesis and tildes indicates evaluation under the null hypothesis. The six test statistics presented above are asymptotically equivalent under the null hypothesis. When H_0 is true, they are all asymptotically distributed as χ^1.

It is possible to use, multiplying by the sign function to avoid negative occurrences, $T_{LR}^{1/2}$, ${T^*_{LR}}^{1/2}$, $T_W^{1/2}$, ${T^*_W}^{1/2}$, $T_{LM}^{1/2}$, ${T^*_{LM}}^{1/2}$ to test $H_0:\ \varepsilon = 0$ (Gumbel) against $H_2:\ \varepsilon < 0$ (R. Weibull) or $H_0:\ \varepsilon = 0$ (Gumbel) against $H_3:\ \varepsilon > 0$ (Fréchet), which makes possible to select the exact extreme value distribution. Under null hypothesis, such test statistics have a limiting standard normal distribution. Positive deviations are taken as evidence that $\varepsilon > 0$ and negative deviations are taken as evidence that $\varepsilon < 0$. Standard normal one-tailed critical values may be used accordingly.

The simulation results presented by (HOSKING, 1984) show different results for each alternative hypothesis: For H_1, the test based on $sign(T^*_{LR}){T^*_{LR}}^{1/2}$ is the least biased, and the tests based on T^*_{LM}, T^*_W are liberal; for H_2, the tests that use $sign(T^*_{LR}){T^*_{LR}}^{1/2}$ and $sign(T_{LM}^{1/2})T_{LM}^{1/2}$ have almost the same performance. Finally, for H_3, the test based on the statistic $sign(T^*_{LR}){T^*_{LR}}^{1/2}$ has good performance.

1.5 EMPIRICAL APPLICATIONS

In what follows we shall present some empirical applications. As shown in Figure 2, parameter estimation for type II and III distributions can be problematic. Hence, we shall only consider the Gumbel (i.e., type I) distribution. Parameter estimates are obtained as described

in previous sections. Standard errors are obtained using the parametric bootstrap method with 5000 replications and compared, for the maximum likelihood method, to those obtained from the expected (I_f) and observed (∇_a) information matrices. All computations were carried out using the Julia programming language. (BEZANSON *et al.*, 2017). The data used in this Section can be found in (CASTILLO, 1988).

1.5.1 Precipitation data

We use data on the yearly total precipitation in Philadelphia for the last 40 years, measured in inches. The aim of the analysis relates to drought risk determination. Table 1 contains some descriptive statistics. We note the large kurtosis (leptokurtic distribution) and the negative skewness.

Table 1 – Descriptive statistics, precipitation data.

Min	Max	Median	Mean	Variance	Skewness	Kurtosis
29.340	52.130	41.450	41.380	34.261	-0.278	2.180

Table 2 presents the parameter estimates and their respective standard errors. Note that the three estimates for each parameter are similar.

Table 2 – Estimates (standard errors) for the precipitation data.

$\hat{\mu}$	$\tilde{\mu}$	$\ddot{\mu}$
38.423 (0.95595)	38.745 (0.97264)	38.559 (0.81706)
$\hat{\sigma}$	$\tilde{\sigma}$	$\ddot{\sigma}$
5.717 (0.70678)	4.563 (0.92764)	4.886 (0.70564)

For maximum likelihood estimates, it is possible to compare the standard errors computed by bootstrap to those obtained from Fisher's information matrix and observed information matrix. Such results are in Table 3

Table 3 – Standard errors, precipitation data.

	$\hat{\mu}$	$\hat{\sigma}$
I_f	0.95184	0.70484
∇_a	0.95983	0.66818

Figure 3 contains the data histogram and also the estimated densities obtained using the three estimation methods. It is noteworthy that the density estimates obtained using the method of moments and the probability weighted moments are similar.

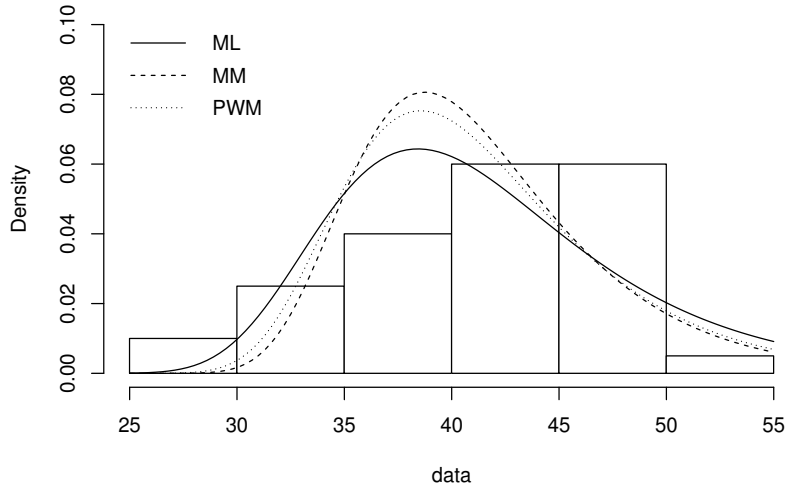

Figure 3 – Histogram of the precipitation data with the fitted densities obtained by the maximum likelihood method (ML), method of moments (MM) and probability weighted moments methods (PWM).

1.5.2 Epicenter data

We shall now model the distances, in miles, to a nuclear power plant of the most recent 8 earthquakes of intensity larger than a given value. The analysis allows one to assess the risk of earthquakes occurring close to the central site. In addition, it is known that a fault is the main cause of earthquakes in the area, and the closest point of the fault is 50 miles.

Table 4 – Descriptive statistics from the epicenter data.

Min	Max	Median	Mean	Variance	Skewness	Kurtosis
58.200	238.900	155.200	144.543	3029.319	-0.034	1.761

Table 4 presents some descriptive statistics. The data variance is large, the skewness is close to zero and the kurtosis is also small. The parameter estimates and their respective standard errors are given in Table 5

Table 5 – Parameter estimates and standard errors, epicenter data.

$\hat{\mu}$	$\tilde{\mu}$	$\ddot{\mu}$
117.286 (6.87559)	119.773 (7.02376)	118.031 (6.40411)
$\hat{\sigma}$	$\tilde{\sigma}$	$\ddot{\sigma}$
49.384 (4.98416)	42.9139 (6.56193)	45.932 (5.37179)

For maximum likelihood estimation, it is possible to obtain standard error from Fisher's information matrix. These results are given in Table 6.

Table 6 – Standard errors for parameters of Gumbel distribution fitting on epicenter data.

	$\hat{\mu}$	$\hat{\sigma}$
I_f	6.71291	4.97092
∇_a	6.75238	4.96471

The data histogram and the three estimated densities are presented in Figure 4. The probability-weighted moments estimated density lies between the other two estimated densities.

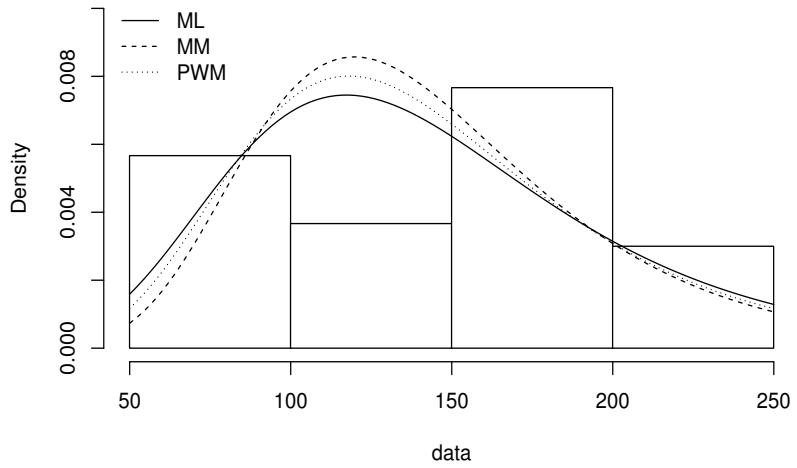

Figure 4 – Histogram of the precipitation data with the fitted densities obtained with maximum likelihood method (ML), method of moments (MM) and probability weighted moments methods (PWM).

1.5.3 Simulated data

In the last application we use simulated data. Using the Mersenne-Twister algorithm implemented in R (R Core Team, 2018), we simulated samples of sizes $n = 20, 40, 60, 80, 100, 150$ and 200 from the standard uniform distribution and them applied the quantile function presented in Section 1.3.3. The parameter values are $\mu = 2$ and $\sigma = 3$.

The seed used was 341. We have used the Julia programming language (BEZANZON *et al.*, 2012) and also the statistical computing enviroment R (R Core Team, 2018) and the object-oriented matrix programming language Ox (DOORNIK, 2007).

For each sample and each language we computed parameter estimates, standard errors using observed and expected Fisher information and time elapsed (in seconds) for each method. In Julia and R we used the L-BFGS-B method, a constrained version of BFGS using a limited amount of computer memory and in Ox we used the BFGS method. The results are presented in Table 7. The initial values are obtained using the method of moments. In all cases, the analytic gradient is used.

In Julia, the algorithms took only two iterations for every sample size to find the maximum likelihood estimate, the time elapsed being higher than that of Ox, but very close. The Ox implementation is the fastest.

The implementation in R is the slowest. For sizes $n = 20, 40, 60$ and 80 the algorithm did not converge. For the other sample sizes, the estimates are similar to those obtained using Julia and Ox.

The values for expected and observed Fisher information matrix are very similar and the estimated standard errors decrease as the sample size increases, improving the accuracy of maximum likelihood method. The precision for Gumbel is higher on scale parameter than the location parameter.

Table 7 – Parameter estimates and standard errors, simulated data using Julia, Ox and R.

		Julia		Ox		R	
		$\hat{\mu}$	$\hat{\sigma}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\mu}$	$\hat{\sigma}$
20	Est	1.25616	3.50121	1.25616	3.50121	-	-
	I_f	0.82433	0.61042	0.82433	0.61042	-	-
	∇_a	0.82768	0.60960	0.82768	0.60960	-	-
	Time	0.005179		0.00057		-	
	Conv	Yes		Yes		No	
40	Est	1.57492	2.70357	1.57492	2.70357	-	-
	I_f	0.45009	0.33329	0.45009	0.33329	-	-
	∇_a	0.45106	0.33033	0.45106	0.33033	-	-
	Time	0.01826		0.00051		-	
	Conv	Yes		Yes		No	
60	Est	1.75830	3.23437	1.75830	3.23437	-	-
	I_f	0.43965	0.32556	0.43965	0.32556	-	-
	∇_a	0.44000	0.32755	0.44000	0.32755	-	-
	Time	0.00143		0.00074		-	
	Conv	Yes		Yes		No	
80	Est	2.18932	3.21092	2.18932	3.21092	-	-
	I_f	0.37799	0.27990	0.37799	0.27990	-	-
	∇_a	0.37887	0.28093	0.37887	0.28093	-	-
	Time	0.00212		0.00137		-	
	Conv	Yes		Yes		No	
100	Est	2.17276	2.91838	2.17276	2.91838	2.18063	2.91635
	I_f	0.30728	0.22754	0.30728	0.22754	0.30707	0.22738
	∇_a	0.30682	0.23125	0.30682	0.23125	0.30661	0.23076
	Time	0.00371		0.00127		0.019	
	Conv	Yes		Yes		52	
150	Est	1.76318	3.20241	1.76318	3.20241	1.79064	3.19094
	I_f	0.27531	0.20387	0.27531	0.20387	0.27433	0.20314
	∇_a	0.27472	0.20874	0.27472	0.20874	0.27376	0.20671
	Time	0.00524		0.00452		0.012	
	Conv	Yes		Yes		52	
200	Est	2.34119	3.00235	2.34119	3.00235	2.34396	2.99850
	I_f	0.22353	0.16552	0.22353	0.16552	0.22324	0.16531
	∇_a	0.22332	0.16529	0.22332	0.16529	0.22304	0.16476
	Time	0.00796		0.00387		0.016	
	Conv	Yes		Yes		52	

2 THE GENERALIZED EXTREME VALUE NONLINEAR REGRESSION MODEL

Let $\boldsymbol{y} = (y_1 \ldots, y_n)^\top$ be a vector of independent random variables such that each y_i, $i = 1, \ldots, n$, follows a Gumbel distribution. Here, $\mathbb{E}(y_i) = \mu_i + \gamma\sigma_i$. (BARRETO-SOUZA; VASCONCELLOS, 2011) proposed the following regression model:

$$g_1(\boldsymbol{\mu}) = \boldsymbol{\eta}_1 = f_1(X, \boldsymbol{\beta}),$$
$$g_2(\boldsymbol{\sigma}) = \boldsymbol{\eta}_2 = f_2(Z, \boldsymbol{\theta}),$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^\top$, $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_n)^\top$, $\boldsymbol{\eta}_1 = (\eta_{11}, \ldots, \eta_{1n})^\top$, $\boldsymbol{\eta}_2 = (\eta_{21}, \ldots, \eta_{2n})^\top$, $g_1(\cdot)$ and $g_2(\cdot)$ are strictly monotonic and twice-differentiable link functions that map $\mathbb{R}$ and $\mathbb{R}^+$, respectively, onto $\mathbb{R}$, $f_1(\cdot;\boldsymbol{\beta})$ and $f_2(\cdot;\boldsymbol{\theta})$ are continuously twice-differentiable functions, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top \in \mathbb{R}^p$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)^\top \in \mathbb{R}^q$ are unknown parameter vectors ($p + q < n$), and X and Z are $n \times p$ and $n \times q$ full column rank matrices not necessarily different.

Estimation of $\boldsymbol{\psi} = (\boldsymbol{\beta}^\top, \boldsymbol{\theta}^\top)^\top$ can be performed by maximum likelihood. The log-likelihood function is given by

$$\ell(\boldsymbol{\psi}) = \ell(\boldsymbol{\beta}, \boldsymbol{\theta}) = -\sum_{i=1}^{n} \log(\sigma_i) - \sum_{i=1}^{n} \frac{y_i - \mu_i}{\sigma_i} - \sum_{i=1}^{n} \exp\left(-\frac{y_i - \mu_i}{\sigma_i}\right).$$

Let $y_i^* = \exp(-y_i/\sigma_i)$, $\mu_i^* = \exp(-\mu_i/\sigma_i)$ and $v_i = -1 + (y_i - \mu_i)[1 - \exp(-(y_i - \mu_i)/\sigma_i)]/\sigma_i$. The components of the score function vector $\boldsymbol{U}_{\boldsymbol{\psi}} = (\boldsymbol{U}_{\boldsymbol{\beta}}^\top, \boldsymbol{U}_{\boldsymbol{\theta}}^\top)^\top$ are given by

$$\boldsymbol{U}_{\beta_j} = \frac{\partial \ell(\boldsymbol{\psi})}{\partial \beta_j} = -\sum_{i=1}^{n} \left(\frac{y_i^* - \mu_i^*}{\mu_i^* \sigma_i}\right) \frac{\partial \mu_i}{\partial \eta_{1i}} \frac{\partial \eta_{1i}}{\partial \beta_j}, \quad j = 1, \ldots, p,$$
$$\boldsymbol{U}_{\theta_j} = \frac{\partial \ell(\boldsymbol{\psi})}{\partial \theta_j} = \sum_{i=1}^{n} \frac{v_i}{\sigma_i} \frac{\partial \sigma_i}{\partial \eta_{2i}} \frac{\partial \eta_{2i}}{\partial \theta_j}, \quad j = 1, \ldots, q.$$

Following (BARRETO-SOUZA; VASCONCELLOS, 2011), we define the $n \times p$ matrix $\tilde{X} = (\partial \eta_{1i}/\partial \beta_j)_{i,j}$, the $n \times q$ matrix $\tilde{S} = (\partial \eta_{2i}/\partial \theta_j)_{i,j}$ and the following $n \times n$ diagonal matrices: $\Omega = -\operatorname{diag}(\mu_i^* \sigma_i)$, $\Sigma = \operatorname{diag}(\sigma_i)$, $M_1 = \operatorname{diag}(d\mu_i/d\eta_{1i})$ and $M_2 = \operatorname{diag}(d\sigma_i/d\eta_{2i})$. We can then write, in matrix notation,

$$\frac{\partial \ell(\boldsymbol{\psi})}{\partial \boldsymbol{\beta}} = \tilde{X}^\top \Omega^{-1} M_1 (\boldsymbol{y}^* - \boldsymbol{\mu}^*) \quad \text{and} \quad \frac{\partial \ell(\boldsymbol{\psi})}{\partial \boldsymbol{\theta}} = \tilde{S}^\top \Sigma^{-1} M_2 \boldsymbol{\nu},$$

where $\boldsymbol{y}^* = (y_1^*, \ldots, y_n^*)^\top$, $\boldsymbol{\mu}^* = (\mu_1^*, \ldots, \mu_n^*)^\top$ and $\boldsymbol{\nu} = (v_1, \ldots, v_n)^\top$.

The maximum likelihood estimator (MLE) of the vector of parameters that index the nonlinear regression model (i.e., of $\boldsymbol{\psi}$), say $\hat{\boldsymbol{\psi}} = (\hat{\boldsymbol{\beta}}^\top, \hat{\boldsymbol{\theta}}^\top)^\top$, is obtained as the solution to $\boldsymbol{U}_{\boldsymbol{\psi}} = \boldsymbol{0}$. The solution to such a system of equations cannot be expressed in closed form.

Parameter estimates can be, however, obtained by numerically maximizing $\ell_{\boldsymbol{\psi}}$ with respect to $\boldsymbol{\psi}$ using a nonlinear optimization algorithm, such as a Newton (e.g., Newton-Raphson) or quasi-Newton (e.g., BFGS) algorithm.

Next, we shall present the model Hessian matrix, $\ell_{\boldsymbol{\psi}\boldsymbol{\psi}}$. Let $d_i = (y_i - \mu_i)/\sigma_i$ and $d_i^* = \exp(-d_i)$. It can be shown that the elements of the Hessian matrix are

$$\frac{\partial^2 \ell(\boldsymbol{\psi})}{\partial \beta_j \partial \beta_l} = \sum_{i=1}^{n} \left(\frac{1-d_i^*}{\sigma_i}\right)\left(\frac{\partial^2 \mu_i}{d^2 \eta_{1i}}\frac{\partial \eta_{1i}}{\partial \beta_l}\frac{\partial \eta_{1i}}{\partial \beta_j} + \frac{\partial \mu_i}{\partial \eta_{1i}}\frac{\partial^2 \eta_{1i}}{\partial \beta_j \partial \beta l}\right) - \frac{d_i^*}{\sigma_i^2}\left(\frac{\partial \mu_i}{\partial \eta_{1i}}\right)^2 \frac{\partial \eta_{1i}}{\partial \beta_j}\frac{\partial \eta_{2i}}{\partial \theta_l},$$

$$\frac{\partial^2 \ell(\boldsymbol{\psi})}{\partial \theta_j \partial \theta_l} = \sum_{i=1}^{n} \left(\frac{d_i - 1 - d_i d_i^*}{\sigma_i}\right)\left(\frac{\partial^2 \sigma_i}{\partial^2 \eta_{2i}}\frac{\partial \eta_{2i}}{\partial \theta_l}\frac{\partial \eta_{2i}}{\partial \theta_j} + \frac{\partial \sigma_i}{\partial \eta_{2i}}\frac{\partial^2 \eta_{2i}}{\partial \theta_j \partial \theta l}\right) + \left(\frac{1 - d_i(2 + d_i^*(d_i - 2\sigma_i))}{\sigma_i^2}\right)$$
$$\times \quad \frac{\partial \sigma_i}{\partial \eta_{2i}}\frac{\partial \eta_{2i}}{\partial \theta_j}\frac{\partial \eta_{2i}}{\partial \theta_l} \quad \text{and}$$

$$\frac{\partial^2 \ell(\boldsymbol{\psi})}{\partial \theta_j \partial \beta_l} = \frac{\partial^2 \ell(\boldsymbol{\psi})}{\partial \beta_l \partial \theta_j} = \sum_{i=1}^{n} \left(\frac{d_i^*(1 + d_i\sigma_i) - 1}{\sigma_i^2}\right)\frac{\partial \mu_i}{\partial \eta_{1i}}\frac{\partial \sigma_i}{\partial \eta_{2i}}\frac{\partial \eta_{1i}}{\partial \beta_l}\frac{\partial \eta_{2i}}{\partial \theta_l}.$$

Let $J_{\boldsymbol{\beta}\boldsymbol{\beta}} = \text{diag}((\partial \mu_i/\partial \eta_{1i})^2/\sigma_i^2)$, $J_{\boldsymbol{\theta}\boldsymbol{\beta}} = J_{\boldsymbol{\beta}\boldsymbol{\theta}} = \text{diag}((\gamma - 1)(\partial \mu_i/\partial \eta_{1i})(\partial \sigma_i/\partial \eta_{2i})/\sigma_i^2)$ and $J_{\boldsymbol{\theta}\boldsymbol{\theta}} = \text{diag}((\Gamma^{(2)}(2) + 1)(\partial \sigma_i/\partial \eta_{2i})^2/\sigma_i^2)$, where $\Gamma^{(2)}(\cdot)$ denotes the gamma function second derivative. (BARRETO-SOUZA; VASCONCELLOS, 2011) showed that Fisher's information matrix is given by

$$I_{\boldsymbol{\psi}\boldsymbol{\psi}} = \begin{bmatrix} \tilde{X}^\top J_{\boldsymbol{\beta}\boldsymbol{\beta}} \tilde{X} & \tilde{X}^\top J_{\boldsymbol{\beta}\boldsymbol{\theta}} \tilde{S} \\ \tilde{S}^\top J_{\boldsymbol{\theta}\boldsymbol{\beta}} \tilde{X} & \tilde{S}^\top J_{\boldsymbol{\theta}\boldsymbol{\theta}} \tilde{S} \end{bmatrix}.$$

When n is large, $\hat{\boldsymbol{\psi}}$ is approximately distributed as $N_{p+q}(\boldsymbol{\psi}, I_{\boldsymbol{\psi}\boldsymbol{\psi}}^{-1})$. Such an approximation can be used when performing interval estimation and hypothesis testing inferences.

2.1 DIAGNOSTIC ANALYSIS

Diagnostics analysis tools allow practitioners to determine whether the fitted regression model is valid, i.e., whether it provides a good representation to the data, and whether the sample contains atypical/influential observations. Ideally, the point estimates should not be highly sensitive to small data perturbations. In particular, such tools can be used to determine whether the relevant model assumptions hold. In what follows, we shall develop a set of diagnostic analysis tools for the generalized extreme value nonlinear regression model.

2.1.1 Residual analysis

Residuals measure the discrepancy between observed response values and the corresponding model fitted values. They can be used for detecting atypical data points. In what

follows we shall propose two different residuals for the GEVNRM. The first residual is similar to that proposed by (FERRARI; CRIBARI-NETO, 2004) for the beta regression model. The standardized residual is given by

$$r_i^s = \frac{y_i - \hat{y}_i}{\sqrt{\widehat{\mathrm{Var}}(y_i)}},$$

where $\hat{y}_i = \hat{\eta}_{1i} + \gamma\hat{\eta}_{2i}$ and

$$\widehat{\mathrm{Var}}(y_i) = \hat{\eta}_{2i}^2 \frac{\pi^2}{6}.$$

Here, $\hat{\eta}_{ji}$ is the ith element of $\boldsymbol{\eta}_j$ evaluated at the maximum likelihood estimates, $i = 1,\ldots,n$ and $j = 1,2$. When the model is correctly specified and provides a good representation to the data, there should be no detectable pattern in the index plot of r_i^s or in the plot of r_i^s against $\hat{y}_i$. The same holds true for any alternative residual.

The second residual, proposed by (FERRARI; CRIBARI-NETO, 2004) for use with beta regressions is based on the deviance. It is based on a measure of discrepancy between the log-likelihood of the saturated model and that of the fitted model. The deviance residual is defined as

$$r_i^d = \mathrm{sgn}(y_i - \hat{y}_i)\sqrt{2(\ell(y_i,\tilde{\mu}_i,\tilde{\sigma}_i) - \ell(y_i,\hat{\mu}_i,\hat{\sigma}_i))},$$

where sgn is the sign function. The ith observation contribution to the deviance equals $(r_i^d)^2$, and hence an observation with a large absolute value of r_i^d can be viewed as discrepant. Here, $\tilde{\boldsymbol{\mu}} = (\tilde{\mu}_1,\ldots,\tilde{\mu}_n)^\top$ and $\tilde{\boldsymbol{\sigma}} = (\tilde{\sigma}_1,\ldots,\tilde{\sigma}_n)^\top$ are the values of $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ that simultaneously solve $\boldsymbol{U}_{\boldsymbol{\mu}} = \mathbf{0}$ and $\boldsymbol{U}_{\boldsymbol{\sigma}} = \mathbf{0}$, respectively. We note that $\boldsymbol{U}_{\boldsymbol{\sigma}} = \mathbf{0}$ does not have a closed form solution, but $\boldsymbol{U}_{\boldsymbol{\mu}} = \mathbf{0}$ yields $\tilde{\boldsymbol{\mu}} = \boldsymbol{y}$. We then decided to set $\tilde{\boldsymbol{\sigma}} = \hat{\boldsymbol{\sigma}}$ and obtained

$$r_i^d = \mathrm{sgn}(y_i - \hat{\eta}_{1i} - \gamma\hat{\eta}_{2i})\sqrt{2(\ell(y_i,y_i,\hat{\sigma}_i) - \ell(y_i,\hat{\mu}_i,\hat{\sigma}_i))}.$$

After some algebra, we arrived at the following expression for the deviance residual:

$$r_i^d = \mathrm{sgn}(y_i - \hat{\eta}_{1i} - \gamma\hat{\eta}_{2i})\sqrt{2\left(\frac{y_i - \hat{\eta}_{1i}}{\hat{\eta}_{2i}} + \exp\left(-\frac{y_i - \hat{\eta}_{1i}}{\hat{\eta}_{2i}}\right)\right)}.$$

The two residuals presented above can be used to construct half-normal plots with simulated envelopes; see (ATKINSON, 1985). Such plots can then be used to identify atypical data points.

2.1.2 Generalized leverage

In what follows we shall consider the generalized leverage proposed by (WEI; HU; FUNG, 1998), which measures the impacts of observed values on predicted values. Let $\hat{\boldsymbol{y}} = (\hat{y}_1, \ldots, \hat{y}_n)^\top$ be the vector of fitted values. The generalized leverage $GL(\boldsymbol{\psi})$ is the $n \times n$ matrix whose (i, j) element is $\partial \hat{y}_i / \partial y_j$, i.e., the instantaneous rate of change of the ith predicted value associated with an infinitesimally small change in the jth response value. The generalized leverage is invariant under reparameterization and observations for which GL_{ii} is large are taken to be leverage points. (WEI; HU; FUNG, 1998) proposed the following leverage matrix:

$$GL(\boldsymbol{\psi}) = D_\psi(-\ell_{\boldsymbol{\psi}\boldsymbol{\psi}})^{-1}\ell_{\boldsymbol{\psi}\boldsymbol{y}},$$

where $\ell_{\boldsymbol{\psi}\boldsymbol{y}} = \partial^2\ell/\partial\boldsymbol{\psi}\partial\boldsymbol{y}^\top$ and $D_{\boldsymbol{\psi}} = \partial\mathbb{E}(Y)/\partial\boldsymbol{\psi}^\top$, both evaluated at the maximum likelihood estimate $\hat{\boldsymbol{\psi}}$. The ith observation is taken to be a leverage point if GL_{ii}, the ith diagonal element $GL(\boldsymbol{\psi})$, is large.

For the GEVNRM, we write $\ell_{\boldsymbol{\psi}\boldsymbol{y}} = (\partial^2\ell/\partial\boldsymbol{\beta}\partial\boldsymbol{y}^\top, \partial^2\ell/\partial\boldsymbol{\theta}\partial\boldsymbol{y}^\top)^\top$ and use the fact that $\mathbb{E}(y_i) = \mu_i + \gamma\sigma_i$ to obtain

$$\frac{\partial\mathbb{E}(y_i)}{\partial\beta_j} = \frac{\partial\mu_i}{\partial\eta_{1i}}\frac{\partial\eta_{1i}}{\partial\beta_j}, \quad j = 1, \ldots, p,$$
$$\frac{\partial\mathbb{E}(y_i)}{\partial\theta_l} = \gamma\frac{\partial\sigma_i}{\partial\eta_{2i}}\frac{\partial\eta_{2i}}{\partial\theta_l}, \quad l = 1, \ldots, q.$$

Let $\tilde{X}^*$ be an $n \times p$ matrix whose (i, j) element is given by $(d\mu_i/d\eta_{1i})(\partial\eta_{1i}/\partial\beta_j)$ and let $\tilde{S}^*$ be an $n \times q$ matrix whose (i, j) element is $\gamma(\partial\sigma_i/\partial\eta_{2i})(\partial\eta_{2i}/\partial\theta_j)$. It follows that $D_{\boldsymbol{\psi}} = [\tilde{X}^* \ \tilde{S}^*]$ and that $\ell_{\boldsymbol{\psi}\boldsymbol{y}}$ can be written as

$$\ell_{\boldsymbol{\psi}\boldsymbol{y}} = \begin{bmatrix} \tilde{X}^\top M_1 \Sigma_2^{-1} D_1 \\ \tilde{S}^\top M_2 \Sigma_2^{-1} D_2 \end{bmatrix},$$

where D_1 and D_2 are $n \times n$ diagonal matrices whose ith diagonal entries are given by d_i^* and $1 - d_i^*(1 - d_i)$, respectively, with d_i and d_i^* as defined in Section 2 and $\Sigma_2 = \mathrm{diag}(\hat{\sigma}_i^2)$.

2.1.3 Generalized Cook's distance

Cook's distance method is recommended for global influence analysis, i.e., when the interest lies in evaluating changes in the parameter estimates caused by the exclusion of a subset of observations from the sample. The generalized Cook distance associated with the ith observation is given by

$$GD_i = (\hat{\boldsymbol{\psi}}_{(i)} - \boldsymbol{\psi})^\top M(\hat{\boldsymbol{\psi}}_{(i)} - \boldsymbol{\psi}),$$

where $\hat{\boldsymbol{\psi}}_{(i)}$ denotes the estimate of ψ obtained after excluding the ith observation from the sample and M is a nonnegative definite matrix, usually taken to be $M = -\ell_{\boldsymbol{\psi}\boldsymbol{\psi}}$, the observed information matrix; see (COOK; WEISBERG, 1982) and (XIE; WEI, 2007).

Computation of $\hat{\boldsymbol{\psi}}_{(i)}$ may be expensive when the sample size is large. To circumvent that, we follow (PREGIBON, 1981) and use the following one-step approximation:

$$\hat{\boldsymbol{\psi}}^1_{(i)} = \hat{\boldsymbol{\psi}} + \{-\ell_{\boldsymbol{\psi}\boldsymbol{\psi}}\}^{-1}\ell_{\boldsymbol{\psi}(i)}, \tag{2.1}$$

where $\ell_{\boldsymbol{\psi}(i)} = \partial\ell_{(i)}(\boldsymbol{\psi})/\partial\boldsymbol{\psi}$, $\ell_{(i)}(\boldsymbol{\psi})$ denoting the log-likelihood function without the ith observation. The terms on the right hand side of Equation (2.1) are evaluated at the maximum likelihood estimates.

Considering the GEVNRM and using the same notation as in (BARRETO-SOUZA; VASCONCELLOS, 2011), we obtained, after some some algebra,

$$\left.\frac{\partial\ell_{(r)}(\boldsymbol{\psi})}{\partial\beta_j}\right|_{\boldsymbol{\psi}=\hat{\boldsymbol{\psi}}} = -\sum_{i\neq r}^{n}(y_i^* - \mu_i^*)\frac{\partial\mu_i}{\partial\eta_{1i}}\frac{\partial\eta_{1i}}{\partial\beta_u}, \quad j = 1,\ldots,p,$$
$$\left.\frac{\partial\ell_{(r)}(\boldsymbol{\psi})}{\partial\theta_v}\right|_{\boldsymbol{\psi}=\hat{\boldsymbol{\psi}}} = \sum_{i\neq r}^{n}\frac{v_i}{\sigma_i}\frac{\partial\sigma_i}{\partial\eta_{2i}}\frac{\partial\eta_{2i}}{\partial\theta_l}, \quad l = 1,\ldots,q.$$

2.1.4 Local influence

The local influence method proposed by (COOK, 1986) and the global influence method proposed by (COOK; WEISBERG, 1982) are recommended when the main interest lies in measuring the impact of a particular set of observations on the resulting parameter estimates. The major difference between the two approaches lies in the fact that the global influence method involves performing parameter estimation without a given set of observations in the sample whereas the local influence method uses a local perturbation scheme with the complete data.

Let $\boldsymbol{\omega}$ be a k-dimensional perturbation vector such that $\boldsymbol{\omega} \in \Omega \subset \mathbb{R}^k$, let $\ell(\hat{\boldsymbol{\psi}}|\boldsymbol{\omega})$ be the perturbed log-likelihood function and assume that that there exists a non-perturbation vector $\boldsymbol{\omega}_0$ such that $\ell(\boldsymbol{\psi},\boldsymbol{\omega}_0) = \ell(\boldsymbol{\psi})$. The impact of minor perturbations on the maximum likelihood estimate $\hat{\boldsymbol{\psi}}$ can be measured using the log-likelihood displacement $LD(\boldsymbol{\omega}) = 2\{\ell(\hat{\boldsymbol{\psi}}) - \ell(\hat{\boldsymbol{\psi}}_{\boldsymbol{\omega}})\}$, where $\hat{\boldsymbol{\psi}}_{\boldsymbol{\omega}}$ denotes the maximum likelihood estimate under $\ell(\boldsymbol{\psi}|\boldsymbol{\omega})$.

Local influence methods analyze the local behavior of the log-likelihood displacement around $\boldsymbol{\omega}_0$ by computing the normal curvature of the plot of $LD(\boldsymbol{\omega}_0 + \alpha\boldsymbol{d})$ against α, where $\alpha \in \mathbb{R}$ and $\boldsymbol{d}$ is a unit norm direction. The main interest lies in $\boldsymbol{d}_{\max}$, the direction corresponding to $C_{\boldsymbol{d}_{\max}}$, the largest curvature. The index plot of $\boldsymbol{d}_{\max}$ reveals which data points can be considered

atypical, i.e., which observations have considerable influence on *LD* under minor perturbations. (COOK, 1986) showed that the normal curvature at $\boldsymbol{d}$ is given by

$$C_{\boldsymbol{d}}(\hat{\boldsymbol{\psi}}) = 2|\boldsymbol{d}^\top \Delta^\top \ell^{-1}_{\boldsymbol{\psi}\boldsymbol{\psi}} \Delta \boldsymbol{d}|,$$

where Δ is a $p+q \times n$ perturbation matrix.

In the framework of the GEVNRM,

$$\Delta = (\{\Delta_{ij}\}, \{\Delta_{il}\})^\top = \left(\left\{ \frac{\partial^2 \ell}{\partial \beta_j \partial \omega_i} \right\}, \left\{ \frac{\partial^2 \ell}{\partial \theta_l \partial \omega_i} \right\} \right)^\top, \quad j = 1, \ldots, p, \ l = 1, \ldots, q \text{ and } i = 1, \ldots, n.$$

This matrix is evaluated at both $\boldsymbol{\psi} = \hat{\boldsymbol{\psi}}$ and $\boldsymbol{\omega} = \boldsymbol{\omega}_0$. In what follows, we shall consider three different perturbation schemes for local influence analysis in the GEVNRM, namely: case-weights perturbation, response perturbation and explanatory variable perturbation. Some details regarding the derivation of the influence measures can be found in the Appendix.

2.1.4.1 Case-weights perturbation

In this scheme, the perturbation is a weight that represents the contribution for each observation to the log-likelihood function. The perturbed log-likelihood function is given by

$$\ell(\boldsymbol{\psi}|\boldsymbol{\omega}) = \sum_{i=1}^{n} \omega_i \ell_i(\boldsymbol{\psi}|\boldsymbol{\omega}),$$

where $\ell_i(\boldsymbol{\psi}|\boldsymbol{\omega}) = -\log(\sigma_i) - (y_i - \mu_i)/\sigma_i - \exp(-(y_i - \mu_i)/\sigma_i)$. The no perturbation vector is given by $\boldsymbol{\omega}_0 = (1, \ldots, 1)^\top$ and it is possible to show that the components of the perturbation matrix are

$$\Delta_{1\boldsymbol{\beta}} = \tilde{X}^\top \Omega^{-1} M_1 R_1,$$
$$\Delta_{1\boldsymbol{\theta}} = \tilde{S}^\top \Sigma^{-1} M_2 R_2,$$

where $R_1 = \text{diag}(y_i^* - \mu_i^*)$ and $R_2 = \text{diag}(v_i)$. The perturbation matrix can be expressed as $\Delta = (\Delta_{1\boldsymbol{\beta}}, \Delta_{1\boldsymbol{\theta}})^\top$, a matrix of dimension $(p+q) \times n$.

2.1.4.2 Response perturbation

We shall now consider perturbations to the dependent variable. It is noteworthy that this perturbation scheme is closely related to the observations leverages. Additionally, (SCHWARZMANN, 1991) showed that it is closely related to Cook's distance measures in the

context of the classic linear regression model. In nonlinear models that belong to the exponential family, (WEI; HU; FUNG, 1998) established a similar result but using the generalized leverage. (LAURENT; COOK, 1993) presented two new leverage measures: the Jacobian and tangent plane leverages. They showed the response variable perturbation scheme can be described by two quadratic forms which use such leverage measures. Their results generalize those of (SCHWARZMANN, 1991).

We consider that each response y_i is perturbed as $y_{iw} = y_i + \omega_i S_y$, where ω_i is the ith components of the perturbation vector $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^\top$ and S_y is a scale factor, usually taken to be the estimated standard deviation of $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$. The no perturbation vector is $\boldsymbol{\omega}_0 = \mathbf{0}_n$, where $\mathbf{0}_n = (0, \ldots, 0)^\top$, i.e., $\mathbf{0}_n$ denotes the n-dimensional vector of zeros. Let $l_{i\boldsymbol{\omega}} = (y_{iw} - \mu_i)/\sigma_i$ and $l^*_{i\boldsymbol{\omega}} = \exp(-l_{i\boldsymbol{\omega}})$. The perturbed log-likelihood function is given by

$$\ell_i(\boldsymbol{\psi}|\boldsymbol{\omega}) = -\log(\sigma_i) - l_{i\boldsymbol{\omega}} - l^*_{i\boldsymbol{\omega}}.$$

After some algebra, we arrived at

$$\Delta_{2\boldsymbol{\beta}} = S_y \tilde{X}^\top M_1 \Sigma_2^{-1} \mathscr{L}_1,$$
$$\Delta_{2\boldsymbol{\theta}} = S_y \tilde{S}^\top M_2 \Sigma_2^{-1} \mathscr{L}_2,$$

where $\mathscr{L}_1 = \mathrm{diag}(l^*_{i\boldsymbol{\omega}})$, $\Sigma_2 = \mathrm{diag}(\hat{\sigma}_i^2)$ and $\mathscr{L}_2 = \mathrm{diag}(1 + l^*_{i\boldsymbol{\omega}}(l_{i\boldsymbol{\omega}} - 1))$. The perturbation matrix is given by $\Delta = (\Delta_{2\boldsymbol{\beta}}, \Delta_{2\boldsymbol{\theta}})^\top$, a matrix of dimension $(p+q) \times n$.

2.1.4.3 Explanatory variable perturbation

This scheme is considered when we are interested in analyzing the impact of small perturbations to the values of a given explanatory variable on the resulting inferences. Since there are submodels for $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$, we shall consider three kinds of perturbation: perturbation on $\boldsymbol{\mu}$ by making $x_{ij\boldsymbol{\omega}} = x_{ij} + \omega_{ix} S_x$, perturbation on $\boldsymbol{\sigma}$ by making $z_{ik\boldsymbol{\omega}} = z_{ik} + \omega_{iz} S_z$, and perturbation on both parameters by adding the same perturbation to $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ simultaneously, i.e., $\omega_{ix} = \omega_{iz} = \omega_i$, where $j \in \{1, \ldots, p\}$, $k \in \{1, \ldots, q\}$ and $i = 1, \ldots, n$. Here, S_x and S_z are scaling factors, usually taken to be the standard deviations of $\boldsymbol{x}_j = (x_{1j}, \ldots, x_{nj})^\top$ and $\boldsymbol{z}_k = (z_{1k}, \ldots, z_{nk})^\top$, respectively. The perturbed predictors as

$$g_1(\mu_{i\boldsymbol{\omega}}) = \eta_{1i\boldsymbol{\omega}} = f_1(X, \boldsymbol{\beta}, \omega_{ix}),$$
$$g_2(\sigma_{i\boldsymbol{\omega}}) = \eta_{2i\boldsymbol{\omega}} = f_2(Z, \boldsymbol{\theta}, \omega_{iz}).$$

The no perturbation vector is $\boldsymbol{\omega}_0 = (0,\ldots,0)^\top$. The perturbation matrix is given by

$$\Delta = (\{\Delta_{ji}\},\{\Delta_{li}\})^\top = \left(\left\{\frac{\partial^2\ell}{\partial\beta_j\partial\omega_{ix}}\right\},\left\{\frac{\partial^2\ell}{\partial\theta_l\partial\omega_{ix}}\right\}\right)^\top,$$

For $j = 1,\ldots,p,\ l = 1,\ldots,q$ and $i = 1,\ldots,n$. In what follows we shall use a notation similar to that used by (LEMONTE; PATRIOTA, 2011):

$$\begin{aligned}
m_{1i} &= \frac{d\mu_i}{d\eta_{1i}}, \quad m_{2i} = \frac{d\sigma_i}{d\eta_{2i}}, \quad \dot{m}_{1i} = \frac{d^2\mu_i}{d\eta_{1i}^2}, \quad \dot{m}_{2i} = \frac{d^2\sigma_i}{d\eta_{2i}^2},\\
m_{1i}^* &= \frac{d\mu_{i\omega}}{d\eta_{1i\omega}}, \quad m_{2i}^* = \frac{d\sigma_{i\omega}}{d\eta_{2i\omega}}, \quad \dot{m}_{1i}^* = \frac{d^2\mu_{i\omega}}{d\eta_{1i\omega}^2}, \quad \dot{m}_{2i}^* = \frac{d^2\sigma_{i\omega}}{d\eta_{2i\omega}^2},\\
\dot{\mu}_{ij} &= \frac{\partial\eta_{1i}}{\partial\beta_j}, \quad \dot{\mu}_{ij}^* = \frac{\partial\eta_{1i\omega}}{\partial\beta_j}, \quad \ddot{\mu}_{ij} = \frac{\partial^2\eta_{1i\omega}}{\partial\beta_j\partial\omega_{ix}}, \quad \dot{\mu}_{i\omega} = \frac{\partial\eta_{1i\omega}}{\partial\omega_{ix}},\\
\dot{\sigma}_{il} &= \frac{\partial\eta_{2i}}{\partial\theta_l}, \quad \dot{\sigma}_{il}^* = \frac{\partial\eta_{2i\omega}}{\partial\theta_l}, \quad \ddot{\sigma}_{il} = \frac{\partial^2\eta_{2i\omega}}{\partial\theta_l\partial\omega_{iz}}, \quad \dot{\sigma}_{i\omega} = \frac{\partial\eta_{2i\omega}}{\partial\omega_{iz}}.
\end{aligned}$$

We start by considering perturbation on $\boldsymbol{\mu}$. Here, $\Delta_{\boldsymbol{\beta}_1}$ and $\Delta_{\boldsymbol{\theta}_1}$ are matrices of dimensions $p\times n$ and $q\times n$ with typical elements Δ_{ij} and Δ_{il}, respectively. Let $v_{i\boldsymbol{\omega}} = (y_i-\mu_{i\omega})/\sigma_i$ and $v_{i\boldsymbol{\omega}}^* = \exp(-v_{i\boldsymbol{\omega}})$. It can be shown that

$$\begin{aligned}
\Delta_{ij} &= \frac{1-v_{i\boldsymbol{\omega}}^*}{\sigma_i}[m_{1i}^*\ddot{\mu}_{ij} + \dot{m}_{1i}^*\dot{\mu}_{ij}\dot{\mu}_{i\omega}] - \frac{v_{i\boldsymbol{\omega}}^*}{\sigma_i}(m_{1i}^*)^2\dot{\mu}_{ij}\dot{\mu}_{i\omega}, \quad j = 1,\ldots,p,\\
\Delta_{il} &= -\frac{1+v_{i\boldsymbol{\omega}}^*(v_{i\boldsymbol{\omega}}+1)}{\sigma_i^2}m_{2i}\dot{\sigma}_{il}m_{1i}^*\dot{\mu}_{i\omega}, \quad l = 1,\ldots,q.
\end{aligned}$$

We now move to perturbation on $\boldsymbol{\sigma}$. Let $t_{i\boldsymbol{\omega}} = (y_i-\mu_i)/\sigma_{i\omega}$, and $t_{i\boldsymbol{\omega}}^* = \exp(-t_{i\boldsymbol{\omega}})$. Also, let $\Delta_{\boldsymbol{\beta}_2}$ and $\Delta_{\boldsymbol{\theta}_2}$ be matrices with typical elements Δ_{ij} and Δ_{il} respectively. After some algebra, we arrive at

$$\begin{aligned}
\Delta_{ij} &= \frac{t_{i\boldsymbol{\omega}}^*(1-t_{i\boldsymbol{\omega}})-1}{\sigma_{i\omega}^2}m_{1i}\dot{\mu}_{ij}m_{2i}^*\dot{\sigma}_{i\omega}, \quad j = 1,\ldots,p,\\
\Delta_{il} &= \frac{1}{\sigma_{i\omega}^2}\left(1+2t_{i\boldsymbol{\omega}}\left(t_{i\boldsymbol{\omega}}^* - \frac{t_{i\boldsymbol{\omega}}^*t_{i\boldsymbol{\omega}}}{2}-1\right)\right)(m_{2i}^*)^2\dot{\sigma}_{il}\dot{\sigma}_{i\omega}\\
&\quad + \frac{t_{i\boldsymbol{\omega}}(1-t_{i\boldsymbol{\omega}}^*)-1}{\sigma_{i\omega}^2}[m_{2i}^*\ddot{\sigma}_{il} + \dot{m}_{2i}^*\dot{\sigma}_{il}^*\dot{\sigma}_{i\omega}], \quad l = 1,\ldots,q.
\end{aligned}$$

Finally, we consider joint perturbation on $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$. Let $j_{i\boldsymbol{\omega}} = (y_i-\mu_{i\omega})/\sigma_{i\omega}$ and $j_{i\boldsymbol{\omega}}^* = \exp(-j_{i\boldsymbol{\omega}})$. Also, let $\Delta_{\boldsymbol{\beta}_3}$ and $\Delta_{\boldsymbol{\theta}_3}$ be matrices with typical elements Δ_{ij} and Δ_{il}, respectively. It can be shown that

$$\begin{aligned}
\Delta_{ij} &= \frac{1-j_{i\boldsymbol{\omega}}^*}{\sigma_{i\omega}}[m_{1i}^*\dot{\mu}_{ij}^*\dot{\mu}_{i\omega}^* + \dot{m}_{1i}^*\ddot{\mu}_{ij}] + \frac{j_{i\boldsymbol{\omega}}^*}{\sigma_{i\omega}^2}[(m_{1i}^*)^2\dot{\mu}_{ij}^*\dot{\mu}_{i\omega}]\\
&\quad - \frac{1}{\sigma_{i\omega}}\left(1+\frac{j_{i\boldsymbol{\omega}}^*(j_{i\boldsymbol{\omega}}-1)}{\sigma_{i\omega}}\right)(m_{1i}^*\dot{\mu}_{ij}^*m_{2i}^*\dot{\sigma}_{i\omega}), \quad j = 1,\ldots,p
\end{aligned}$$

$$\Delta_{il} = \frac{j_{i\boldsymbol{\omega}}(\sigma_{i\omega} - j^*_{i\boldsymbol{\omega}}) - 1}{\sigma_{i\omega}}(m^*_{2i}\dot{\sigma}^*_{il}\dot{\sigma}^*_{i\omega} + \dot{m}^*_{2i}\ddot{\sigma}_{il}) + \frac{1 - j_{i\boldsymbol{\omega}}(\sigma_{i\omega} - j^*_{i\boldsymbol{\omega}}(2 - j_{i\boldsymbol{\omega}}))}{\sigma^2_{i\omega}}[(m^*_{2i})^2\dot{\sigma}^*_{il}\dot{\sigma}_{i\omega}] - \frac{j^*_{i\boldsymbol{\omega}}(1 - j_{i\boldsymbol{\omega}}) - \sigma_{i\omega}}{\sigma^2_{i\omega}}(m^*_{2i}\dot{\sigma}^*_{il}m^*_{1i}\dot{\mu}_{i\omega}), \quad l = 1, \ldots, q.$$

2.2 MISSPECIFICATION AND NONNESTED HYPOTHESIS TESTS FOR THE GEVNRM

We shall now focus on testing inferences for the GEVNRM. Our interest is twofold. First, we wish to test the null hypothesis that a given fitted GEVNRM is correctly specified. Such a null hypothesis is tested against the alternative hypothesis that the model specification is in error. Misspecification can be due to the use of incorrect link functions, omitted covariates, neglected nonlinearity, etc. Second, we wish to be able to choose one model from a set of nonnested GEVNRMs. For instance, we may have at our disposal two plausible models that differ in the link functions they use in the submodels for $\boldsymbol{\sigma}$. How can one of such models be selected with the aid of a hypothesis test? We shall address this issue.

2.2.1 A misspecification test for the GEVNRM

Our interest now lies in testing whether the model specification is in error. To that end, we shall consider the misspecification test proposed by (RAMSEY, 1969) for the classical linear regression model; see also (RAMSEY; GILBERT, 1972). Such a test is known as the RESET test (Regression Specification Error Test). The null hypothesis is that the model is correctly specified which is tested against the alternative hypothesis that the model is incorrectly specified. The structure of the GEVNRM is more complex than that of the classical linear model, since it contains two submodels with possibly nonlinear predictors, each submodel using a link function. Additionally, none of the parameters modelled separately is the mean response. In what follows, we shall outline a modified version of the RESET test that can be used with the GEVNRM. The underlying idea is that the model is augmented using some testing variables and one then test their exclusion. If the testing variables prove to be statistically significant, then there is evidence of model misspecification and the null hypothesis is rejected. The testing variables can be taken to be powers of the fitted values. Consider the GEVNRM with two submodels, i.e., suppose that one models $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$, such that both submodels contain regressors. The test can be performed as follows:

1. Estimate the parameters of the GEVNRM and obtain $\hat{\boldsymbol{\eta}}_1$ and $\hat{\boldsymbol{\eta}}_2$ (the estimated predictors).

2. Estimate the parameters of the augmented GEVNRM obtained by adding $\hat{\boldsymbol{\eta}}_1^2$ as an additional regressor to the submodel for $\boldsymbol{\mu}$ and adding $\hat{\boldsymbol{\eta}}_2^2$ as an additional regressor to the submodel for $\boldsymbol{\sigma}$.
3. Test the joint exclusion of the two artificial regressors (i.e., the joint exclusion of $\hat{\boldsymbol{\eta}}_1$ and $\hat{\boldsymbol{\eta}}_2$).
4. If the null hypothesis is rejected, reject the model under evaluation. Otherwise, conclude that the model is correctly specified.

When $\boldsymbol{\sigma}$ ($\boldsymbol{\mu}$) is taken to be constant (i.e., not affected by covariates), only the submodel for $\boldsymbol{\mu}$ ($\boldsymbol{\sigma}$) using $\hat{\boldsymbol{\eta}}_1^2$ ($\hat{\boldsymbol{\eta}}_2^2$), and only one restriction is tested.

In the test procedure outlined above, we only use the squared estimated predictors of both submodels as testing variables, i.e., we augment the submodels for $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ using $\hat{\boldsymbol{\eta}}_1^2$ and $\hat{\boldsymbol{\eta}}_2^2$, respectively. It is possible to use additional powers of the estimated predictors as additional testing variables. For instance, one can augment the submodel for $\boldsymbol{\mu}$ using $\hat{\boldsymbol{\eta}}_1^2$ and $\hat{\boldsymbol{\eta}}_1^3$ and augment the submodel for $\boldsymbol{\sigma}$ using $\hat{\boldsymbol{\eta}}_2^2$, $\hat{\boldsymbol{\eta}}_2^3$ and $\hat{\boldsymbol{\eta}}_2^4$. In that case, one would test the joint exclusion of five variables.

2.2.2 Nonnested hypothesis tests for the GEVNRM

Oftentimes the practitioner has at his/her disposal more than one candidate model and such models are nonnested. That is, no model can be obtained as a particular case of some other model. At the outset, we shall consider the standard case of nested models. For example, consider Model M_1:

$$g_1^{(1)}(\mu_i) = \beta_1 + \beta_2 x_{1i} + \beta_3 x_{2i},$$
$$g_2^{(1)}(\sigma_i) = \theta_1 + \theta_2 z_{1i} + \theta_3 z_{2i}.$$

Suppose that the practitioner suspects that the covariates x_2 and z_2 are not important for explaining the phenomenon of interest. He/she can then test $\mathscr{H}_0 : (\beta_3, \theta_3)^\top = (0,0)^\top$ against $\mathscr{H}_1 : (\beta_3, \theta_3)^\top \neq (0,0)^\top$. If the null hypothesis is rejected, the following simpler model, say Model M_2, can be used:

$$g_1^{(2)}(\mu_i) = \beta_1 + \beta_2 x_{1i},$$
$$g_2^{(2)}(\sigma_i) = \theta_1 + \theta_2 z_{1i}.$$

(The superscripts in parentheses in the link functions index the model.) Such an approach is valid because the two models are nested: the latter is a special case of the former. When the models

are nonnested such standard testing approach cannot be used.

Two GEVNRMs, say M_1 and M_2, are said to be nonnested if it is not possible to obtain M_1 by imposing restrictions on the parameters that index M_2, and vice-versa. The concept can be easily extended to more than two models. A test for nonnested classical linear regression models was introduced by (DAVIDSON; MACKINNON, 1981). In what follows we shall adapt their test, which is known as the J test, for use with GEVNRMs. The main idea is to consider an artificial model that encompasses all nonnested models, and then replace the parameters of the models that are not being tested by their respective maximum likelihood estimates, which are consistent when each of such models is the true model.

Suppose there are $N \geq 2$ nonnested models, say, $M_1, \ldots, M_N$:

$$M_a : g_1(\boldsymbol{\mu}) = \boldsymbol{\eta}_1 = f_1(X, \boldsymbol{\beta}),$$
$$g_2(\boldsymbol{\sigma}) = \boldsymbol{\eta}_2 = f_2(Z, \boldsymbol{\theta}),$$

$a = 1, \ldots, N$. Notice that the models can differ in at least one of the following: (i) regressors used in the submodel for $\boldsymbol{\mu}$, (ii) regressors used in the submodel for $\boldsymbol{\sigma}$, (iii) link function used in the submodel for $\boldsymbol{\mu}$, (iv) link function used in the submodel for $\boldsymbol{\sigma}$, (v) any combination of the aforementioned factors.

Let N_p and N_q be the number of $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ submodels that differ from the respective submodels in Model M_a. That is, N_p (N_q) denotes the number of models whose submodels for $\boldsymbol{\mu}$ ($\boldsymbol{\sigma}$) differ from the submodel for $\boldsymbol{\mu}$ ($\boldsymbol{\sigma}$) used in Model a. In order to test M_a using the J test, one estimates the parameters that index the remaining models by maximum likelihood, and then includes the corresponding estimated predictors that differ from M_a as additional covariates in the corresponding submodels of M_a. Next, one tests the exclusion of extra $N_p + N_q$ regressors. The J test statistic for testing the joint exclusion of the added testing variables (i.e., for testing Model a) is

$$J_a = 2\{\ell(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) - \ell(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\theta}})\},$$

where $\ell(\boldsymbol{\beta}, \boldsymbol{\theta})$ is the log-likelihood function and $(\hat{\boldsymbol{\beta}}^\top, \hat{\boldsymbol{\theta}}^\top)^\top$ and $(\tilde{\boldsymbol{\beta}}^\top, \tilde{\boldsymbol{\theta}}^\top)^\top$ are, respectively, the unrestricted and restricted maximum likelihood estimators of $(\boldsymbol{\beta}^\top, \boldsymbol{\theta}^\top)^\top$. Note that N_p and N_q additional covariates are included in the augmented submodels for $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$, respectively, and that the test statistic above is the standard likelihood ratio test statistic. Model M_a is rejected at significant level α if $J_a > \chi^2_{1-\alpha, N_p+N_q}$.

Notice that the model under test is not rejected if it fit is not noticeably improved when fitted values from the competing models are used as additional explanatory variables, i.e., when the coefficients of such regressors are not significantly different than zero. Otherwise, the model is rejected. The test needs to be sequentially applied to each candidate model. For instance, when there are two nonnested models, each model has to be tested against the other model. As a result, one model may be accepted and the other model rejected, both models may be accepted, and both models may turn out to be rejected.

A related test was introduced by (HAGEMANN, 2012): the *MJ* test. As with the *J* test, it was designed for use with the classical linear regression model. We shall now present a modified version of that test that can be used with the GEVNRM. The underlying idea is that if Model a is the correct model, the associated J statistic has a well-defined asymptotic distribution and the J statistics corresponding to all other candidate models diverge to ∞ as the sample size increases. Therefore, the model that corresponds to the smallest J statistic is a natural candidate to be taken as the correct model. When the smallest J statistic is large one can safely reject all models. The hypotheses used in the *MJ* test differ slightly from those used in the J test. Given a set of N candidate models, for each model one tests, using the J test, $\mathscr{H}_0$: the model is correct vs. $\mathscr{H}_1$: the model is not correct. In contrast, the null hypothesis used in the *MJ* test is $\mathscr{H}_0$: one of the candidate models is the true model; it is tested against $\mathscr{H}_1$: none of the candidate models is the true model. The *MJ* test statistic equals the minimum of all J statistics, i.e.,

$$MJ = \min\{J_1, \ldots, J_N\}.$$

The null hypothesis is rejected at significance level α if $MJ > \chi^2_{1-\alpha,N-1}$.

When the null hypothesis is not rejected, all that can be concluded is that one of the N models can be taken as the correct model. The test is not informative as to which model is the true model. It is noteworthy, however, that there is a model selection procedure associated to the test, namely: If the null hypothesis is not rejected, one can select the model that corresponds to the minimal J test statistic and discard the alternative $N-1$ models.

2.3 NUMERICAL EVALUATION

In this section we shall present results from numerical evaluations that were carried out using the tests and diagnostic tools presented in the previous sections. All numerical

evaluations were performed using the JULIA programming language (BEZANSON *et al.*, 2017). JULIA is a free open source, high-level, high-performance, dynamic programming language that uses a JIT (just-in-time) compiler and was developed for numerical computing; see <https://julialang.org>. Uniform random number generation was carried out using the Mersenne Twister algorithm. Gumbel random number generation was performed using the inversion method. Log-likelihood maximization was carried out using the L-BFGS-B quasi-Newton nonlinear optimization algorithm with analytical first derivatives. The initial value used for each parameter in the numerical optimization scheme was 1.0. There were no convergence failures. All computations were performed on a computer with an Intel i7 processor and 16 GB of memory running the LINUX operating system. All plots were produced using the R statistical computing environment (R Core Team, 2018).

2.3.1 Local influence and residual analysis

The analysis we shall now perform will be based on a simulated data set. We randomly generated $y_i \sim \text{Gumbel}(\mu_i, \sigma_i)$, where

$$\mu_i = \beta_1 - \exp(\beta_2 x_{2i}), \tag{2.2}$$

$$\log(\sigma_i) = \theta_1 + \theta_2 x_{2i}, \tag{2.3}$$

$i = 1, \ldots, 30$. The true values of β_1, β_2, θ_1 and θ_2 are 1.7, 2.2, 0.5 and -1.3, respectively. The values of x_{2i} were obtained as random draws from the beta distribution with parameters $\alpha = 2$ and $\beta = 7$. We shall attempt to identify atypical observations and then investigate whether they are influential, i.e., whether they noticeably impact the parameter estimates.

Overall, observations #9,#13,#14 and #18 appear to be most atypical data points. In order to check whether they are influential, we removed them from the sample (individually and jointly) and estimated the model's parameters again. In Table 8 we present the parameter estimates obtained using each data subset and also the relative changes in such estimates (as percentages) relative to the estimates computed using the complete data. The largest changes take place in the estimates of β_1 and β_2, the parameters that index the location submodel. Indeed, such changes are quite large. It is also clear that observation #18 is more influential than observations #9, #13 and #14. For instance, the relative changes in the estimates of β_1, β_2, θ_1 and θ_2 caused by the removal of observation #18 are, respectively, 109.03%, 76.17%, 4.70% and 17.18%. Also, the relative changes in $\hat{\beta}_1$ and $\hat{\beta}_2$ that follow from simultaneously removing data points #14 and #18 from the data are approximately equal to 111% and 69%, respectively. The removal of these

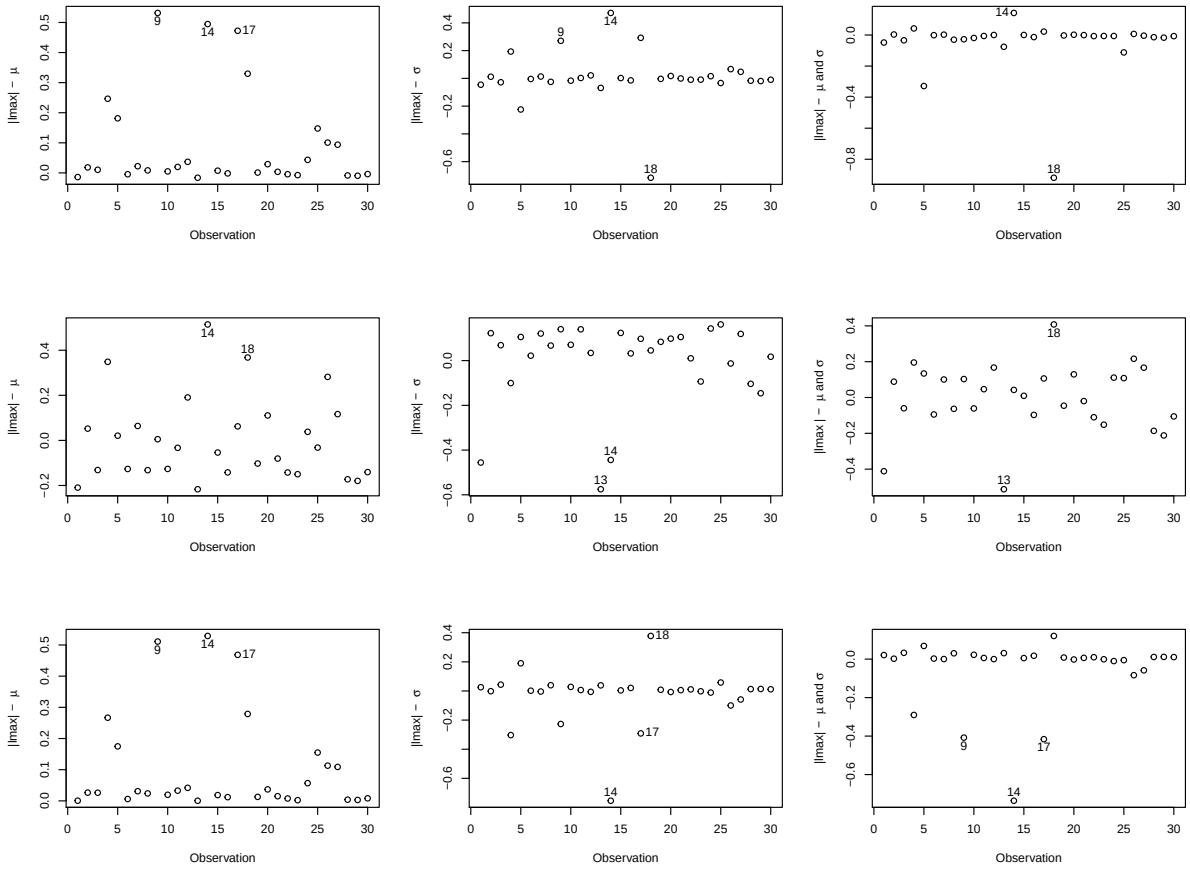

Figure 5 – Local influence measures for μ (first column), σ (middle column) and for μ and σ simultaneously (third column); the first, second and third rows relate to perturbations to the covariate, cases and response values.

two observations from the data also lead to the largest relative changes in $\hat{\theta}_1$ and $\hat{\theta}_2$, which are 23% and 50% (approximately). Hence, the diagnostic tools correctly indicated that these data points were worthy of further investigation.

We have also computed the standardized and deviance residuals and produced half normal plots with simulated envelopes which are presented in Figures 8 (standardized residual) and 9 (deviance residual). All points lie inside the envelope bands when the standardized residual is used. When the deviance residual is used, only one point falls outside the simulated envelope bands.

Finally, we shall manually introduce a data perturbation and then examine how the diagnostic measures respond to such perturbation. At the outset, we multiply the largest response value (which is the first value, y_1) by a positive scalar, which we take to be 1.25,1.50 and 2.00. Figures 10, 11 and 12 present the new Cook's distance values and Figures 13, 14 and 15 present the local influence measures computed using the modified data. As before, the first, second and third rows in the local influence plots relate to perturbations to the covariate, cases and response

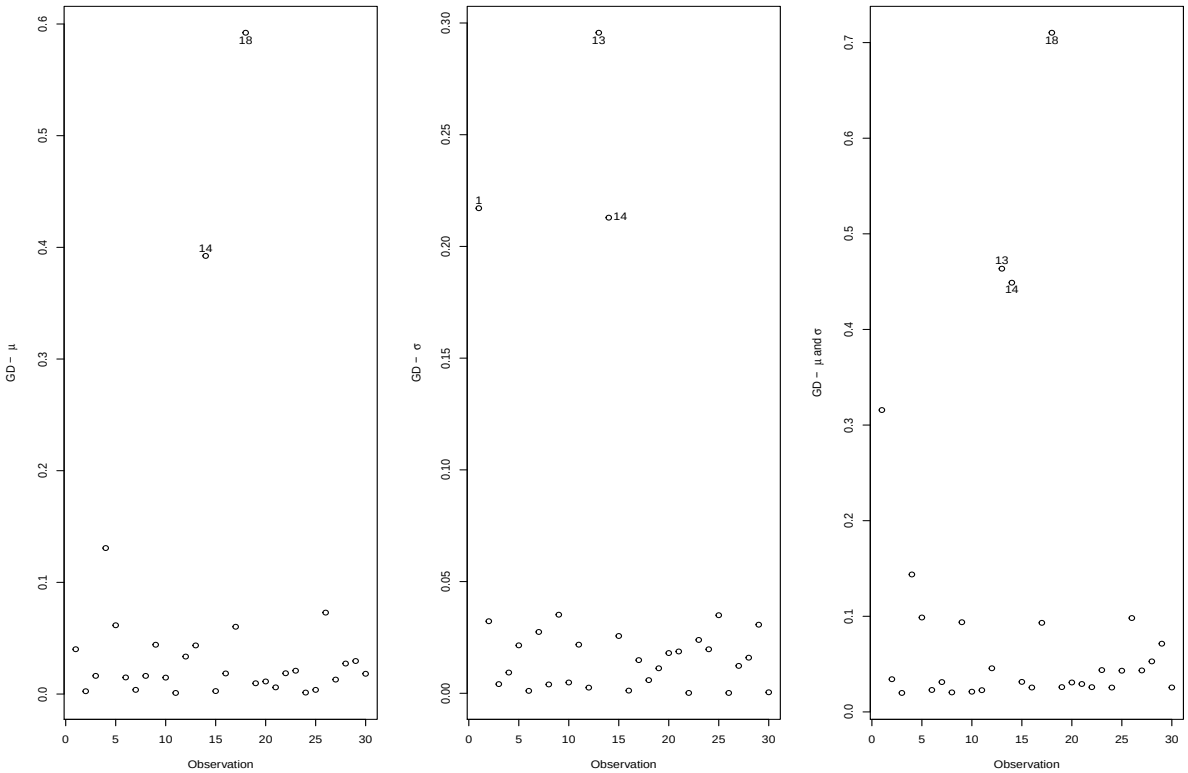

Figure 6 – Generalized Cook's distances for each observation of the data for μ, for σ, and for μ and σ simultaneously.

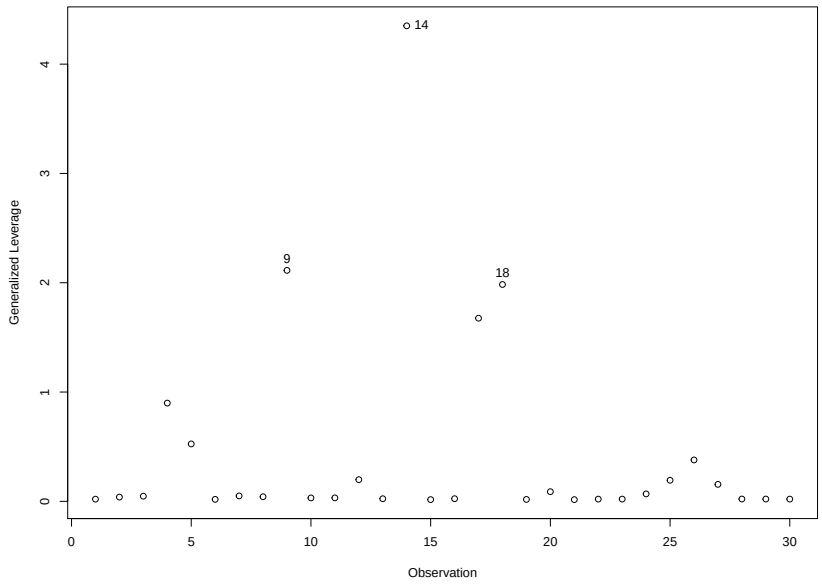

Figure 7 – Generalized leverages for μ and σ simultaneously.

Table 8 – Parameter estimates and relative changes (%) in such estimates based on incomplete data; simulated data.

Observation(s)	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\theta}_1$	$\hat{\theta}_2$	RC_{β_1}	RC_{β_2}	RC_{θ_1}	RC_{θ_2}
9	1.9772	1.8451	0.5768	−2.4012	89.55	44.46	3.54	9.26
13	1.7240	1.6807	0.6219	−3.1756	65.28	31.58	11.63	44.50
14	1.9690	1.8103	0.6859	−3.2469	88.77	41.73	23.11	47.74
18	2.1804	2.2502	0.5833	−2.5751	109.03	76.17	4.70	17.18
9, 14	1.9921	1.7654	0.6833	−3.1692	90.98	38.21	22.65	44.21
9, 18	2.2504	2.3396	0.5060	−2.0840	115.74	83.17	−9.17	−5.17
9, 13	1.7986	1.7081	0.5634	−2.8078	72.43	33.73	1.14	27.77
13, 14	2.0004	2.0118	0.5162	−3.1157	91.77	57.50	−7.35	41.78
13, 18	2.0360	2.1487	0.5359	−2.8434	95.19	68.22	−3.81	29.38
14, 18	2.2053	2.1659	0.6898	−3.3055	111.42	69.57	23.82	50.41
9, 14, 18	2.2413	2.1694	0.5971	−2.6938	114.88	69.84	7.18	22.58
9, 13, 14	1.8120	1.6367	0.6683	−3.6570	73.72	28.14	19.95	66.41
9, 13, 18	2.1162	2.2537	0.4633	−2.3859	102.87	76.44	−16.83	8.57
13, 14, 18	2.0518	2.0580	0.6508	−3.6975	96.71	61.12	16.83	68.25
9, 13, 14, 18	2.0889	2.0612	0.5725	−3.1618	100.26	61.37	2.76	43.87

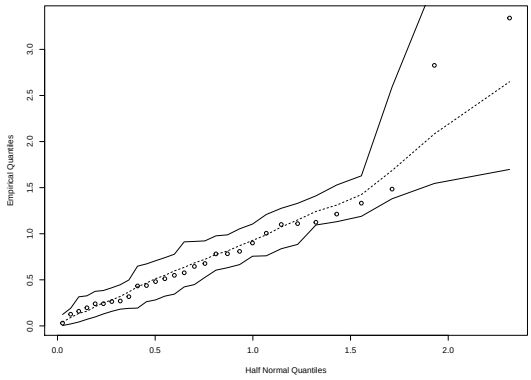

Figure 8 – Half normal plot, standardized residual; simulated data

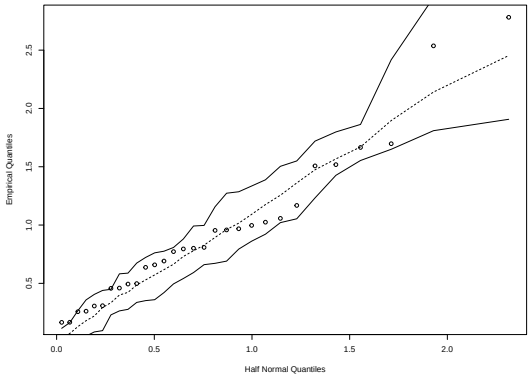

Figure 9 – Half normal plot, deviance residual; simulated data.

values. It is clear that the new outlying data point is singled out as influential by Cook's distance (schemes for $\boldsymbol{\sigma}$ and also for $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ jointly). It is noteworthy that the magnitudes of the Cook's distances for observation #1 increase with the value of the multiplying constant used to alter the value of the first response. We also note that observation #1 is singled out as influential by the local influence measures (cases perturbation for $\boldsymbol{\sigma}$ and for $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ simultaneously).

Next, we modify the largest covariate value, $x_{2,18}$, and then assess how the diagnostic

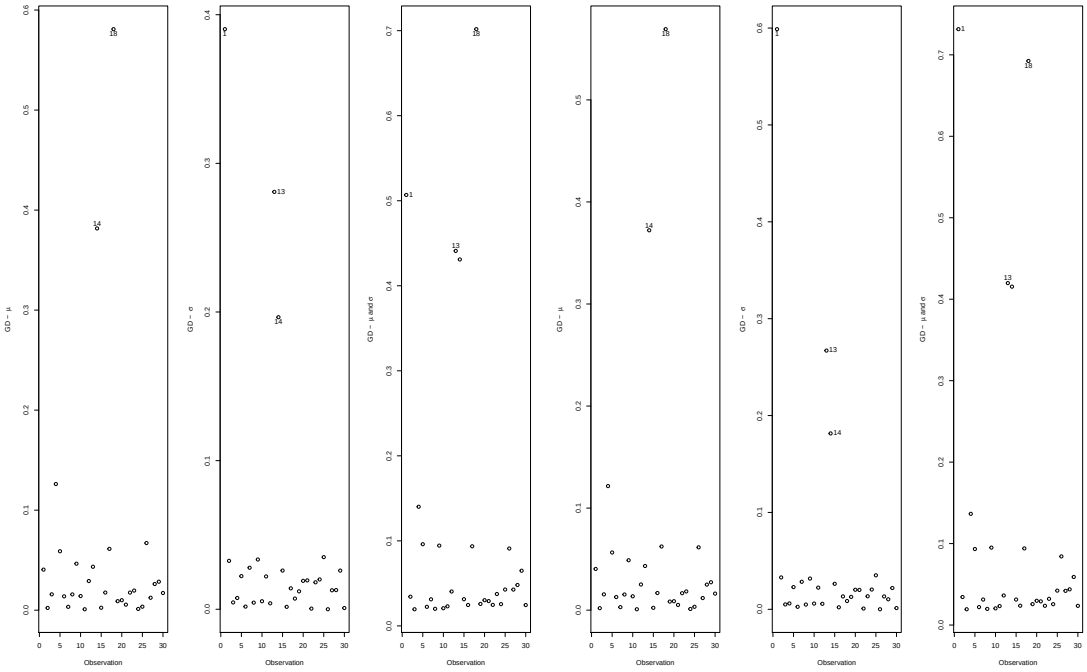

Figure 10 – Cook's distance $(y_1 \times 1.25)$**.** **Figure 11 – Cook's distance** $(y_1 \times 1.50)$**.**

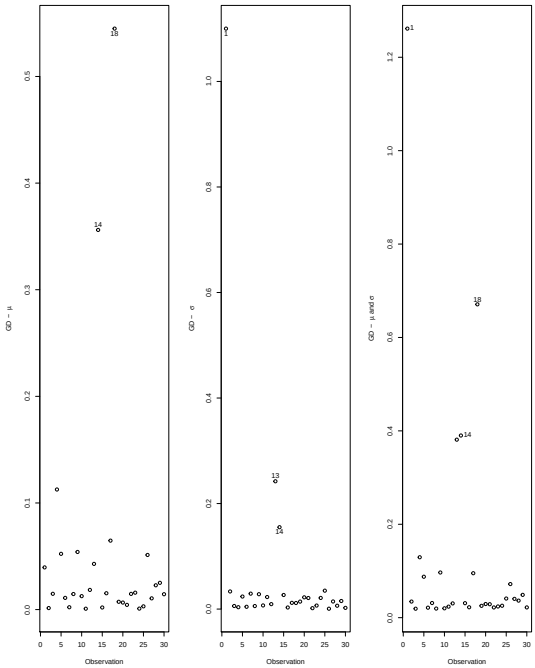

Figure 12 – Cook's distance $(y_1 \times 2.00)$**.**

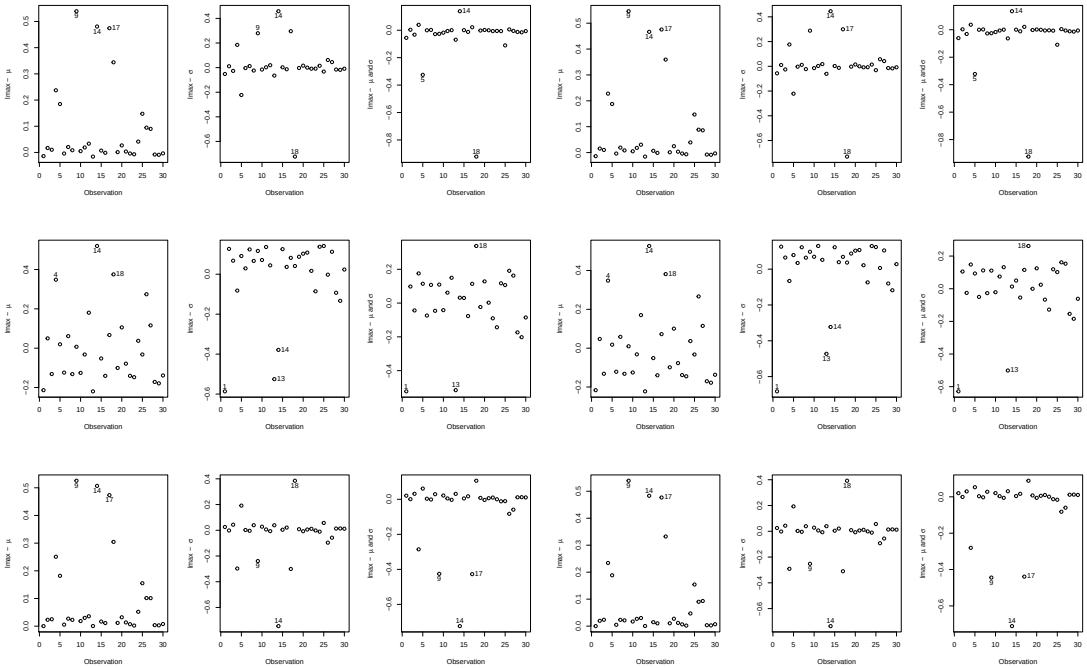

Figure 13 – Local influence measures $(y_1 \times 1.25)$.

Figure 14 – Local influence measures $(y_1 \times 1.50)$.

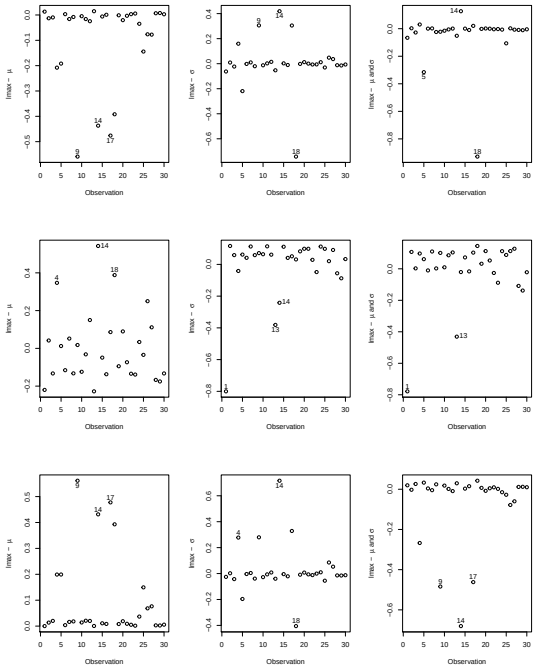

Figure 15 – Local influence measures $(y_1 \times 2.00)$.

measures are impacted. We sequentially multiply the largest covariate value by 1.10, 1.20 and 1.30 after generating the responses, thus making such data point progressively more atypical. For brevity, we do not present the diagnostic plots. We note that the generalized leverages of observation # 18 corresponding to the original data and to the three aforementioned altered data are, respectively, 1.984, 2.525, 3.466, 101.197, It is thus clear that the generalized leverage measure we derived is capable of detecting data points that are atypical in the regressors dimension. It is also worthy noting that the local influence measure relative to the covariate perturbation scheme is also quite sensitive to the increase in the largest independent variable value. For the original data, we obtain the following values for the 18th local influence measure for perturbation on $\boldsymbol{\mu}$, on $\boldsymbol{\sigma}$ and on both parameters simultaneously: 0.329, -0.715 and -0.919, respectively. When the largest covariate value is multiplied by 1.10 (1.20) [1.30], we obtain 0.456, -0.846 and -0.972 (0.673, -0.891 and -0.985) [-0.985, -0.930 and -0.996]. It is clear that the local influence analysis indicates that observation #18 is atypical with progressively more emphasis as it becomes progressively more atypical.

2.3.2 Misspecification detection

We performed a set of Monte Carlo simulations to evaluate the finite sample performance of the RESET misspecification test. The number of Monte Carlo replications is 10,000 and the sample sizes are $n \in \{50, 100, 150, 200, 250\}$. We implement the misspecification test using the generalized likelihood ratio (LR), Wald, Score (S) and gradient (G) tests. Due to the poor performance of the Wald test, we chose to use instead the modified Wald (MW) test proposed by (LEMONTE, 2016). The significance levels are 10% and 5%. In each Monte Carlo replication, we randomly generated $y_i \sim \text{Gumbel}(\mu_i, \sigma_i)$ using the model structure given in Equations (2.2) and (2.3). Here, the values of x_{2i} are selected as random standard uniform draws. The tests null rejection rates are presented in Table 9. We note that the tests are considerably size-distorted when the sample size is small ($n = 50$); such size distortions become smaller as the sample size increases. The tests that use the generalized likelihood ratio, modified Wald and gradient criteria are liberal whereas and that based on the score statistic is slightly conservative. The latter is the most accurate in small samples. The worst performer is the modified Wald test for which an extremely large sample size is needed for the test size distortion to become small. Even though we do not present results for very large sample sizes, we note that the modified Wald test null rejection rate at the 5% is 6.5% when $n = 2500$. That is, the test null rejection rates slowly

converge to the test significance levels.

Table 9 – Misspecification test, null rejection rates.

n	LR		S		MW		G	
	10%	5%	10%	5%	10%	5%	10%	5%
50	0.1516	0.0826	0.0857	0.0521	0.2105	0.1993	0.1445	0.0917
100	0.1320	0.0637	0.0912	0.0526	0.2090	0.1745	0.1366	0.0795
150	0.1201	0.0640	0.0848	0.0530	0.1745	0.1612	0.1054	0.0689
200	0.1146	0.0682	0.0975	0.0479	0.1586	0.1264	0.1076	0.0637
250	0.1037	0.0578	0.0933	0.0457	0.1487	0.1141	0.1049	0.0590

We have also performed simulations to estimate the tests powers under different types of model misspecification. Here, we seek to determine whether the test is capable of detecting that the model is incorrectly specified. We consider four different schemes for data generation mechanisms, denoted by $MS1$, $MS2$, $MS3$ and $MS4$:

$MS1$ $\mu_i = \beta_1 - \exp(\beta_2 x_{2i}) + \beta_3(x_{2i} \times x_{3i})$ and $\log(\sigma_i) = \theta_1 + \theta_2 x_{2i} + \theta_3(x_{2i} \times x_{3i})$;

$MS2$ $\mu_i = \beta_1 - \exp(\beta_2 x_{2i}) + \beta_3 x_{4i}$ and $\log(\sigma_i) = \theta_1 + \theta_2 x_{2i} + \theta_3 x_{4i}$;

$MS3$ $\mu_i = \beta_1 - \beta_2 x_{2i}$ and $\log(\sigma_i) = \theta_1 + \theta_2 \sin(x_{2i})$;

$MS4$ $\mu_i = \beta_1 - \exp(\beta_2 x_{3i})$ and $\log(\sigma_i) = \theta_1 + \theta_2 x_{3i}$;

$MS5$ $\mu_i = (\beta_1 - \exp(\beta_2 x_{2i}))^{\phi}$ and $\log(\sigma_i) = \theta_1 + \theta_2 x_{2i}$.

We use $\beta_3 = 2.5$, $\theta_3 = -0.7$ and $\phi = 1.5$. The values of x_{3i} and x_{4i} are selected as random $U(0, 1.5)$ and $U(1, 2)$ draws, respectively. Schemes $MS6$ and $MS7$ are similar to Scheme $MS5$ with the single difference that now $\phi = 2.0$ and $\phi = 2.5$, respectively. The final scheme, $MS8$, uses Equation (2.2) and (2.3) but estimation is carried out by taking σ to be constant.

We note that regardless of the scheme used for data generation, we fitted the model given in Equations (2.2) and (2.3). Since some tests are liberal, all tests are now performed using exact (estimated in the size simulations) critical values. We shall thus compare powers of tests that are adjusted to have the correct size.

The tests nonnull rejection rates are presented in Table 10. The tests are more (less) powerful under Scheme $MS1$ ($MS2$). It is also noteworthy that the tests are typically considerably more powerful when they are carried after augmenting both submodels.

Table 10 – Misspecification test, nonnull rejection rates.

Scheme	n	LR		S		MW		G	
		10%	5%	10%	5%	10%	5%	10%	5%
	50	0.8230	0.7318	0.8076	0.7141	0.8026	0.6692	0.8046	0.6849
	100	0.8875	0.8186	0.8722	0.7956	0.9013	0.8369	0.8888	0.8196
$MS1$	150	0.5488	0.4190	0.5609	0.4348	0.5391	0.3972	0.5485	0.4185
	200	0.9087	0.8425	0.9120	0.8477	0.9073	0.8381	0.9100	0.8463
	250	0.9917	0.9817	0.9937	0.9853	0.9899	0.9744	0.9917	0.9818
	50	0.3830	0.2654	0.3662	0.2441	0.3891	0.2768	0.3728	0.2546
	100	0.4911	0.3819	0.5331	0.4272	0.4536	0.3484	0.4933	0.3838
$MS2$	150	0.6439	0.4909	0.5176	0.3251	0.5352	0.4309	0.4916	0.3433
	200	0.7071	0.6878	0.7947	0.6230	0.7853	0.7188	0.7234	0.6129
	250	0.9700	0.9456	0.9709	0.9472	0.9733	0.9559	0.9737	0.9548
	50	0.5348	0.3967	0.4736	0.3380	0.5735	0.4357	0.5152	0.3708
	100	0.6099	0.4686	0.5531	0.4048	0.6597	0.5274	0.6073	0.4658
$MS3$	150	0.6553	0.5230	0.6090	0.4703	0.7052	0.5888	0.6578	0.5247
	200	0.6834	0.5492	0.6287	0.4812	0.7403	0.6249	0.6874	0.5545
	250	0.8476	0.7470	0.7947	0.6631	0.8890	0.8144	0.8491	0.7495
	50	0.4715	0.3619	0.5106	0.3703	0.4726	0.3747	0.4384	0.3254
	100	0.5582	0.4483	0.5878	0.4761	0.5585	0.4457	0.5470	0.4339
$MS4$	150	0.6108	0.5107	0.6462	0.5470	0.6167	0.5132	0.6045	0.5037
	200	0.6370	0.5429	0.6798	0.5829	0.6333	0.5485	0.6261	0.5357
	250	0.6655	0.5671	0.7035	0.6112	0.6449	0.5520	0.6539	0.5590
	50	0.4980	0.3491	0.4032	0.2530	0.6656	0.5845	0.5705	0.4578
	100	0.7327	0.6074	0.6901	0.5495	0.8112	0.7325	0.7627	0.6566
$MS5$	150	0.8944	0.8167	0.8795	0.7956	0.9296	0.8881	0.9096	0.8506
	200	0.9431	0.8900	0.9374	0.8745	0.9622	0.9339	0.9527	0.9113
	250	0.9835	0.9650	0.9824	0.9591	0.9892	0.9792	0.9865	0.9712
	50	0.7512	0.6893	0.7723	0.7619	0.8432	0.8623	0.8982	0.7665
	100	0.7952	0.7623	0.7923	0.7814	0.8954	0.8893	0.8593	0.7791
$MS6$	150	0.8268	0.8074	0.8125	0.8012	0.9028	0.9102	0.8125	0.8001
	200	0.9202	0.8890	0.9647	0.9430	0.9317	0.9085	0.9186	0.8895
	250	1.0000	0.9997	0.9999	0.9997	1.0000	0.9999	1.0000	0.9998
	50	0.4937	0.3451	0.3962	0.2473	0.6584	0.5790	0.5653	0.4522
	100	0.7155	0.5855	0.6696	0.5177	0.7933	0.7165	0.7465	0.6337
$MS7$	150	0.8553	0.7573	0.8332	0.7257	0.9032	0.8485	0.8762	0.8004
	200	0.8899	0.8086	0.8781	0.7886	0.9222	0.8718	0.9053	0.8348
	250	0.9762	0.9495	0.9731	0.9416	0.9839	0.9685	0.9803	0.9580
	50	0.2720	0.1852	0.2956	0.2121	0.2255	0.1274	0.2487	0.1574
	100	0.4243	0.3072	0.4195	0.2979	0.3928	0.2618	0.4008	0.2729
$MS8$	150	0.4318	0.3024	0.4441	0.3151	0.4009	0.2580	0.4114	0.2734
	200	0.6274	0.5122	0.6590	0.5460	0.6077	0.4752	0.6165	0.4893
	250	0.6964	0.5738	0.7210	0.6044	0.6828	0.5464	0.6867	0.5572

2.3.3 Nonnested hypothesis tests

We performed a set of Monte Carlo simulations to evaluate the finite sample performance of the J and MJ nonnested hypothesis tests. The number of Monte Carlo replications is 10,000 and the sample sizes are $n \in \{50, 100, 150, 200, 250\}$. All testing inferences are carried out using the likelihood ratio criterion. The significance levels are 10% and 5%. The first simulation study was performed to evaluate the J test null behavior. The following models are used:

$M1$ $\mu_i = \beta_1 + \beta_2 x_{2i}$ and $\log(\sigma_i) = \theta_1$;

$S1$ $\mu_i = \beta_1$ and $\log(\sigma_i) = \theta_1 + \theta_2 z_{2i}$;

$MS1$ $\mu_i = \beta_1 + \beta_2 x_{2i}$ and $\log(\sigma_i) = \theta_1 + \theta_2 z_{2i}$;

$M2$ $\mu_i = \beta_1 + \beta_2 x_{3i}$ and $\log(\sigma_i) = \theta_1$;

$S2$ $\mu_i = \beta_1$ and $\log(\sigma_i) = \theta_1 + \theta_2 z_{3i}$;

$MS2$ $\mu_i = \beta_1 + \beta_2 x_{3i}$ and $\log(\sigma_i) = \theta_1 + \theta_2 z_{3i}$.

Models $M1$ and $M2$ only contain regressors in the submodels for $\boldsymbol{\mu}$, Models $S1$ and $S2$ only contain regressors in the submodels for $\boldsymbol{\sigma}$, and Models $MS1$ and $MS2$ contain regressors in both submodels.

The values of x_{2i}, x_{3i}, z_{2i} and z_{3i} are obtained, respectively, as random draws from the following distributions: Gamma$(3,2)$, $N(0,1)$, t_4, $U(1,2)$, Beta$(2,7)$ and $U(0.5,1.5)$. The values of $\beta_1, \beta_2, \theta_1$ and θ_2 are $2.7, 1.3, 1.8, 1.2$, respectively. In the size simulations, the true data generating processes are models $M1, S1, MS1$ for each scheme. We consider three separate situations, namely: (i) we test $M1$ vs. $M2$ (Case 1), (ii) we test $S1$ vs. $S2$ (Case 2) and (iii) we test $MS1$ vs. $MS2$ (Case 3).

The test null rejection rates are presented in Table 11. We note that the test is considerably size-distorted when the sample size is small ($n = 50$) under Cases 1 and 2; such distortions become smaller as the sample size increases. The test is most accurate when only the models for $\boldsymbol{\sigma}$ contain regressors (Case 2). For instance, under Case 2, with $n = 50$ ($n = 100$) and at the 10% significance level, the test null rejection rate is 10.4% (10.1%).

We performed similar simulations to evaluate the test power, i.e., the test ability to detect that the model under test is not correct. We consider the same cases as before, but we now use Models $M2$ (Case 1), $S2$ (Case 2) and $MS2$ (Case 3) as the true data generating mechanisms. Since the test is sometimes liberal, all tests are now carried out using exact (estimated from the size simulations) critical values. Thus, we evaluate the powers of tests that have the correct size.

Table 11 – J test, null rejection rates.

n	Case 1		Case 2		Case 3	
	10%	5%	10%	5%	10%	5%
50	0.137	0.064	0.104	0.042	0.150	0.082
100	0.086	0.040	0.101	0.068	0.128	0.054
150	0.098	0.054	0.100	0.052	0.119	0.062
200	0.080	0.043	0.121	0.064	0.101	0.054
250	0.096	0.032	0.104	0.051	0.106	0.051

The test nonnull rejection rates are presented in Table 12. The test power increases with the sample size in all three scenarios. For $n = 200$ and at the 10% significance level, the nonnull rejection rate for Case 1 (Case 2) [Case 3] is 90.03% (89.25%) [95.32%].

Table 12 – J test, nonnull rejection rates.

n	Case 1		Case 2		Case 3	
	10%	5%	10%	5%	10%	5%
50	0.4248	0.3084	0.4123	0.3059	0.4837	0.3686
100	0.6774	0.5547	0.6617	0.5487	0.7455	0.6437
150	0.8131	0.7249	0.7981	0.7072	0.8880	0.8225
200	0.9003	0.8391	0.8925	0.8299	0.9532	0.9162
250	0.9520	0.9104	0.9405	0.8980	0.9823	0.9648

We shall now move to MJ testing inference and consider three candidate models, i.e., $N = 3$. The three models only differ in the independent variable they use. The following models are considered:

$m1$ $\mu_i = \beta_1 + \beta_2 x_{2i}$ and $\log(\sigma_i) = \theta_1$;

$s1$ $\mu_i = \beta_1$ and $\log(\sigma_i) = \theta_1 + \theta_2 z_{2i}$;

$ms1$ $\mu_i = \beta_1 + \beta_2 x_{2i}$ and $\log(\sigma_i) = \theta_1 + \theta_2 z_{2i}$;

$m2$ $\mu_i = \beta_1 + \beta_2 x_{3i}$ and $\log(\sigma_i) = \theta_1$;

$s2$ $\mu_i = \beta_1$ and $\log(\sigma_i) = \theta_1 + \theta_2 z_{3i}$;

$ms2$ $\mu_i = \beta_1 + \beta_2 x_{3i}$ and $\log(\sigma_i) = \theta_1 + \theta_2 z_{3i}$;

$m3$ $\mu_i = \beta_1 + \beta_2 x_{4i}$ and $\log(\sigma_i) = \theta_1$;

$s3$ $\mu_i = \beta_1$ and $\log(\sigma_i) = \theta_1 + \theta_2 z_{4i}$;

$ms3$ $\mu_i = \beta_1 + \beta_2 x_{4i}$ and $\log(\sigma_i) = \theta_1 + \theta_2 z_{4i}$.

The true values for $\beta_1, \beta_2, \theta_1, \theta_2$ are $1.7, 2.3, 1.8$ and -0.7 respectively. The val-

ues of the covariates are selected as in the previous set of simulations, i.e., as in in the simulations that dealt with the J test. We test $\mathscr{H}_0$: One of the candidate models is correct vs. $\mathscr{H}_1$: No candidate model is correct. Under Cases 1, 2 and 3, the correct model is, respectively, $m1$, $s1$ and $ms1$. The test null rejection rates are presented in Table 13. The only configuration in which the test does not work well in when $n = 50$ (small sample) under Case 3; in that setting, the test is considerably liberal. In all other cases, the test displays good control of the type I error frequency. For instance, when $n = 50$ ($n = 100$) [$n = 250$] and at the 10% significance level under Case 1 (i.e., $m1$ is the true model), the test null rejection rate is 10.8% (10.2%) [9.6%]. The test null rejection rates converge to the corresponding significance levels as the sample increases.

Table 13 – *MJ* test, null rejection rates.

n	Case 1		Case 2		Case 3	
	10%	5%	10%	5%	10%	5%
50	0.108	0.064	0.078	0.036	0.194	0.118
100	0.102	0.050	0.104	0.066	0.110	0.056
150	0.116	0.054	0.086	0.056	0.130	0.068
200	0.091	0.046	0.102	0.046	0.100	0.034
250	0.096	0.051	0.096	0.054	0.096	0.054

As noted earlier, the *MJ* statistic is the minimal J statistic. When the null hypothesis is not rejected, it can be used as a model selection criterion. We computed the frequencies of correct model selection using the *MJ* statistic as a model selection criterion (given that $\mathscr{H}_0$ was not rejected). The selected model is that with the smallest J test statistic, i.e., the model that corresponds to the *MJ* test statistic. We only consider the Monte Carlo replications in which the null hypothesis was not rejected. The results are presented in Table 14 (entries are percentages). The results show that model selection works quite well when performed using the *MJ* test statistic. Under Cases 1 and 3, the correct model was always selected. Under Case 2, the rate of success range from 70.23% ($n = 50$) to 99.27% ($n = 250$).

Simulations were also performed to evaluate the power of the *MJ* test. Here, the data are generated according to a model that does not belong to the set of candidate models. We performed data generation by replacing the predictor variables with $x_{5i} = x_{2i} \times x_{4i}$ and $z_{5i} = z_{2i} \times z_{4i}$. That is, the following models were used for data generation (Cases 1, 2 and 3, respectively):

$$m4 : \mu_i = \beta_1 + \beta_2 x_{5i}$$

Table 14 – Frequencies (%) of correct model selection using the *MJ* statistic as a model selection criterion (conditional on the null hypothesis not being rejected).

n	Case 1	Case 2	Case 3
50	100.00	70.23	100.00
100	100.00	89.04	100.00
150	100.00	95.49	100.00
200	100.00	98.30	100.00
250	100.00	99.27	100.00

$$\begin{aligned} &\log(\sigma_i) = \theta_1; \\ s4: &\mu_i = \beta_1 \\ &\log(\sigma_i) = \theta_1 + \theta_2 z_{5i}; \\ ms4: &\mu_i = \beta_1 + \beta_2 x_{4i} \\ &\log(\sigma_i) = \theta_1 + \theta_2 z_{5i}. \end{aligned}$$

The test nonnull rejection rates are presented in Table 15. When the sample is quite small ($n = 50$), the test is less powerful under Case 2 (i.e., data generated from Model $s4$). For sample sizes ranging from $n = 100$ to $n = 250$ and at the 10% (5%) significance level, the estimated nonnull rejection rates range from approximately 61% (approximately 53%) to 100% (nearly 100%). That is, as long as the sample size is not too small, the test seems to be adequately powerful.

Table 15 – *MJ* test, nonnull rejection rates.

n	Case 1		Case 2		Case 3	
	10%	5%	10%	5%	10%	5%
50	0.5125	0.4371	0.3546	0.2252	0.6969	0.5865
100	0.6137	0.5370	0.7189	0.6029	0.9322	0.8854
150	0.7350	0.6783	0.8824	0.8084	0.9882	0.9762
200	0.7928	0.7432	0.9541	0.9175	0.9983	0.9957
250	0.8416	0.8029	0.9844	0.9659	1.0000	0.9996

The numerical evidence presented above indicates that the two nonnested hypothesis tests considered in this dissertation can be used to select a GEVNRM from a set of nonnested models. Provided that the sample size is not too small, such inferences are typically accurate and trustworthy.

2.4 EMPIRICAL APPLICATION

We shall now use the GEVNRM in an analysis of real, observed data. To that end, we shall consider the data presented by (HUET *et al.*, 2004). Such data relate to the growth of winter wheat and highlight the differences in dry weights of the wheat tillers and stems. The explanatory variable x is measured on a cumulative degree-days scale which is an integral over time of all temperatures at which the wheat is exposed; only temperatures that exceed the minimum temperature at which wheat can develop are considered. Temperatures are measured in degrees Celsius and time is measured in days, the initial time being determined by the physiological state of the wheat. Plants growing on $n = 18$ randomly chosen small areas of about $0.15m^2$ are harvested each week and the dry weights of the tillers for plants harvested from each area are measured in milligrams. Our interest lies in modeling the behavior of dry weight (y) of tillers which we assume that follows an extreme value distribution. We shall use the same model as (BARRETO-SOUZA; VASCONCELLOS, 2011):

$$\mu_i = \beta_1 + \exp(\beta_2 + \beta_3 x_i), \tag{2.4}$$

$$\log(\sigma_i) = \theta_1 x_i. \tag{2.5}$$

The maximum likelihood parameter estimates (standard errors in parentheses) are $\hat{\beta}_1 = 81.5575$ (13.5103), $\hat{\beta}_2 = -2.7405$ (0.9088), $\hat{\beta}_3 = 0.0140$ (0.0013) and $\hat{\theta}_1 = 0.0066$ (0.0002). The residual half-normal plots with simulated envelopes are given in Figures 16 (standardized residual) and 17 (deviance residual). It is clear that most residuals lie inside the confidence bands, thus indicating that the distributional assumptions hold.

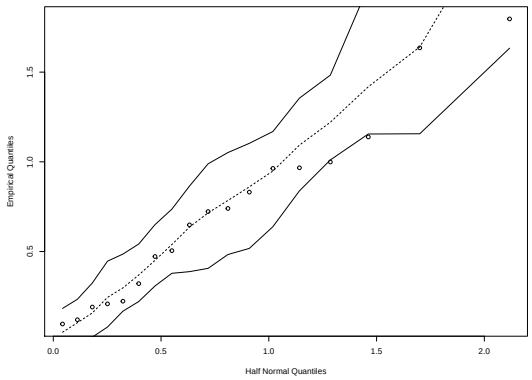

Figure 16 – Half normal plot, standardized residual; growth of winter wheat data.

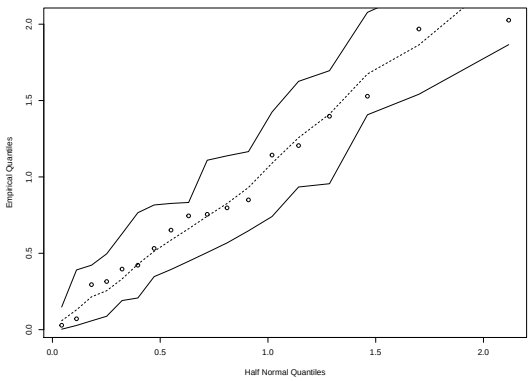

Figure 17 – Half normal plot, deviance residual; growth of winter wheat data.

We also performed the RESET misspecification test for model using the square of the predicted values ($\hat{\eta}_1^2$ and $\hat{\eta}_2^2$) as testing variables. The test p-values based on the likelihood ratio, score, modified Wald and gradient criteria are, respectively, 0.1793, 0.3758, 0.2349 and 0.3296. Hence, the null hypothesis of correct model specification is not rejected at the usual significance levels. That is, there is no evidence of model misspecification.

Figure 18 contains local influence plots constructed using the three perturbation schemes discussed earlier. It is noteworthy that the most (locally) influential data points are observations #2 and #17. We computed the generalized Cook distance for each observation; see Figure 19. The results indicate that observations #2, #16 and #17 are atypical. However, observation #16 does not appear to be atypical when both parameters are considered simultaneously. The generalized leverage measures are presented in Figure 20. They indicate that observation #2 is the most atypical data point.

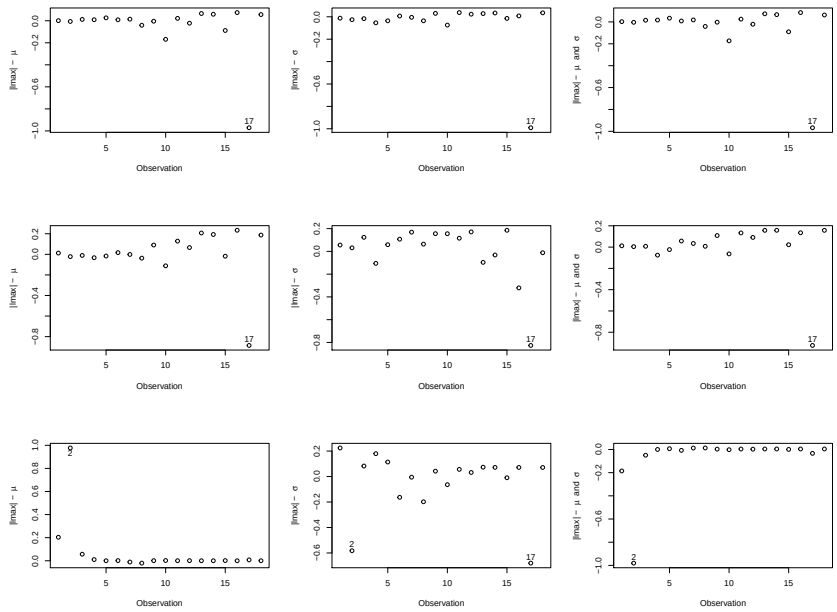

Figure 18 – Local influence measures for μ (first column), σ (middle column), and for μ and σ simultaneously (third column); the first, second and third rows relate to perturbations to the covariate, cases and response values.

Overall, observations #2 and #17 stand out. We note that their response values are considerably smaller than those of other data points with similar covariate values. The 2nd and 17th response values are considerably different from the corresponding fitted values: the fitted value for observation #2 exceeds the observed value by 20.67%, the corresponding figure for observation #17 being 21.81%. Such data points are worthy of further investigation.

We removed the atypical data points (#2 and #17, individually and jointly) from the

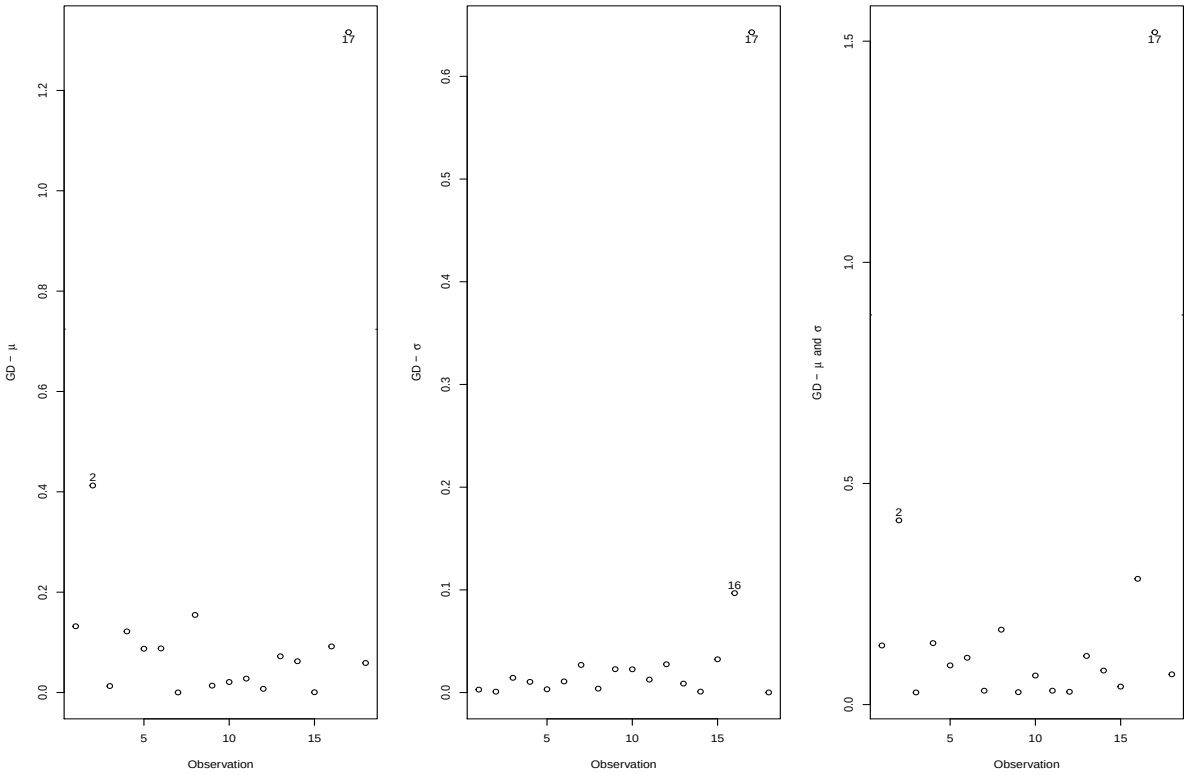

Figure 19 – Generalized Cook's distances for μ, for σ, and for μ and σ simultaneously.

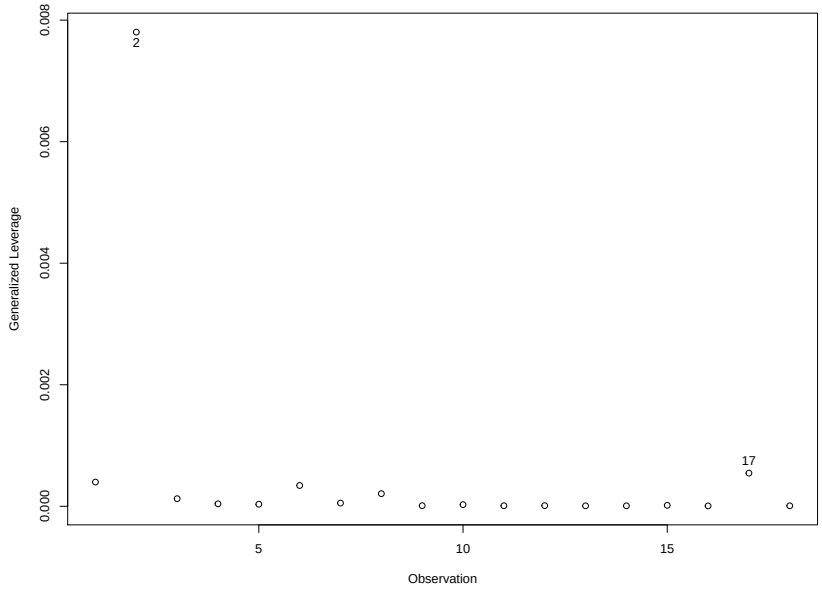

Figure 20 – Generalized leverage for μ, for σ, and for μ and σ simultaneously.

data and estimated the model parameters again. Table 16 presents the point estimates together with the relative changes (%) in such estimates. It is noteworthy that the estimates of β_1 and β_2 change considerably when the atypical points are not in the data. Data point #2 appears to be particularly influential. For example, when observation #2 is removed from the data, $\hat{\beta}_1$ and $\hat{\beta}_2$

increase by nearly 26% and by approximately 22.5%, respectively. When the data do not contain observations #2 and #17, such relative changes jump to nearly 30% and over 41%.

Table 16 – Parameter estimates and relative changes (%) in such estimates based on incomplete data; growth of winter wheat data.

Observation(s)	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\theta}_1$	RC_{β_1}	RC_{β_2}	RC_{β_3}	RC_{θ_1}
2	102.64297	−3.35762	0.01491	0.00657	25.85	22.52	6.21	−0.47
17	84.48671	−3.25823	0.01486	0.00616	3.59	18.89	5.82	−6.75
2, 17	105.63374	−3.87697	0.01573	0.006074	29.52	41.47	12.04	−8.08

We performed the RESET misspecification test. Recall that we test the null hypothesis that the model is correctly specified against the alternative hypothesis that the model specification is in error. Since the sample size is small, we used parametric bootstrap resampling to obtain bootstrap p-values on the basis of 1,000 artificial samples. The null hypothesis of correct model model specification is not rejected at the usual significance levels. The bootstrap likelihood ratio test p-value equals 0.19020; when the most influential observation (#2) is not in the data, such a p-value jumps to 0.41078. That is, the presence of such an observation in the data increases the evidence against the null hypothesis, but not to the extent of reversing the conclusion that the model is correctly specified.

Finally, we considered a second model for the data at hand, namely: $\mu_i = \beta_1 + \exp(\beta_2 + \beta_3 x_i)$ and $\sqrt{\sigma_i} = \theta_1 x_i$. Notice that the alternative model uses a different link function in the second submodel. The two models are thus nonnested. We implemented the J test using the likelihood ratio criterion. Since the sample size is small, the tests were performed using parametric bootstrap resampling. The number of bootstrap replications is 1,000. The p-value of the test of the model given in Equations (2.4) and (2.5) against the alternative model are very large regardless of whether we use the complete data (p-value: 0.99800) or the incomplete data (p-value when observation 2 is not in the data: 0.99900; p-value when observation 17 is not in the data: 0.99701). Hence, such a model is not rejected. In contrast, the alternative model is rejected at the usual significance levels since the p-value for testing such a model is small: 0.00100. The same p-value is obtained when the two incomplete data sets are used.

2.5 CONCLUDING REMARKS

Extreme value theory is quite useful for modeling extremal events. The behavior of such events are oftentimes impacted by other variables and such dependence is captured using a regression framework. In this paper we considered the generalized extreme value nonlinear regression model. Our focus was on influence diagnostics and model validation. That is, we developed and presented tools that can be used by practitioners after parameter estimation has been carried out. First, we defined two residuals for use with the general extreme value regression model. Such residuals can be used, e.g., to construct residual half-normal plots with simulated envelopes. Second, we derived a generalized leverage measure, which can be used to detect atypical data points, especially observations that are atypical in the regressors dimension. Third, we obtained an expression for Cook's distance which can be used to detect influential observations using a global influence analysis. Fourth, we developed local influence analysis based on three different local perturbation schemes. Such results can also be used to detect influential observations. Fifth, we presented a test that can be used to determine whether a fitted general extreme value regression model is correctly specified. Sixth, we presented two tests that can be used to select a model from a set of nonnested general extreme value regression models. Additionally, we presented the results of several numerical evaluations that were carried out to evaluate the performances of the different tools considered in our paper. In particular, we applied the diagnostic tools developed in the paper to a simulated dataset. It was clear that such diagnostic tools were able to identify influential observations and also observations that are atypical in the regressors dimension. We reported Monte Carlo evidence on the finite sample performances of the misspecification and nonnested hypothesis tests. Finally, we presented and discussed an empirical application.

REFERENCES

ATKINSON, A. C. **Plots, Transformations and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis**. New York: Oxford University Press, 1985. v. 1. (Oxford Statistical Science Series, v. 1). ISBN 978-0198533719.

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

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.

BEZANSON, J.; EDELMAN, A.; KARPINSKI, S.; SHAH, V. B. Julia: A fresh approach to numerical computing. **SIAM Review**, v. 59, n. 1, p. 65–98, 2017.

BEZANZON, J.; KARPINSKI, S.; SHAH, V.; EDELMAN, A. Julia: A fast dynamic language for technical computing. In: **Lang.NEXT**. [s.n.], 2012. Disponível em: <http://julialang.org/images/lang.next.pdf>.

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**, Taylor & Francis, v. 40, n. 6, p. 1172–1188, 2013.

CASTILLO, E. **Extreme Value Theory in Engineering, Statistical Modeling and Decision Science**. Boston: Academic Press, 1988. v. 1.

CHERNOFF, H.; GASTWIRTH, J. L.; JOHNS, M. V. Asymptotic distribution of linear combinations of functions of order statistics with applications to estimation. **Annals of Mathematical Statistics**, The Institute of Mathematical Statistics, v. 38, n. 1, p. 52–72, 1967.

COLES, S.; PERICCHI, L. R.; SISSON, S. A fully probabilistic approach to extreme rainfall modeling. **Journal of Hydrology**, Elsevier, v. 273, n. 1-4, p. 35–50, 2003.

COOK, D. R. Assessment of local influence. **Journal of the Royal Statistical Society B**, Royal Statistical Society, v. 48, n. 2, p. 133–169, 1986.

COOK, R. D.; WEISBERG, S. **Residuals and Influence in Regression**. New York: Chapman & Hall, 1982. ISBN 9780412242809.

CORSINI, G.; GINI, F.; GRECO, M. V.; VERRAZZANI, L. Cramer-Rao bounds and estimation of the parameters of the Gumbel distribution. **IEEE Transactions on Aerospace and Electronic Systems**, IEEE, v. 31, n. 3, p. 1202–1204, 1995.

DAVIDSON, R.; MACKINNON, J. G. Several tests for model specification in the presence of alternative hypotheses. **Econometrica**, v. 49, n. 3, p. 781–793, 1981.

DEKKERS, A. L. M.; EINMAHL, J. H. J.; HAAN, L. D. A moment estimator for the index of an extreme-value distribution. **Annals of Statistics**, JSTOR, v. 17, n. 4, p. 1833–1855, 1989.

DOORNIK, J. A. **Object-Oriented Matrix Programming Using Ox**. London, UK, 2007.

ESCALANTE-SANDOUAL, C. Multivariate extreme value distribution with mixed Gumbel marginals. **JAWRA Journal of the American Water Resources Association**, John Wiley & Sons, v. 34, n. 2, p. 321–333, 1998.

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

GREENWOOD, J. A.; LANDWEHR, J. M.; MATALAS, N. C.; WALLIS, J. R. Probability weighted moments: Definition and relation to parameters of several distributions expressable in inverse form. **Water Resources Research**, John Wiley & Sons, v. 15, n. 5, p. 1049–1054, 1979.

GUMBEL, E. J. Les valeurs extrêmes des distributions statistiques. **Annales de l'institut Henri Poincaré**, institut henri poincaré et les presses universitaires de france, v. 5, n. 2, p. 115–158, 1935.

HAGEMANN, A. A simple test for regression specification with non-nested alternatives. **Journal of Econometrics**, v. 166, n. 2, p. 247–254, 2012.

HILL, B. M. A simple general approach to inference about the tail of a distribution. **Annals of Statistics**, JSTOR, v. 3, n. 5, p. 1163–1174, 1975.

HOSKING, J. R. M. Testing whether the shape parameter is zero in the generalized extreme-value distribution. **Biometrika**, JSTOR, v. 71, n. 2, p. 367–374, 1984.

HOSKING, J. R. M.; WALLIS, J. R.; WOOD, E. F. Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. **Technometrics**, JSTOR, v. 27, n. 3, p. 251–261, 1985.

HUET, S.; BOUVIER, A.; POURSAT, M. A.; JOLIVET, E. **Statistical Tools for Nonlinear Regression: A Practical Guide with S-PLUS and R Examples**. 2. ed. New York: Springer, 2004.

III, J. P. Statistical inference using extreme order statistics. **Annals of Statistics**, JSTOR, v. 3, n. 1, p. 119–131, 1975.

JENKINS, A. F. The frequency distribution of the annual maximum (or minimum) values of meteorological events. **Quarterly Journal of the Royal Meteorological Society**, v. 81, p. 158–172, 1955.

KOTZ, S.; NADARAJAH, S. **Extreme Value Distributions: Theory and Applications**. London: Imperial College Press, 2000.

LAURENT, R. T. S.; COOK, R. D. Leverage, local influence and curvature in nonlinear regression. **Biometrika**, Oxford University Press, v. 80, n. 1, p. 99–106, 1993.

LEMONTE, A. J. Robust inference for Birnbaum-Saunders regressions. **Journal of Statistical Computation and Simulation**, Taylor & Francis, v. 86, n. 3, p. 611–622, 2016.

LEMONTE, A. J.; PATRIOTA, A. G. Influence diagnostics in Birnbaum-Saunders nonlinear regression models. **Journal of Applied Statistics**, Taylor & Francis, v. 38, n. 5, p. 871–884, 2011.

LITTELL, R. C. **Encyclopedia of Environmetrics**. [S.l.]: John Wiley & Sons, 2006.

MARTINS, E. S.; STEDINGER, J. R. Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. **Water Resources Research**, John Wiley & Sons, v. 36, n. 3, p. 737–744, 2000.

OLIVEIRA, J. T. Decision results for the parameters of the extreme value (Gumbel) distribution based on the mean and the standard deviation. **Trabajos de Estadística y de Investigación Operativa**, Springer, v. 14, n. 1-2, p. 61–81, 1963.

PARK, H. W.; SOHN, H. Parameter estimation of the generalized extreme value distribution for structural health monitoring. **Probabilistic Engineering Mechanics**, Elsevier, v. 21, n. 4, p. 366–376, 2006.

PREGIBON, D. Logistic regression diagnostics. **Annals of Statistics**, JSTOR, v. 9, n. 4, p. 705–724, 1981.

PRESCOTT, P.; WALDEN, A. T. Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. **Biometrika**, v. 67, n. 3, p. 723–724, 1980.

PRESCOTT, P.; WALDEN, A. T. Maximum likelihood estimation of the parameters of the three-parameter generalized extreme-value distribution from censored samples. **Journal of Statistical Computation and Simulation**, Taylor & Francis, v. 16, n. 3-4, p. 241–250, 1983.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2018. Disponível em: <https://www.R-project.org/>.

RAMSEY, J. B. Tests for specification errors in classical linear least-squares regression analysis. **Journal of the Royal Statistical Society B**, v. 31, n. 2, p. 350–371, 1969.

RAMSEY, J. B.; GILBERT, R. A Monte Carlo study of some small sample properties of tests for specification error. **Journal of the American Statistical Association**, Taylor & Francis, v. 67, n. 337, p. 180–186, 1972.

ROSSI, F.; FIORENTINO, M.; VERSACE, P. Two-component extreme value distribution for flood frequency analysis. **Water Resources Research**, John Wiley & Sons, v. 20, n. 7, p. 847–856, 1984.

SCHWARZMANN, B. A connection between local-influence analysis and residual diagnostics. **Technometrics**, Taylor & Francis, v. 33, n. 1, p. 103–104, 1991.

SEN, P. K.; SINGER, J. M.; LIMA, A. C. P. **From Finite Sample to Asymptotic Methods in Statistics**. [S.l.]: Cambridge University Press, 2009. v. 1.

SHI, D. Multivariate extreme value distribution and its Fisher information matrix. **Acta Mathematicae Applicatae Sinica**, Springer, v. 11, n. 4, p. 421–428, 1995.

SIMIU, E.; HECKERT, N. A.; FILLIBEN, J. J.; JOHNSON, S. K. Extreme wind load estimates based on the Gumbel distribution of dynamic pressures: An assessment. **Structural Safety**, Elsevier, v. 23, n. 3, p. 221–229, 2001.

SINGH, V. P. **Entropy-based parameter estimation in hydrology**. 1. ed. [S.l.]: Springer, 1998. v. 30. 368 p. (Water Science and Technology Library, v. 30).

WANG, X.; DEY, D. K. Generalized extreme value regression for binary response data: An application to B2B electronic payments system adoption. **The Annals of Applied Statistics**, JSTOR, v. 4, n. 4, p. 2000–2023, 2010.

WAYLEN, P.; WOO, M. K. Prediction of annual floods generated by mixed processes. **Water Resources Research**, Wiley Online Library, v. 18, n. 4, p. 1283–1286, 1982.

WEI, B. C.; HU, Y. Q.; FUNG, W. K. Generalized leverage and its applications. **Scandinavian Journal of Statistics**, Blackwell Publishers, v. 24, p. 25–37, 1998.

XIE, F.-C.; WEI, B.-C. Diagnostics analysis for log-Birnbaum-Saunders regression models. **Computational Statistics & Data Analysis**, Elsevier, v. 51, n. 9, p. 4692–4706, 2007.

YUE, S.; OUARDA, T. B. M. J.; BOBéE, B.; LEGENDRE, P.; BRUNEAU, P. The Gumbel mixed model for flood frequency analysis. **Journal of Hydrology**, Elsevier, v. 226, n. 1-2, p. 88–100, 1999.

APPENDIX A – LOCAL INFLUENCE MEASURES

CASE-WEIGHTS PERTURBATION

In this scheme, the perturbation is a weight that represents the contribution for each observation to the log-likelihood function.The no perturbation weight is given by $\omega_0 = 1$ due the construction of the log-likelihood function it is easy to see that the derivatives are given by

$$\frac{\partial^2 \ell}{\partial \beta_j \partial \omega_i} = \frac{\partial \ell}{\partial \beta_j},$$
$$\frac{\partial^2 \ell}{\partial \theta_l \partial \omega_i} = \frac{\partial \ell}{\partial \theta_l}.$$

Which simply reduces to the score function.

Response perturbation

We consider that each response y_i is perturbed as $y_{iw} = y_i + \omega_i S_y$, The no perturbation weight is given by $\omega_0 = 0$. The derivatives are given by

$$\frac{\partial^2 \ell}{\partial \beta_j \partial \omega_i} = \sum_{i=1}^{n} \exp\left(-\frac{y_{iw} - \mu_i}{\sigma_i}\right) \frac{S_y}{\sigma_i^2} \frac{d\mu_i}{d\eta_{1i}} \frac{d\eta_{1i}}{d\beta_j},$$
$$\frac{\partial^2 \ell}{\partial \theta_l \partial \omega_i} = \sum_{i=1}^{n} \frac{S_y}{\sigma_i^2} \left\{1 + \exp\left(-\frac{y_{iw} - \mu_i}{\sigma_i}\right) \left[\left(-\frac{y_{iw} - \mu_i}{\sigma_i^2}\right) \sigma_i - 1\right]\right\} \frac{d\sigma_i}{d\eta_{2i}} \frac{d\eta_{2i}}{d\theta_l}$$

Explanatory variable perturbation

Perturbation on X

Considering the perturbation on μ by making $x_{ij\omega} = x_{ij} + \omega_{ix} S_x$. The no perturbation weight is given by $\omega_{ix} = 0$. S_x is scale factor, usually taken to be the standard deviations of $\boldsymbol{x}_j = (x_{1j}, \ldots, x_{nj})^\top$. The derivatives are given by

$$\frac{\partial^2 \ell}{\partial \beta_j \partial \omega_i} = \sum_{i=1}^{n} -1 \frac{1}{\sigma_i} \left[1 + \exp\left(-\frac{y_i - \mu_{iw}}{\sigma_i}\right)\right] \left[\frac{d\mu_{iw}}{d\eta_{1iw}} \frac{d\eta_{1iw}}{d\beta_j d\omega_i} + \frac{d^2\mu_{iw}}{d\eta_{1iw}^2} \frac{d\eta_{1iw}}{d\beta_j} \frac{d\eta_{1iw}}{d\omega_i}\right]$$
$$- \exp\left(-\frac{y_i - \mu_{iw}}{\sigma_i}\right) \frac{1}{\sigma_i^2} \left(\frac{d\mu_{iw}}{d\eta_{1iw}}\right)^2 \frac{d\eta_{1iw}}{d\beta_j} \frac{d\eta_{1iw}}{d\omega_i},$$
$$\frac{\partial^2 \ell}{\partial \theta_l \partial \omega_i} = \sum_{i=1}^{n} -\frac{1}{\sigma_i^2} \left\{1 + \exp\left(-\frac{y_i - \mu_{iw}}{\sigma_i}\right) [(y_i - \mu_{iw}) - 1]\right\} \frac{d\mu_{iw}}{d\eta_{1iw}} \frac{d\eta_{1iw}}{d\omega_i} \frac{d\sigma_i}{d\eta_{2i}} \frac{d\eta_{2i}}{d\theta_j}.$$

Perturbation on Z

Here, we consider the perturbation on σ by making $z_{ik\omega} = z_{ik} + \omega_{iz}S_z$. The no perturbation weight is given by $\omega_{iz} = 0$. S_z is scale factor, usually taken to be the standard deviations of $\boldsymbol{z}_j = (z_{1k}, \ldots, z_{nk})^\top$. The derivatives are given by

$$\frac{\partial^2 \ell}{\partial \beta_j \partial \omega_i} = \sum_{i=1}^{n} -\frac{1}{\sigma_{iw}^2}\left\{1 - \exp\left(-\frac{y_i-\mu_i}{\sigma_{iw}}\right)\left[1-\left(\frac{y_i-\mu_i}{\sigma_{iw}}\right)\right]\right\}\frac{d\mu_i}{d\eta_{1i}}\frac{d\eta_{1i}}{d\beta_j}\frac{d\sigma_{iw}}{d\eta_{2iw}}\frac{d\eta_{2iw}}{d\omega_i},$$

$$\begin{aligned}\frac{\partial^2 \ell}{\partial \theta_l \partial \omega_i} = \sum_{i=1}^{n} &-\frac{1}{\sigma_{iw}}\left\{1-\left(\frac{y_i-\mu_i}{\sigma_{iw}}\right)\left[1-\exp\left(-\frac{y_i-\mu_i}{\sigma_{iw}}\right)\right]\right\}\left[\frac{d\sigma_{iw}}{d\eta_{2iw}}\frac{d\eta_{2iw}}{d\theta_l d\omega_i}+\frac{d^2\sigma_{iw}}{d\eta_{2iw}^2}\frac{d\eta_{2iw}}{d\theta_l}\frac{d\eta_{2iw}}{d\omega_i}\right]\\
&+\left[\frac{1}{\sigma_{iw}^2}-2\left(\frac{y_i-\mu_i}{\sigma_{iw}^3}\right)+2\exp\left(-\frac{y_i-\mu_i}{\sigma_{iw}}\right)\left(\frac{y_i-\mu_i}{\sigma_{iw}^3}\right)-\exp\left(-\frac{y_i-\mu_i}{\sigma_{iw}}\right)\left(\frac{y_i-\mu_i}{\sigma_{iw}^2}\right)^2\right]\\
&\left(\frac{d\sigma_{iw}}{d\eta_{2iw}}\right)^2\frac{d\eta_{2iw}}{d\theta_l}\frac{d\eta_{2iw}}{d\omega_i}.\end{aligned}$$

Perturbation on X and Z

Here we consider both perturbations presented on previous subsections. The no perturbation weight is given by $\omega_{iz} = \omega_{ix} = 0$. The derivatives are given by

$$\begin{aligned}\frac{\partial^2 \ell}{\partial \beta_j \partial \omega_i} = \sum_{i=1}^{n} &-1\frac{1}{\sigma_i}\left[1+\exp\left(-\frac{y_i-\mu_{iw}}{\sigma_i}\right)\right]\left[\frac{d\mu_{iw}}{d\eta_{1iw}}\frac{d\eta_{1iw}}{d\beta_j d\omega_i}+\frac{d^2\mu_{iw}}{d\eta_{1iw}^2}\frac{d\eta_{1iw}}{d\beta_j}\frac{d\eta_{1iw}}{d\omega_i}\right]\\
&-\exp\left(-\frac{y_i-\mu_{iw}}{\sigma_i}\right)\frac{1}{\sigma_i^2}\left(\frac{d\mu_{iw}}{d\eta_{1iw}}\right)^2\frac{d\eta_{1iw}}{d\beta_j}\frac{d\eta_{1iw}}{d\omega_i}\\
&-\frac{1}{\sigma_{iw}^2}\left\{1-\exp\left(-\frac{y_i-\mu_i}{\sigma_{iw}}\right)\left[1-\left(\frac{y_i-\mu_i}{\sigma_{iw}}\right)\right]\right\}\frac{d\mu_i}{d\eta_{1i}}\frac{d\eta_{1i}}{d\beta_j}\frac{d\sigma_{iw}}{d\eta_{2iw}}\frac{d\eta_{2iw}}{d\omega_i},\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 \ell}{\partial \theta_l \partial \omega_i} = \sum_{i=1}^{n} &-\frac{1}{\sigma_i^2}\left\{1+\exp\left(-\frac{y_i-\mu_{iw}}{\sigma_i}\right)[(y_i-\mu_{iw})-1]\right\}\frac{d\mu_{iw}}{d\eta_{1iw}}\frac{d\eta_{1iw}}{d\omega_i}\frac{d\sigma_i}{d\eta_{2i}}\frac{d\eta_{2i}}{d\theta_j}\\
&-\frac{1}{\sigma_{iw}}\left\{1-\left(\frac{y_i-\mu_i}{\sigma_{iw}}\right)\left[1-\exp\left(-\frac{y_i-\mu_i}{\sigma_{iw}}\right)\right]\right\}\left[\frac{d\sigma_{iw}}{d\eta_{2iw}}\frac{d\eta_{2iw}}{d\theta_l d\omega_i}+\frac{d^2\sigma_{iw}}{d\eta_{2iw}^2}\frac{d\eta_{2iw}}{d\theta_l}\frac{d\eta_{2iw}}{d\omega_i}\right]\\
&+\left[\frac{1}{\sigma_{iw}^2}-2\left(\frac{y_i-\mu_i}{\sigma_{iw}^3}\right)+2\exp\left(-\frac{y_i-\mu_i}{\sigma_{iw}}\right)\left(\frac{y_i-\mu_i}{\sigma_{iw}^3}\right)-\exp\left(-\frac{y_i-\mu_i}{\sigma_{iw}}\right)\left(\frac{y_i-\mu_i}{\sigma_{iw}^2}\right)^2\right]\\
&\left(\frac{d\sigma_{iw}}{d\eta_{2iw}}\right)^2\frac{d\eta_{2iw}}{d\theta_l}\frac{d\eta_{2iw}}{d\omega_i}.\end{aligned}$$