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

Rodney Vasconcelos Fonseca

Bimodal Birnbaum-Saunders statistical modeling

Recife
2017

Rodney Vasconcelos Fonseca

Bimodal Birnbaum-Saunders statistical modeling

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.

Advisor: *Francisco Cribari-Neto*

Recife
2017

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

F676b Fonseca, Rodney Vasconcelos
Bimodal Birnbaum-Saunders statistical modeling / Rodney Vasconcelos Fonseca. – 2017.
78 f.: il., fig., tab.

Orientador: Francisco Cribari Neto.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CCEN, Estatística, Recife, 2017.
Inclui referências.

1. Análise de regressão. 2. Verossimilhança. I. Cribari Neto, Francisco (orientador). II. Título.

519.536 CDD (23. ed.) UFPE- MEI 2017-46

RODNEY VASCONCELOS FONSECA

BIMODAL BIRNBAUM-SAUNDERS STATISTICAL MODELING

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: 10 de fevereiro de 2017.

BANCA EXAMINADORA

Prof. Francisco Cribari Neto
UFPE

Prof. Raydonal Ospina Martínez
UFPE

Prof. Aluísio de Souza Pinheiro
Unicamp

ACKNOWLEDGEMENTS

I would like to thank my family, especially my parents, Giovania and Ricardo, and all my friends for supporting me.
I am very grateful to professor Francisco Cribari and all his invaluable advices, confidence and guidance.
I would also like to thank all professors at UFPE, especially prof. Alex Dias, prof. Audrey Cysneiros, prof. Francisco Cysneiros and prof. Gauss Cordeiro.
I would like to thank my former professors at Universidade Federal do Ceará (UFC), especially prof. Juvêncio Nobre and João Maurício, who helped me immensely during my undergraduate studies.
I am also indebted to my former teachers, especially my math teacher Luciano Nunes.
I am thankful to all colleagues and friends I made at Recife, especially Alejandro, Cristina, Olivia, Vinícius and Yuri. I would not like to be unfair with anyone, so I will not cite more names, but I would like to thank everyone I spent time with at this great city.
I am also indebted to Jucelino and Jonas, who were like brothers to me. Thank you guys very much for all help and support in these two years.
I would like to thank all staff members at CCEN, especially Valeria and her competent assistance.
I would like to thank CNPq for the financial support.

ABSTRACT

The Birnbaum-Saunders distribution has been widely studied in the literature, being used to analyze different kind of data, like the lifetime of objects being exposed to fatigue activity and quantities corresponding to air pollution, for example. It was also extended to more general settings by several authors. This dissertation is composed of two main and independent chapters. In the first work, we investigate problems related to maximum likelihood estimation in a bimodal extension of the Birnbaum-Saunders distribution. We propose a penalization scheme that, when applied to the log-likelihood function, greatly reduces the frequency of convergence failures. Hypothesis testing inference based on the penalized log-likelihood function is investigated. In the second essay, we develop a bimodal Birnbaum-Saunders regression model. We discuss point estimation, interval estimation and hypothesis testing inference. We also propose two residuals and develop local influence analyses. Bootstrap-based prediction intervals are also presented and different model selection criteria for the proposed model are discussed. Additionally, we present results from Monte Carlo simulations and from some empirical applications.

Keywords: Bimodal Birnbaum-Saunders distribution. Diagnostic methods. Nonnested hypothesis test. Penalized likelihood.

RESUMO

A distribuição Birnbaum-Saunders tem sido amplamente estudada na literatura, sendo utilizada para analisar diferentes tipos de dados, como tempo de vida de materias sujeitos à fadiga e quantidades referentes a poluição atmosférica, por exemplo. Essa distribuição foi estendida para modelos mais gerais por diversos autores. Essa dissertação é composta de dois capítulos principais e independentes. No primeiro trabalho, investigamos problemas relacionados com estimação por máxima verossimilhança em uma extensão bimodal da distribuição Birnbaum-Saunders. Propomos um esquema de penalização que, quando aplicado à função de log-verossimilhança, reduz bastante a frequência de falhas de convergência. Inferência através de testes de hipóteses baseada na função de log-verossimilhança penalizada é investigada. No segundo ensaio, desenvolvemos um modelo de regressão Birnbaum-Saunders bimodal. Discutimos sobre inferências por estimação pontual, estimação intervalar e testes de hipóteses. Também propomos dois resíduos e desenvolvemos análise de influência local. Intervalos de predição baseados em reamostragem bootstrap também são apresentados e diferentes critérios de seleção de modelos para o modelo proposto são discutidos. Adicionalmente, apresentamos resultados de simulação de Monte Carlo e algumas aplicações empíricas.

Palavras-chave: Distribuição Birnbaum-Saunders Bimodal. Métodos de diagnóstico. Testes de hipóteses não-encaixados. Verossimilhança penalizada.

LIST OF FIGURES

LIST OF TABLES

CONTENTS

Preliminaries

In this chapter we present a brief outline of this dissertation, which is composed by two main chapters with independent works. The subject of the dissertation is bimodal Birnbaum-Saunders modeling, where we present and develop some new methodologies for bimodal extensions of the distribution initially proposed by Birnbaum and Saunders (1969a). Many variations of the Birnbaum-Saunders distribution have been proposed in the last decades, likewise its applications in different fields of research through a variety of statistical models, as Leiva (2015) presents in a extensive discussion on Birnbaum-Saunders models. Our motivation is to give a contribution on the literature of Birnbaum-Saunders modeling by investigating two bimodal versions of this distribution recently proposed in the literature.

In Chapter 2 we address the issue of performing inference on the parameters of a bimodal extension of the Birnbaum-Saunders distribution proposed by Olmos et al. (2016). We show that maximum likelihood point estimation can be problematic since the standard nonlinear optimization algorithms may fail to converge. To deal with this problem, we penalize the log-likelihood function of this model. The numerical evidence we present show that maximum likelihood estimation based on such penalized function is made considerably more reliable. We also consider hypothesis testing inference based on the penalized log-likelihood function. In particular, we consider likelihood ratio, signed likelihood ratio, score and Wald tests. Bootstrap-based testing inference is also considered and we derive analytical corrections to some tests statistics. Moreover, we use a nonnested hypothesis test to distinguish between two bimodal Birnbaum-Saunders laws.

In Chapter 3 we introduce an extension of the log-linear Birnbaum-Saunders model based on another version of the Birnbaum-Saunders distribution, discussed by Owen and Ng (2015) which is more flexible than the standard Birnbaum-Saunders law since its density may assume both unimodal and bimodal shapes. We show how to perform point estimation, interval estimation and hypothesis testing inferences on the parameters of the regression model we propose. We also present some diagnostic tools, such as residual analysis, local influence, generalized leverage, generalized Cook's distance and model misspecification tests. Additionally, we investigate the usefulness of model selection criteria and the accuracy of prediction intervals for the proposed model.

Each chapter contains results of Monte Carlo simulations and empirical applications. Moreover, the notation and terminology used is consistent within each chapter. The programming routines for Monte Carlo simulations were carried out in the Ox matrix programming language (Doornik, 2009) and all figures in this dissertation were generated in the R programming language (R Core Team, 2016).

Inference in a bimodal Birnbaum-Saunders model

2.1 Introduction

The Birnbaum-Saunders distribution was proposed by Birnbaum and Saunders (1969a) to model failure time due to fatigue under cyclic loading. In such a model, failure follows from the development and growth of a dominant crack. Based on that setup, the authors obtained the following distribution function:

$$F(x) = \Phi\left[\frac{1}{\alpha}\left(\sqrt{\frac{x}{\beta}} - \sqrt{\frac{\beta}{x}}\right)\right], \quad x > 0, \tag{2.1}$$

where $\alpha > 0$ and $\beta > 0$ are shape and scale parameters, respectively, and $\Phi(\cdot)$ is the standard normal distribution function. We write $X \sim \mathcal{BS}(\alpha, \beta)$.

Maximum likelihood estimation of the parameters that index the $\mathcal{BS}$ distribution was first investigated by Birnbaum and Saunders (1969b). Bias-corrected estimators were obtained by Lemonte et al. (2007) and Lemonte et al. (2008). Improved maximum likelihood estimation of the $\mathcal{BS}$ parameters was developed by Cysneiros et al. (2008). Ng et al. (2003) compared the finite-sample performance of maximum likelihood estimators (MLEs) to that of estimators obtained using the modified method of moments. For details on the $\mathcal{BS}$ distribution, its main properties and applications, readers are referred to Leiva (2015).

Several extensions of the $\mathcal{BS}$ distribution have been proposed in the literature aiming at making the model more flexible. For instance, Díaz-García and Leiva (2005) and Sanhueza et al. (2008) used non-Gaussian kernels to extend the $\mathcal{BS}$ model. The $\mathcal{BS}$ distribution was also extended through the inclusion of additional parameters; see, e.g., Díaz-García and Domınguez-Molina (2006), Owen (2006) and Owen and Ng (2015). More recently, extensions of the $\mathcal{BS}$ model were proposed by Bourguignon et al. (2014), Cordeiro et al. (2013), Cordeiro and Lemonte (2014) and Zhu and Balakrishnan (2015). Alternative approaches are the use of scale-mixture of normals, as discussed by Balakrishnan et al. (2009) and Patriota (2012), for example, and the use of mixtures of $\mathcal{BS}$ distributions, as in Balakrishnan et al. (2011). Again, details can be found in Leiva (2015).

A bimodal $\mathcal{BS}$ distribution was proposed by Olmos et al. (2016). The authors used the approach described in Gómez et al. (2011) to obtain a variation of the $\mathcal{BS}$ model that can assume bimodal shapes. Another variant of the $\mathcal{BS}$ distribution that exhibits bimodality was discussed by Díaz-García and Domınguez-Molina (2006) and Owen and Ng (2015), which the latter authors denoted by $\mathcal{GBS}_2$. In their model, bimodality takes place when two parameter values exceed certain thresholds. In what follows we shall work with the $\mathcal{BBS}$ model instead

of the $\mathcal{GBS}_2$ distribution because in the former bimodality is controlled by a single parameter. Even though we shall focus on the $\mathcal{BBS}$ distribution, in some parts of this dissertation we shall consider the $\mathcal{GBS}_2$ law as an alternative model.

A problem with the $\mathcal{BBS}$ distribution we detected is that log-likelihood maximizations based on Newton or quasi-Newtons methods oftentimes fail to converge. In this chapter we analyze some possible solutions to such a problem, such as the use of resampling methods and the inclusion of a penalization term in the log-likelihood function.

As a motivation, consider the data provided by Folks and Chhikara (1978) that consist of 25 observations on runoff amounts at Jug Bridge, in Maryland. Figure 2.1a shows log-likelihood contour curves obtained by varying the values of α and γ while keeping the value of β fixed. Notice that there is a region apparently flat of the profile log-likelihood function, which might be making the optimization process fail to converge. In Figure 2.1b we present similar contour curves for a penalized version of the log-likelihood function. It can be seen that plausible estimates are obtained. We shall return to this application in Section 2.7.

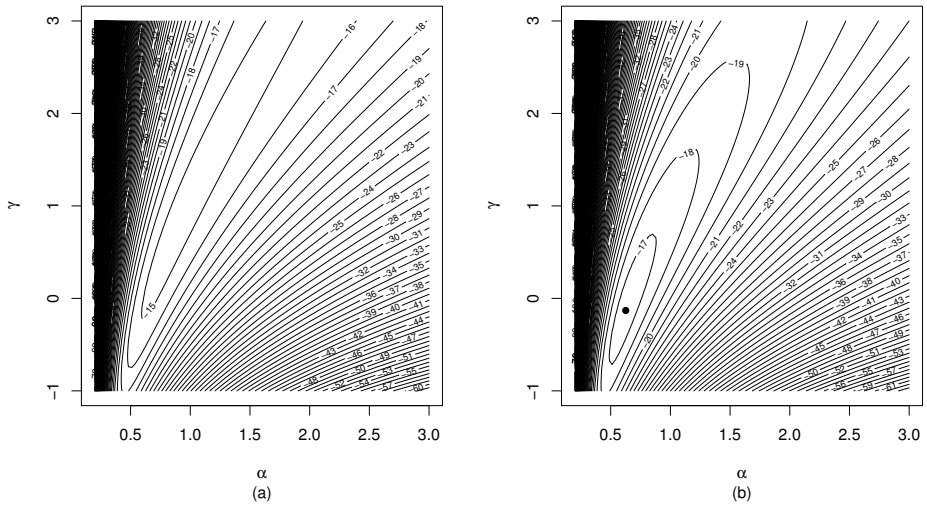

Figure 2.1 Contour curves of the profile log-likelihood of α and γ with $\beta = 0.69$ (fixed) for the runoff amounts data. Panel (a) corresponds to no penalization and panel (b) follows from penalizing the log-likelihood function.

The chief goal of this chapter is to provide a solution to the convergence failure and implausible parameter estimates associated with log-likelihood maximization in the $\mathcal{BBS}$ model. We compare different estimation procedures and propose a penalization term in the log-likelihood function. In particular, regions of the parameter space where the likelihood is flat or nearly flat are heavily penalized. That approach considerably improves maximum likelihood parameter estimation. We also focus on hypothesis testing inference based on the penalized log-likelihood function. For instance, a one-sided hypothesis test is used to test whether the variate follows the $\mathcal{BBS}$ law with two modes. Analytical and bootstrap corrections are proposed to improve the finite sample performances of such test. Moreover, we present nonnested hypothesis tests that can be used to distinguish between two bimodal extensions of the $\mathcal{BS}$ distribution,

the $\mathscr{BBS}$ and $\mathscr{GBS}_2$ models. The finite sample performances of all tests are numerically evaluated using Monte Carlo simulations.

This chapter unfolds as follows. Section 2.2 presents the $\mathscr{BBS}$ distribution and its main properties. We outline some possible solutions to the numerical difficulties associated with $\mathscr{BBS}$ log-likelihood maximization on Section 2.3. Two-sided hypothesis tests in the $\mathscr{BBS}$ model are discussed in Section 2.4. In Section 2.5 we focus on one-sided tests where the main interest lies in detecting bimodality. Section 2.6 describes nonnested hypothesis testing inference. Empirical applications are presented and discussed in Section 2.7. Finally, some concluding remarks are offered in Section 2.8.

2.2 The bimodal Birnbaum-Saunders distribution

The Birnbaum-Saunders distribution proposed by Olmos et al. (2016) can be used to model positive data and is more flexible than the original $\mathscr{BS}$ distribution since it can accommodate bimodality. A random variable X is $\mathscr{BBS}(\alpha, \beta, \gamma)$ distributed if its probability density function (pdf) is given by

$$f(x) = \frac{x^{-3/2}(x+\beta)}{4\alpha\beta^{1/2}\Phi(-\gamma)}\phi(|t|+\gamma), \quad x > 0, \tag{2.2}$$

where $\alpha, \beta > 0$, $\gamma \in \mathbb{R}$, $t = \alpha^{-1}(\sqrt{x/\beta} - \sqrt{\beta/x})$ and $\phi(\cdot)$ is the standard normal pdf. Figure 2.2 shows plots of the density in (2.2) for some parameter values. We note that when $\gamma < 0$ the density is bimodal.

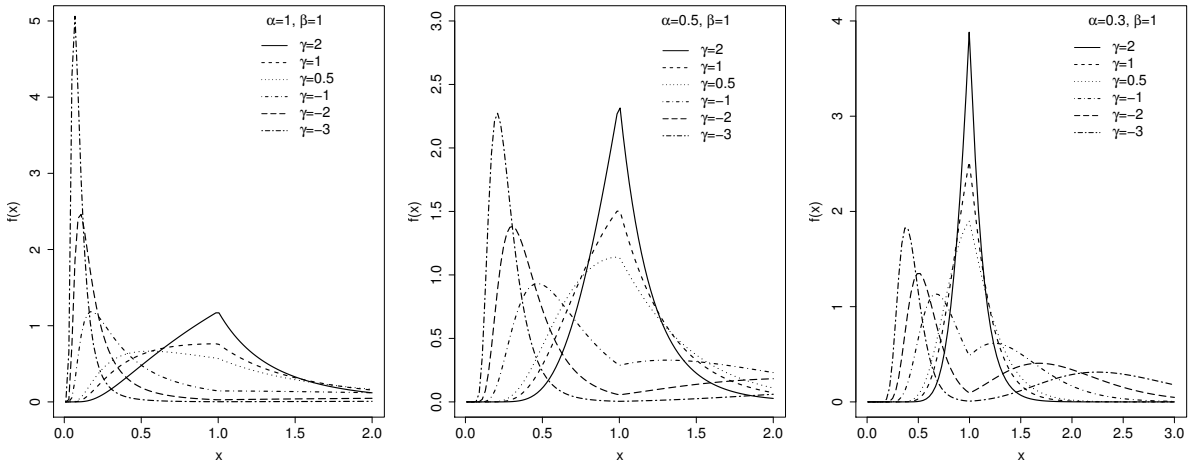

Figure 2.2 $\mathscr{BBS}(\alpha, \beta, \gamma)$ densities for some parameter values.

The cumulative distribution function (cdf) of X is

$$F(x) = \left[\frac{\Phi(t-\gamma)}{2\Phi(-\gamma)}\right]^{I(x,\beta)} \left[\frac{1}{2} + \frac{\Phi(t-\gamma) - \Phi(\gamma)}{2\Phi(-\gamma)}\right]^{1-I(x,\beta)}, \quad x > 0, \tag{2.3}$$

where

$$I(x,\beta) = \begin{cases} 1 & \text{if } x < \beta \\ 0 & \text{if } x \geq \beta \end{cases},$$

and $\Phi(\cdot)$ denotes the standard normal cdf. Some key properties of the $\mathcal{BS}$ distribution also hold for $\mathcal{BBS}$ model, such as proportionality and reciprocity closure, i.e., $aX \sim \mathcal{BBS}(\alpha, a\beta, \gamma)$ and $X^{-1} \sim \mathcal{BBS}\left(\alpha, \beta^{-1}, \gamma\right)$, respectively, where a is a positive scalar.

An expression for the rth ordinary moment of X is

$$\mathbb{E}(X^r) = \frac{\beta^r}{\Phi(-\gamma)} \sum_{k=0}^{r} \sum_{j=0}^{k} \sum_{s=0}^{m} \binom{2r}{2k} \binom{k}{j} \binom{m}{s} \left(\frac{\alpha}{2}\right)^m (-\gamma)^{m-s} d_s(\gamma), \tag{2.4}$$

where $r \in \mathbb{N}$ and $d_a(r)$ is the rth standard normal incomplete moment:

$$d_r(a) = \int_a^\infty t^r \phi(t) dt.$$

Using Equation (2.4) we computed the skewness and kurtosis coefficients of the $\mathcal{BBS}$ distribution to evaluate how γ affects the shape of the $\mathcal{BBS}$ density. Plots of these two measures are presented in Figure 2.3. Note that the largest values of the skewness coefficient correspond to values of γ close to zero. Moreover, we also note that positive values of γ lead to large values of the kurtosis coefficient, which are associated with leptokurtic densities.

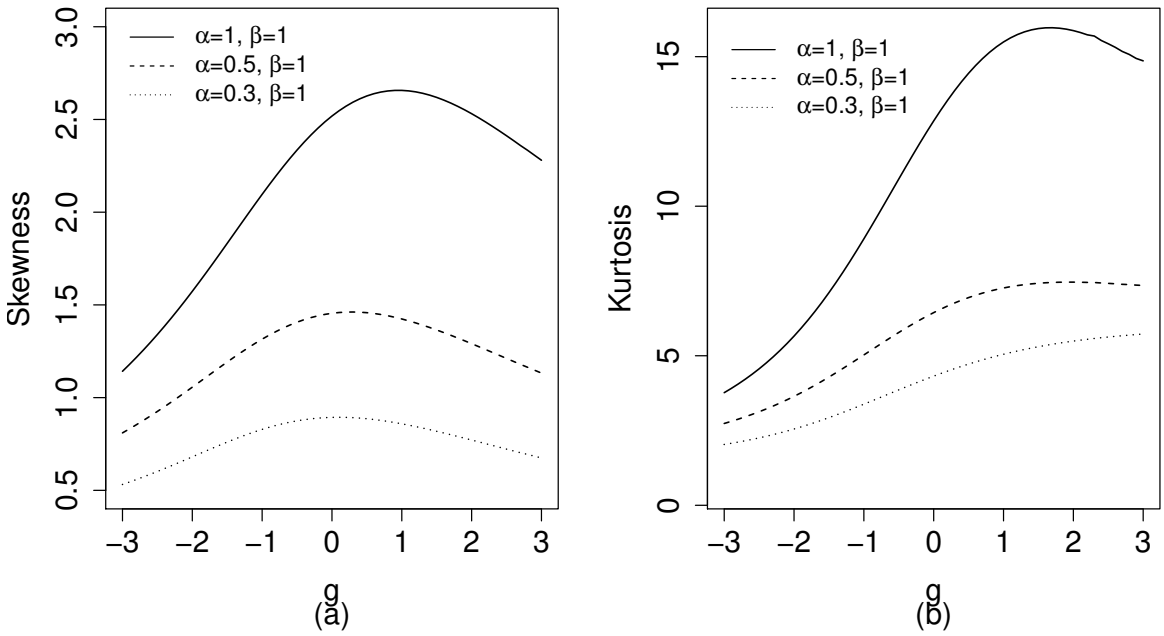

Figure 2.3 $\mathcal{BBS}(\alpha,\beta,\gamma)$ skewness (a) and kurtosis (b) coefficients as a function of γ for some fixed values of α and β.

A useful stochastic representation is $Y = |T| + \gamma$. Here, Y follows the truncated standard normal distribution with support in (γ, ∞), $T = (\sqrt{X/\beta} - \sqrt{\beta/X})/\alpha$ and $X \sim \mathcal{BBS}(\alpha, \beta, \gamma)$. This relationship can be used to compute moments of the $\mathcal{BBS}$ distribution.

2.3 Log-likelihood functions

Consider a row vector $\mathbf{x} = (x_1, \ldots, x_n)$ of independent and identically distributed (iid) observations from the $\mathcal{BBS}(\alpha, \beta, \gamma)$ distribution. Let $\theta = (\alpha, \beta, \gamma)$ be the vector of unknown parameters to be estimated. The log-likelihood function is

$$\ell(\theta) = -n \log\left\{4\alpha\beta^{1/2}\Phi(-\gamma)(2\pi)^{1/2}\right\} - \frac{3}{2}\sum_{i=1}^{n}\log(x_i) + \sum_{i=1}^{n}\log(x_i+\beta) - \frac{1}{2}\sum_{i=1}^{n}(|t_i|+\gamma)^2.$$

Differentiating the log-likelihood function with respect to each parameter we obtain the score function $U_\theta = (U_\alpha, U_\beta, U_\gamma)$, where

$$U_\alpha = \frac{\partial \ell(\theta)}{\partial \alpha} = -\frac{n}{\alpha} + \frac{1}{2}\sum_{i=1}^{n} t_i^2 + \frac{\gamma}{\alpha}\sum_{i=1}^{n}|t_i|,$$

$$U_\beta = \frac{\partial \ell(\theta)}{\partial \beta} = -\frac{n}{2\beta} + \sum_{i=1}^{n}\frac{1}{x_i+\beta} + \sum_{i=1}^{n}\frac{\operatorname{sign}(t_i)(|t_i|+\gamma)}{2\alpha\beta^{3/2}}\left(x_i^{1/2} + \frac{\beta}{x_i^{1/2}}\right),$$

$$U_\gamma = \frac{\partial \ell(\theta)}{\partial \gamma} = n\frac{\phi(\gamma)}{\Phi(-\gamma)} - n\gamma - \sum_{i=1}^{n}|t_i|,$$

and $\operatorname{sign}(\cdot)$ represents the sign function.

The parameters maximum likelihood estimators, namely $\hat{\theta} = (\hat{\alpha}, \hat{\beta}, \hat{\gamma})$, can be obtained by solving $U_\theta = 0$. They cannot be expressed in closed-form and parameter estimates are obtained by numerically maximizing the log-likelihood function using a Newton or quasi-Newton algorithm. To that end, one must specify an initial point for the iterative scheme. We propose using as starting values for α and β their modified method of moments estimates (Ng et al., 2003), and also using $\gamma = 0$ as a starting value; the latter means that the algorithm starts at the $\mathcal{BS}$ law. We used such starting values in the numerical evaluations, and they proved to work well.

Based on several numerical experiments we noted a serious shortcoming: iterative numerical maximization of the $\mathcal{BBS}$ log-likelihood function may fail to converge and may yield implausible parameter estimates. Indeed, that is very likely to happen, especially when $\gamma > 0$. It is not uncommon to obtain very large (thus implausible) $\mathcal{BBS}$ parameter estimates, which is indicative that the likelihood function may be monotone; see Pianto and Cribari-Neto (2011). We shall address this problem in the sections that follow.

2.3.1 Log-likelihood function penalized by the Jeffreys prior

An interesting estimation procedure was proposed by Firth (1993), where the score function is modified in order to reduce the bias of the maximum likelihood estimator. An advantage of this method is that maximum likelihood estimates need not be finite since the correction is

applied in a preventive fashion. For models in the canonical exponential family, the correction can be applied directly to the likelihood function:

$$L^*(\theta|\mathbf{x}) = L(\theta|\mathbf{x})|K|^{1/2},$$

where $|K|$ is the determinant of the expected information matrix. Thus, penalization of the likelihood function entails multiplying the likelihood function by the Jeffreys invariant prior.

Even though the $\mathscr{BBS}$ distribution is not a member of the canonical exponential family, we shall consider the above penalization scheme. In doing so, we follow Pianto and Cribari-Neto (2011) who used the same approach in speckled imagery analysis. We seek to prevent cases of monotone likelihood function that might lead to frequent optimization nonconvergences and implausible estimates. The $\mathscr{BBS}(\alpha,\beta,\gamma)$ expected information matrix was obtained by Olmos et al. (2016). Its determinant is

$$|K| = \left[L_{\beta\beta} + \frac{1}{\alpha^2\beta^2} + \frac{\gamma(\gamma-\omega)}{4\beta^2}\right]\left[\frac{(\gamma-\omega)\omega(3-\gamma\omega-\gamma^2)+2}{\alpha^2}\right],$$

where $\omega = \phi(\gamma)/\Phi(-\gamma)$ and $L_{\beta\beta} = \mathbb{E}\left[(X+\beta)^{-2}\right]$. Thus, the log-likelihood function penalized by the Jeffreys prior can be written as

$$\begin{aligned}\ell^*(\theta) = {} & -n\log\left\{4\alpha\beta^{1/2}\Phi(-\gamma)(2\pi)^{1/2}\right\} - \frac{3}{2}\sum_{i=1}^{n}\log(x_i) + \sum_{i=1}^{n}\log(x_i+\beta) \\ & - \frac{1}{2}\sum_{i=1}^{n}(t_i^2 + 2|t_i|\gamma+\gamma^2) + \frac{1}{2}\log\left[L_{\beta\beta} + \frac{1}{\alpha^2\beta^2} + \frac{\gamma(\gamma-\omega)}{4\beta^2}\right] \\ & + \frac{1}{2}\log\left[\frac{(\gamma-\omega)\omega(3+\gamma(\gamma-\omega))+2}{\alpha^2}\right].\end{aligned}$$

If the likelihood function is monotone, the function becomes very flat for large parameter values and the Jeffreys penalization described above essentially eliminates such parameter range from the estimation. The frequency of nonconvergences taking place and implausible estimates being obtained should be greatly reduced.

2.3.2 Log-likelihood function modified by the better bootstrap

An alternative approach uses the method proposed by Cribari-Neto et al. (2002), where bootstrap samples are used to improve maximum likelihood estimation similarly to the approach introduced by Efron (1990) and known as 'the better bootstrap'. The former, however, does not require the estimators to have closed-form expressions. Based on the sample $\mathbf{x} = (x_1,\ldots,x_n)$ of n observations, we obtain pseudo-samples $\mathbf{x}^*$ of the same size by sampling from $\mathbf{x}$ with replacement. Let P_i^* denote the proportion of times that observation x_i is selected, $i = 1,\ldots,n$. We obtain the row vector $\mathbf{P}^{*b} = (P_1^{*b},\cdots,P_n^{*b})$ for the bth pseudo-sample, $b = 1,\ldots,B$. Now compute

$$\mathbf{P}^*(\cdot) = \frac{1}{B}\sum_{b=1}^{B} P_n^{*b},$$

i.e., compute the vector of mean selection frequencies using the B bootstrap samples. The vector $\mathbf{P}^*(\cdot)$ is then used to modify the log-likelihood function in the following manner:

$$\ell(\theta) = -n\log\left\{4\alpha\beta^{1/2}\Phi(-\gamma)(2\pi)^{1/2}\right\} - \frac{3n}{2}\mathbf{P}^*(\cdot)\log(\mathbf{x})^\top + n\mathbf{P}^*(\cdot)\log(\mathbf{x}+\beta)^\top - \frac{n}{2}\mathbf{P}^*(\cdot)\mathbf{t}_\gamma^\top,$$

where $\log(\mathbf{x}) = (\log(x_1),\ldots,\log(x_n))$, $\log(\mathbf{x}+\beta) = (\log(x_1+\beta),\ldots,\log(x_n+\beta))$ and $\mathbf{t}_\gamma = ((|t_1|+\gamma)^2,\ldots,(|t_n|+\gamma)^2)$ are row vectors. The motivation behind the method is to approximate the ideal bootstrap estimates (which corresponds to $B=\infty$) faster than with the usual nonparametric bootstrap approach. In this chapter we shall investigate whether this method is able to attenuate the numerical difficulties associated with $\mathscr{BBS}$ log-likelihood function maximization.

2.3.3 Log-likelihood function with a modified Jeffreys prior penalization

Monotone likelihood cases can arise with considerable frequency in models based on the asymmetric normal distribution, with some samples leading to situations where maximum likelihood estimates of the asymmetry parameter may not be finite, as noted by Liseo (1990). A solution to such problem was proposed by Sartori (2006), who used the score function transformation proposed by Firth (1993) in the asymmetric normal and Student-t models. A more general solution was proposed by Azzalini and Arellano-Valle (2013), who penalized the log-likelihood function as follows:

$$\ell^*(\theta) = \ell(\theta) - Q,$$

where $\ell(\theta)$ and $\ell^*(\theta)$ denote the log-likelihood function and its modified version, respectively. The authors imposed some restrictions on Q, namely: (i) $Q \geq 0$; (ii) $Q = 0$ when the asymmetry parameter equals zero (values close to zero can lead to monotone likelihood cases in the asymmetric normal model); (iii) $Q \to \infty$ when the asymmetry parameter in absolute value tends to infinity. Additionally, Q should not depend on the data or, at least, be $O_p(1)$. According to Azzalini and Arellano-Valle (2013), when these conditions are satisfied, the estimators obtained using $\ell^*(\theta)$ are finite and have the same asymptotic properties as standard MLEs, such as consistency and asymptotic normality.

We shall now use a similar approach for the $\mathscr{BBS}$ model. In particular, we propose modifying the Jeffreys penalization term so that the new penalization satisfies the conditions listed by Azzalini and Arellano-Valle (2013). Since the numerical problems are mainly associated with α and γ, only terms involving these parameters were used. We then arrive at the following penalization term:

$$Q = Q_\gamma + Q_\alpha = -\frac{1}{2}\log\left\{\frac{(\gamma-\omega)\omega[3+\gamma(\gamma-\omega)]}{2}+1\right\} + \frac{1}{2}\log(1+\alpha^2),$$

where, as before, $\omega = \phi(\gamma)/\Phi(-\gamma)$.

We note that $Q_\alpha \geq 0$. Additionally, $Q_\alpha \to 0$ when $\alpha \to 0$, and $Q_\alpha \to \infty$ when $\alpha \to \infty$. It can be shown that $Q_\gamma \to \infty$ when $\gamma \to \infty$, $Q_\gamma \to 0$ when $\gamma \to -\infty$ and that $Q_\gamma \geq 0$. Figure 2.4 shows the penalization terms as a function of the corresponding parameters. The quantities Q_γ and Q_α penalize large positive values of γ and α, helping avoid estimates that are unexpectedly

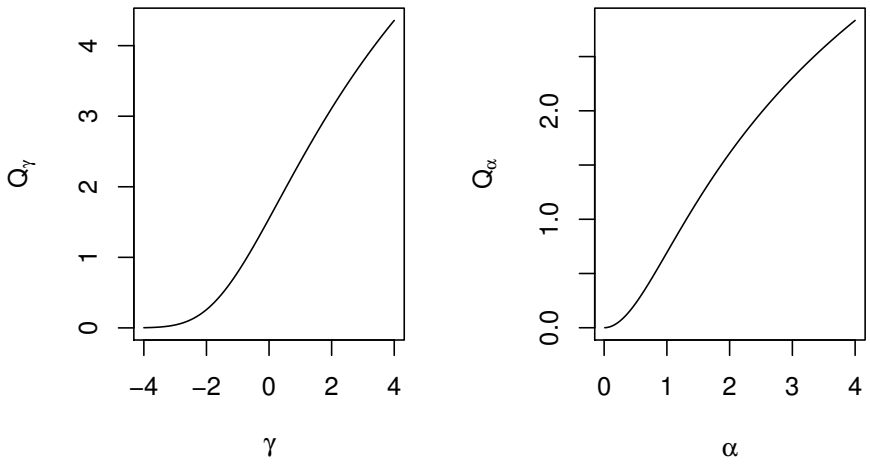

Figure 2.4 Q_α and Q_γ, modified Jeffreys penalization.

large. Therefore, the proposed penalization satisfies the conditions indicated by Azzalini and Arellano-Valle (2013). An advantage of the penalization scheme we propose is that, unlike the Jeffreys penalization, it does not require the computation of $L_{\beta\beta}$. In what follows we shall numerically evaluate the effectiveness of the proposed correction when performing point estimation.

2.3.4 Numerical evaluation

A numerical evaluation of the methods described in this Section was performed. We considered different $\mathscr{BBS}$ estimation strategies. In what follows we shall focus on the estimation of the bimodality parameter γ.

The Monte Carlo simulations were carried out using the Ox matrix programming language (Doornik, 2009). Numerical maximizations were performed using the the BFGS quasi-Newton method. We considered alternative nonlinear optimization algorithms such as Newton-Raphson and Fisher's scoring, but they did not outperform the BFGS algorithm. We then decided to employ the BFGS method, which is typically regarded as the best performing method (Mittelhammer et al., 2000, Section 8.13). The results are based on 5,000 Monte Carlo replications for values of γ ranging from -2 to 2 and samples of size $n = 50$. In each replication, maximum likelihood estimates were computed and it was verified whether the nonlinear optimization algorithm converged. At the end of the experiment, the frequency of nonconvergences (proportion of samples for which there was no convergence) was computed for each method (denoted by pnf). Figure 2.5 shows the proportion of nonconvergences corresponding to the standard MLEs, the MLEs obtained using the better bootstrap ($\text{MLE}_{\text{bboot}}$) and the MLEs obtained from the log-likelihood function penalized using the Jeffreys prior (MLE_{jp}) and its modified version (MLE_{p}) as a function of γ. Notice that the MLE and the $\text{MLE}_{\text{bboot}}$ are the worst performers

when $\gamma > 0$; they display the largest rates of nonconvergence. The methods based on penalized log-likelihood function display the smallest values of pnf, with slight advantage for MLE_{jp}.

In order to evaluate the impact of the sample size on nonconvergence rates, a numerical study similar to the previous one was performed, but now with the value of the bimodality parameter fixed at $\gamma = 1$. The samples sizes are $n \in \{30, 45, 60, 75, \ldots, 300\}$. The number of Monte Carlo replications was 5,000 for each value of n. The results are displayed in Figure 2.6. We note that the sample size does not seem to influence the MLE and $\text{MLE}_{\text{bboot}}$ nonconvergence rates. The corresponding optimizations failed in approximately 40% of the samples regardless of the sample size. In contrast, the MLE_{jp} and MLE_{p} failure rates display a slight increase and then stabilize as n increases. Recall that one of the conditions imposed by Azzalini and Arellano-Valle (2013) on the penalization term is that it should remain $O_p(1)$ as $n \to \infty$, i.e., the penalization influence seems to decrease as larger sample sizes are used, which leads to slightly larger nonconvergence frequencies in larger samples.

A second set of Monte Carlo simulations was carried out, this time only considering the estimator that uses the better bootstrap resampling scheme and also estimators based on the two penalized likelihood functions, i.e., we now only consider $\text{MLE}_{\text{bboot}}$, MLE_{jp} and MLE_{p}. Again, 5,000 Monte Carlo replications were performed. We estimated the bias (denoted by B) and mean squared errors (denoted by MSE) of the three estimators. The number of nonconvergences is denoted by nf. Tables 2.1, 2.2 and 2.3 contain the results for the estimates of α, β and γ, respectively, with these estimators. Overall, MLE_{p} outperforms MLE_{jp}. For instance, when $n = 30$ in the last combination of parameter values, the MSEs of $\hat{\alpha}_{\text{jp}}$, $\hat{\beta}_{\text{jp}}$ and $\hat{\gamma}_{\text{jp}}$ are, respectively, 0.0211, 0.0022 and 2.6671, whereas the corresponding values for $\hat{\alpha}_{\text{p}}$, $\hat{\beta}_{\text{p}}$ and $\hat{\gamma}_{\text{p}}$ are 0.0167, 0.002 and 2.0471. $\text{MLE}_{\text{bboot}}$ is typically less biased when it comes to the estimation of α and γ, but there are more convergence failures when computing better bootstrap estimates. Overall, the estimator based on the log-likelihood function that uses the penalization term we proposed typically yields more accurate estimates than MLE_{jp} and outperforms $\text{MLE}_{\text{bboot}}$ in terms of convergence rates.

Next, we shall evaluate how changes in the penalization term impact the frequency of nonconvergences when computing MLE_{p}. In particular, we consider the following penalized log-likelihood function:

$$\ell^*_\phi(\theta) = \ell(\theta) - Q^\phi,$$

with $\phi > 0$ fixed. This additional quantity controls for the penalization strength, with $\phi = 1$ resulting in MLE_{p} and different values of ϕ leading to stronger or weaker penalizations. A Monte Carlo study was performed to evaluate the accuracy of the parameter estimates for $\gamma \in \{0, 0.1, \ldots, 2.0\}$ and $\phi \in \{0.1, 0.2, \ldots, 2.1\}$. The parameter values are $\alpha = 0.5$ and $\beta = 1$, and the sample size is $n = 50$. Again, 5,000 replications were performed for each combination of values of ϕ and γ. Samples for which there was convergence failure were discarded. Figure 2.7 shows the estimated MSEs and the number of nonconvergences for each combination of γ and ϕ. Figures 2.7a and 2.7b show that estimates of α are less accurate than those of β, both being considerably more accurate than the estimates of γ (Figure 2.7c). The MSE of the estimator of γ tends to be smaller when the value of ϕ is between 0.4 and 1, especially for larger values of γ. Visual inspection of Figure 2.7d shows that larger values of γ lead to more nonconvergences, which was expected in light of our previous results. Furthermore, nf tends to decrease when

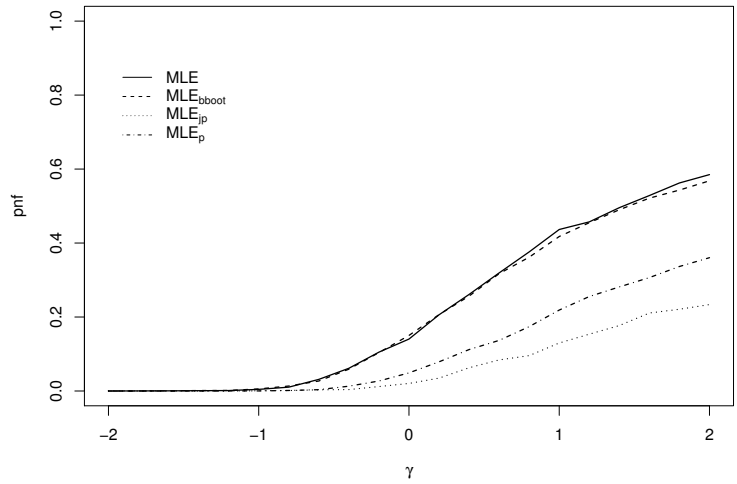

Figure 2.5 Nonconvergence proportions (pnf) for different estimation methods using different values of γ.

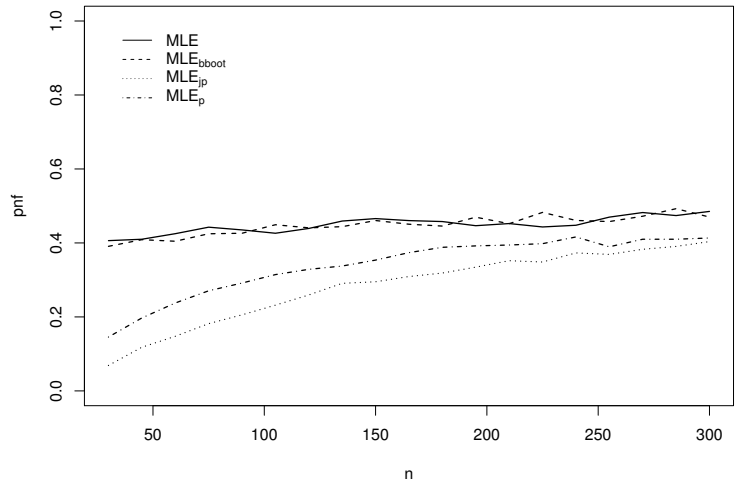

Figure 2.6 Nonconvergence proportions (pnf) for different estimation methods using different sample sizes.

Table 2.1 Bias and mean squared error of MLE_{jp}, MLE_{p} and $\text{MLE}_{\text{bboot}}$ for estimates of α in some combinations of parameter values

	MLE_{jp}			MLE_{p}			$\text{MLE}_{\text{bboot}}$		
n	$\widehat{\text{B}}(\hat{\alpha}_{\text{jp}})$	$\widehat{\text{MSE}}(\hat{\alpha}_{\text{jp}})$	nf	$\widehat{\text{B}}(\hat{\alpha}_{\text{p}})$	$\widehat{\text{MSE}}(\hat{\alpha}_{\text{p}})$	nf	$\widehat{\text{B}}(\hat{\alpha}_{\text{bboot}})$	$\widehat{\text{MSE}}(\hat{\alpha}_{\text{bboot}})$	nf
$\alpha = 0.5$, $\beta = 1$ and $\gamma = -1$									
30	−0.0851	0.0134	1	−0.0575	0.0114	12	−0.0283	0.0141	126
50	−0.0567	0.0079	7	−0.0364	0.0072	4	−0.013	0.0086	38
100	−0.0303	0.0038	2	−0.019	0.0035	0	−0.0075	0.004	2
150	−0.0208	0.0025	0	−0.0128	0.0025	0	−0.0041	0.0027	0
$\alpha = 0.5$, $\beta = 1$ and $\gamma = 0$									
30	−0.1532	0.0303	67	−0.1163	0.0236	214	−0.0811	0.0281	1494
50	−0.1103	0.0184	124	−0.0804	0.0151	240	−0.0476	0.0177	1224
100	−0.0692	0.0095	185	−0.0477	0.0082	287	−0.0215	0.0093	860
150	−0.0484	0.0061	181	−0.0323	0.0057	318	−0.0138	0.0061	657
$\alpha = 0.5$, $\beta = 1$ and $\gamma = 1$									
30	−0.2295	0.0585	373	−0.1965	0.0474	857	−0.1609	0.0498	4640
50	−0.1865	0.0403	729	−0.1535	0.0321	1398	−0.1203	0.038	5096
100	−0.1332	0.0229	1520	−0.1058	0.0185	2200	−0.0735	0.0228	5877
150	−0.107	0.0164	2115	−0.0815	0.0136	2688	−0.0562	0.0154	5926
$\alpha = 0.3$, $\beta = 1$ and $\gamma = -1$									
30	−0.052	0.0048	5	−0.0326	0.0042	14	−0.0159	0.005	150
50	−0.0353	0.0029	2	−0.0197	0.0025	7	−0.009	0.0031	43
100	−0.0174	0.0013	2	−0.0102	0.0013	1	−0.004	0.0015	2
150	−0.0119	0.0009	1	−0.0065	0.0009	0	−0.0032	0.0009	0
$\alpha = 0.3$, $\beta = 1$ and $\gamma = 0$									
30	−0.0914	0.0109	86	−0.0682	0.0085	252	−0.0472	0.0101	1543
50	−0.0663	0.0067	127	−0.0474	0.0053	298	−0.028	0.007	1252
100	−0.0401	0.0033	205	−0.0277	0.0029	349	−0.0127	0.0034	813
150	−0.0298	0.0022	208	−0.019	0.0020	349	−0.0090	0.0023	698
$\alpha = 0.3$, $\beta = 1$ and $\gamma = 1$									
30	−0.1380	0.0211	419	−0.1161	0.0167	979	−0.0983	0.0179	4528
50	−0.1114	0.0144	745	−0.0920	0.0116	1475	−0.0716	0.0131	5128
100	−0.0808	0.0084	1652	−0.0624	0.0067	2491	−0.0428	0.0094	5844
150	−0.0649	0.0059	2211	−0.0478	0.0049	3066	−0.0328	0.0059	6142

Table 2.2 Bias and mean squared error of MLE_{jp}, MLE_{p} and $\text{MLE}_{\text{bboot}}$ for estimates of β in some combinations of parameter values

	MLE_{jp}			MLE_{p}			$\text{MLE}_{\text{bboot}}$		
n	$\widehat{\text{B}}(\hat{\beta}_{\text{jp}})$	$\widehat{\text{MSE}}(\hat{\beta}_{\text{jp}})$	nf	$\widehat{\text{B}}(\hat{\beta}_{\text{p}})$	$\widehat{\text{MSE}}(\hat{\beta}_{\text{p}})$	nf	$\widehat{\text{B}}(\hat{\beta}_{\text{bboot}})$	$\widehat{\text{MSE}}(\hat{\beta}_{\text{bboot}})$	nf
$\alpha = 0.5$, $\beta = 1$ and $\gamma = -1$									
30	−0.0048	0.0118	1	0.0078	0.0121	12	0.0091	0.0121	126
50	−0.0032	0.0068	7	0.0006	0.0068	4	0.0035	0.007	38
100	−0.0025	0.0032	2	0.0018	0.0032	0	0.0017	0.0033	2
150	−0.0016	0.0021	0	0.0004	0.0021	0	0.0018	0.0021	0
$\alpha = 0.5$, $\beta = 1$ and $\gamma = 0$									
30	−0.0037	0.0103	67	0.0035	0.01	214	0.0056	0.0093	1494
50	−0.0017	0.0057	124	0.0042	0.0056	240	0.0038	0.0054	1224
100	−0.0002	0.0026	185	0.0015	0.0026	287	0.0001	0.0025	860
150	−0.0006	0.0017	181	0.0009	0.0017	318	0.0001	0.0016	657
$\alpha = 0.5$, $\beta = 1$ and $\gamma = 1$									
30	0.0001	0.0064	373	0.0035	0.0057	857	0.003	0.0053	4640
50	−0.0018	0.0031	729	0.0002	0.0029	1398	0.0026	0.0028	5096
100	−0.0009	0.0012	1520	0.0003	0.0011	2200	0.0003	0.0012	5877
150	−0.0005	0.0008	2115	0.0011	0.0007	2688	0.0001	0.0007	5926
$\alpha = 0.3$, $\beta = 1$ and $\gamma = -1$									
30	0.0004	0.005	5	0.0012	0.0049	14	0.0016	0.0049	150
50	−0.0005	0.0028	2	0.0005	0.0029	7	0.0028	0.0029	43
100	−0.0013	0.0013	2	0.0005	0.0013	1	0.0008	0.0013	2
150	−0.0003	0.0009	1	0.0002	0.0009	0	0.0008	0.0008	0
$\alpha = 0.3$, $\beta = 1$ and $\gamma = 0$									
30	−0.0023	0.0039	86	0.0012	0.0036	252	0.0024	0.0034	1543
50	−0.0018	0.0022	127	0.002	0.0021	298	0.0000	0.002	1252
100	−0.0003	0.001	205	0.0012	0.0009	349	0.0003	0.0009	813
150	0.0000	0.0006	208	0.0005	0.0006	349	0.0007	0.0006	698
$\alpha = 0.3$, $\beta = 1$ and $\gamma = 1$									
30	0.0002	0.0022	419	0.0005	0.0020	979	0.0007	0.0019	4528
50	0.0001	0.0011	745	−0.0005	0.0010	1475	0.0002	0.0010	5128
100	−0.0009	0.0004	1652	0.0000	0.0004	2491	0.0004	0.0004	5844
150	0.0000	0.0003	2211	0.0005	0.0003	3066	0.0000	0.0003	6142

Table 2.3 Bias and mean squared error of MLE_{jp}, MLE_{p} and $\text{MLE}_{\text{bboot}}$ for estimates of γ in some combinations of parameter values

	MLE_{jp}			MLE_{p}			$\text{MLE}_{\text{bboot}}$		
n	$\widehat{\text{B}}(\hat{\gamma}_{\text{jp}})$	$\widehat{\text{MSE}}(\hat{\gamma}_{\text{jp}})$	nf	$\widehat{\text{B}}(\hat{\gamma}_{\text{p}})$	$\widehat{\text{MSE}}(\hat{\gamma}_{\text{p}})$	nf	$\widehat{\text{B}}(\hat{\gamma}_{\text{bboot}})$	$\widehat{\text{MSE}}(\hat{\gamma}_{\text{bboot}})$	nf
$\alpha = 0.5$, $\beta = 1$ and $\gamma = -1$									
30	−0.4369	0.377	1	−0.3129	0.3151	12	−0.1575	0.3221	126
50	−0.2844	0.2022	7	−0.1918	0.173	4	−0.0761	0.1898	38
100	−0.1498	0.0876	2	−0.0999	0.0795	0	−0.043	0.0861	2
150	−0.1018	0.0574	0	−0.067	0.0546	0	−0.023	0.0568	0
$\alpha = 0.5$, $\beta = 1$ and $\gamma = 0$									
30	−0.8487	0.9668	67	−0.6441	0.7143	214	−0.4596	0.7186	1494
50	−0.5798	0.5201	124	−0.4358	0.4199	240	−0.2658	0.4071	1224
100	−0.3523	0.2476	185	−0.2505	0.2069	287	−0.1192	0.2052	860
150	−0.2412	0.151	181	−0.1697	0.1369	318	−0.078	0.1365	657
$\alpha = 0.5$, $\beta = 1$ and $\gamma = 1$									
30	−1.5358	2.6802	373	−1.3012	2.0986	857	−1.0607	1.9644	4640
50	−1.1795	1.636	729	−0.9703	1.2726	1398	−0.7657	1.2973	5096
100	−0.8061	0.8366	1520	−0.6446	0.6702	2200	−0.4606	0.7213	5877
150	−0.6337	0.5793	2115	−0.4877	0.4739	2688	−0.3442	0.4782	5926
$\alpha = 0.3$, $\beta = 1$ and $\gamma = -1$									
30	−0.4429	0.3812	5	−0.2940	0.3119	14	−0.1554	0.3204	150
50	−0.2874	0.2068	2	−0.1725	0.1672	7	−0.0820	0.1927	43
100	−0.1400	0.0875	2	−0.0909	0.0846	1	−0.0384	0.0883	2
150	−0.0970	0.0578	1	−0.0590	0.0556	0	−0.0292	0.0551	0
$\alpha = 0.3$, $\beta = 1$ and $\gamma = 0$									
30	−0.8477	0.9736	86	−0.6436	0.7375	252	−0.4434	0.7141	1543
50	−0.5873	0.5309	127	−0.4288	0.4085	298	−0.2633	0.4315	1252
100	−0.3406	0.2366	205	−0.2424	0.2034	349	−0.115	0.2104	813
150	−0.2472	0.1512	208	−0.1686	0.134	349	−0.0823	0.1416	698
$\alpha = 0.3$, $\beta = 1$ and $\gamma = 1$									
30	−1.5331	2.6671	419	−1.281	2.0471	979	−1.0801	1.9779	4528
50	−1.1799	1.6382	745	−0.9769	1.2924	1475	−0.7586	1.2703	5128
100	−0.8149	0.8538	1652	−0.6381	0.6761	2491	−0.4492	0.7788	5844
150	−0.645	0.5863	2211	−0.4844	0.4685	3066	−0.3410	0.5082	6142

larger values of ϕ are used.

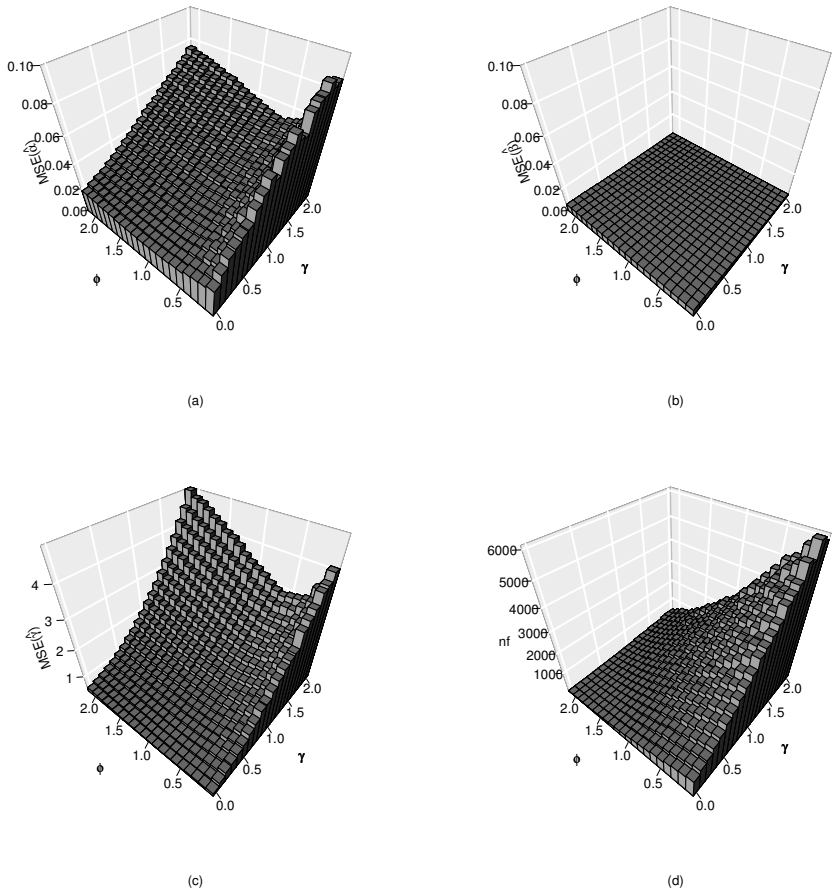

Figure 2.7 Mean squared errors of the estimators of α (a), β (b) and γ (c), and the number of nonconvergences nf (d), for different values of γ and ϕ.

We note from Figure 2.7 that the estimates of γ are the ones most sensitive to changes in the values of γ and ϕ. Figure 2.8 presents the number of nonconvergences (right vertical axis) and the MSE of $\hat{\gamma}$ (left vertical axis) as a function of γ for three different values of ϕ. Figure 2.8a shows that, although $\phi = 0.5$ yields more accurate estimates it also leads to more nonconvergences. For $\phi = 1.5$, the number of nonconvergences did not exceed 1400, but $\widehat{\mathrm{MSE}}(\hat{\gamma})$ was larger relative to other values of ϕ. Overall, $\phi = 1$ seems to balance well accuracy and the likelihood of convergence. In what follows we shall use $\phi = 1$.

2.4 Two-sided hypothesis tests

In this section we consider two-sided hypothesis tests in the $\mathscr{BBS}$ model. Our interest lies in investigating the finite-sample performances of tests based on the penalized log-likelihood function by the modified Jeffreys prior. The first test we consider is the penalized likelihood

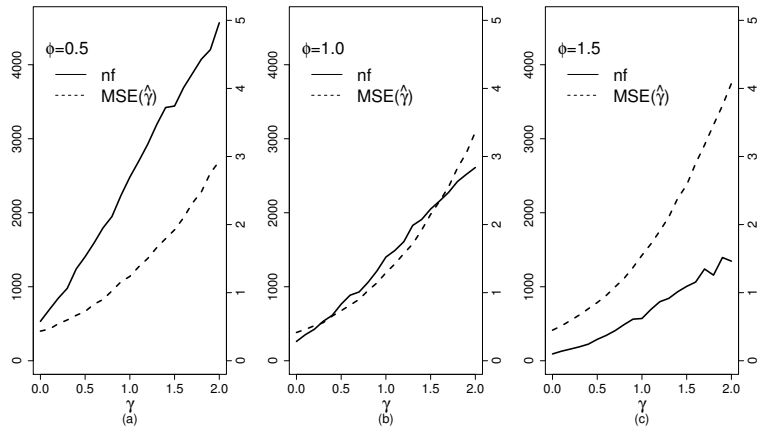

Figure 2.8 Number of nonconvergences (solid line) and MSE of $\hat{\gamma}$ (dashed line) for different values of γ with $\phi \in \{0.5, 1.0, 1.5\}$. The nf values are shown in the left vertical axis and the values of $\widehat{\mathrm{MSE}}(\hat{\gamma})$ are shown in the right vertical axis.

ratio test, denoted by LR. Consider a model parametrized by $\theta = (\psi, \lambda)$, where ψ is the parameter of interest and λ is a nuisance parameter vector. Our interest lies in testing $H_0 : \psi = \psi_0$ against a two-sided alternative hypothesis. The LR test statistic is

$$W = 2\{\ell^*(\hat{\theta}) - \ell^*(\tilde{\theta})\},$$

where $\hat{\theta}$ is the unrestricted MLE$_\mathrm{p}$ of θ, i.e., $\hat{\theta}$ is obtained by maximizing $\ell^*(\theta)$ without imposing restrictions on the parameters, and $\tilde{\theta}$ is the restricted MLE$_\mathrm{p}$, which follows from the maximization of $\ell^*(\theta)$ subject to the restrictions in H_0. Critical values at the $\varepsilon \times 100\%$ significance level are obtained from the null distribution of W which, based on the results in Azzalini and Arellano-Valle (2013), can be approximated by χ^2_1 when ψ is scalar. When ψ is a vector of dimension q (≤ 3), the test is performed in similar fashion with the single difference that the critical value is obtained from χ^2_q.

It is also possible to test $H_0 : \psi = \psi_0$ against $H_1 : \psi \neq \psi_0$ using the score and Wald tests. To that end, we use the score function and the expected information matrix obtained using the penalized log-likelihood function. The score and Wald test statistics are given, respectively, by

$$W_S = U^*(\tilde{\theta})^\top K^*(\tilde{\theta})^{-1} U^*(\tilde{\theta}),$$
$$W_W = (\hat{\psi} - \psi_0)^2 / K^*(\hat{\theta})^{\psi\psi},$$

where $U^*(\theta)$ and $K^*(\theta)$ denote the score function and the expected information, respectively, obtained using the penalized log-likelihood function and $K^*(\theta)^{\psi\psi}$ is the diagonal element of the inverse of $K^*(\theta)$ corresponding to ψ. Both test statistics are asymptotically distributed as χ^2_1 under the null hypothesis.

A Monte Carlo simulation study was performed to evaluate the finite sample performances of the LR, score (denoted by S) and Wald tests in the $\mathscr{BBS}$ model. Log-likelihood maximizations were carried out using the BFGS quasi-Newton method. The number of Monte Carlo

replications is 5,000 replications, the sample sizes are $n \in \{30, 50, 100, 150\}$ and the significance levels are $\varepsilon \in \{0.1, 0.05, 0.01\}$. The tests were performed for each parameter of the model $\mathscr{BBS}(0.5, 1, 0)$. It is noteworthy that by testing $H_0 : \gamma = 0$ against $H_1 : \gamma \neq 0$ we test whether the data follows the $\mathscr{BS}$ law, i.e., the original version of the Birnbaum-Saunders distribution. The data were generated according to the model implied by the null hypothesis and samples for which convergence did not take place were discarded.

Table 2.4 Null rejection rates of the LR, score, Wald, LR_{pb}, LR_{bbc} and S_{pb} tests for testing of $H_0 : \alpha = 0.5$ against $H_1 : \alpha \neq 0.5$ in the $\mathscr{BBS}(0.5, 1, 0)$ model

n	LR	S	Wald	LR_{pb}	LR_{bbc}	S_{pb}
			$\varepsilon = 0.1$			
30	0.2660	0.2166	0.4460	0.1010	0.0934	0.0994
50	0.2008	0.1656	0.3374	0.0990	0.0970	0.0964
100	0.1472	0.1288	0.2392	0.1006	0.1026	0.0988
150	0.1346	0.1202	0.2078	0.0950	0.0970	0.0940
			$\varepsilon = 0.05$			
30	0.1632	0.1116	0.3850	0.0542	0.0388	0.0500
50	0.1156	0.0820	0.2836	0.0484	0.0420	0.0466
100	0.0844	0.0664	0.1914	0.0510	0.0520	0.0488
150	0.0716	0.0574	0.1524	0.0474	0.0484	0.0486
			$\varepsilon = 0.01$			
30	0.0508	0.0104	0.2706	0.0100	0.0038	0.0082
50	0.0352	0.0100	0.1822	0.0096	0.0058	0.0108
100	0.0230	0.0094	0.1094	0.0106	0.0102	0.0106
150	0.0166	0.0084	0.0766	0.0096	0.0096	0.0110

Tables 2.4 to 2.6 contain the null rejection rates of the tests of $H_0 : \alpha = 0.5$, $H_0 : \beta = 1$ and $H_0 : \gamma = 0$, respectively, against two-sided alternative hypotheses. We note that all tests are considerably liberal when the sample size is small (30 or 50). We also note that the score test outperforms the competitors. The Wald test was the worst performer.

The tests null rejection rates converge to the corresponding nominal levels as $n \to \infty$. Such convergence, however, is rather slow. More accurate testing inference can be achieved by using bootstrap resampling; see Davison and Hinkley (1997). The tests employ critical values that are estimated in the bootstrapping scheme instead of asymptotic (approximate) critical values. A number of B bootstrap samples are generated imposing the null hypothesis and the test statistic is computed for each artificial sample. The critical value of level $\varepsilon \times 100\%$ is obtained as the $1 - \varepsilon$ upper quantile of the B test statistics, i.e., of the test statistics computed using the bootstrap samples. The bootstrap tests are indicated by the subscript 'pb'. We also use bootstrap resampling to estimate the Bartlett correction factor to the likelihood ratio test as proposed by Rocke (1989). The bootstrap Bartlett corrected test is indicated by the subscript 'bbc'. For details on bootstrap tests, Bartlett-corrected tests and Bartlett corrections based on the bootstrap, the reader is referred to Cordeiro and Cribari-Neto (2014). Since the Wald test proved to be considerably unreliable, we shall not consider it.

Table 2.5 Null rejection rates of the LR, score, Wald, LR_{pb}, LR_{bbc} and S_{pb} tests for testing of $H_0 : \beta = 1$ against $H_1 : \beta \neq 1$ in the $\mathscr{BBS}(0.5,1,0)$ model

n	LR	S	Wald	LR_{pb}	LR_{bbc}	S_{pb}
			$\varepsilon = 0.1$			
30	0.1692	0.0832	0.2728	0.1088	0.1088	0.1104
50	0.1416	0.0826	0.2116	0.1064	0.1054	0.1096
100	0.1190	0.0894	0.1650	0.1016	0.1002	0.1164
150	0.1170	0.0934	0.1494	0.1048	0.1058	0.1140
			$\varepsilon = 0.05$			
30	0.1040	0.0340	0.2162	0.0496	0.0490	0.0514
50	0.0802	0.0370	0.1480	0.0532	0.0528	0.0538
100	0.0620	0.0442	0.1090	0.0538	0.0524	0.0574
150	0.0558	0.0438	0.0808	0.0510	0.0510	0.0554
			$\varepsilon = 0.01$			
30	0.0298	0.0048	0.1094	0.0108	0.0100	0.0114
50	0.0218	0.0080	0.0734	0.0092	0.0094	0.0112
100	0.0112	0.0076	0.0354	0.0102	0.0112	0.0102
150	0.0112	0.0090	0.0268	0.0098	0.0106	0.0104

Table 2.6 Null rejection rates of the LR, score, Wald, LR_{pb}, LR_{bbc} and S_{pb} tests for testing $H_0 : \gamma = 0$ against $H_1 : \gamma \neq 0$ in the $\mathscr{BBS}(0.5,1,0)$ model

n	LR	S	Wald	LR_{pb}	LR_{bbc}	S_{pb}
			$\varepsilon = 0.1$			
30	0.2570	0.2434	0.3704	0.1034	0.0948	0.1022
50	0.1974	0.1920	0.2896	0.1040	0.1028	0.1014
100	0.1368	0.1320	0.2014	0.0980	0.1012	0.0976
150	0.1364	0.1326	0.1774	0.0986	0.0992	0.1014
			$\varepsilon = 0.05$			
30	0.1678	0.1522	0.2992	0.0522	0.0392	0.0492
50	0.1208	0.1076	0.2134	0.0532	0.0488	0.0514
100	0.0860	0.0762	0.1454	0.0464	0.0488	0.0458
150	0.0734	0.0660	0.1140	0.0468	0.0488	0.0456
			$\varepsilon = 0.01$			
30	0.0520	0.0334	0.1632	0.0094	0.0030	0.0106
50	0.0336	0.0234	0.1100	0.0098	0.0062	0.0106
100	0.0218	0.0150	0.0658	0.0082	0.0078	0.0088
150	0.0154	0.0114	0.0474	0.0082	0.0080	0.0082

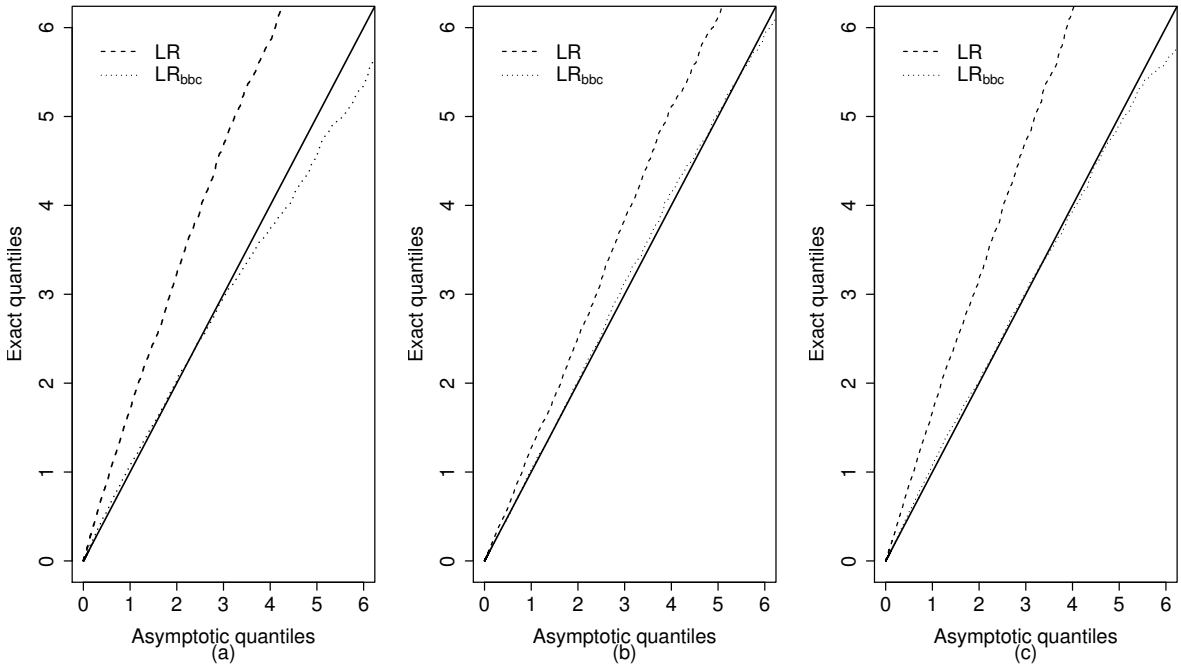

Figure 2.9 Quantile-quantile plots for the LR and LR$_{bbc}$ test statistics with $n = 50$, for the tests on α (a), on β (b) and on γ (c).

Next we shall numerically evaluate the finite sample performances of the LR$_{pb}$, LR$_{bbc}$ and S$_{pb}$ tests under the same scenarios considered in the previous simulation. The number of Monte Carlo replications is as before. Samples for which the optimization methods failed to reach convergence were discard, even for bootstrap samples. The same 1,000 bootstrap samples ($B = 1,000$) were used in all tests. The null rejection rates of the tests for making inferences on α, β and γ are also presented in Tables 2.4, 2.5 and 2.6, respectively. It is noteworthy that the tests size distortions are now considerably smaller. For instance, when making inference on α based on a sample of size $n = 30$ and $\varepsilon = 0.05$, the LR and score null rejection rates are 16.32% and 11.16% (Table 2.4), whereas the corresponding figures for their bootstrap versions LR$_{\text{bp}}$, LR$_{\text{bbc}}$ and S$_{\text{bp}}$ are 5.42%, 3.88% and 5%, respectively, which are much more closer to 5%. When testing restrictions on β with $n = 30$ and $\varepsilon = 0.10$, the LR and score null rejection rates are, respectively, 16.92% and 8.32% (Table 2.5) whereas their bootstrap versions, LR$_{\text{bp}}$, LR$_{\text{bbc}}$ and S$_{\text{bp}}$, display null rejection rates of 10.88%, 10.88% and 11.04%, respectively. Finally, when the interest lies in making inferences on γ with $n = 30$ and $\varepsilon = 0.01$, the LR and score null rejection rates are 5.20% and 3.34% (Table 2.6); for the bootstrap-based tests LR$_{\text{bp}}$, LR$_{\text{bbc}}$ and S$_{\text{bp}}$ we obtain 0.94%, 0.3% and 1.06%, respectively. Figure 2.9 shows the quantile-quantile (QQ) plots of the LR and LR$_{bbc}$ test statistics for samples of size $n = 50$. It is noteworthy that the empirical quantiles of the LR$_{bbc}$ test statistic are much closer of the corresponding asymptotic quantiles than those of W. Hence, we note that testing inference in small samples can be made considerably more accurate by using bootstrap resampling.

2.5 One-sided hypothesis tests

One-sided tests on a scalar parameter can be performed using the signed likelihood ratio (SLR) test, which is particularly useful in the $\mathscr{BBS}$ model since it allows practitioners to make inferences on γ in a way that makes it possible to detect bimodality. The signed penalized likelihood ratio test statistic is

$$R = \text{sign}(\hat{\psi} - \psi_0)\sqrt{W} = \text{sign}(\hat{\psi} - \psi_0)\sqrt{2\{\ell^*(\hat{\theta}) - \ell^*(\tilde{\theta})\}}. \tag{2.5}$$

The statistic R is asymptotically distributed as standard normal under the null hypothesis. An advantage of the SLR test over the tests described in Section 2.4 is that it can be used to perform two-sided and one-sided tests. In this section, we shall focus on one-sided hypothesis testing inference. Our interest lies in detecting bimodality. The null hypothesis is $H_0 : \gamma \geq 0$ which is tested against $H_1 : \gamma < 0$. Rejection of H_0 yields evidence that the data came from a bimodal distribution. On the other hand, when H_0 is not rejected, there is evidence that the data follows a distribution with a single mode.

Consider the sample $\mathbf{x} = (x_1, \ldots, x_n)$ for a model with vector parameter $\theta = (\psi, \lambda)$ with dimension $1 \times p$, where the parameter of interest ψ is a scalar and the vector of nuisance parameters λ has dimension $1 \times (p-1)$. The test statistic R, given in Equation (2.5), is asymptotically distributed as standard normal with error of order $O(n^{-1/2})$ when the null hypothesis is true. Such an approximation may not be accurate when the sample size is small. Some analytical corrections for R were proposed in the literature. They can be used to improve the test finite sample behavior.

An important contribution was made by Barndorff-Nielsen (1986, 1991). The author proposed a correction term $\mathscr{U}$ of the form

$$R^* = R + \log(\mathscr{U}/R)/R,$$

where R represents the SLR statistic and R^* is its corrected version. Let $\ell(\theta)$ be the log-likelihood function of the parameters. Its derivatives will be denoted here by

$$\ell_\theta(\theta) = \frac{\partial \ell(\theta)}{\partial \theta} \quad \text{and} \quad \ell_{\theta\theta}(\theta) = \frac{\partial^2 \ell(\theta)}{\partial \theta \partial \theta^\top}.$$

The observed information matrix is given by $J_{\theta\theta}(\theta) = -\ell_{\theta\theta}(\theta)$. To obtain the correction proposed by Barndorff-Nielsen (1986), the sufficient statistic has to be of the form $(\hat{\theta}, a)$, where $\hat{\theta}$ is the MLE of θ and a is an ancillary statistic. Additionally, it is necessary to compute sample space derivatives of the log-likelihood, such as

$$\ell_{;\hat{\theta}}(\theta) = \frac{\partial \ell(\theta)}{\partial \hat{\theta}} \quad \text{and} \quad \ell_{\theta;\hat{\theta}}(\theta) = \frac{\partial \ell_{;\hat{\theta}}(\theta)}{\partial \theta^\top},$$

where derivatives are taken with respect to some functions of the sample while keeping other terms fixed, as explained in Severini (2000). The quantity $\mathscr{U}$ is given by

$$\mathscr{U} = \frac{\begin{vmatrix} \ell_{;\hat{\theta}}(\hat{\theta}) - \ell_{;\hat{\theta}}(\tilde{\theta}) \\ \ell_{\lambda;\hat{\theta}}(\tilde{\theta}) \end{vmatrix}}{\left|J_{\lambda\lambda}(\tilde{\theta})\right|^{1/2} \left|J_{\theta\theta}(\hat{\theta})\right|^{1/2}},$$

where $\tilde{\theta}$ is the restricted MLE of θ and the indices indicate which components are being used in each vector or matrix.

The null distribution of R^* is standard normal with error of order $O(n^{-3/2})$. Although the null distribution of R^* is better approximated by the limiting distribution than that of R, the computation of $\mathscr{U}$ is restricted to some specific classes of models, such as exponential family and transformation models (Severini, 2000).

Some alternatives to R^* were proposed in the literature. They approximate the sample space derivatives used in $\mathscr{U}$. For instance, approximations were obtained by DiCiccio and Martin (1993), Fraser et al. (1999) and Severini (1999). They were computed by Wu and Wong (2004) for the $\mathscr{BS}$ model and by Lemonte and Ferrari (2011) for a Birnbaum-Saunders regression model. Other recent contributions are made by Ferrari and Pinheiro (2016) and Smith et al. (2015).

In this section, we apply the approximations proposed by Severini (1999) and Fraser et al. (1999) using the log-likelihood without penalization. Our interest is to evaluate the performance of the corrections when applied for the statistic R, calculated using the penalized log-likelihood function, comparing the performance of the corrected tests with the SLR and its bootstrap version.

Using the same notation as Lemonte and Ferrari (2011), the approximation proposed by Fraser et al. (1999) (denoted by SLR_{c1}) for $\mathscr{U}$ can be written as

$$\mathscr{U}_1 = \frac{\left| \begin{array}{c} \Gamma_\theta \\ \Psi_{\lambda\theta} \end{array} \right|}{\left|J_{\lambda\lambda}(\tilde{\theta})\right|^{1/2} \left|J_{\theta\theta}(\hat{\theta})\right|^{1/2}},$$

where

$$\Gamma_\theta = [\ell_{;\mathbf{x}}(\hat{\theta}) - \ell_{;\mathbf{x}}(\tilde{\theta})]V(\hat{\theta})[\ell_{\theta;\mathbf{x}}(\hat{\theta})V(\hat{\theta})]^{-1}J_{\theta\theta}(\hat{\theta}),$$

with

$$\Psi_{\theta\theta} = \left[\begin{array}{c} \Psi_{\psi\theta} \\ \Psi_{\lambda\theta} \end{array} \right] = \ell_{\theta;\mathbf{x}}(\tilde{\theta})V(\hat{\theta})[\ell_{\theta;\mathbf{x}}(\hat{\theta})V(\hat{\theta})]^{-1}J_{\theta\theta}(\hat{\theta}),$$

where $\ell_{;\mathbf{x}}(\theta) = \partial l(\theta)/\partial \mathbf{x}$ is a $1 \times n$ vector, $\ell_{\theta;\mathbf{x}}(\tilde{\theta}) = \partial^2 \ell(\theta)/\partial\theta^\top \partial \mathbf{x}$ is a $p \times n$ matrix and

$$V(\theta) = -\left[\frac{\partial \mathbf{z}(\mathbf{x};\theta)}{\partial \mathbf{x}}\right]^{-1}\left[\frac{\partial \mathbf{z}(\mathbf{x};\theta)}{\partial \theta^\top}\right]$$

is an $n \times p$ matrix, $\mathbf{z}(\mathbf{x};\theta)$ being a vector of pivotal quantities.

The corrected SLR statistic obtained using the approximation given by Fraser et al. (1999) is $R_{c1} = R + \log(\mathscr{U}_1/R)/R$, which has asymptotic standard normal distribution with error of order $O(n^{-3/2})$ under the null hypothesis. We computed the quantities needed to obtain $\mathscr{U}_1$ in the $\mathscr{BBS}$ model, which are presented below.

Consider the variable $Y = |T| + \gamma$, where $T = \alpha^{-1}(\sqrt{X/\beta} - \sqrt{\beta/X})$ and $X \sim \mathscr{BBS}(\alpha, \beta, \gamma)$. We used the distribution function of $Y = |T| + \gamma$ to obtain SLR_{c1} in the $\mathscr{BBS}$ model. The

distribution of Y is truncated standard normal distribution with support (γ,∞), its distribution function being given by

$$F_Y(y) = \begin{cases} 0 & \text{if } y < \gamma, \\ \frac{\Phi(y)-\Phi(\gamma)}{1-\Phi(\gamma)} & \text{if } y \geq \gamma. \end{cases}$$

Therefore, $Z = F_Y(Y)$ is uniformly distributed in the standard interval $(0,1)$. Hence, it is a pivotal quantity that can be used for obtaining the approximations to sample space derivatives proposed by Fraser et al. (1999). Let $\mathbf{x} = (x_1,\ldots,x_n)$ be a random $\mathscr{BBS}(\alpha,\beta,\gamma)$ sample. It follows that $\partial z_i/\partial x_j = 0$ when $i \neq j$ and $\partial z_i/\partial x_i = \phi(y_i)\,\mathrm{sign}(t_i)(x_i+\beta)/[\Phi(-\gamma)2\alpha\beta^{1/2}x_i^{3/2}]$, with $t_i = \alpha^{-1}(\sqrt{x_i/\beta} - \sqrt{\beta/x_i})$. Moreover, $\partial z_i/\partial\alpha = -\phi(y_i)\,\mathrm{sign}(t_i)t_i/\Phi(-\gamma)\alpha$, $\partial z_i/\partial\beta = -\phi(y_i)\,\mathrm{sign}(t_i)\,[\sqrt{x_i/\beta} + \sqrt{\beta/x_i}]/[\Phi(-\gamma)2\alpha\beta]$ and $\partial z_i/\partial\gamma = [\Phi(y_i)-\Phi(\gamma)]\phi(\gamma)/\Phi^2(-\gamma) + [\phi(y_i)-\phi(\gamma)]/\Phi(-\gamma)$, where $y_i = |t_i| + \gamma$ and $z_i = F_Y(y_i)$. Therefore, $v_{\alpha i} = 2\beta^{1/2}x_i^{3/2}t_i/(x_i+\beta)$, $v_{\beta i} = x_i/\beta$ and $v_{\gamma i} = -2\alpha\beta^{1/2}x_i^{3/2}\{[\Phi(y_i)-\Phi(\gamma)]\phi(\gamma)/\Phi(-\gamma) + \phi(y_i) - \phi(\gamma)\}/[\phi(y_i)\,\mathrm{sign}(t_i)(x_i+\beta)]$. The vectors $\mathbf{v}_\alpha$, $\mathbf{v}_\beta$ and $\mathbf{v}_\gamma$ are used to form the matrix $V(\theta)$. For instance, $\mathbf{v}_\alpha = (v_{\alpha 1},\ldots,v_{\alpha n})^\top$ is a $n\times 1$ vector. Here, $V(\theta) = \left[\mathbf{v}_\alpha\ \mathbf{v}_\beta\ \mathbf{v}_\gamma\right]$. Furthermore, we have that

$$\begin{aligned}
\ell_{;x_i}(\theta) &= \frac{-3}{2x_i} + \frac{1}{x_i+\beta} - (|t_i|+\gamma)\mathrm{sign}(t_i)\frac{(x_i+\beta)}{2\alpha\beta^{1/2}x_i^{3/2}}, \\
\ell_{\alpha;x_i}(\theta) &= \frac{\mathrm{sign}(t_i)(x_i+\beta)}{2\beta^{1/2}x_i^{3/2}\alpha^2}(2|t_i|+\gamma), \\
\ell_{\beta;x_i}(\theta) &= \frac{(-1)}{(x_i+\beta)^2} + \frac{(x_i+\beta)}{4\alpha^2\beta^{3/2}x_i^{3/2}}\left(\frac{x_i^{1/2}}{\beta^{1/2}} + \frac{\beta^{1/2}}{x_i^{1/2}}\right) + \frac{\mathrm{sign}(t_i)(|t_i|+\gamma)(x_i-\beta)}{4\alpha\beta^{3/2}x_i^{3/2}}, \\
\ell_{\gamma;x_i}(\theta) &= -\frac{\mathrm{sign}(t_i)(x_i+\beta)}{2\alpha\beta^{1/2}x_i^{3/2}}.
\end{aligned}$$

The method proposed by Severini (1999)(denoted by SLR_{c2}) approximates the sample space derivatives by covariances of the log-likelihood function. The main idea is to use the sample to obtain the covariance values empirically. Using again the notation of Lemonte and Ferrari (2011), the approximation of $\mathscr{U}$ proposed by Severini (1999) is given by

$$\mathscr{U}_2 = \frac{\left|\begin{array}{c}\Delta_\theta \\ \Sigma_{\lambda\theta}\end{array}\right|}{\left|J_{\lambda\lambda}(\tilde\theta)\right|^{1/2}\left|J_{\theta\theta}(\hat\theta)\right|^{1/2}},$$

with

$$\Delta_\theta = [Q(\hat\theta;\hat\theta) - Q(\tilde\theta;\hat\theta)]I(\hat\theta;\hat\theta)^{-1}J_{\theta\theta}(\hat\theta)$$

and

$$\Sigma_{\theta\theta} = \left[\begin{array}{c}\Sigma_{\psi\theta} \\ \Sigma_{\lambda\theta}\end{array}\right] = I(\tilde\theta;\hat\theta)I(\hat\theta;\hat\theta)^{-1}J_{\theta\theta}(\hat\theta),$$

where $Q(\theta;\theta_0)=\sum_{i=1}^{n}\ell^{(i)}(\theta)\ell_{\theta}^{(i)}(\theta_0)^{\top}$ is a $1\times p$ vector and $I(\theta;\theta_0)=\sum_{i=1}^{n}\ell_{\theta}^{(i)}(\theta)\ell_{\theta}^{(i)}(\theta_0)^{\top}$ is a $p\times p$ matrix, the index (i) indicating that the quantity corresponds to the ith sample observation. The corrected statistic proposed by Severini (1999) is $R_{c2}=R+\log(\mathscr{U}_2/R)/R$. Its null distribution is standard normal with error of order $O(n^{-1})$. The score function and the observed information matrix, which can be found in Olmos et al. (2016), are used to obtain $\mathscr{U}_2$ in the $\mathscr{BBS}$ model.

Alternatively, bootstrapping resampling can be used to obtain critical values for the SLR test. Since we test $H_0:\gamma\geq 0$ against $H_1:\gamma<0$, the critical value of level $\varepsilon\times 100\%$ is obtained as the ε quantile of the B test statistics computed using the bootstrap samples.

Table 2.7 Null rejection rates of the SLR, SLR_{c1}, SLR_{c2} and SLR_{bp} tests of $H_0:\gamma\geq 0$ against $H_1:\gamma<0$ in a sample of size 30 of the model $\mathscr{BBS}(0.5,1,\gamma)$

ε	SLR	SLR_{c1}	SLR_{c2}	SLR_{bp}
		$\gamma=-1$		
0.10	0.7488	0.5768	0.6310	0.4960
0.05	0.6300	0.4276	0.4728	0.3488
0.01	0.3560	0.1910	0.2100	0.1376
		$\gamma=-0.5$		
0.10	0.4634	0.2766	0.3300	0.2248
0.05	0.3242	0.1762	0.2042	0.1326
0.01	0.1328	0.0588	0.0652	0.0376
		$\gamma=0$		
0.10	0.2614	0.1334	0.1678	0.1042
0.05	0.1658	0.0746	0.0892	0.0498
0.01	0.0486	0.0202	0.0210	0.0106
		$\gamma=0.5$		
0.10	0.1546	0.0722	0.0928	0.0526
0.05	0.0890	0.0378	0.0442	0.0222
0.01	0.0214	0.0098	0.0102	0.0046
		$\gamma=1$		
0.10	0.1144	0.0488	0.0640	0.0340
0.05	0.0606	0.0246	0.0292	0.0150
0.01	0.0144	0.0050	0.0046	0.0022

A simulation study was performed to evaluate the sizes and powers of the SLR, SLR_{c1}, SLR_{c2} and SLR_{bp} tests. We tested $H_0:\gamma\geq 0$ against $H_1:\gamma<0$. The true parameter values are $\gamma\in\{-1,-0.5,0,0.5,1\}$. The most reliable tests are those with large power (i.e., higher probability of rejecting H_0 when $\gamma<0$) and small size distortions. Again, 5,000 Monte Carlo replications were performed. The SLR_{bp} test is based on 1,000 bootstrap samples. The simulation results are presented in Table 2.7. The most powerful tests are SLR, SLR_{c2} and SLR_{c1}, in that order, whereas the tests with the smallest size distortions are SLR_{bp}, SLR_{c1} and SLR_{c2}. We recommend testing inference to be based on either SLR_{c1} or SLR_{c2}, since these tests display a good balance between size and power.

2.6 Nonnested hypothesis tests

In the previous section we presented a test that is useful for detecting whether the data came from a bimodal $\mathcal{BBS}$ law or not. That was done by testing a restriction on γ. In this section we shall present tests that are useful for distinguishing between $\mathcal{BBS}$ model and another extension of the $\mathcal{BS}$ distribution that can display bimodality.

As noted in the Introduction, another variation of the $\mathcal{BS}$ distribution that can exhibit bimodality is the model recently discussed by Owen and Ng (2015), which the authors denoted by $\mathcal{GBS}_2$. Let $X \sim \mathcal{GBS}_2(\alpha, \beta, \nu)$. Its probability density function is given by

$$g(x) = \frac{\nu}{\alpha x}\left[\left(\frac{x}{\beta}\right)^{\nu} + \left(\frac{\beta}{x}\right)^{\nu}\right]\phi\left(\frac{1}{\alpha}\left[\left(\frac{x}{\beta}\right)^{\nu} - \left(\frac{\beta}{x}\right)^{\nu}\right]\right), \quad x > 0,$$

where $\alpha > 0$, $\beta > 0$ and $\nu > 0$. According to Owen and Ng (2015), the $\mathcal{GBS}_2$ density is bimodal when $\alpha > 2$ and $\nu > 2$ (simultaneously).

Therefore, when bimodality is detected, the data analysis may be carried out with either the $\mathcal{BBS}$ distribution or the $\mathcal{GBS}_2$ model. It would then be useful to have a hypothesis test that could be used to distinguish between the two models. Obviously, the tests discussed so far cannot be used to that end. Model selection criteria for the $\mathcal{BS}$ distribution were considered by Leiva (2015) and Leiva et al. (2015). Model selection is usually based on the Bayes factor and also on the Schwarz and Akaike information criteria. We shall use a different approach, developing tests for nonnested hypotheses. Notice that the $\mathcal{GBS}_2$ distribution cannot be obtained from the $\mathcal{BBS}$ distribution by imposing restrictions on the model parameters, and vice-versa. Hence, the two models are not nested.

The literature of nonnested models began with Cox (1961, 1962). The author introduced likelihood ratio tests for some nonnested models. His main results were generalized by Vuong (1989), who considered nested, nonnested and overlapping models and derived the required asymptotics. For nonnested models, Vuong (1989) established a relationship between the likelihood ratio statistic and the Kullback-Leibler information. Let F and G be competing nonnested models. The author presented a test of the null hypothesis H_0 that both models are equivalent, the alternative hypotheses being H_f: model F is better and H_g: model G is better. An alternative approach for testing of nonnested models was considered by Williams (1970) and Lewis et al. (2011). The authors only considered tests of the hypothesis H_f and H_g. They proposed to consider H_f and H_g sequentially.

We shall consider the hypothesis involving the $\mathcal{BBS}$ and $\mathcal{GBS}_2$ models as:

- H_f - the data came from the $\mathcal{BBS}$ distribution,
- H_g - the data came from the $\mathcal{GBS}_2$ distribution.

The test statistic we considered is the following likelihood ratio statistic:

$$W_{ne} = \log\left(\frac{\hat{f}}{\hat{g}}\right) = \hat{\ell}_f - \hat{\ell}_g,$$

where $\hat{f}$ and $\hat{g}$ denote the likelihood functions of the $\mathcal{BBS}$ and $\mathcal{GBS}_2$ models, respectively, evaluated at their respective MLE estimates, ℓ representing the log-likelihood function of the

model indicated by its index. Then, for a given sample $\mathbf{x}$, a large positive value of W_{ne} yields evidence in favor H_f and against H_g; on the other hand, a large negative value of W_{ne} favors H_g. The $\mathscr{BBS}$ parameters are estimated using the penalized log-likelihood function and those of $\mathscr{GBS}_2$ are estimated using the standard log-likelihood function.

In the test discussed by Vuong (1989) for nonnested models, the asymptotic null distribution is standard normal. On some Monte Carlo simulations not reported here, this test, based on asymptotic critical values, was observed to indicate equivalence of the models at a large rate. Since this test is based on large sample approximations, a bootstrap resampling method could be used as an alternative to provide more accurate critical values and to indicate only one of the models with higher frequency. On the other hand, application of bootstrap resampling is not straightforward for this kind of test because we would have to define a model equivalent to $\mathscr{BBS}$ and $\mathscr{GBS}_2$ to generate pseudo-samples under the null distribution. Thus, an approach similar to the one taken by Lewis et al. (2011) will be used in this section, considering only the hypothesis H_f and H_g in the test. Then, taking H_f as null hypothesis, the bootstrap test is done using the following procedures:

1. Calculate the value of W_{ne} for the sample $\mathbf{x}$;
2. With the MLE estimates of the parameters from the $\mathscr{BBS}$ model, generate a bootstrap sample $\mathbf{x}^*$, and then compute W_{ne}^* for this sample;
3. Execute step 2 for B times and obtain the p-value bootstrap: $p_b = \frac{\#\{W_{ne}^* < W_{ne}\}+1}{B+1}$.

Hence, at the significance level of $\varepsilon \times 100\%$, H_f is rejected if $p_b < \varepsilon$, i.e., we reject that the data was originated from a $\mathscr{BBS}$ and accept that the $\mathscr{GBS}_2$ distribution is more adequate. The same can be done taking H_g as null hypothesis. In this case, the procedures are the following:

1. Calculate the value of W_{ne} for the sample $\mathbf{x}$;
2. With the MLE estimates of the parameters from the $\mathscr{GBS}_2$ model, generate a bootstrap sample $\mathbf{x}^*$, and then compute W_{ne}^* for this sample;
3. Execute step 2 for B times and obtain the p-value bootstrap: $p_b = \frac{\#\{W_{ne}^* > W_{ne}\}+1}{B+1}$.

It is noteworthy that the step 3 is different from the corresponding step in the first procedure, since the rejection region changes when we consider H_g as null hypothesis. Again, at the significance level of $\varepsilon \times 100\%$, the null hypothesis considered is rejected if $p_b < \varepsilon$, but in this case it means that we reject the hypothesis H_g that the data comes from a $\mathscr{GBS}_2$ distribution and accept that $\mathscr{BBS}$ model is more adequate.

The problem with this approach is that four inferences results are possible to happen, as cited by Williams (1970) and Lewis et al. (2011):

*R*1- Both hypothesis H_f and H_g are not rejected, indicating that both models are adequate;

*R*2- We do not reject H_f, but H_g is rejected, indicating that the $\mathscr{BBS}$ model is more adequate;

$R3$- We do not reject H_g, but H_f is rejected, indicating that the $\mathcal{GBS}_2$ model is more adequate;

$R4$- We reject both H_f and H_g, indicating that both models are not adequate;

Under some regularity conditions, Vuong (1989) has shown that, in nonnested models, an adjusted likelihood ratio tends to infinity under H_f when $n \to \infty$ and that under H_g it tends to minus infinity when $n \to \infty$. This way, the test statistic tends to indicate the correct model as the sample size increases. Therefore, for a situation where we obtain the result $R1$, a way to choose between the two models is to use the statistic W_{ne} calculated for the sample, opting for the $\mathcal{BBS}$ distribution in case $W_{ne} > 0$ and for the $\mathcal{GBS}_2$ distribution in case $W_{ne} < 0$.

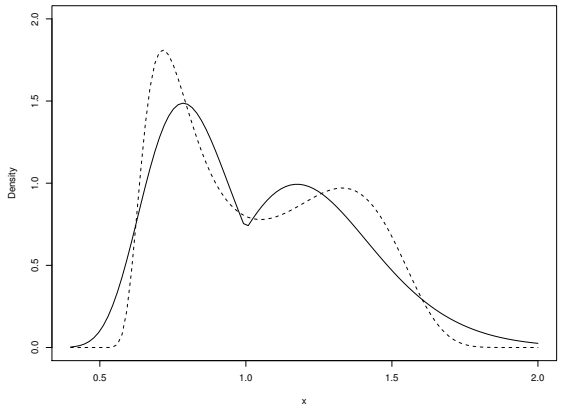

Figure 2.10 Densities $\mathcal{BBS}(0.2,1,-1)$ (solid line) and $\mathcal{GBS}_2(5,1,5)$ (dashed line).

A simulation study was done to evaluate the performance of the nonnested hypothesis tests involving the $\mathcal{BBS}$ and $\mathcal{GBS}_2$ distributions. The models considered were $\mathcal{BBS}(0.2,\ 1,\ -1)$ and $\mathcal{GBS}_2(5,1,5)$. Figure 2.10 shows their density plots, where we can see they have close density shapes. The number of Monte Carlo replications used was $5,000$. First, we considered the case when the true distribution is $\mathcal{BBS}(0.2,1,-1)$; in each replication, $B = 1,000$ bootstrap samples were generated under H_f and another $B = 1,000$ under H_g, obtaining then, one of the results previously cited ($R1$, $R2$, $R3$ or $R4$) for the test in each replication analyzed. Table 2.8 contains the proportion of times each result occurred for this simulation and the proportion of time each distribution is chosen as the most suitable model, which is: the $\mathcal{BBS}$ model, when we obtain result R_1 and $W_{ne} > 0$ or we obtain result R_2; the $\mathcal{GBS}_2$ model, when we obtain result R_1 and $W_{ne} < 0$ or we obtain result R_3; none of the distributions, when we obtain result R_4. The same procedures were done when the true model is the $\mathcal{GBS}_2(5,1,5)$ distribution, with results presented in Table 2.9.

From the results contained in Table 2.8, we can note that the null rejection rates of the true hypothesis (H_f), are close of the nominal levels considered. For instance, when $n = 30$ and $\varepsilon = 0.10$, adding the cells corresponding to $R3$ and $R4$, we see that the rejection rate of H_f is 9.46%, a rate close to the nominal level considered. We can also note that for small or moderate sample sizes the tests tend to indicate equivalence of both models, but when the value

Table 2.8 Proportions of outcomes of the test of H_f against H_g when the data generating function is $\mathcal{BBS}(0.2,1,-1)$ (first four columns), and proportions of $\mathcal{BBS}$ and $\mathcal{GBS}_2$ model selection and porportion of no model selected, with the test significance level being ε

ε	$R1$	$R2$	$R3$	$R4$	$\mathcal{BBS}$	$\mathcal{GBS}_2$	None
			$n=30$				
0.10	0.6676	0.2378	0.0946	0.0000	0.4354	0.5646	0.0000
0.05	0.8284	0.1322	0.0394	0.0000	0.4354	0.5646	0.0000
0.01	0.9650	0.0292	0.0058	0.0000	0.4354	0.5646	0.0000
			$n=50$				
0.10	0.5454	0.3578	0.0968	0.0000	0.5574	0.4426	0.0000
0.05	0.7230	0.2326	0.0444	0.0000	0.5572	0.4428	0.0000
0.01	0.9174	0.0754	0.0072	0.0000	0.5572	0.4428	0.0000
			$n=100$				
0.10	0.3420	0.5538	0.1036	0.0006	0.6952	0.3042	0.0006
0.05	0.5344	0.4138	0.0518	0.0000	0.6940	0.3060	0.0000
0.01	0.8024	0.1884	0.0092	0.0000	0.6940	0.3060	0.0000
			$n=150$				
0.10	0.1920	0.6908	0.1138	0.0034	0.7682	0.2284	0.0034
0.05	0.3772	0.5646	0.0582	0.0000	0.7656	0.2344	0.0000
0.01	0.6720	0.3118	0.0162	0.0000	0.7654	0.2346	0.0000

Table 2.9 Proportions of outcomes of the test of H_f against H_g when the data generating function is $\mathcal{GBS}_2(5,1,5)$ (first four columns), and proportions of $\mathcal{BBS}$ and $\mathcal{GBS}_2$ model selection and porportion of no model selected, with the test significance level being ε

ε	$R1$	$R2$	$R3$	$R4$	$\mathcal{BBS}$	$\mathcal{GBS}_2$	None
			$n=30$				
0.10	0.4438	0.1080	0.4482	0.0000	0.1784	0.8216	0.0000
0.05	0.6658	0.0538	0.2804	0.0000	0.1784	0.8216	0.0000
0.01	0.9040	0.0112	0.0848	0.0000	0.1784	0.8216	0.0000
			$n=50$				
0.10	0.2204	0.1010	0.6786	0.0000	0.1270	0.8730	0.0000
0.05	0.4502	0.0466	0.5032	0.0000	0.1238	0.8762	0.0000
0.01	0.7624	0.0090	0.2286	0.0000	0.1238	0.8762	0.0000
			$n=100$				
0.10	0.0146	0.0662	0.8804	0.0388	0.0672	0.8940	0.0388
0.05	0.1068	0.0560	0.8350	0.0022	0.0712	0.9266	0.0022
0.01	0.4000	0.0112	0.5888	0.0000	0.0674	0.9326	0.0000
			$n=150$				
0.10	0.0002	0.0158	0.8968	0.0872	0.0158	0.8970	0.0872
0.05	0.0136	0.0310	0.9344	0.0210	0.0320	0.9470	0.0210
0.01	0.1610	0.0122	0.8268	0.0000	0.0322	0.9678	0.0000

of n increases, the tests tend to indicate the $\mathcal{BBS}$ model as the most suitable with higher frequency. When $\varepsilon = 0.05$, for example, in the column corresponding to $R2$, this happens 13.22% of the time for $n = 30$, while for $n = 150$, in 56.46% of the time the $\mathcal{BBS}$ distribution is considered the most adequate model. This can be observed also in the fifth column, where we can see that as n increases, the $\mathcal{BBS}$ is chosen more frequently.

Table 2.9 contains the results when H_g is the true hypothesis. Once again, the null rejection rates stayed close of the nominal levels. When $n = 30$ and $\varepsilon = 0.05$, adding the cells relative to $R2$ and $R4$, we see that the null rejection rate of H_g is 5,38%, a value close of the nominal level considered. Moreover, we could observe that the results of Table 2.9 were better than when H_f is the true hypothesis. In the column corresponding to $R3$ are presented the proportion of time that only the $\mathcal{GBS}_2$ is chosen as the most adequate model without considering the sign of W_{ne}. When $\varepsilon = 0.05$, for $n = 30$ this happens 28.04% of time and for $n = 150$ it happens 93.44% of time, a superior performance of when H_f is true. This can also be seen in the sixth column, where we can observe that the proportion of time the $\mathcal{GBS}_2$ distribution is correctly chosen are higher than the corresponding values in the fifth column of Table 2.8.

Therefore, we can note from these results that the nonnested hypothesis tests with bootstrap proposed for the $\mathcal{BBS}$ and $\mathcal{GBS}_2$ models presented satisfactory performances. For both distributions the null rejection rates are close of the nominal levels when the generation is made under the correct null hypothesis. We can also observe that, as n increases, the tests tend to indicate the true distribution at higher rates. We also noted that, for the models considered, the tests presented better performance when the true model is the $\mathcal{GBS}_2$ distribution. Hence, we recommend the use of the nonnested hypothesis test presented above when both $\mathcal{BBS}$ and the $\mathcal{GBS}_2$ laws are plausible for the application at hand since the test was shown to be a reliable tool for distinguishing between the two models, especially when the sample size is not small.

2.7 Empirical applications

2.7.1 Runoff amounts

We shall now return to the data set remarked on Section 2.1, which we used to illustrate how the problem of nonconvergence of optimization processes during the parameters estimation can occur in practical situations with the $\mathcal{BBS}$ model. The data, provided by Folks and Chhikara (1978), consists of 25 runoff amounts at Jug Bridge, in Maryland. Table 2.10 contains some descriptive statistics of this data set. We can observe that the data has a large kurtosis coefficient, i.e., it has a leptokurtic distribution and low variance, which might indicate the data is concentrated around the mean and median values; these characteristics might be an indicative we are dealing with a unimodal data set.

Table 2.10 Descriptive statistics from the runoff data

min	max	median	mean	variance	asymmetry	kurtosis
0.17	2.92	0.7	0.84	0.3459	1.7953	6.7493

The models fitted were the $\mathscr{BS}(\alpha,\beta)$ and $\mathscr{BBS}(\alpha,\beta,\gamma)$ distributions. The parameter estimates of the first model were $\hat{\alpha}=0.66\ (0.0936)$ and $\hat{\beta}=0.69\ (0.0865)$, with standard errors in parenthesis. For the second model, the maximum likelihood estimates were not possible to be obtained because the optimization process failed to reach convergence. As was showed in Figure 2.1a, the log-likelihood have a region apparently flat for some values of the parameter α and γ, with the value of β being fixed at 0.69. Nonetheless, using the estimator MLE_p the process reaches convergence and, as we can see in Figure 2.1b, the penalization modified the log-likelihood function in a way that a maximum point is possible to be reached by the optimization process. The parameters estimates of the $\mathscr{BBS}$ model using the MLE_p were $\hat{\alpha}=0.63\ (0.2287)$, $\hat{\beta}=0.69\ (0.0817)$ and $\hat{\gamma}=-0.13\ (0.8449)$. We can observe that the standard error of $\hat{\gamma}$ is large relative to the point estimate and indicates the $\mathscr{BS}$ model is more adequate.

Figure 2.11 contains the histogram of the data set and the fitted densities. We can note that the estimates obtained for the models lead to very similar densities. Since the $\mathscr{BS}$ distribution is simpler than the $\mathscr{BBS}$ distribution, it seems more suitable than the bimodal distribution. To evaluate this, we tested the hypothesis $H_0:\gamma=0$ against a two-sided alternative. The p-values of the tests LR, score, Wald, LR_{pb}, LR_{bbc} and S_{pb} were 0.85, 0.81, 0.87, 0.92, 0.89 and 0.91, respectively. Therefore, we have strong evidences that the fit obtained with the $\mathscr{BS}$ distribution is better.

In summary, this example illustrates how is possible to find cases where the optimization process being used to obtain maximum likelihood estimates in the $\mathscr{BBS}$ model fails to reach convergence, but that might be possible to solve this by using a penalized log-likelihood function, allowing to carry on further analysis that might indicate whether the $\mathscr{BBS}$ is suitable or the $\mathscr{BS}$ provides a better fit.

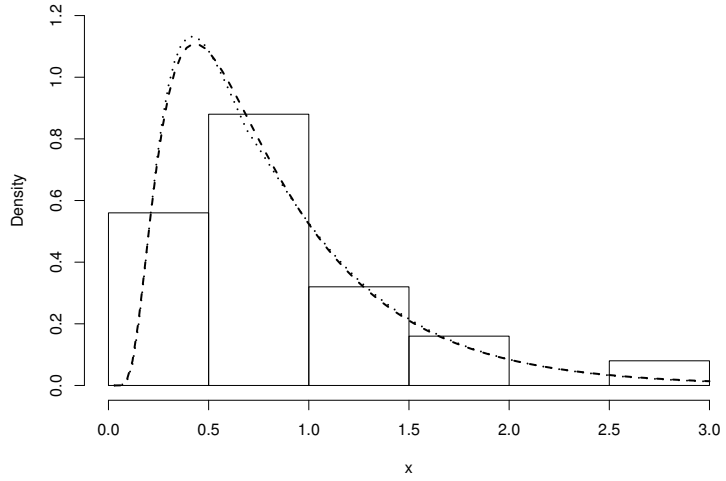

Figure 2.11 Histogram of the runoff data with the fitted densities obtained with $\mathscr{BS}$(0.66, 0.69) (dashed line) and $\mathscr{BBS}$(0.63,0.69,-0.13) (dotted line).

2.7.2 Depressive condition data

The second application example is a data set about 134 children's emotional condition, with their measures of depression. The data was analyzed, for example, by Leiva et al. (2010) and Balakrishnan et al. (2011), where both works involved mixture of distribution in the analysis.

Table 2.11 presents some descriptive statistics of this data set, where we can see the data is right-skewed, has leptokurtic distribution and quite large variance. The fitted model were $\mathscr{BS}$, $\mathscr{BBS}$ and $\mathscr{GBS}_2$, obtaining the Akaike (AIC) and Schwarz (BIC) information criteria for each fit. The estimates, with standard error in parenthesis, of the parameters from the $\mathscr{BS}$ model were $\hat{\alpha} = 0.603$ (0.0368) and $\hat{\beta} = 7.58$ (0.3773), providing AIC and BIC values of 780.09 and 785.89, respectively. The parameters estimates from the $\mathscr{BBS}$ model were $\hat{\alpha} = 0.42$ (0.0481), $\hat{\beta} = 7.54$ (0.2645) and $\hat{\gamma} = -0.85$ (0.2569), with AIC and BIC values of 776.26 and 784.95, respectively. For the $\mathscr{GBS}_2$ model the estimates were $\hat{\alpha} = 2.38$ (0.6290), $\hat{\beta} = 7.74$ (0.3045) and $\hat{\nu} = 1.53$ (0.2525), with AIC and BIC values of 771.78 and 780.47, respectively. The histogram of the data set with the fitted densities of these models are presented in Figure 2.12.

From the information criteria, we would conclude that the most adequate model is the $\mathscr{GBS}_2$ distribution, followed by the $\mathscr{BBS}$ distribution. To test which of these two models is the most suitable for this data, a nonnested hypothesis test was performed, providing a test statistic of value $W_{ne} = -0.0167$ with p-value of 0.0189 under H_f (supposing the $\mathscr{BBS}$ distribution is the true model) and p-value of 0.6453 under H_g (supposing the $\mathscr{GBS}_2$ is the true model). Then, we conclude that the $\mathscr{GBS}_2$ distribution is more adequate for the emotional condition data set.

Table 2.11 Descriptive statistics of the depressive condition data

min	max	median	mean	variance	asymmetry	kurtosis
3	28	8	8.96	28.73	1.11	3.88

2.7.3 Adhesive strength

The third data set analyzed is provided by Ehsani et al. (1996) and was also analyzed by Olmos et al. (2016), who used the $\mathscr{BBS}$ distribution in the analysis. The data is consisted of 48 observations about adhesive strength to concrete of bars reinforced with glass fibre. Some descriptive statistics are presented in Table 2.12, where we can observe that the data presents high kurtosis coefficient, with value greater than 5, is right-skewed and has a variance value much greater than the mean and median.

Table 2.12 Descriptive statistics of the adhesive strength data

min	max	median	mean	variance	asymmetry	kurtosis
3.4	25.5	5.95	8.08	23.7017	1.448	5.0345

Once more, the fitted models were $\mathscr{BS}$, $\mathscr{BBS}$ and $\mathscr{GBS}_2$. For the first model the parameters estimates were $\hat{\alpha} = 0.54$ (0.0553) and $\hat{\beta} = 7.05$ (0.5316), providing AIC and BIC

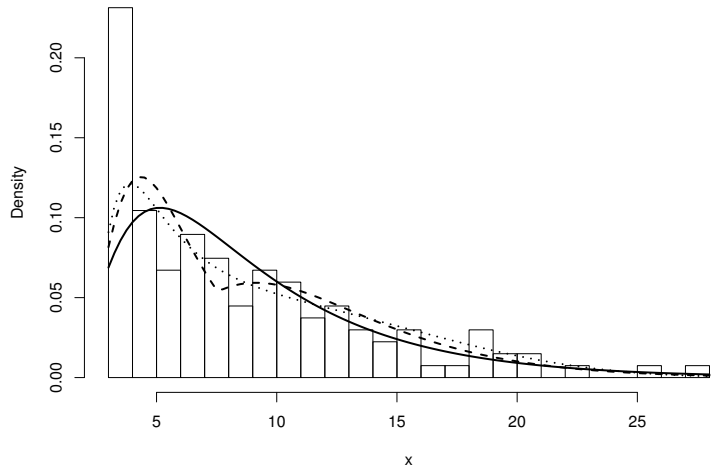

Figure 2.12 Histogram of the depressive condition data with the fitted densities obtained with $\mathscr{BS}$(0.60,7.58) (solid line), $\mathscr{BBS}$(0.42, 7.54, -0.85) (dashed line) and $\mathscr{GBS}_2$(2.38, 7.74, 1.53) (dotted line).

values of 264.52 and 268.26, respectively. The parameter estimates of the second model were $\hat{\alpha} = 0.31\ (0.0460)$, $\hat{\beta} = 7.39\ (0.3162)$ and $\hat{\gamma} = -1.38\ (0.3525)$, providing AIC and BIC values of 260.06 and 265.67, respectively. For the $\mathscr{GBS}_2$ model, the parameter estimates were $\hat{\alpha} = 3.19\ (1.5536)$, $\hat{\beta} = 8.05\ (0.5371)$ and $\hat{\nu} = 1.99\ (0.5203)$, providing AIC and BIC values of 262.26 and 267.88, respectively. The histogram of the fitted densities is shown in Figure 2.13.

For this data set, the best fit according to the information criteria was the $\mathscr{BBS}$ distribution, followed by the $\mathscr{GBS}_2$ model. The nonnested hypothesis test of these two models provided a test statistic of value $W_{ne} = 0.0229$, with p-value of 0.6543 under H_f and with p-value of 0.0489 under H_g, and then, we can see that there is evidence at the 5% significance level that the $\mathscr{BBS}$ distribution is the most adequate model for this data.

Adopting the $\mathscr{BBS}$ as the most suitable model, tests to verify if the data has a bimodal distribution were done. The hypothesis tested were $H_0 : \gamma \geq 0$ and $H_1 : \gamma < 0$. The p-values of the tests SLR, SLR_{c1}, SLR_{c2} and SLR_{bp} were 0.0002, 0.0007, 0.0006 and 0.002, respectively, i.e., in all tests we reject H_0 in favor of H_1. Hence, there is strong evidences that $\gamma < 0$, which indicates the data comes from a bimodal distribution.

2.8 Conclusion

Optimization processes might fail to reach convergence with considerable frequency during maximum likelihood estimation in the $\mathscr{BBS}$ model. A penalization in the log-likelihood was proposed using a modification of the Jeffreys prior. Different methods were tried to solve the nonconvergence problems and the ones that use some penalization in the log-likelihood presented the lowest rates of nonconvergence. In general, the log-likelihood penalized with the modified Jeffreys prior presented the best performance, considering the trade off between quality of the estimates and the nonconvergence rates of the methods analyzed.

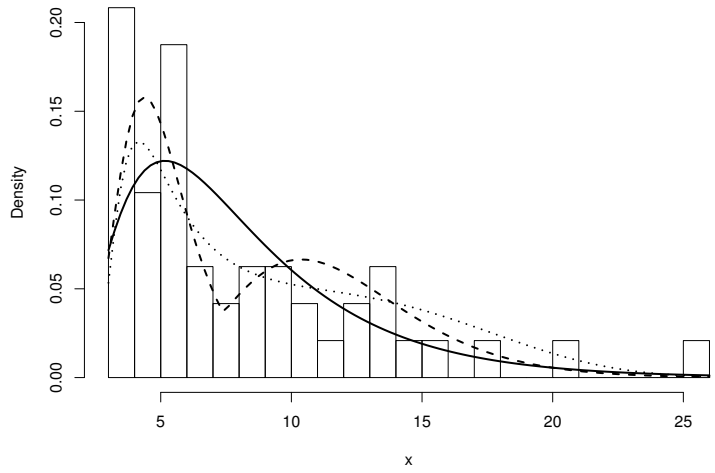

Figure 2.13 Histogram of the adhesive strength data with the fitted densities $\mathscr{BS}$(0.54,7.05) (solid line), $\mathscr{BBS}$(0.31, 7.39, -1.38) (dashed line) and $\mathscr{GBS}_2$(3.19, 8.05, 1.99) (dotted line).

Hypothesis tests with the $\mathscr{BBS}$ model were studied with the penalized log-likelihood, using the modified Jeffreys prior penalization. We analyzed the likelihood ratio, score and Wald tests, observing that the tests are quite liberal, but that their null rejection rates tend to the nominal levels as the sample size increases. The first two tests had the best performances, and we could note that bootstrap resampling can provide much better results. One-sided tests using the signed likelihood ratio were also investigated. We compared the sizes and powers of the test without corrections, with analytical and with bootstrap corrections; the tests with analytical corrections presented the best performances in terms of size and power. Nonnested tests with bootstrap were also considered, providing a way to choose between the $\mathscr{BBS}$ model and another version of the Birnbaum-Saunders distribution that exhibits bimodality. Since in this case there are two distributions, the test was held considering two null hypothesis at time. Besides this difficulty, we could note that null rejection rates of the true hypothesis gets closer to the true nominal level as the sample size increases and that the test tends to indicate the true model more frequently.

An example with real data where the optimization process fails to reach convergence during estimation with the $\mathscr{BBS}$ model was presented, illustrating how this problem can appear in practice and that plausible estimates can be obtained using a penalized log-likelihood. Another two examples with real data were discussed for the nonnested hypothesis tests with distinct results. For the data where the $\mathscr{BBS}$ distribution is chosen as the most suitable model, unilateral tests were performed to verify if their distribution is unimodal, proving results compatible with the empirical distribution of the data set.

Log-linear bimodal Birnbaum-Saunders regression model

3.1 Introduction

The Birnbaum-Saunders distribution ($\mathscr{BS}$) was proposed by Birnbaum and Saunders (1969a) to analyze fatigue lifetime data. It has been widely discussed in the literature and was used as the baseline for several related probability distributions. A concise review on the Birnbaum-Saunders distribution and its extensions can be found in Leiva (2015).

Regression models based on the Birnbaum-Saunders distribution were also discussed in the literature, the first model being introduced by Rieck and Nedelman (1991), who proposed the log-linear Birnbaum-Saunders regression model (BSRM). Their model was later analyzed by Galea et al. (2004), who developed some diagnostic tools for the model, and by Xie and Wei (2007), who provided additional tools for detecting atypical observations.

The BSRM was extended in several different directions. For instance, Barros et al. (2008) developed a model using the Student-*t* Birnbaum-Saunders distribution and Lemonte and Cordeiro (2009) proposed a nonlinear version of the BSRM. A mixed model for censored data based on the Birnbaum-Saunders distribution was introduced by Villegas et al. (2011) and Lemonte (2013) proposed a log-linear model based on a extended Birnbaum-Saunders distribution. The log-linear Birnbaum-Saunders power regression model was proposed by Martínez-Flórez et al. (2016), who also introduced the nonlinear sinh-power-normal regression model. Model misspecification tests for the BSRM were proposed by Santos and Cribari-Neto (2015). Bayesian inference for the BSRM was developed by Tsionas (2001). More recently, Vilca et al. (2016) introduced the nonlinear sinh-normal/independent regression model, which encompasses several other $\mathscr{BS}$ regression models and developed Bayesian inference for such a model.

The chief goal of this chapter is to propose a log-linear regression model based on a bimodal version of the Birnbaum-Saunders distribution that has been recently introduced by Owen and Ng (2015). Such a distribution is more flexible than the original $\mathscr{BS}$ law and we use it as the basis for developing a regression model that is more general than the BSRM introduced by Rieck and Nedelman (1991). Parameter estimation and standard inferential strategies are presented. A second goal of this chapter is to provide diagnostic tools for the proposed regression model, thus allowing practitioners to verify whether the model assumptions are satisfied and making it possible for them to detect atypical observations, much in the same spirit as done for the BSRM by Galea et al. (2004) and Xie and Wei (2007). We also present a RESET-like misspecification test which is similar to the one introduced by Santos and Cribari-Neto (2015) for the log-linear Birnbaum-Saunders model. It can be used to check whether the model's

functional form is correctly specified. In addition, we consider the issue of performing model selection based on model selection criteria and the construction of prediction intervals for the regression model we propose similarly to what was done by Bayer and Cribari-Neto (2015) and Espinheira et al. (2014), respectively, for beta regressions. In summary, we propose a new regression model which is more flexible than the classic BSRM model and develop model selection strategies, prediction intervals, residuals and diagnostic tools for the proposed model.

The remainder of the chapter is organized as follows. In Section 3.2, the bimodal Birnbaum-Saunders distribution is presented as well as some of its key properties. The log-linear model for responses that follow such a bimodal Birnbaum-Saunders distribution is proposed in Section 3.3. A broad variety of diagnostic tools for the proposed model are presented in Section 3.4. In Section 3.5 we address the issue of constructing prediction intrevals for non-observed response values and in Section 3.6 we consider different model selection strategies. Results from Monte Carlo simulations are reported in Section 3.7, and an empirical application is presented and discussed in Section 3.8. Finally, Section 3.9 offers some concluding remarks.

3.2 The bimodal Birnbaum-Saunders distribution

The generalized Birnbaum-Saunders distribution considered in this chapter was introduced by Díaz-García and Domınguez-Molina (2006). They obtained it by adding a second shape parameter to the $\mathscr{BS}$ distribution function. More recently, Owen and Ng (2015) analyzed the distribution, which the authors denoted by $\mathscr{GBS}2$, in a paper where they investigated the relationship among the inverse Gaussian, $\mathscr{BS}$ and $\mathscr{GBS}_2$ distributions.

A random variable T is said to follow the $\mathscr{GBS}2(\alpha,\eta,\nu)$ law if its distribution function is given by

$$F_T(t|\alpha,\eta,\nu) = \Phi\left(\frac{1}{\alpha}\left[\left(\frac{t}{\eta}\right)^{\nu} - \left(\frac{\eta}{t}\right)^{\nu}\right]\right), \quad t > 0, \tag{3.1}$$

where $\Phi(\cdot)$ denotes the standard normal distribution function, $\alpha > 0$, $\eta > 0$ and $\nu > 0$. Here, η is the scale parameter whereas α and ν are shape parameters. The probability density function of T is

$$f_T(t|\alpha,\eta,\nu) = \frac{\nu}{t\alpha\sqrt{2\pi}}\left[\left(\frac{t}{\eta}\right)^{\nu} + \left(\frac{\eta}{t}\right)^{\nu}\right]\exp\left\{-\frac{1}{2\alpha^2}\left[\left(\frac{t}{\eta}\right)^{\nu} - \left(\frac{\eta}{t}\right)^{\nu}\right]^2\right\}, \quad t > 0.$$

The $\mathscr{GBS}_2$ distribution has a noteworthy advantage over the original $\mathscr{BS}$ distribution: the density of the former can be unimodal and bimodal whereas that of the latter does not allow for more than one mode. According to Owen and Ng (2015), the $\mathscr{GBS}_2$ density is bimodal whenever $\alpha > 2$ and $\nu > 2$ simultaneously. Figure 3.1 contains $\mathscr{GBS}_2$ density plots for different values of α, η and ν. It is noteworthy (Figure 3.1a) that the $\mathscr{GBS}_2$ density becomes more symmetric as the value of ν increases and the values of the remaining parameters are held constant. The value of α also impacts the distribution asymmetry; the distribution becomes less asymmetric as the value of α decreases (Figure 3.1b). In Figure 3.1c we see an example of a bimodal density when we take $\alpha > 2$ and $\nu > 2$. Notice that the $\mathscr{GBS}_2$ density is quite flexible, since it may assume a variety of different shapes.

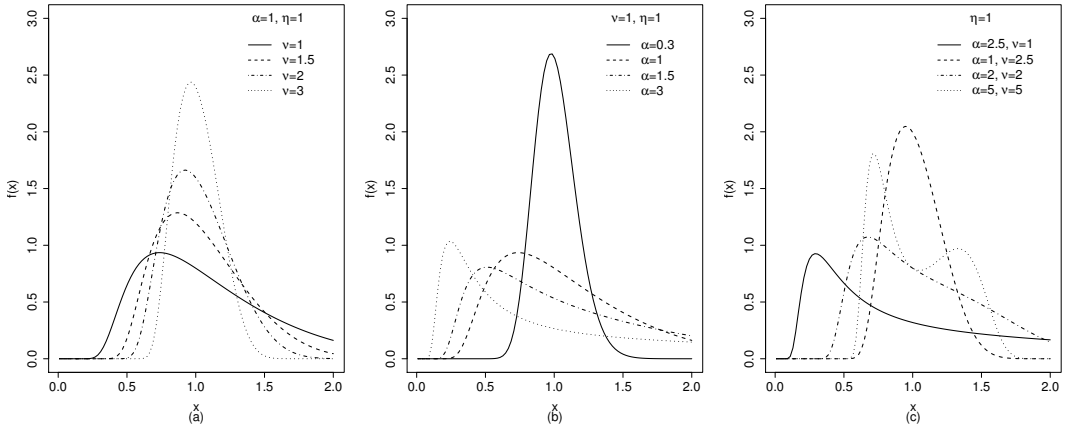

Figure 3.1 $\mathcal{GBS}_2(\alpha,\eta,\nu)$ densities for some parameter values.

Several useful properties of the $\mathcal{GBS}_2$ distribution were obtained by Owen and Ng (2015), some of them also holding for the $\mathcal{BS}$ distribution. For instance, it was shown that η is the distribution median, which can be easily verified from Equation (3.1). Additionally, the $\mathcal{GBS}_2$ distribution is closed under reciprocity and proportionality, i.e., $T^{-1} \sim \mathcal{GBS}_2(\alpha,\eta^{-1},\nu)$ and $aT \sim \mathcal{GBS}_2(\alpha,a\eta,\nu)$, for $a>0$. Pseudo-random number generation from $T \sim \mathcal{GBS}_2$ can be performed using the following stochastic representation:

$$T = \eta\left[\frac{\alpha Z}{2} + \sqrt{\left(\frac{\alpha Z}{2}\right)^2 + 1}\right]^{1/\nu},$$

where Z follows the standard normal distribution.

Another stochastic representation for the $\mathcal{GBS}_2$ distribution, as noted by Owen and Ng (2015), is as follows: if $T \sim \mathcal{GBS}_2(\alpha,\eta,\nu)$, then $Y = \log(T)$ is distributed as hyperbolic sine normal ($\mathcal{SHN}$), whose distribution function is

$$F_Y(y|\alpha,\mu,\sigma) = \Phi\left(\frac{2}{\alpha}\sinh\left(\frac{y-\mu}{\sigma}\right)\right), \quad y \in \mathbb{R},$$

where $\mu = \log(\eta)$, $\sigma = 1/\nu$ and $\sinh(\cdot)$ denotes the hyperbolic sine function. We shall write $Y \sim \mathcal{SHN}(\alpha,\mu,\sigma)$, α being a shape parameter, σ being a scale parameter and μ being a location parameter and mean of the distribution.

The relationship between the $\mathcal{BS}$ and $\mathcal{SHN}$ distributions was established by Rieck and Nedelman (1991), who noted that the log-Birnbaum-Saunders is a particular case of the $\mathcal{SHN}$ distribution: the latter reduces to the former when $\sigma = 2$. The authors have also presented several properties of the $\mathcal{SHN}$ distribution and used the relationship between the two distributions to propose the log-linear Birnbaum-Saunders regression model.

Using the results developed by Rieck (1999), who obtained closed-form expressions for integer and fractional $\mathcal{BS}$ moments, Owen and Ng (2015) obtained an expression for the rth

$\mathcal{GBS}_2$ moment:

$$\mathbb{E}(T^r) = \eta^r \frac{\exp(\alpha^{-2})}{\alpha\sqrt{2\pi}} \left[K_{(r/\nu+1)/2}(\alpha^{-2}) + K_{(r/\nu-1)/2}(\alpha^{-2})\right],$$

where $K_\omega(z)$ denotes the modified Bessel function of the third kind of order ω, i.e.,

$$K_\omega(z) = \frac{1}{2}\int_{-\infty}^{\infty} \exp[-z\cosh(x) - \omega x]dx,$$

$\cosh(\cdot)$ denoting the hyperbolic cosine function.

3.3 Log-linear $\mathcal{GBS}_2$ regression model

We use the relationship between the $\mathcal{GBS}_2$ and $\mathcal{SHN}$ distributions to propose a log-linear $\mathcal{GBS}_2$ regression model, the response variable being $Y = \log(T)$, where T follows the $\mathcal{GBS}_2$ law. Consider n independent random variables $T_1, \ldots, T_n$, where $T_i \sim \mathcal{GBS}_2(\alpha, \eta_i, \nu)$, $i = 1, \ldots, n$. The log-linear $\mathcal{GBS}_2$ regression model (GBS2RM) is defined by

$$y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \ldots, n,$$

where $y_i = \log(t_i)$, with $t_1, \ldots, t_n$ representing observations on the random variables $T_1, \ldots, T_n$. Here, $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^\top$ is a vector of explanatory variables associated with the response variable, $\boldsymbol{\beta}$=$(\beta_1, \ldots, \beta_p)^\top$ is a p-vector of unknown parameters and $\varepsilon_i \overset{\text{iid}}{\sim} \mathcal{SHN}(\alpha, 0, \nu^{-1})$, iid indicating that the random variables are independent and identically distributed. Note that $t_i = \exp(\mathbf{x}_i^\top \boldsymbol{\beta})e^{\varepsilon_i}$, where $e^{\varepsilon_i} \sim \mathcal{GBS}_2(\alpha, 1, \nu)$. Since $\mathcal{GBS}_2$ distribution is closed under proportionality, it follows that T_i is $\mathcal{GBS}_2(\alpha, \exp(\mathbf{x}_i^\top \boldsymbol{\beta}), \nu)$ distributed.

Estimation of $\boldsymbol{\theta} = (\boldsymbol{\beta}, \alpha, \nu)^\top$, the GBS2RM parameter vector, can be carried out by maximum likelihood. The log-likelihood function is given by

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left\{ \log\left(\frac{2}{\sqrt{2\pi}}\right) + \log(\xi_{i1}) - \frac{1}{2}\xi_{i2}^2 \right\},$$

where $\xi_{i1} = \nu\alpha^{-1}\cosh[\nu(y_i - \mu_i)]$ and $\xi_{i2} = 2\alpha^{-1}\sinh[\nu(y_i - \mu_i)]$, $\mu_i = \mathbf{x}_i^\top \boldsymbol{\beta}$ being the linear predictor, $i = 1, \ldots, n$. The first derivatives of $\ell(\boldsymbol{\theta})$ with respect to the model's parameters are

$$\begin{aligned}
\ell_{\boldsymbol{\beta}} &= \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} = X^\top \mathbf{a}, \\
\ell_\alpha &= \frac{\partial \ell(\boldsymbol{\theta})}{\partial \alpha} = \sum_{i=1}^{n} \left\{ \frac{1}{\alpha}\left(\xi_{i2}^2 - 1\right) \right\}, \\
\ell_\nu &= \frac{\partial \ell(\boldsymbol{\theta})}{\partial \nu} = \sum_{i=1}^{n} \left\{ \frac{1}{\nu} + \frac{\xi_{i2}}{\xi_{i1}}\frac{\nu}{2}(y_i - \mu_i) - \xi_{i1}\xi_{i2}\frac{2}{\nu}(y_i - \mu_i) \right\},
\end{aligned}$$

where $X = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^\top$ is an $n \times p$ full column rank matrix and $\mathbf{a} = (a_1, \ldots, a_n)^\top$, where

$$a_i = 2\xi_{i1}\xi_{i2} - \frac{\nu^2}{2}\frac{\xi_{i2}}{\xi_{i1}}.$$

The maximum likelihood estimator (MLE) $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{\alpha}, \hat{\nu})^\top$ of the parameters that index the model are obtained by solving $\ell_{\boldsymbol{\theta}} = 0$, where $\ell_{\boldsymbol{\theta}} = \partial\ell(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$. They cannot be expressed in closed-form. Estimates can be obtained by numerically maximizing $\ell(\boldsymbol{\theta})$ using a Newton (e.g., Newton-Raphson) or quasi-Newton (e.g., BFGS) nonlinear optimization algorithm. To that end, one must select a starting point for the iterative scheme. We suggest using the least squares estimate $\boldsymbol{\beta}_0 = (X^\top X)^{-1}X^\top \mathbf{y}$ as a starting value for $\hat{\boldsymbol{\beta}}$, where $\mathbf{y} = (y_1, \ldots, y_n)^\top$, along with $\nu_0 = 0.5$ as a starting point for ν, which is the value of ν that corresponds to the $\mathscr{BS}$ distribution. For $\hat{\alpha}$ we propose using

$$\alpha_0 = \sqrt{\frac{4}{n}\sum_{i=1}^{n}\sinh^2[\nu_0(y_i - \mathbf{x}_i^\top\boldsymbol{\beta}_0)]},$$

which corresponds to the solution of $\ell_\alpha|_{(\boldsymbol{\beta},\nu)=(\boldsymbol{\beta}_0,\nu_0)} = 0$.

The model Hessian matrix is

$$\ell_{\boldsymbol{\theta}\boldsymbol{\theta}} = \frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} = \begin{bmatrix} X^\top VX & X^\top\mathbf{d} & X^\top\mathbf{g} \\ \mathbf{d}^\top X & \operatorname{tr}[D(\mathbf{b})] & \operatorname{tr}[D(\mathbf{c})] \\ \mathbf{g}^\top X & \operatorname{tr}[D(\mathbf{c})] & \operatorname{tr}[D(\mathbf{f})] \end{bmatrix},$$

where $V = \operatorname{diag}\{v_1, \ldots, v_n\}$, $D(\mathbf{b}) = \operatorname{diag}\{b_1, \ldots, b_n\}$, $D(\mathbf{c}) = \operatorname{diag}\{c_1, \ldots, c_n\}$, $D(\mathbf{f}) = \operatorname{diag}\{f_1, \ldots, f_n\}$, $\mathbf{d} = (d_1, \ldots, d_n)^\top$ and $\mathbf{g} = (g_1, \ldots, g_n)^\top$, with diag denoting a diagonal matrix and tr denoting the trace operator. The components of these vectors and matrices are

$$\begin{aligned}
v_i &= \nu^2\left\{\frac{1}{\cosh^2[\nu(y_i-\mu_i)]} - \frac{4}{\alpha^2}\cosh[2\nu(y_i-\mu_i)]\right\},\\
b_i &= -\frac{3}{\alpha^2}\xi_{i2}^2 + \frac{1}{\alpha^2},\\
c_i &= \frac{4}{\alpha\nu}\xi_{i1}\xi_{i2}(y_i-\mu_i),\\
d_i &= -\frac{4}{\alpha}\xi_{i1}\xi_{i2},\\
f_i &= -\frac{1}{\nu^2} + \left[\frac{(y_i-\mu_i)\nu}{\alpha\xi_{i1}}\right]^2 - \frac{4}{\alpha^2}(y_i-\mu_i)^2\cosh[2\nu(y_i-\mu_i)],\\
g_i &= \frac{-\nu(y_i-\mu_i)}{\cosh^2[\nu(y_i-\mu_i)]} - \tanh[\nu(y_i-\mu_i)] + \frac{4\nu}{\alpha^2}\cosh[2\nu(y_i-\mu_i)](y_i-\mu_i) + \frac{2}{\alpha^2}\sinh[2\nu(y_i-\mu_i)].
\end{aligned}$$

Under mild regularity conditions (Severini, 2000), it can be shown that $\hat{\boldsymbol{\theta}}$ is asymptotically distributed as $\mathscr{N}_{p+2}(\boldsymbol{\theta}, \Sigma_{\boldsymbol{\theta}})$. The asymptotic covariance of $\hat{\boldsymbol{\theta}}$, $\Sigma_{\boldsymbol{\theta}}$, can be approximated by $-\hat{\ell}_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1}$, where $\hat{\ell}_{\boldsymbol{\theta}\boldsymbol{\theta}}$ denotes $\ell_{\boldsymbol{\theta}\boldsymbol{\theta}}$ evaluated at $\hat{\boldsymbol{\theta}}$. Hence, based on the asymptotic normality of $\hat{\boldsymbol{\theta}}$,

it is possible to obtain an approximated $100 \times (1-\gamma)\%$ confidence region for $\boldsymbol{\theta}$, $0 < \gamma < 1$, which is given by the set of values of $\boldsymbol{\theta}$ such that

$$(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^\top (-\hat{\ell}_{\boldsymbol{\theta}\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \leq \chi^2_{p+2}(\gamma),$$

where $\chi^2_{p+2}(\gamma)$ denotes the $1-\gamma$ quantile of the chi-square distribution with $p+2$ degrees of freedom.

Consider now the following partition of the parameter vector: $\boldsymbol{\theta} = (\boldsymbol{\psi}, \boldsymbol{\lambda})^\top$, where $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_r)^\top$ is the vector of parameters of interest and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_s)^\top$ is the vector of nuisance parameters, with $r + s = p + 2$. We shall focus on the test of $\mathscr{H}_0 : \boldsymbol{\psi} = \boldsymbol{\psi}_0$ against $\mathscr{H}_1 : \boldsymbol{\psi} \neq \boldsymbol{\psi}_0$ in the context the GBS2RM model. The likelihood ratio statistic is given by

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

where $\tilde{\boldsymbol{\theta}}$ denotes the restricted maximum likelihood estimator of $\boldsymbol{\theta}$, which is obtained by maximizing $\ell(\boldsymbol{\theta})$ subject to $\boldsymbol{\psi} = \boldsymbol{\psi}_0$, i.e., imposing the null hypothesis. Under standard regularity conditions, the asymptotic distribution of W under $\mathscr{H}_0$ is chi-squared with r degrees of freedom. Therefore, the null hypothesis is rejected at significance level γ if $W > \chi^2_r(\gamma)$.

The coefficient of determination, R^2, is widely used in the classic linear regression to measure how well the model fits the data. It assumes values in the interval [0,1] and the larger the R^2, the better the model fit. A generalization of such a measure was proposed by Nagelkerke (1991), which we shall denote by R^2_N. It can be computed for different regression models, including the GBS2RM, where R^2_N can be viewed as a pseudo-R^2. Denoting the full model likelihood function by $L(\hat{\boldsymbol{\theta}})$ and the likelihood function obtained only using the intercept in the linear predictor by $L(0)$, Nagelkerke's pseudo R^2 is given by

$$R^2_N = \frac{1 - \{L(0)/L(\hat{\boldsymbol{\theta}})\}^{2/n}}{1 - L(0)^{2/n}},$$

where n is the sample size. Similarly to R^2, the closer to 1 the coefficient R^2_N is, the better will be the fitted model.

3.4 Diagnostic methods

Diagnostic analysis tools allow practitioners to verify whether a fitted regression model represents well the data at hand. In particular, such tools can be used to verify whether the model assumptions are satisfied and also whether parameter estimation is considerably affected by a few atypical observations. In what follows we shall develop some diagnostic analysis tools for the GBS2RM model.

3.4.1 Residual analysis

We propose two different residuals for the GBS2RM model. They can both be easily computed. The first residual we introduce is based on the stochastic relationship between the normal

and $\mathcal{SHN}$ distributions, namely: if $Y \sim \mathcal{SHN}(\alpha, \mu, \nu^{-1})$, then $Z = 2\alpha^{-1}\sinh[\nu(Y-\mu)]$ follows the standard normal distribution. Let $\hat{\boldsymbol{\mu}} = (\hat{\mu}_1, \ldots, \hat{\mu}_n)^\top$ be the estimated linear predictor. The first residual is given by

$$r_{SHN_i} = \frac{2}{\hat{\alpha}}\sinh(\hat{\nu}(y_i - \hat{\mu}_i)) = \hat{\xi}_{i2}, \quad i = 1, \ldots, n.$$

It is standard normally distributed if the model's distributional assumptions are correct.

The second proposed residual is the generalized Cox-Snell residual, which in the GBS2RM is given by

$$r_{CSG_i} = -\log\left(1 - F_Y(y_i|\hat{\boldsymbol{\theta}})\right) = -\log\left(1 - \Phi(\hat{\xi_{i2}})\right).$$

This residual is expected to be exponentially distributed with unit mean if the model's distributional assumptions hold true.

A common practice is to use the idea outlined by Atkinson (1985) when performing a residual analysis. He suggested constructing confidence bands for quantile-quantile (QQ) plots of the residuals. That can be easily done in the context of the GBS2RM model, as we shall now explain. Given a vector $\hat{\boldsymbol{\theta}}$ of estimates and given a vector of residuals $\mathbf{r} = (r_1, \ldots, r_n)^\top$ for which we want to obtain confidence bands, we can proceed as follows:

1. Generate $\boldsymbol{\varepsilon}^* = (\varepsilon_1^*, \ldots, \varepsilon_n^*)^\top$ from $\mathcal{SHN}(\hat{\alpha}, 0, 1/\hat{\nu})$ and obtain a simulated sample: $\mathbf{y}^* = X\hat{\boldsymbol{\beta}} + \boldsymbol{\varepsilon}^*$.

2. Fit a GBS2RM model using $\mathbf{y}^*$ as the response variable and X as the matrix of explanatory variables, and obtain the corresponding residuals: $\mathbf{r}^b = (r_1^b, \ldots, r_n^b)^\top$.

3. Execute steps 1 and 2 a large number of times, say B. Then, obtain the ordered residuals $\mathbf{r}^b_{ord} = (r^b_{(1)}, \ldots, r^b_{(n)})^\top$, where $r^b_{(i)}$ is the ith order statistic of $\mathbf{r}^b$, $b = 1, \ldots, B$. Obtain a $(1-\rho)$ confidence interval for $r_{(i)}$, the ith order statistic of $\mathbf{r}$, by computing the $\rho/2$ and $1-\rho/2$ quantiles of $r^1_{(i)}, \ldots, r^B_{(i)}$, the residuals obtained from the simulation, $i = 1, \ldots, n$. Then, the $(1-\rho)$ confidence bands for the residuals are obtained by plotting the empirical quantiles of $\mathbf{r}$ along with its respective $1-\rho$ confidence intervals against the theoretical quantiles of residuals' reference distribution.

One can then compute confidence bands for the residuals r_{SHN} and r_{CSG} and use them to check whether the fitted model represents well the data.

3.4.2 Local influence

Practitioners are oftentimes interested in measuring the impact of different observations on the resulting parameter estimates. That can be accomplished using the local influence method proposed by Cook (1986), which is based on the likelihood displacement $LD(\boldsymbol{\omega}) = 2[\ell(\hat{\boldsymbol{\theta}}) - \ell(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega})]$ as a function of a given perturbation vector $\boldsymbol{\omega} \in \Omega$. Here, $\boldsymbol{\omega}$ is a $q \times 1$ vector, Ω is an open subset of $\mathbb{R}^q$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^\top$ is the model parameter vector. The no perturbation vector $\boldsymbol{\omega}_0$ yields the minimal likelihood displacement and is such that $\ell(\hat{\boldsymbol{\theta}}) = \ell(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0)$. The

interest lies in evaluating the behavior of $LD(\boldsymbol{\omega})$ around $\boldsymbol{\omega}_0$ by analyzing the normal curvature of the plot of $LD(\boldsymbol{\omega}_0 + a\mathbf{l})$ against a, where $a \in \mathbb{R}$ and $\mathbf{l}$ is the unit norm direction. According to Cook (1986), such a curvature is given by

$$C_{\mathbf{l}}(\hat{\boldsymbol{\theta}}) = 2|\mathbf{l}^\top \Delta^\top \ell_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1} \Delta \mathbf{l}|,$$

where Δ is the perturbation matrix, whose (i, j) element is

$$\Delta_{ij} = \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \theta_i \partial \omega_j}, \quad i = 1, \ldots, p \text{ and } j = 1, \ldots, q.$$

This matrix is evaluated at both $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ and $\boldsymbol{\omega} = \boldsymbol{\omega}_0$.

According to Cook (1986), the main interest lies in the maximal curvature, C_{max}, which is given by the largest eigenvalue of the matrix $B = \Delta^\top \ell_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1} \Delta$. The direction $\mathbf{l}_{max}$ is the eigenvector of B corresponding to C_{max}. The index plot of $\mathbf{l}_{max}$ may reveal which data points lead to the largest changes in $LD(\boldsymbol{\omega})$. It can thus be used to detect influential observations.

Consider the partition of the parameter vector as $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)^\top$ and suppose the interest lies in evaluating the influence on $\boldsymbol{\theta}_1$. Cook (1986) showed that the normal curvature is $C_{\mathbf{l}}(\hat{\boldsymbol{\theta}}_1) = 2|\mathbf{l}^\top \Delta^\top (\ell_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1} - B_1) \Delta \mathbf{l}|$, where

$$B_1 = \begin{bmatrix} 0 & 0 \\ 0 & \ell_{\boldsymbol{\theta}_2\boldsymbol{\theta}_2}^{-1} \end{bmatrix},$$

where $\ell_{\boldsymbol{\theta}_2\boldsymbol{\theta}_2} = \partial^2 \ell(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}_2 \boldsymbol{\theta}_2^\top$. Hence, an analysis of influence can be based on the index plot of the eigenvector corresponding to the largest eigenvalue of $\Delta^\top (\ell_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1} - B_1) \Delta$. Similarly, when the interest lies $\boldsymbol{\theta}_2$, the normal curvature is given by $C_{\mathbf{l}}(\hat{\boldsymbol{\theta}}_2) = 2|\mathbf{l}^\top \Delta^\top (\ell_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1} - B_2) \Delta \mathbf{l}|$, where

$$B_2 = \begin{bmatrix} \ell_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}^{-1} & 0 \\ 0 & 0 \end{bmatrix}.$$

Here, $\ell_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1} = \partial^2 \ell(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}_1 \boldsymbol{\theta}_1^\top$. Again, the analysis is based on the index plot of the eigenvector corresponding to the largest eigenvalue of $\Delta^\top (\ell_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1} - B_2) \Delta$.

We shall consider three different perturbation schemes for the local influence analysis in the GBS2RM, namely: case-weights perturbation, response variable perturbation, and explanatory variables perturbation. In the following subsections, we shall provide closed-form expressions for the perturbation matrix in such perturbation schemes.

3.4.2.1 Case-weights perturbation

In this scheme, the weight ω_i represents the contribution of y_i to the log-likelihood, $i = 1, \ldots, n$. The perturbed log-likelihood function is thus given by

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

where $\ell_i(\boldsymbol{\theta}|\boldsymbol{\omega}) = \log(2/\sqrt{2\pi}) + \log(\xi_{i1}) - \xi_{i2}^2/2$. The no perturbation vector is $\boldsymbol{\omega}_0 = (1,\ldots,1)^\top$ and it is possible to show that the components of the perturbation matrix are

$$\begin{aligned}
\Delta_{\boldsymbol{\beta}} &= X^\top D(\mathbf{a}),\\
\Delta_{\alpha i} &= (\xi_{i2}^2 - 1)/\alpha,\\
\Delta_{\nu i} &= \frac{1}{\nu} + \frac{\xi_{i2}}{\xi_{i1}}\frac{\nu}{2}(y_i - \mu_i) - \xi_{i2}\xi_{i1}\frac{2}{\nu}(y_i - \mu_i),
\end{aligned}$$

where $\mathbf{a}$ is as before. Let $\Delta_\alpha = (\Delta_{\alpha 1}\cdots\Delta_{\alpha n})$ and $\Delta_\nu = (\Delta_{\nu 1}\cdots\Delta_{\nu n})$ be row vectors. The perturbation matrix can be expressed as $\Delta = (\Delta_{\boldsymbol{\beta}}^\top, \Delta_\alpha^\top, \Delta_\nu^\top)^\top$, a matrix of dimension $(p+2)\times n$.

3.4.2.2 Response variable perturbation

In this perturbation scheme, a modified response variable of the form $y_{i\omega} = y_i + \omega_i S_y$ is considered, $i = 1,\ldots,n$, where ω_i is the ith component of the perturbation vector $\boldsymbol{\omega} = (\omega_1,\ldots,\omega_n)^\top$ and S_y is a scaling factor, usually taken to be the standard deviation of $\mathbf{y} = (y_1,\ldots,y_n)^\top$. The no perturbation vector is $\boldsymbol{\omega}_0 = (0,\ldots,0)^\top$. After some algebra, we obtained

$$\begin{aligned}
\Delta_{\boldsymbol{\beta}} &= S_y X^\top D(\mathbf{m}),\\
\Delta_{\alpha i} &= 4\xi_{i1}\xi_{i2}S_y/\alpha,\\
\Delta_{\nu i} &= \frac{\nu S_y(y_i - \mu_i)}{\cosh^2[\nu(y_i - \mu_i)]} + S_y\tanh[\nu(y_i - \mu_i)] - \frac{4\nu S_y}{\alpha^2}(y_i - \mu_i)\cosh[2\nu(y_i - \mu_i)]\\
&\quad - \frac{4S_y}{\alpha^2}\sinh[\nu(y_i - \mu_i)]\cosh[\nu(y_i - \mu_i)],
\end{aligned}$$

where $D(\mathbf{m}) = \mathrm{diag}\{m_1,\ldots,m_n\}$, with $m_i = \nu^2\left[4\alpha^{-2}\cosh[2\nu(y_i - \mu_i)] - \cosh^{-2}[\nu(y_i - \mu_i)]\right]$. The perturbation matrix is given by $\Delta = (\Delta_{\boldsymbol{\beta}}^\top, \Delta_\alpha^\top, \Delta_\nu^\top)^\top$.

3.4.2.3 Explanatory variables perturbation

This scheme is considered when we are interested in analyzing the impact of a perturbation on a specific explanatory variable, of index j say. The perturbation here is of the form $x_{kj\omega} = x_{kj} + \omega_k S_x$, where $j \in \{1,\ldots,p\}$, $k = 1,\ldots,n$ and S_x is a scaling factor, usually equal to the standard deviation of $(x_{1j},\ldots,x_{nj})^\top$. The no perturbation vector is $\boldsymbol{\omega}_0 = (0,\ldots,0)^\top$. It is possible to show that the perturbation matrix is composed by

$$\begin{aligned}
\Delta_{\boldsymbol{\beta}} &= S_x\beta_j X^\top D(\mathbf{o}_1) + S_x\mathbf{q}^{(j)}\mathbf{o}_2^\top,\\
\Delta_{\alpha i} &= -4S_x\beta_j\xi_{i1}\xi_{i2}/\alpha,\\
\Delta_{\nu i} &= -S_x\beta_j\left\{\frac{(y_i - \mu_i)\nu^3}{\alpha^2\xi_{i1}^2} + \frac{\nu\xi_{i2}}{2\xi_{i1}} - \frac{4\nu}{\alpha^2}(y_i - \mu_i)\cosh[2\nu(y_i - \mu_i)] - \frac{2}{\nu}\xi_{i2}\xi_{i1}\right\},
\end{aligned}$$

where $\mathbf{o}_1 = (o_{11}, \ldots, o_{1n})^\top$ and $\mathbf{o}_2 = (o_{21}, \ldots, o_{2n})^\top$, with

$$o_{1i} = \nu^2 \left\{ \frac{1}{\cosh^2[\nu(y_i - \mu_i)]} - \frac{4}{\alpha^2} \cosh[2\nu(y_i - \mu_i)] \right\},$$
$$o_{2i} = \nu \left\{ \frac{2}{\alpha^2} \sinh[2\nu(y_i - \mu_i)] - \tanh[\nu(y_i - \mu_i)] \right\},$$

and $\mathbf{q}^{(j)}$ is a vector with one in the jth position and zero elsewhere. The perturbation matrix is once again given by $\Delta = (\Delta_{\boldsymbol{\beta}}^\top, \Delta_\alpha^\top, \Delta_\nu^\top)^\top$.

3.4.3 Generalized leverage

The generalized leverage method was proposed by Wei et al. (1998) and aims at measuring the influence of observed values on predicted values. Let $\tilde{\mathbf{y}} = (\tilde{y}_1, \ldots, \tilde{y}_n)^\top$ be the vector of predicted values. The generalized leverage is given by $\partial \tilde{y}_i / \partial y_j$, i.e., it is the change in the ith predicted value induced by the jth response value. The leverage matrix proposed by the authors is given by

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

where $\ell_{\boldsymbol{\theta}\mathbf{y}} = \partial^2 \ell(\boldsymbol{\theta}) / \partial \boldsymbol{\theta} \partial \mathbf{y}^\top$ and $D_{\boldsymbol{\theta}} = \partial \boldsymbol{\mu} / \partial \boldsymbol{\theta}^\top$, $\boldsymbol{\mu}$ denoting the expected value of $\mathbf{y}$. Such a matrix is evaluated at the maximum likelihood estimate $\hat{\boldsymbol{\theta}}$ and the leverage points are those observations with large values of GL_{ii}, the ith diagonal element of $GL(\theta)$, $i = 1, \ldots, n$.

In the GBS2RM we have that $\boldsymbol{\mu} = X\boldsymbol{\beta}$, thus $D_{\boldsymbol{\theta}} = [X \; \mathbf{0} \; \mathbf{0}]$, an $n \times (p+2)$ matrix, where $\mathbf{0}$ denotes an n-vector of zeros. Additionally,

$$\ell_{\boldsymbol{\theta}\mathbf{y}} = (-1) \begin{bmatrix} X^\top V \\ \mathbf{d}^\top \\ \mathbf{g}^\top \end{bmatrix},$$

where the expressions of V, $\mathbf{d}$ and $\mathbf{g}$ are given in Section 3.3.

When we only focus on the vector $\boldsymbol{\beta}$, we obtain $D_{\boldsymbol{\beta}} = X$, $\ell_{\boldsymbol{\beta}\mathbf{y}} = -X^\top V$ and $\ell_{\boldsymbol{\beta}\boldsymbol{\beta}} = X^\top V X$. Consequently, the leverage matrix for the regression parameters is given by

$$GL(\boldsymbol{\beta}) = X \left(X^\top V X \right)^{-1} X^\top V.$$

3.4.4 Generalized Cook's distance

According to Xie and Wei (2007) and Cook and Weisberg (1982), the generalized Cook distance is given by

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

where $\hat{\boldsymbol{\theta}}_{(i)}$ denotes the estimate of $\boldsymbol{\theta}$ obtained after excluding the ith observation from the sample and M is a nonnegative definite matrix, usually taken to be $M = -\ell_{\boldsymbol{\theta}\boldsymbol{\theta}}$, the observed information matrix.

Obtaining $\hat{\boldsymbol{\theta}}_{(i)}$ may be computationally cumbersome when the sample size is large. An alternative is to use the one-step approximation to $\hat{\boldsymbol{\theta}}_{(i)}$ as proposed by Xie and Wei (2007) for the log-linear Birnbaum-Saunders model. The one-step approximation is given by

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

where $\ell_{\boldsymbol{\theta}(i)} = \partial\ell_{(i)}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$, with $\ell_{(i)}$ denoting the log-likelihood function of model without the ith observation. The terms on the right hand side of the equality are evaluated at the maximum likelihood estimates. The index 1 in $\hat{\boldsymbol{\theta}}^1_{(i)}$ indicates that we are using the one-step approximation.

Following Xie and Wei (2007) and using the fact that $\ell_{\boldsymbol{\theta}}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = 0$, it can be shown that the components of $\ell_{\boldsymbol{\theta}(i)}$ are

$$\begin{aligned}
\left.\frac{\partial\ell_{(i)}(\boldsymbol{\theta})}{\partial\beta_l}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} &= x_{il}\left\{\frac{\nu^2}{2}\frac{\xi_{i2}}{\xi_{i1}} - 2\xi_{i2}\xi_{i1}\right\}\Bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \quad l = 1,\ldots,p,\\
\left.\frac{\partial\ell_{(i)}(\boldsymbol{\theta})}{\partial\alpha}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} &= \left\{\frac{1-\xi_{i2}^2}{\alpha}\right\}\Bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}},\\
\left.\frac{\partial\ell_{(i)}(\boldsymbol{\theta})}{\partial\nu}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} &= \left\{\xi_{i2}\xi_{i1}\frac{2(y_i-\mu_i)}{\nu} - \frac{1}{\nu} - \frac{\xi_{i2}}{\xi_{i1}}\frac{\nu}{2}(y_i-\mu_i)\right\}\Bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}.
\end{aligned}$$

Since $\hat{\boldsymbol{\theta}}^1_{(i)} - \hat{\boldsymbol{\theta}} = \{-\ell_{\boldsymbol{\theta}\boldsymbol{\theta}}\}^{-1}\ell_{\boldsymbol{\theta}(i)}$, using $M = \{-\ell_{\boldsymbol{\theta}\boldsymbol{\theta}}\}$ we obtain that the generalized Cook distance can be approximated by

$$GD^1_i = \ell^\top_{\boldsymbol{\theta}(i)}\{-\ell_{\boldsymbol{\theta}\boldsymbol{\theta}}\}^{-1}\ell_{\boldsymbol{\theta}(i)}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}.$$

In order to evaluate the impact of the ith observation on $\boldsymbol{\beta}$, α or ν, we approximate GD_i in each of these cases in the following manner, respectively:

$$\begin{aligned}
GD^1_i(\boldsymbol{\beta}) &= \ell^\top_{\boldsymbol{\beta}(i)}\{[-\ell_{\boldsymbol{\theta}\boldsymbol{\theta}}]^{-1}\}^{\boldsymbol{\beta}\boldsymbol{\beta}}\ell_{\boldsymbol{\beta}(i)}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}},\\
GD^1_i(\alpha) &= \ell^\top_{\alpha(i)}\{[-\ell_{\boldsymbol{\theta}\boldsymbol{\theta}}]^{-1}\}^{\alpha\alpha}\ell_{\alpha(i)}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}},\\
GD^1_i(\nu) &= \ell^\top_{\nu(i)}\{[-\ell_{\boldsymbol{\theta}\boldsymbol{\theta}}]^{-1}\}^{\nu\nu}\ell_{\nu(i)}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}},
\end{aligned}$$

where $\{\cdot\}^{\boldsymbol{\theta}_j\boldsymbol{\theta}_j}$ represents the diagonal block corresponding to $\boldsymbol{\theta}_j$ in the matrix.

3.4.5 A misspecification test

A key assumption of the GBS2RM is that the variable $\mathbf{y}$ is linearly related to the vector of regression parameters $\boldsymbol{\beta}$, which may not hold true in some applications. Other misspecifications may take place, such as the omission of an important covariate or of interactions between covariates. Therefore, it is important to test whether the functional form of a fitted GBS2RM is adequate. In short, we wish to test whether the GBS2RM is misspecified.

The effect of model misspecification on the residuals of classic linear regression models was investigated by Ramsey (1969), who proposed the RESET test (Regression Specification Error Test) that can be used to determine whether a given classic linear model is correctly specified. Given a classic linear model, the test procedure consists of augmenting the model using one or more testing variables and then testing their exclusion. Rejection of the null hypothesis indicates that the model specification is in error. The testing variables can be taken as powers of the model fitted values, as proposed by Ramsey and Gilbert (1972).

A RESET-type test for the log-linear Birnbaum-Saunders model was considered by Santos and Cribari-Neto (2015), who investigated the test size distortions in small samples and its power under different types of misspecification. The authors noted that the test is capable of detecting model misspecification, especially when the sample size is large. The same approach can be applied for the GBS2RM. The misspecification test for the GBS2RM model can be carried out as follows:

1. Estimate the parameters of the GBS2RM

$$y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \cdots + x_{ip}\beta_p + \varepsilon_i, \quad i = 1, \ldots, n,$$

   and obtain the predicted values $\hat{\boldsymbol{\mu}} = (\hat{\mu}_1, \ldots, \hat{\mu}_n)^\top$.

2. Estimate the parameters of the augmented GBS2RM, given by

$$y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \cdots + x_{ip}\beta_p + \gamma_1\hat{\mu}_i^2 + \cdots + \gamma_{k-1}\hat{\mu}_i^k + \varepsilon_i, \quad i = 1, \ldots, n,$$

   where k is an integer greater or equal to 2.

3. Test $\mathscr{H}_0 : \gamma_1 = \cdots = \gamma_{k-1} = 0$ (correct model specification) against $\mathscr{H}_1$ that $\gamma_j \neq 0$ for at least one $j \in \{1, \ldots, k-1\}$ (model misspecification).

4. If the null hypothesis is rejected, reject the model under evaluation. However, if $\mathscr{H}_0$ is not rejected, there is evidence that the model functional form is adequate.

3.5 Prediction intervals

In this section we address the issue of obtaining prediction intervals for a non-observed response value. We shall use an approach similar to that developed by Stine (1985) for linear models. Stine's proposal involves the use bootstrap resampling to estimate the prediction error distribution, which is then used to obtain the prediction intervals. In similar fashion, Davison and Hinkley (1997) provide an algorithm to compute bootstrap prediction intervals for generalized linear models, which was recently extended by Espinheira et al. (2014) for beta regression models. In fact, building upon the work of Mojirsheibani and Tibshirani (1996) on confidence intervals for parameters based on future samples, Espinheira et al. (2014) proposed a method that can be used to compute BC_a (bias-corrected and accelerated) prediction intervals, which we shall now apply to the GBS2RM.

Consider a sample $y_1, \ldots, y_n$ of the response variable and let X be the corresponding matrix of covariates. The goal lies in computing a prediction interval for a non-observed response

value y_+ based on a new observation of the covariates, denoted by $\mathbf{x}_+$. We must consider a prediction error function $\mathscr{R}(y;\mu)$, which is a monotonic function of y, has constant variance and whose ρth quantile is denoted by δ_ρ. Here, μ denotes the mean of y. The lower and upper prediction limits of a $1-\rho$ prediction interval for y_+ are, respectively, $y_{+,\rho/2}$ and $y_{+,1-\rho/2}$, such that $\mathscr{R}(y_{+,\rho/2};\mu)=\delta_{\rho/2}$ and $\mathscr{R}(y_{+,1-\rho/2};\mu)=\delta_{1-\rho/2}$. Since the distribution of $\mathscr{R}(y_+;\mu)$ is usually unknown, we make use of resampling methods to estimate it. Espinheira et al. (2014) propose a method to construct BC_a (Efron, 1987) prediction intervals by using the estimate

$$\tilde{\rho}=\Phi\left(\hat{z}_0+\frac{\hat{z}_0+z_\rho}{1-\hat{a}(\hat{z}_0+z_\rho)}\right),$$

where $\hat{z}_0$ is a bias correction constant, $\hat{a}$ is a factor known as acceleration constant and z_ρ is the ρth standard normal quantile. The following estimates for z_0 and a are used for constructing the prediction intervals for new response observations:

$$\hat{z}_0=\Phi^{-1}\left(\frac{\#\{\mathscr{R}_+^*<\mathscr{R}_m\}}{B}\right)\quad\text{and}\quad a=\frac{\sqrt{n}\,\mathbb{E}(\dot{\ell}_+^3)}{6\mathrm{Var}(\dot{\ell}_+)^{3/2}},$$

where $\mathscr{R}_m$ is the median of the prediction errors $\mathscr{R}_1,\ldots,\mathscr{R}_n$ of the fitted model, $\mathscr{R}_+^*$ denotes the bootstrap estimate of the prediction error for the non-observed response value, B is the number of bootstrap replications used and $\dot{\ell}=\partial\log f_Y(y_+|\boldsymbol{\theta})/\partial\mu$.

In the GBS2RM, $a=0$. To see that notice that $\dot{\ell}=2\xi_1\xi_2-\nu^2\xi_2/(2\xi_1)$, where $\xi_1=\nu\cosh[\nu(y-\mu)]/\alpha$ and $\xi_2=2\sinh\nu(y-\mu)]/\alpha$. Therefore, the numerator of a is proportional to

$$\mathbb{E}(\dot{\ell}^3)=-\frac{\nu^6}{8}\mathbb{E}\left(\frac{\xi_2^3}{\xi_1^3}\right)+\frac{3\nu^4}{2}\mathbb{E}\left(\frac{\xi_2^3}{\xi_1}\right)-6\nu^2\mathbb{E}\left(\xi_1\xi_2^3\right)+8\nu^2\mathbb{E}\left(\xi_1^3\xi_2^3\right).$$

Using the fact that $\xi_1=\nu\sqrt{1+(\alpha\xi_2/2)^2}/2$ and that expected values of odd functions of a standard normal random variable equal zero, it is possible to show that $\mathbb{E}(\dot{\ell}^3)=0$ and, as a consequence, $a=0$. Hence, in the GBS2RM the BC_a method reduces to the BC (bias-corrected) method proposed by Efron (1981), where the estimate of ρ turns out to be

$$\tilde{\rho}=\Phi\left(2\hat{z}_0+z_\rho\right). \tag{3.2}$$

We consider the r_{SHN} residual as the prediction error function for the GBS2RM, where $\mathscr{R}_i=r_{SHN_i}$, $i=1,\ldots,n$. Moreover, we notice that such a prediction error function is monotonic in y and has constant variance, since the reference distribution of the r_{SHN} residual is standard normal.

For a given estimate $\hat{\boldsymbol{\theta}}$, a new observation of the covariates $\mathbf{x}_+$ and the residuals $\mathbf{r}_{SHN}=(r_{SHN_1},\ldots,r_{SHN_n})^\top$ from the fitted model, the algorithm for computing prediction intervals for y_+ in the GBS2RM model can be outlined as follows:

1. Draw a random sample from $\mathbf{r}_{SHN}$ with replacement to construct $\mathbf{r}^*=(r_1^*,\ldots,r_n^*)^\top$, and then, obtain a bootstrap sample of the response variable $y_i^*=\sinh^{-1}(\hat{\alpha}r_i^*/2)/\hat{\nu}+\hat{\mu}_i$, where $\hat{\mu}_i=\mathbf{x}_i^\top\hat{\boldsymbol{\beta}}$, $i=1,\ldots,n$.

2. Fit the GBS2RM using $\mathbf{y}^* = (y_1^*, \ldots, y_n^*)^\top$ and X, obtaining $\widehat{\boldsymbol{\theta}}^* = (\widehat{\boldsymbol{\beta}}^*, \widehat{\alpha}^*, \widehat{\nu}^*)^\top$.

3. Randomly draw r^* from $\mathbf{r}_{SHN}$ and then compute $y_+^* = \sinh^{-1}(\hat{\alpha} r^*/2)/\hat{\nu} + \hat{\mu}_+$, where $\hat{\mu}_+ = \mathbf{x}_+^\top \hat{\boldsymbol{\beta}}$. Now compute the bootstrap prediction error $\mathscr{R}_+^* = 2\sinh(\widehat{\nu}^*(y_+^* - \widehat{\mu}_+^*))/\widehat{\alpha}^*$, where $\widehat{\mu}_+^* = \mathbf{x}_+^\top \widehat{\boldsymbol{\beta}}^*$.

4. Execute steps 1, 2 and 3 for B times. Considering the order statistics of the bootstrap prediction errors $\mathscr{R}_{+(1)}^* \leq \cdots \leq \mathscr{R}_{+(B)}^*$, compute the BC$_a$ quantiles $\delta_{+(\tilde{\rho}/2)}^{\mathrm{BC}_a} = \mathscr{R}_{+(\lceil B\tilde{\rho}/2 \rceil)}^*$ and $\delta_{+(1-\tilde{\rho}/2)}^{\mathrm{BC}_a} = \mathscr{R}_{+(\lceil B(1-\tilde{\rho}/2) \rceil)}^*$, where $\tilde{\rho}$ is given in Equation (3.2). Finally, the lower and upper limits of the $1-\rho$ BC$_a$ prediction interval are, respectively,

$$y_{+,\tilde{\rho}/2} = \frac{1}{\hat{\nu}} \sinh^{-1}\left(\frac{\hat{\alpha}\delta_{+(\tilde{\rho}/2)}^{\mathrm{BC}_a}}{2}\right) + \hat{\mu}_+ \quad \text{and} \quad y_{+,1-\tilde{\rho}/2} = \frac{1}{\hat{\nu}} \sinh^{-1}\left(\frac{\hat{\alpha}\delta_{+(1-\tilde{\rho}/2)}^{\mathrm{BC}_a}}{2}\right) + \hat{\mu}_+.$$

This algorithm can also be used to compute a $1-\rho$ confidence percentile prediction interval, which is obtained setting $\hat{z}_0 = 0$ in Equation (3.2) in the step 4. Prediction intervals for several unobserved response values can be obtained by using the above algorithm for each missing data value.

3.6 Model selection criteria

Model specification is of paramount importance in regression analysis. It is oftentimes based on model selection criteria, i.e., on criteria that seek to identify the best fitting model. The most commonly used model selection criteria are the Akaike information criteria (AIC), proposed by Akaike (1973), and the Schwarz information criteria (SIC), introduced by Schwarz (1978). For a detailed account of the different model selection criteria and their use in regression modeling, readers are referred to McQuarrie and Tsai (1998). Our goal in what follows is to present some model selection criteria that can be used with the GBS2RM. We follow Bayer and Cribari-Neto (2015), who considered model selection in the class of varying dispersion beta regressions.

Several model selection criteria were developed as extensions of previously existing criteria, such as the AIC. Akaike (1973) derived the AIC as an estimator of the quantity $\Delta(\theta_0, k) = \mathbb{E}_0\left[-2\log f(\mathbf{y}|\boldsymbol{\theta})\right]\big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$, a measure of the discrepancy between the true model $f(\mathbf{y}|\boldsymbol{\theta}_0)$ and a fitted candidate model $f(\mathbf{y}|\hat{\boldsymbol{\theta}})$, where $\mathbb{E}_0$ indicates that the expectation is computed with respect to the true model and k is the dimension of $\boldsymbol{\theta}$. The AIC is given by $-2\log f(\mathbf{y}|\hat{\boldsymbol{\theta}}) + 2k$, the term $-2\log f(\mathbf{y}|\hat{\boldsymbol{\theta}})$ being Akaike's estimator of $\Delta(\theta_0, k)$ and $2k$ being an asymptotic bias correction. The AIC is the most commonly used model selection criterion. Nonetheless, it may perform poorly in small samples, as pointed out by Hurvich and Tsai (1989). This occurs because the AIC becomes progressively more negatively biased as k becomes larger relative to n, which leads the AIC to often select over-specified models. As a consequence, bias-adjusted variants of the AIC have been proposed and investigated in the literature. They are typically obtained by deriving bias corrections that are superior to $2k$. The AIC$_c$ of Sugiura (1978) and Hurvich

and Tsai (1989) is given by

$$\mathrm{AIC}_c = -2\log f(\mathbf{y}|\hat{\boldsymbol{\theta}}) + 2k\frac{n}{n-k-2}.$$

Furthermore, some authors suggested using bootstrap resampling to estimate the bias of the term $-2\log f(\mathbf{y}|\hat{\boldsymbol{\theta}})$; see, e.g., Ishiguro and Sakamoto (1991), Cavanaugh and Shumway (1997) and Shibata (1997). Using the same notation as in Bayer and Cribari-Neto (2015), some of the bootstrap-based variants of the AIC can be expressed as:

$$\mathrm{EIC}_1 = -2\log f(\mathbf{y}|\hat{\boldsymbol{\theta}}) + \frac{1}{B}\sum_{b=1}^{B}\left[2\log f(\mathbf{y}^b|\hat{\boldsymbol{\theta}}^b) - 2\log f(\mathbf{y}|\hat{\boldsymbol{\theta}}^b)\right],$$

$$\mathrm{EIC}_2 = -2\log f(\mathbf{y}|\hat{\boldsymbol{\theta}}) + \frac{1}{B}\sum_{b=1}^{B}\left[4\log f(\mathbf{y}|\hat{\boldsymbol{\theta}}) - 4\log f(\mathbf{y}|\hat{\boldsymbol{\theta}}^b)\right],$$

$$\mathrm{EIC}_3 = -2\log f(\mathbf{y}|\hat{\boldsymbol{\theta}}) + \frac{1}{B}\sum_{b=1}^{B}\left[4\log f(\mathbf{y}^b|\hat{\boldsymbol{\theta}}^b) - 4\log f(\mathbf{y}^b|\hat{\boldsymbol{\theta}})\right],$$

$$\mathrm{EIC}_4 = -2\log f(\mathbf{y}|\hat{\boldsymbol{\theta}}) + \frac{1}{B}\sum_{b=1}^{B}\left[4\log f(\mathbf{y}^b|\hat{\boldsymbol{\theta}}) - 4\log f(\mathbf{y}|\hat{\boldsymbol{\theta}}^b)\right],$$

$$\mathrm{EIC}_5 = -2\log f(\mathbf{y}|\hat{\boldsymbol{\theta}}) + \frac{1}{B}\sum_{b=1}^{B}\left[4\log f(\mathbf{y}^b|\hat{\boldsymbol{\theta}}^b) - 4\log f(\mathbf{y}|\hat{\boldsymbol{\theta}})\right],$$

where B denotes the number of bootstrap replications used, $\mathbf{y}^b$ denotes the bth bootstrap response value and $\hat{\boldsymbol{\theta}}^b$ is the bth bootstrap estimate, $b = 1,\ldots,B$, obtained by maximizing the function $\log f(\mathbf{y}^b|\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$. The criteria EIC_1 and EIC_2 were proposed by Ishiguro and Sakamoto (1991) and by Cavanaugh and Shumway (1997), respectively. Shibata (1997) showed that the EIC_1 and EIC_2 are asymptotically equivalent and proposed the EIC_3, EIC_4 and EIC_5 criteria.

The five bootstrap-based information criteria listed above differ in the way they estimate the bias term, the goodness of fit factor $-2\log f(\mathbf{y}|\hat{\boldsymbol{\theta}})$ being the same for all of them. Following Pan (1999), Bayer and Cribari-Neto (2015) proposed a bootstrap model selection criterion that focuses on the goodness of fit term rather than on the bias. The authors introduced the BQCV (bootstrap quasi cross-validation) criterion, which is given by

$$\mathrm{BQCV} = \frac{1}{B}\sum_{b=1}^{B} -2\log f(\mathbf{y}|\hat{\boldsymbol{\theta}}^b),$$

where the bootstrap is carried out parametrically. As noted by Bayer and Cribari-Neto (2015), the BQCV criterion estimates the discrepancy between the true model and a candidate model directly, but can overestimate $\Delta(\theta_0, k)$. Hence, the authors proposed using the following the criterion:

$$632\mathrm{QCV} = 0.368 \times [-2\log f(\mathbf{y}|\hat{\boldsymbol{\theta}})] + 0.632 \times \mathrm{BQCV}$$

which follows from the results in Pan (1999) and from the fact that $-2\log f(\mathbf{y}|\hat{\boldsymbol{\theta}})$ and BQCV underestimate and overestimate $\Delta(\theta_0,k)$, respectively. Thus, 632QCV aims at balancing these two terms.

Several other (not AIC-based) criteria were proposed in the literature, such as the aforementioned SIC, the SIC_c derived by McQuarrie (1999), the HQ proposed by Hannan and Quinn (1979) and its corrected version presented by McQuarrie and Tsai (1998) for the normal linear model. Such criteria are given by:

$$\begin{aligned}
\text{SIC} &= -2\log f(\mathbf{y}|\hat{\boldsymbol{\theta}}) + k\log(n),\\
\text{SIC}_c &= -2\log f(\mathbf{y}|\hat{\boldsymbol{\theta}}) + k\log(n)n/(n-k-2),\\
\text{HQ} &= -2\log f(\mathbf{y}|\hat{\boldsymbol{\theta}}) + 2k\log(\log(n)),\\
\text{HQ}_c &= -2\log f(\mathbf{y}|\hat{\boldsymbol{\theta}}) + 2k\log(\log(n))n/(n-k-2).
\end{aligned}$$

The finite sample performances of the criteria listed above were investigated in the literature for a wide variety of models, such as regression, time series and nonparametric models; for details, see McQuarrie and Tsai (1998). In the next section we shall numerically evaluate how well model selection schemes based on model selection criteria work in the log-linear $\mathscr{GBS}_2$ model.

3.7 Numerical evaluations

In this section we shall present results from Monte Carlo simulation studies that were conducted to numerically evaluate the finite performances of maximum likelihood estimators of the parameters that index the GBS2RM model and to evaluate the quality of the approximation of the r_{SHN} and r_{CSG} residuals distributions by their respective reference distributions. Additionally, we shall numerically evaluate the finite sample performance of the RESET-type test for misspecification of the GBS2RM model, prediction intervals for non-observed response values and model selection schemes. The simulations were performed using the Ox matrix programming language (Doornik, 2009). Log-likelihood maximizations were carried out using the BFGS quasi-Newton nonlinear optimization algorithm with analytical first derivatives.

3.7.1 Maximum likelihood estimation

In our simulation study, the parameter vector is $\boldsymbol{\theta} = (\beta_1,\beta_2,\alpha,\nu)^\top$ and the regression model is

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i, \quad i = 1,\ldots,n, \tag{3.3}$$

where $\varepsilon_i \overset{\text{iid}}{\sim} \mathscr{SHN}(\alpha,0,\nu^{-1})$. The parameter values are $\beta_1 = 1$, $\beta_2 = 0.5$, $\alpha = 1$ and $\nu = 1$ and the starting points for the log-likelihood maximization were obtained by using the proposal made in Section 3.3.

The covariates values were randomly generated from three different distributions: uniform $\mathscr{U}(0,1)$, exponential with unit mean, and standard normal. We shall denote such data generation schemes by $E1$, $E2$ and $E3$, respectively. The values of all covariates were kept constant during the simulations.

The results we report are based on $10,000$ Monte Carlo replications, the sample sizes being $n \in \{30,60,90\}$. We computed the mean squared error (MSE) of the MLE of each parameter and also the corresponding relative biases: $[\mathbb{E}(\hat{\boldsymbol{\theta}}_j) - \boldsymbol{\theta}_j]/\boldsymbol{\theta}_j$, $j = 1,\ldots,4$. The results for the different covariates generation schemes are presented in Table 3.1. We note that the MSEs tend to decrease as larger sample sizes are used, as expected. The estimates of the regression parameters β_1 and β_2 are more accurate than those of α and ν. All estimators perform well when $n = 60$ and $n = 90$. The results for the three covariates generation schemes are similar.

Table 3.1 Relative bias (RB) and mean squared error (MSE), Model (3.3)

Generation	n	Statistic	β_1	β_2	α	ν
$E1$	30	RB	0.0003	−0.0052	1.5198	0.9057
		MSE	0.1711	0.3156	2.3377	1.2659
	60	RB	0.0019	−0.0077	0.6301	0.4362
		MSE	0.1093	0.2078	1.0530	0.7104
	90	RB	0.0017	−0.0061	0.3922	0.2829
		MSE	0.0934	0.1650	0.7314	0.5285
$E2$	30	RB	−0.0016	0.0020	1.4943	0.8875
		MSE	0.1354	0.0999	2.3341	1.2546
	60	RB	−0.0006	0.0015	0.6219	0.4318
		MSE	0.0900	0.0567	1.0415	0.7050
	90	RB	−0.0003	0.0014	0.3859	0.2789
		MSE	0.0705	0.0498	0.7232	0.5242
$E3$	30	RB	−0.0004	0.0004	1.5218	0.8988
		MSE	0.0875	0.0926	2.3787	1.2661
	60	RB	0.0002	−0.0009	0.6344	0.4387
		MSE	0.0606	0.0586	1.0576	0.7101
	90	RB	0.0002	−0.0002	0.3919	0.2836
		MSE	0.0484	0.0458	0.7236	0.5234

3.7.2 Empirical distribution of the residuals

The second simulation study was performed to evaluate how well the distributions of the residuals r_{SHN} and r_{CSG} are approximated by the corresponding reference distributions. The results are for Model (3.3) and the sample size is $n = 60$. The covariates values are obtained as in the previous simulation. Based on all $10,000$ replications, we computed the means, standard deviations, asymmetries and kurtosis of the two residuals. For the r_{SHN} residual, whose distribution is expected to be approximately standard normal, we expect such statistics to be close to 0, 1, 0 and 3, respectively. For the r_{CSG} residual, whose reference distribution is exponential with unit mean, such statistics are expected to be close to 1, 1, 2 and 6, respectively. The results

are presented in Table 3.2. They show that the distribution of r_{SHN} is better approximated by its reference distribution than that of r_{CSG}. Nonetheless, the results for the residual r_{CSG} were also satisfactory, the means and standard deviations being quite close to one. It is then possible to conclude that the distributions of the proposed residuals are well approximated by the respective reference distributions. Practitioners can then compare the quantiles of residuals obtained from a fitted GBS2RM with those of the corresponding reference distributions.

Table 3.2 Means, standard deviations (SD), asymmetries and kurtosis of the residuals r_{SHN} and r_{CSG}, Model (3.3)

Generation	Residual	Mean	SD	Asymmetry	Kurtosis
$E1$	r_{SHN}	-2.11×10^{-5}	0.9989	0.0003	2.9700
	r_{CSG}	1.0000	0.9928	1.7866	6.7611
$E2$	r_{SHN}	-8.85×10^{-5}	0.9989	0.0007	2.9720
	r_{CSG}	0.9999	0.9931	1.7905	6.7936
$E3$	r_{SHN}	0.0003	0.9988	0.0001	2.9712
	r_{CSG}	1.0003	0.9929	1.7871	6.7705

3.7.3 RESET-type misspecification test

Next, we performed a set of Monte Carlo simulations to evaluate the finite sample performance of the RESET-type misspecification test. Since the previous results were similar for the different schemes of covariates values generation, we shall only report results obtained using standard uniform draws. The number of Monte Carlo replications is $10,000$ and the sample sizes are $n \in \{30,60,90\}$. The testing variable is the vector of squared predicted values and the test was performed using the likelihood ratio test criterion. The significance levels are 10%, 5% and 1%.

The first simulation study was carried out to compute the RESET-type test null rejection rates. In each Monte Carlo replication we generated $y_i = 1 - 0.5x_{2i} + 1.3x_{3i} + \varepsilon_i$, where $\varepsilon_i \overset{\text{iid}}{\sim} \mathcal{SHN}(1,0,1)$, $i = 1,\ldots,n$, fitted the GBS2RM and obtained the predicted values. We then fitted the model $y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \gamma\hat{\mu}_i^2 + \varepsilon_i$ and tested the null hypothesis $\mathcal{H}_0 : \gamma = 0$ against a two-sided alternative hypothesis. The test null rejection rates are presented in Table 3.3. We note that the test is considerably size-distorted when the sample size is small ($n = 30$); such distortions become much smaller when the sample size increases to $n = 60$ and $n = 90$. With 90 observations, the test null rejection rate at the 5% nominal level is 6.9%.

Table 3.3 Null rejection rates of the RESET-type test for the GBS2RM

n	Significance level		
	10%	5%	1%
30	0.2107	0.1367	0.0437
60	0.1388	0.0808	0.0218
90	0.1268	0.0690	0.0170

Other simulations were performed to evaluate the power of the RESET-type test under different sources of model misspecification. We consider four different data generation processes (schemes), denoted by $P1$, $P2$, $P3$ and $P4$. In scheme $P1$, $y_i = 1 + 0.5x_{2i} + 1.8x_{3i} + 1.8x_{2i} \times x_{3i} + \varepsilon_i$; notice the interaction between the two covariates. In scheme $P2$, $y_i = 1 + 0.5x_{2i} + 1.8x_{3i} + 1.8x_{i2}^2 + \varepsilon_i$; notice that the linear predictor includes a squared regressor. In scheme $P3$, $y_i = 1 + 0.5x_{2i} + 1.8x_{3i} + 1.5x_{4i} + \varepsilon_i$; notice the covariate $\mathbf{x}_4$. In scheme $P4$, the data generating process is nonlinear: $y_i = (1 + 0.5x_{2i} + 1.8x_{3i})^{\varphi} + \varepsilon_i$, $\varphi \in \mathbb{R}$. In the simulations, we used $\varphi = 1.7$. In all four schemes, $\varepsilon_i \overset{\text{iid}}{\sim} \mathcal{SHN}(1,0,1)$, $i = 1,\ldots,n$, and the following (misspecified) model was estimated: $y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i$. Since the test is liberal, testing inference here was based on exact critical values which were estimated in the previous simulations (size simulations). We thus compute the power of a size-corrected test. The test nonnull rejection rates are presented in Table 3.4. We note that the RESET-type test for the GBS2RM displays good power, with nonnull rejection rates increasing when larger sample sizes are considered. The test was more powerful under scheme $P4$, where linearity is incorrectly assumed.

Table 3.4 Nonnull rejection rates of the RESET-type test for the GBS2RM model under different schemes of model misspecification

Scheme	n	Significance level		
		10%	5%	1%
$P1$	30	0.2456	0.1573	0.0496
	60	0.4778	0.3405	0.1407
	90	0.6471	0.5183	0.2856
$P2$	30	0.1845	0.1103	0.0316
	60	0.3226	0.2083	0.0718
	90	0.4428	0.3234	0.1522
$P3$	30	0.2090	0.1243	0.0354
	60	0.5286	0.3635	0.1276
	90	0.7724	0.6314	0.3428
$P4$	30	0.3601	0.2470	0.0959
	60	0.6571	0.5212	0.2631
	90	0.8352	0.7320	0.5104

3.7.4 Prediction intervals

A simulation study was performed to evaluate the performances of the percentile and BC_a prediction intervals for a non-observed occurrence y_+ of the response variable given a new observation of the covariate in a GBS2RM model. The data generating process is $y_i = 1 + 0.5x_{2i} + \varepsilon_i$, where $\varepsilon_i \overset{\text{iid}}{\sim} \mathcal{SHN}(\alpha,0,1/\nu)$, $i = 1,\ldots,n$. The results are presented in Table 3.5. The sample sizes considered were as in the previous simulation studies, the number of Monte Carlo replications used was $5,000$ and the covariate values were obtained as random draws from $\mathcal{U}(0,1)$. We used $B = 1,000$ bootstrap replications for constructing 95% confidence prediction intervals. In Table 3.5 we present the empirical coverages of the percentile and BC_a prediction intervals and the proportions of the replications for which y_+ was smaller (larger) than the

lower (upper) prediction intervals. The figures in Table 3.5 show that both prediction intervals display empirical coverages close to 95% and that the non-coverages are well distributed in both sides. The percentile intervals perform slightly better than the BC_a intervals. Additionally, the performances of both intervals improve as the sample size increases, as expected. It is thus noteworthy that bootstrap prediction intervals perform well when used with the GBS2RM, especially the percentile method.

Table 3.5 Empirical coverages and left and right non-coverages of 95% prediction intervals in a GBS2RM for different values of α and ν

n	Percentile			BC_a		
	Left	Coverage	Right	Left	Coverage	Right
			$\boldsymbol{\theta} = (1, 0.5, 1, 1)^\top$			
30	0.027	0.942	0.031	0.029	0.939	0.032
60	0.026	0.946	0.028	0.025	0.945	0.030
90	0.026	0.947	0.027	0.027	0.945	0.028
			$\boldsymbol{\theta} = (1, 0.5, 0.5, 1.5)^\top$			
30	0.028	0.941	0.031	0.029	0.937	0.034
60	0.025	0.950	0.025	0.025	0.948	0.027
90	0.024	0.950	0.026	0.025	0.950	0.025
			$\boldsymbol{\theta} = (1, 0.5, 1.5, 0.5)^\top$			
30	0.029	0.945	0.026	0.029	0.941	0.030
60	0.026	0.949	0.025	0.026	0.949	0.025
90	0.028	0.949	0.023	0.028	0.949	0.023

3.7.5 Model selection criteria

We performed a simulation study to evaluate the finite sample performances of the model selection criteria presented in Section 3.6. The simulation is similar to that conducted by Hurvich and Tsai (1989) for linear models. The true model is $y_i = -1 + x_{2i} + x_{3i} + x_{4i} + \varepsilon_i$, where $\varepsilon_i \overset{\text{iid}}{\sim} \mathcal{SHN}(1,0,1)$, $i = 1,\ldots,n$, and the true vector parameter $\boldsymbol{\theta}_0$ has dimension $k_0 = 6$. We consider the $n \times 6$ matrix of candidate regressors $\tilde{X}$, whose first column is a vector of ones. The first four columns of such a matrix contain the regressors that are present in the true data generating process. The models were fitted in nested fashion using as covariates the first $2,3,\ldots,6$ columns of $\tilde{X}$. The number of Monte Carlo and bootstrap replications are as before, as well as the sample sizes considered. All covariate values in $\tilde{X}$ were obtained as random standard uniform draws. The bootstrap extensions of the AIC criteria as well as BQCV and 632QCV were computed using parametric resampling. Nonparametric bootstrap resampling leads to very similar results which are not presented for brevity. We computed the proportion of model under-specification, over-specification and correct specification for each criterion. The results are presented in Table 3.6. They show that AIC tends to select over-specified models, especially for small sample sizes. The bootstrap extensions of the AIC performed well when $n = 60$ and $n = 90$. For instance, for $n = 60$ the largest frequency of correct model selection was achieved by the EIC_1. The performances of the EIC_2 and EIC_4 were also very good for $n = 90$, with

proportions of correct specification in excess of 97%. Overall and on balance, the SIC_c was the best performer. It outperformed all competing criteria when $n = 30$ and was very competitive with the bootstrap criteria when $n = 60$ and $n = 90$. Additionally, SIC_c does not entail the computational burden of performing data resampling. We thus recommend that model selection be based on such a criterion when performing GBS2RM modeling.

Table 3.6 Proportions of model under-specification ($k < k_0$), correct specification ($k = k_0$) and over-specification ($k > k_0$) of a GBS2RM using the selection criteria discussed in Section 3.6

	$n = 30$			$n = 60$			$n = 90$		
	$< k_0$	$= k_0$	$> k_0$	$< k_0$	$= k_0$	$> k_0$	$< k_0$	$= k_0$	$> k_0$
AIC	0.018	0.454	0.527	0.000	0.676	0.324	0.000	0.723	0.277
AIC_c	0.052	0.658	0.290	0.000	0.762	0.238	0.000	0.775	0.225
SIC	0.052	0.625	0.323	0.001	0.887	0.112	0.000	0.929	0.071
SIC_c	0.151	0.741	0.107	0.004	0.936	0.060	0.000	0.960	0.039
HQ	0.031	0.517	0.451	0.001	0.790	0.209	0.000	0.844	0.156
HQ_c	0.083	0.705	0.212	0.001	0.866	0.133	0.000	0.890	0.110
EIC_1	0.909	0.088	0.003	0.021	0.945	0.034	0.000	0.941	0.059
EIC_2	0.953	0.045	0.002	0.099	0.887	0.014	0.000	0.978	0.022
EIC_3	0.059	0.576	0.365	0.001	0.781	0.219	0.000	0.790	0.210
EIC_4	0.954	0.044	0.002	0.099	0.888	0.012	0.000	0.976	0.024
EIC_5	0.059	0.607	0.334	0.001	0.778	0.221	0.000	0.766	0.234
BQCV	0.898	0.098	0.004	0.013	0.930	0.057	0.000	0.886	0.114
BQCV632	0.830	0.161	0.009	0.002	0.890	0.107	0.000	0.776	0.224

3.8 Empirical applications

We shall now use the GBS2RM to analyze real (not simulated) data. The data contain information on patients who suffered from an acute type of leukemia and are provided by Feigl and Zelen (1965). The variable of interest is the patient's lifetime (in weeks) since the disease diagnosis. One of the covariates is the patient's white blood cells count upon leukemia diagnosis. The other covariate marks the presence or absence of a specific factor in the patient white cells, which can be classified as AG positive or AG negative, respectively.

Let $y_i = \log(t_i)$ be the logarithm of the ith patient lifetime, $x_{i2} = \log(W_i)$ be the logarithm of the patient's white blood cells count and x_{i3} be a dummy variable that equals one if the ith patient is AG positive and zero otherwise, $i = 1, \ldots, 33$. Figure 3.2 contains boxplots of $\mathbf{y}$ and $\mathbf{x}_2$ for different levels of $\mathbf{x}_3$ and a scatterplot of $\mathbf{y}$ vs. $\mathbf{x}_2$. We note from Figure 3.2a that lifetime tends to decrease as the white cells count increases and that AG negative patients ($x_3 = 0$) seem to display lower lifetimes than AG positive patients ($x_3 = 1$), which can also be observed in Figure 3.2b. Figure 3.2c shows that there is no clear relationship between $\mathbf{x}_2$ and $\mathbf{x}_3$.

We estimated the following log-linear model $\mathscr{GBS}_2$:

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i, \tag{3.4}$$

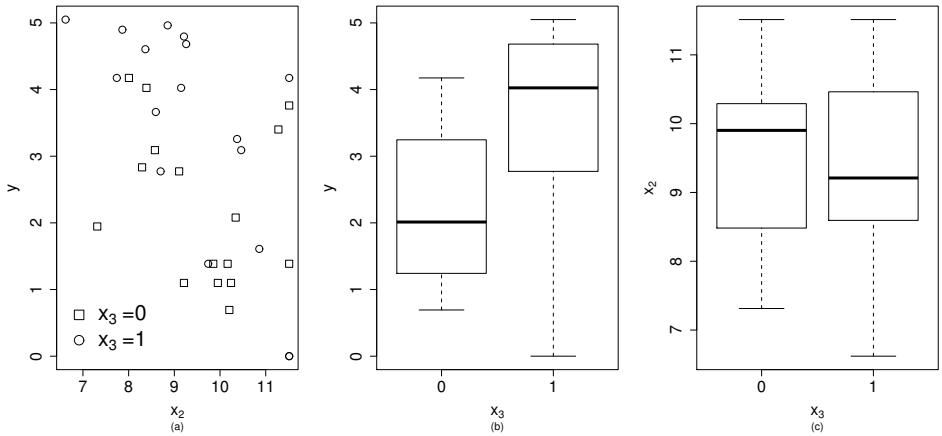

Figure 3.2 Scatterplot of the response variable vs. $\mathbf{x}_2$ (a) and boxplots for different levels of $\mathbf{x}_3$ (b and c).

where $\varepsilon_i \overset{\text{iid}}{\sim} \mathcal{SHN}(\alpha, 0, \nu^{-1})$. The maximum likelihood estimates of the parameters (standard errors in parentheses) are $\hat{\beta}_1 = 6.159\ (0.8280)$, $\hat{\beta}_2 = -0.360\ (0.0828)$, $\hat{\beta}_3 = 0.055\ (0.2786)$, $\hat{\alpha} = 6.914\ (3.8980)$ and $\hat{\nu} = 1.272\ (0.2794)$. It is noteworthy that the standard error of β_3 is large, which might be an indicative that the variable $\mathbf{x}_3$ is not relevant for the analysis at hand. Indeed, for the the model in Equation (3.4) $\text{SIC}_c = 120.97$ whereas when the model is fitted without the covariate $\mathbf{x}_3$ $\text{SIC}_c = 115.92$. Simultaneous removal of $\mathbf{x}_2$ and of the intercept yielded a larger value of the model selection criterion. The other selection criteria also led to the same conclusions. We shall address the significance of the factor AG later in this section. The likelihood ratio test statistic for testing $\mathscr{H}_0 : \nu = 0.5$ against $\mathscr{H}_1 : \nu \neq 0.5$ equals 4.65, the corresponding p-value being 0.0309. That is, there is evidence (at the 5% significance level) that the GBS2RM is superior to the BSRM for analyzing these data.

The value of Nagelkerke's pseudo-R^2 for the GBS2RM model was $R^2_N = 0.3501$. Figure 3.3 contains residual (r_{SHN} and r_{CSG}) plots. In Figures 3.3a and 3.3b we see plots of residuals against predicted values $\hat{\boldsymbol{\mu}}$ with 95% confidence bands. Notice that most residuals lie inside the intervals and that there is no noticeable pattern in the residuals. Visual inspection of Figures 3.3c and 3.3d reveal that all residuals are inside the confidence regions, thus indicating that the distributional assumptions hold. We also performed the RESET-type test for model misspecification using the square of the predicted values ($\hat{\boldsymbol{\mu}}$) as testing variable. The test statistic equals 0.73, the corresponding p-value being 0.3916. That is, there is no evidence of model misspecification at the usual significance levels.

Local influence analysis for the GBS2RM model should be able to identify data points that might be largely influencing the parameters estimates. Figure 3.4 contains local influence plots relative to the regression and shape parameters using the three perturbation schemes discussed earlier. It is noteworthy that the most influential data points are observations 14, 15 and 17.

We computed the generalized Cook distance for each observation. The *GD* measures for the model parameters are presented in Figure 3.5. Notice that such results agree with those obtained using local influence analysis, i.e., observations 14, 15 and 17 are singled out as atypical.

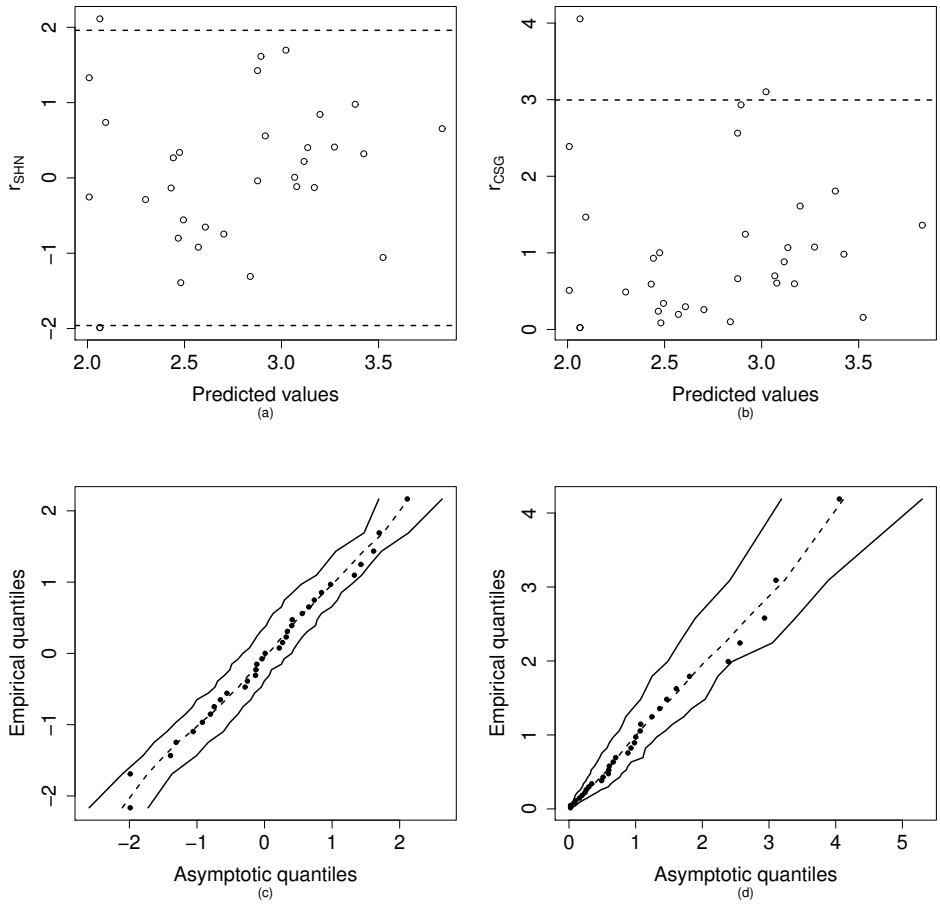

Figure 3.3 Predicted values $\hat{\boldsymbol{\mu}}$ against the residuals r_{SHN} (a) and r_{CSG} (b), and simulated envelopes with bands of 95% of confidence for the residuals r_{SHN} (c) and r_{CSG} (d). The dashed lines in the panels (a) and (b) indicate approximate confidence regions (95% confidence).

Generalized leverage measures are presented in Figure 3.6. Observations 2 and 21 stand out. The former corresponds to the patient with the lowest white cells count among all AG positive patients, whereas the latter corresponds to the patient with the lowest white cells count among all AG negative patients.

We sequentially removed each atypical observation from the data and fitted the model after each data point removal. In each case, we computed the absolute relative change in the estimates, i.e., we computed $|(\hat{\boldsymbol{\theta}}_{j(i)} - \hat{\boldsymbol{\theta}}_j)/\hat{\boldsymbol{\theta}}_j|$, where $\hat{\boldsymbol{\theta}}_j$ represents the jth parameter estimate obtained using the complete data and $\hat{\boldsymbol{\theta}}_{j(i)}$ represents the corresponding estimate obtained after the ith observation removal. Additionally, we tested the significance of each regressor when the reduced data was used. We also tested whether ν was significantly different from 0.5, i.e., whether the testing inference still suggested that the GBS2RM was superior to the log-linear Birnbaum-Saunders model. The main goal is to determine whether any relevant inferential decision was reversed after the atypical observations were removed from the data. The relative changes in the parameter estimates and in the tests $p-$values are presented in Table 3.7.

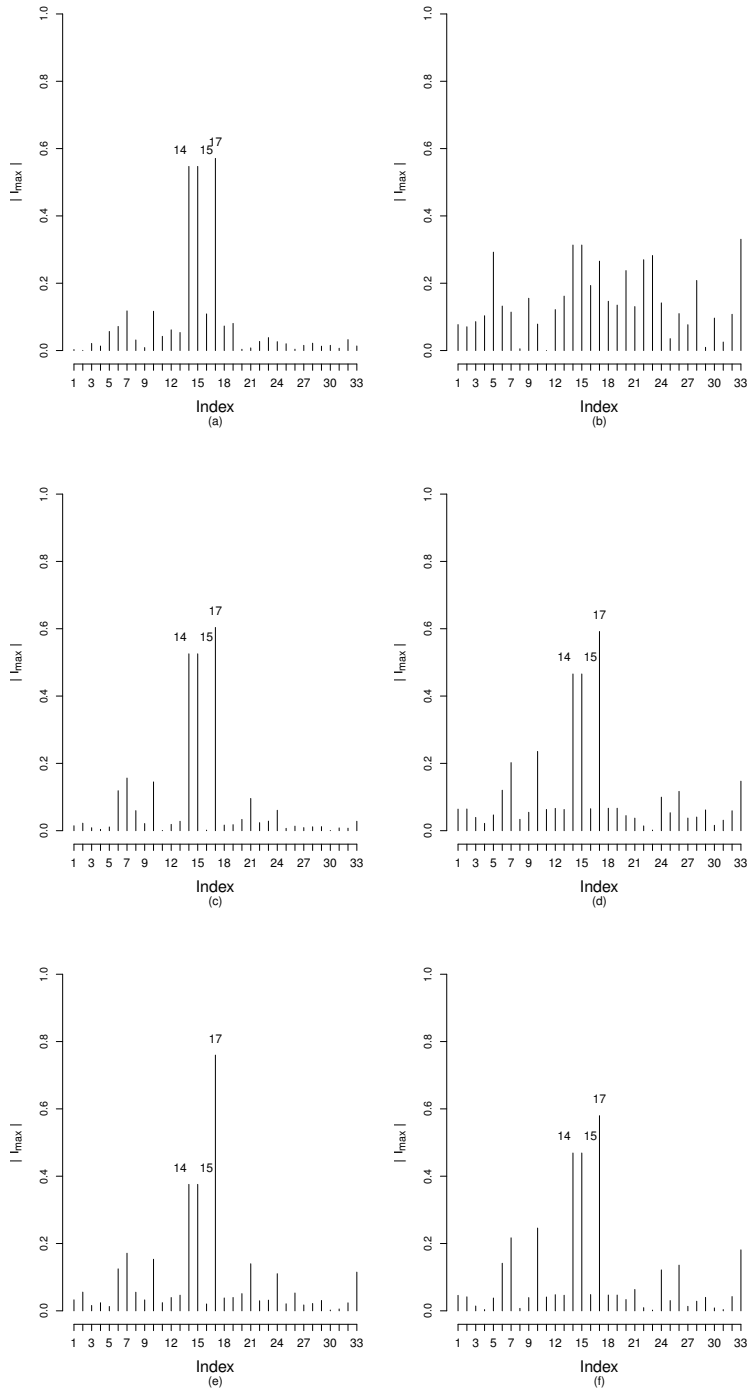

Figure 3.4 Local influence measures for the regression parameters (first column) and for the shape parameters α and ν (second column). The perturbation schemes considered were: case-weights perturbation, in panels (a) and (b); response variable perturbation, in panels (c) and (d); perturbation on the covariate $\mathbf{x}_2$, panels (e) and (f).

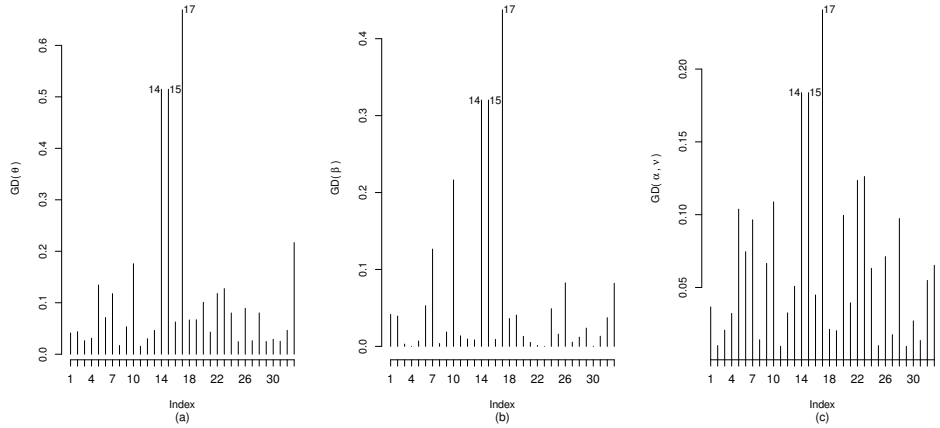

Figure 3.5 Generalized Cook distance for each observation of the data for the vector parameter $\boldsymbol{\theta}$ (a), the vector $\boldsymbol{\beta}$ (b) and the shape parameters α and ν (c).

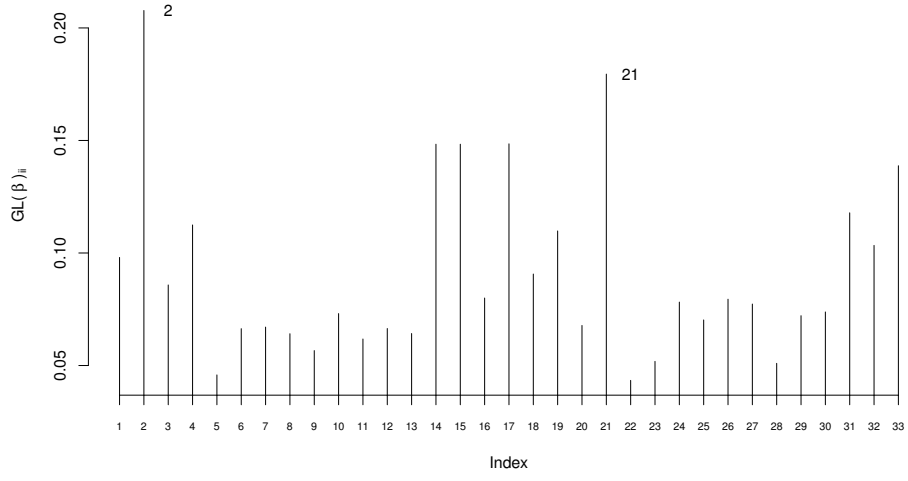

Figure 3.6 Generalized leverage for $\boldsymbol{\beta}$.

The figures in Table 3.7 show that the intercept is statistically different from zero and the regressor that accounts for white cells count is statistically significant in all scenarios at the usual significance levels. We also note that the $\mathscr{GBS}_2$-based model remains superior to that based on the $\mathscr{BS}$ law at the 10% significance level, thus strengthening the evidence in favor of the GBS2RM. The most intriguing result relates to $\mathbf{x}_3$. Such a regressor is not statistically significant at the usual significance levels when all observations are used in the model fit, as noted earlier. However, after observation 14 or observation 15 is removed from the data the estimate of β_2 changes considerably. Such observations were detected as influential data points and correspond to AG positive patients with short lifetimes but that present high counts of white blood cells, contrary to what is implied by the fitted model. When the model is fitted without these two observations in the data the covariate $\mathbf{x}_3$ becomes statistically significant. Additionally, the SIC_c for the model without $\mathbf{x}_3$ becomes 104.68 whereas the model fitted with such a regressor yields $\mathrm{SIC}_c = 101.01$, i.e., the SIC_c also provides indication that the AG factor impacts the patients' lifetime. Most of the other selection criteria also lead to the same conclusion. Simultaneous removal of cases 14, 15 and 17 leads to the same testing inference, i.e., $\mathbf{x}_3$ is found to be statistically significant. Hence, observations 14 and 15 are the cases responsible for the reversal in the inference decision regarding the statistical significance of $\mathbf{x}_3$.

Table 3.7 Absolute relative changes in the parameter estimates of Model (3.4) with tests p-values in parentheses after removal of the indicated data point(s), where the null hypotheses are $\mathscr{H}_0 : \beta_j = 0$, $j = 1,2,3$, and $\mathscr{H}_0 : \nu = 0.5$

Deleted	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\alpha}$	$\hat{\nu}$
None	— (< 0.001)	— (< 0.001)	— (0.8428)	—	— (0.0309)
obs. 2	0.025 (< 0.001)	0.044 (< 0.001)	0.417 (0.7808)	0.053	0.023 (0.0316)
obs. 14	0.041 (< 0.001)	0.058 (0.0006)	2.503 (0.4781)	0.119	0.018 (0.0554)
obs. 15	0.041 (< 0.001)	0.058 (< 0.001)	2.503 (0.4781)	0.119	0.018 (0.0554)
obs. 17	0.100 (< 0.001)	0.178 (< 0.001)	0.842 (0.9695)	0.035	0.056 (0.0245)
obs. 21	0.022 (< 0.001)	0.037 (< 0.001)	0.187 (0.8202)	0.117	0.048 (0.0819)
obs. 14 and 15	0.315 (< 0.001)	0.504 (0.0175)	10.669 (0.0029)	0.610	0.419 (0.0014)
obs. 14, 15 and 17	0.304 (< 0.001)	0.487 (0.0288)	10.587 (0.0067)	0.444	0.367 (0.0033)

We decided to consider the model fitted without the atypical cases (observations 14 and 15) in the data. The parameter estimates (standard errors in parentheses) are $\hat{\beta}_1 = 4.219$ (0.7047), $\hat{\beta}_2 = -0.179$ (0.0691), $\hat{\beta}_3 = 0.643$ (0.1701), $\hat{\alpha} = 11.135$ (6.2724) and $\hat{\nu} = 1.807$ (0.3499). The pseudo-R^2 value is $R^2_N = 0.4176$, i.e., the fit seems to be superior to that obtained using the complete data. The correct model specification is not rejected by the RESET-type test at the usual nominal levels. In Figure 3.7 are presented prediction intervals obtained with this model for values of x_2 ranging between 6 and 12, for each level of x_3. We note that patients for whom the presence of the AG factor was detected tend to live longer than AG negative patients, thus corroborating the evidence in Figure 3.2. Moreover, we note that higher white blood cells counts in the patients are significantly associated with lower lifetimes, as can be observed in Figure 3.2. Furthermore, nearly all points in Figure 3.7 lie inside the respective prediction intervals, but the observations 14 and 15 are clearly atypical data points.

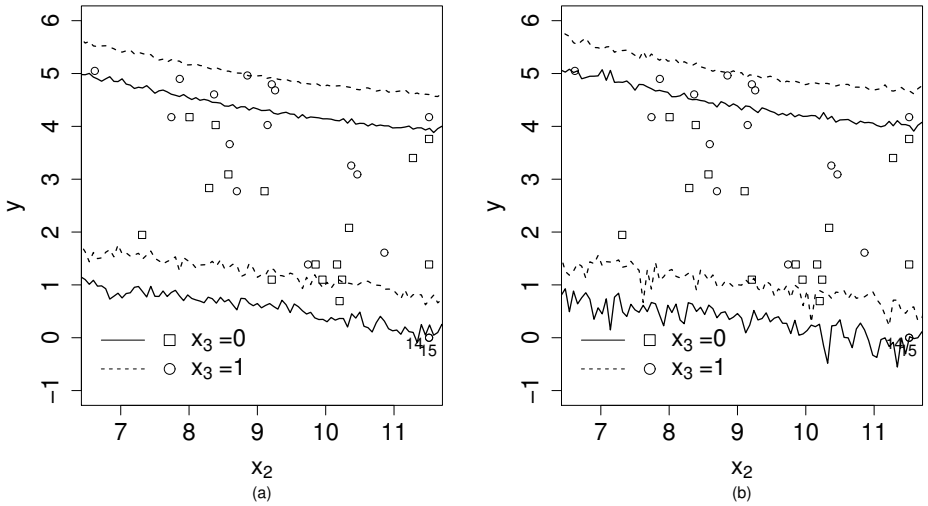

Figure 3.7 95% percentile (a) and BC$_a$ (b) prediction intervals for **y**. Solid lines indicate the intervals for AG negative patients ($x_3 = 0$) and dashed lines indicate intervals for AG positive patients ($x_3 = 1$).

3.9 Conclusion

Log-linear Birnbaum-Saunders regression models have been frequently used in the literature. The model is based on the standard Birnbaum-Saunders distribution. In this chapter, we propose a log-linear model based on a bimodal version of the Birnbaum-Saunders law. The log-linear Birnbaum-Saunders regression model is a particular case of our model. Parameter estimation is carried out by maximum likelihood. We provide an expression for the observed information matrix, discuss hypothesis testing inference and explain how a pseudo-R^2 can be easily computed. Several different diagnostic tools for the proposed model were discussed and two different residuals were introduced. We explained how to perform local influence analyses under three different perturbation schemes (case-weights perturbation, response variable perturbation and explanatory variable perturbation), derived generalized leverage measures and the generalized Cook distance, and outlined a model misspecification test. We also provide an algorithm that can be used to construct prediction intervals for out of sample responses. In addition, we investigated the finite sample performances of different model selection criteria for the proposed model. Simulation results and a empirical application were presented and discussed.

Concluding Remarks

In this thesis we investigated two bimodal versions of the Birnbaum-Saunders distribution that have been recently discussed in the literature, consisting in a new branch of the Birnbaum-Saunders modeling topic. In Chapter 2 we discussed some numerical difficulties that might arise when performing maximum likelihood estimation on the parameters of a bimodal Birnbaum-Saunders distribution and how this problem can be handled by using a penalized log-likelihood function. We also evaluated how different hypothesis tests can be performed using the estimates obtained from the solution proposed and verified that they tend to perform well for finite samples. In Chapter 3 we conducted an analysis on another bimodal Birnbaum-Saunders law, which we used to propose a new log-linear regression model. Along with this regression model, we also provided different tools to perform diagnostic analysis, model selection and to construct prediction intervals.

In future work, we shall address the following issues:

- Consider the use of a log-linear $\mathscr{GBS}2$ regression model with censored data.
- Develop a nonlinear GBS2RM.
- Propose an autoregressive bimodal Birnbaum-Saunders model for time series data.

References

Akaike, H. Information theory and an extension of the maximum likelihood principle. In Petrov, B. N. and Csáki, F., editors, *Proceedings of the second international symposium on information theory*, pages 267–281, Budapest, 1973.

Atkinson, A. C. *Plots, Transformations, and Regression*. Clarendon Press, Oxford, 1985.

Azzalini, A. and Arellano-Valle, R. B. Maximum penalized likelihood estimation for skew-normal and skew-t distributions. *Journal of Statistical Planning and Inference*, 143(2):419–433, 2013.

Balakrishnan, N., Leiva, V., Sanhueza, A., and Vilca, F. Estimation in the Birnbaum–Saunders distribution based on scale-mixture of normals and the EM-algorithm. *Sort*, 33(2):171–192, 2009.

Balakrishnan, N., Gupta, R. C., Kundu, D., Leiva, V., and Sanhueza, A. On some mixture models based on the Birnbaum–Saunders distribution and associated inference. *Journal of Statistical Planning and Inference*, 141(7):2175–2190, 2011.

Barndorff-Nielsen, O. E. Inference on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika*, 73(2):307–322, 1986.

Barndorff-Nielsen, O. E. Modified signed log likelihood ratio. *Biometrika*, 78(3):557–563, 1991.

Barros, M., Paula, G. A., and Leiva, V. A new class of survival regression models with heavy-tailed errors: robustness and diagnostics. *Lifetime Data Analysis*, 14(3):316–332, 2008.

Bayer, F. M. and Cribari-Neto, F. Bootstrap-based model selection criteria for beta regressions. *TEST*, 24(4):776–795, 2015.

Birnbaum, Z. W. and Saunders, S. C. A new family of life distributions. *Journal of Applied Probability*, 6:319–327, 1969a.

Birnbaum, Z. W. and Saunders, S. C. Estimation for a family of life distributions with applications to fatigue. *Journal of Applied Probability*, 6:328–347, 1969b.

Bourguignon, M., Silva, R. B., and Cordeiro, G. M. A new class of fatigue life distributions. *Journal of Statistical Computation and Simulation*, 84(12):2619–2635, 2014.

Cavanaugh, J. E. and Shumway, R. H. A bootstrap variant of AIC for state-space model selection. *Statistica Sinica*, 7:473–496, 1997.

Cook, R. D. Assessment of local influence. *Journal of the Royal Statistical Society B*, 48(2): 133–169, 1986.

Cook, R. D. and Weisberg, S. *Residuals and Influence in Regression*. Chapman & Hall, New York, 1982.

Cordeiro, G. M. and Cribari-Neto, F. *An Introduction to Bartlett Correction and Bias Reduction*. Springer, New York, 2014.

Cordeiro, G. M. and Lemonte, A. J. The exponentiated generalized Birnbaum–Saunders distribution. *Applied Mathematics and Computation*, 247:762–779, 2014.

Cordeiro, G. M., Lemonte, A. J., and Ortega, E. M. An extended fatigue life distribution. *Statistics*, 47(3):626–653, 2013.

Cox, D. R. Tests of separate families of hypotheses. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 105–123, London, 1961. Birckbeck College.

Cox, D. R. Further results on tests of separate families of hypotheses. *Journal of the Royal Statistical Society B*, 24:406–424, 1962.

Cribari-Neto, F., Frery, A. C., and Silva, M. F. Improved estimation of clutter properties in speckled imagery. *Computational Statistics & Data Analysis*, 40(4):801–824, 2002.

Cysneiros, A. H., Cribari-Neto, F., and Araújo, C. A. On Birnbaum-Saunders inference. *Computational Statistics & Data Analysis*, 52(11):4939–4950, 2008.

Davison, A. C. and Hinkley, D. V. *Bootstrap Methods and Their Application*. Cambridge University Press, New York, 1997.

Díaz-García, J. A. and Domınguez-Molina, J. R. Some generalisations of Birnbaum-Saunders and sinh-normal distributions. *International Mathematical Forum*, 1(35):1709–1727, 2006.

Díaz-García, J. A. and Leiva, V. A new family of life distributions based on the elliptically contoured distributions. *Journal of Statistical Planning and Inference*, 128(2):445–457, 2005.

DiCiccio, T. J. and Martin, M. A. Simple modifications for signed roots of likelihood ratio statistics. *Journal of the Royal Statistical Society B*, 55:305–316, 1993.

Doornik, J. A. *An Object-Oriented Matrix Programming Language Ox 6.* Timberlake Consultants Press, London, 2009.

Efron, B. Nonparametric standard errors and confidence intervals. *Canadian Journal of Statistics*, 9(2):139–158, 1981.

Efron, B. Better bootstrap confidence intervals. *Journal of the American statistical Association*, 82(397):171–185, 1987.

Efron, B. More efficient bootstrap computations. *Journal of the American Statistical Association*, 85(409):79–89, 1990.

Ehsani, M. R., Saadatmanesh, H., and Tao, S. Design recommendations for bond of gfrp rebars to concrete. *Journal of Structural Engineering*, 122(3):247–254, 1996.

Espinheira, P. L., Ferrari, S. L., and Cribari-Neto, F. Bootstrap prediction intervals in beta regressions. *Computational Statistics*, 29(5):1263–1277, 2014.

Feigl, P. and Zelen, M. Estimation of exponential survival probabilities with concomitant information. *Biometrics*, 21(4):826–838, 1965.

Ferrari, S. L. and Pinheiro, E. C. Small-sample one-sided testing in extreme value regression models. *Advances in Statistical Analysis*, 100:79–97, 2016.

Firth, D. Bias reduction of maximum likelihood estimates. *Biometrika*, 80(1):27–38, 1993.

Folks, J. and Chhikara, R. The inverse gaussian distribution and its statistical application–a review. *Journal of the Royal Statistical Society B*, 40(3):263–289, 1978.

Fraser, D. A. S., Reid, N., and Wu, J. A simple general formula for tail probabilities for frequentist and Bayesian inference. *Biometrika*, 86(2):249–264, 1999.

Galea, M., Leiva-Sánchez, V., and Paula, G. Influence diagnostics in log-Birnbaum–Saunders regression models. *Journal of Applied Statistics*, 31(9):1049–1064, 2004.

Gómez, H. W., Elal-Olivero, D., Salinas, H. S., and Bolfarine, H. Bimodal extension based on the skew-normal distribution with application to pollen data. *Environmetrics*, 22(1):50–62, 2011.

Hannan, E. J. and Quinn, B. G. The determination of the order of an autoregression. *Journal of the Royal Statistical Society. Series B*, 41(2):190–195, 1979.

Hurvich, C. M. and Tsai, C.-L. Regression and time series model selection in small samples. *Biometrika*, 76(2):297–307, 1989.

Ishiguro, M. and Sakamoto, Y. WIC: An estimation-free information criterion. *Research Memorandum of the Institute of Statistical Mathematics*, 410, 1991.

Leiva, V. *The Birnbaum-Saunders Distribution*. Academic Press, London, 2015.

Leiva, V., Sanhueza, A., Kotz, S., and Araneda, N. A unified mixture model based on the inverse gaussian distribution. *Pakistan Journal of Statistics*, 26(3):445–460, 2010.

Leiva, V., Tejo, M., Guiraud, P., Schmachtenberg, O., Orio, P., and Marmolejo-Ramos, F. Modeling neural activity with cumulative damage distributions. *Biological cybernetics*, 109: 421–433, 2015.

Lemonte, A. J. A new extended Birnbaum–Saunders regression model for lifetime modeling. *Computational Statistics & Data Analysis*, 64:34–50, 2013.

Lemonte, A. J. and Cordeiro, G. M. Birnbaum–Saunders nonlinear regression models. *Computational Statistics & Data Analysis*, 53(12):4441–4452, 2009.

Lemonte, A. J. and Ferrari, S. L. Signed likelihood ratio tests in the Birnbaum–Saunders regression model. *Journal of Statistical Planning and Inference*, 141:1031–1040, 2011.

Lemonte, A. J., Cribari-Neto, F., and Vasconcellos, K. L. P. Improved statistical inference for the two-parameter Birnbaum–Saunders distribution. *Computational Statistics & Data Analysis*, 51(9):4656–4681, 2007.

Lemonte, A. J., Simas, A. B., and Cribari-Neto, F. Bootstrap-based improved estimators for the two-parameter Birnbaum–Saunders distribution. *Journal of Statistical Computation and Simulation*, 78(1):37–49, 2008.

Lewis, F., Butler, A., and Gilbert, L. A unified approach to model selection using the likelihood ratio test. *Methods in Ecology and Evolution*, 2(2):155–162, 2011.

Liseo, B. La classe delle densita normali sghembe: aspetti inferenziali da un punto di vista Bayesiano. *Statistica*, 50(1):71–79, 1990.

Martínez-Flórez, G., Bolfarine, H., and Gómez, H. W. The log-linear birnbaum-saunders power model. *Methodology and Computing in Applied Probability*, 2016. doi: 10.1007/s11009-016-9526-3.

McFadden, D. Conditional logit analysis of qualitative choice behavior. In Zarembka, P., editor, *Frontiers in econometrics*, pages 105–142. Academic Press, New York, 1973.

McQuarrie, A. D. A small-sample correction for the schwarz SIC model selection criterion. *Statistics & Probability Letters*, 44:79–86, 1999.

McQuarrie, A. D. and Tsai, C.-L. *Regression and time series model selection*. World Scientific, Singapore, 1998.

Mittelhammer, R. C., Judge, G. G., and Miller, D. J. *Econometric Foundations*. Cambridge University Press, New York, 2000.

Mojirsheibani, M. and Tibshirani, R. Some results on bootstrap prediction intervals. *Canadian Journal of Statistics*, 24(4):549–568, 1996.

Nagelkerke, N. J. A note on a general definition of the coefficient of determination. *Biometrika*, 78(3):691–692, 1991.

Ng, H., Kundu, D., and Balakrishnan, N. Modified moment estimation for the two-parameter Birnbaum–Saunders distribution. *Computational Statistics & Data Analysis*, 43(3):283–298, 2003.

Olmos, N., Martinez-Florez, G., and Bolfarine, H. Bimodal Birnbaum-Saunders distribution with applications to non-negative measurements. *Communications in Statistics–Theory and Methods*, 2016. doi: 10.1080/03610926.2015.1133824.

Owen, W. J. A new three-parameter extension to the Birnbaum–Saunders distribution. *IEEE Transactions on Reliability*, 55(3):475–479, 2006.

Owen, W. J. and Ng, H. K. T. Revisit of relationships and models for the Birnbaum–Saunders and inverse-gaussian distributions. *Journal of Statistical Distributions and Applications*, 2(1): 1–23, 2015.

Pan, W. Bootstrapping likelihood for model selection with small samples. *Journal of Computational and Graphical Statistics*, 8(4):687–698, 1999.

Patriota, A. G. On scale-mixture Birnbaum–Saunders distributions. *Journal of Statistical Planning and Inference*, 142(7):2221–2226, 2012.

Pianto, D. M. and Cribari-Neto, F. Dealing with monotone likelihood in a model for speckled data. *Computational Statistics & Data Analysis*, 55(3):1394–1409, 2011.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016. URL `https://www.R-project.org/`.

Ramsey, J. and Gilbert, R. A Monte Carlo study of some small sample properties of tests for specification error. *Journal of the American Statistical Association*, 67(337):180–186, 1972.

Ramsey, J. B. Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society B*, 31(2):350–371, 1969.

Rieck, J. R. A moment-generating function with application to the Birnbaum–Saunders distribution. *Communications in Statistics-Theory and Methods*, 28(9):2213–2222, 1999.

Rieck, J. R. and Nedelman, J. R. A log-linear model for the Birnbaum–Saunders distribution. *Technometrics*, 33(1):51–60, 1991.

Rocke, D. M. Bootstrap Bartlett adjustment in seemingly unrelated regression. *Journal of the American Statistical Association*, 84(406):598–601, 1989.

Sanhueza, A., Leiva, V., and Balakrishnan, N. The generalized Birnbaum–Saunders distribution and its theory, methodology, and application. *Communications in Statistics–Theory and Methods*, 37(5):645–670, 2008.

Santos, J. and Cribari-Neto, F. Hypothesis testing in log-Birnbaum–Saunders regressions. *Communications in Statistics-Simulation and Computation*, 2015. doi: 10.1080/03610918.2015.1080834.

Sartori, N. Bias prevention of maximum likelihood estimates for scalar skew normal and skew t distributions. *Journal of Statistical Planning and Inference*, 136(12):4259–4275, 2006.

Schwarz, G. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978.

Severini, T. A. An empirical adjustment to the likelihood ratio statistic. *Biometrika*, 86(2): 235–247, 1999.

Severini, T. A. *Likelihood Methods in Statistics*. Oxford University Press, New York, 2000.

Shibata, R. Bootstrap estimate of Kullback-Leibler information for model selection. *Statistica Sinica*, 7:375–394, 1997.

Smith, B., Wang, S., Wong, A., and Zhou, X. A penalized likelihood approach to parameter estimation with integral reliability constraints. *Entropy*, 17(6):4040–4063, 2015.

Stine, R. A. Bootstrap prediction intervals for regression. *Journal of the American Statistical Association*, 80(392):1026–1031, 1985.

Sugiura, N. Further analysts of the data by Akaike's information criterion and the finite corrections: Further analysts of the data by Akaike's. *Communications in Statistics-Theory and Methods*, 7(1):13–26, 1978.

Tsionas, E. G. Bayesian inference in Birnbaum–Saunders regression. *Communications in Statistics-Theory and Methods*, 30(1):179–193, 2001.

Vilca, F., Azevedo, C. L., and Balakrishnan, N. Bayesian inference for sinh-normal/independent nonlinear regression models. *Journal of Applied Statistics*, 2016. doi: 10.1080/02664763.2016.1238058.

Villegas, C., Paula, G. A., and Leiva, V. Birnbaum–Saunders mixed models for censored reliability data analysis. *IEEE Transactions on Reliability*, 60(4):748–758, 2011.

Vuong, Q. H. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57:307–333, 1989.

Wei, B. C., Hu, Y. Q., and Fung, W. K. Generalized leverage and its applications. *Scandinavian Journal of Statistics*, 25:25–37, 1998.

Williams, D. A. Discrimination between regression models to determine the pattern of enzyme synthesis in synchronous cell cultures. *Biometrics*, 26(1):23–32, 1970.

Wu, J. and Wong, A. C. Improved interval estimation for the two-parameter Birnbaum–Saunders distribution. *Computational Statistics & Data Analysis*, 47(4):809–821, 2004.

Xie, F. C. and Wei, B. C. Diagnostics analysis for log-Birnbaum–Saunders regression models. *Computational Statistics & Data Analysis*, 51(9):4692–4706, 2007.

Zhu, X. and Balakrishnan, N. Birnbaum–Saunders distribution based on Laplace kernel and some properties and inferential issues. *Statistics & Probability Letters*, 101:1–10, 2015.