Log-symmetric regressions models under the presence of uncensored, left- or right-censored observations: a semi-parametric approach

Luis Hernando Vanegas Penagos

Thesis presented
to the
Institute of Mathematics and Statistics
of the
University of São Paulo
to
fulfill the reuirements of the degree
of
Doctor of Science

Program: Statistics

Supervisor: Prof. Dr. Gilberto A. Paula

During the developed of this work, the author received financial support from the Brazilian PEC-PG program through the agency CAPES

São Paulo, May 2015

Log-symmetric regressions models under the presence of uncensored, left- or right-censored observations: a semi-parametric approach

This is the original version of the thesis written by the candidate Luis Hernando Vanegas Penagos, as submitted to the examining board.

RESUMO

VANEGAS, L.H. **Modelos de regressão log-simétricos na presença de observações com e sem censura: uma abordagem semiparamétrica**. 2015. Tese (Doutorado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2015.

O principal objetivo deste trabalho é proporcionar ferramentas estatísticas no contexto da regressão semiparamétrica para analisar dados estritamente positivos e assimétricos na presença de observações censuradas (a esquerda ou direita mas não informativa) e não censuradas. São discutidos processos iterativos de estimação de parâmetros com base nos algoritmos de esperança/maximização, a esperança/maximização restrita e backfitting. Propriedades assintóticas dos estimadores de máxima verossimilhança penalizada são estudados analiticamente e usando experimentos de simulação. Discussões sobre estimação de graus de liberdade e intervalos de confiança simultâneos são dadas e alguns procedimentos de diagnóstico, tais como resíduos de tipo desvio e medidas de influência local, são derivadas. O pacote **ssym** do R é desenvolvido. Este pacote implementa as metodologias estatísticas abordados neste trabalho.

Palavras-chave: respostas assimétricas, estimativas de máxima verossimilhança penalizada, modelos semi-paramétricos, estimativas robustas, splines cúbicos naturais, B-splines, observações censuradas.

ABSTRACT

VANEGAS, L.H. **Log-symmetric regressions models under the presence of uncensored, left- or right-censored observations: a semi-parametric approach**. 2015. Tese (Doutorado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2015.

The main objective of this work is to provide statistical tools in the context of semiparametric regression to analyze strictly positive and asymmetric data under the presence of uncensored and non-informative left- or right-censored observations. Iterative processes of parameter estimation based on the expectation/maximization, expectation/constrained-maximization and backfitting algorithms are discussed. Asymptotic properties of the maximum penalized likelihood estimators are studied analytically and by using simulation experiments. Discussions on effective degrees of freedom estimation and simultaneous confidence intervals are given and some diagnostic procedures, such as deviance-type residuals and local influence measures, are derived. The R package **ssym** is developed. This package implements the statistical methodologies addressed in this work.

Keywords: skewness, asymmetric responses, maximum penalized likelihood estimates, semiparametric models, robust estimates, natural cubic splines, B-splines, censored observations.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Linear normal models are of central interest in statistical modeling and various extensions have been proposed in the sense of relaxing the error assumptions and the systematic component form. In particular, the assumption of symmetric errors has been largely investigated, especially in the last decade. The main attractive of this approach is the possibility of allowing both light- and heavy-tailed error distributions. Other possible forms of extensions are concerned with the assumption of nonlinear functions (parametric or nonparametric) in the systematic component as well as for modeling the dispersion. In addition, linear normal models may be applied for estimating the parameters of nonlinear multiplicative models with log-normal errors. The log-normal is a well known asymmetric distribution for positive values, where its two parameters may be interpreted as the median and the skewness (or the relative dispersion) of the distribution. Then, by applying a logarithm transformation to the response of the multiplicative model one may obtain a linear normal model from which the original parameters may be estimated. This approach is the main motivation for studying the log-symmetric class and to developing this work.

So that, applying the exponential transformation to the symmetric variables, a class of asymmetric distributions for positive values is generated with the log-normal distribution being a particular case. The former family of distributions, so-called log-symmetric class, includes bimodal distributions and covers light- and heavy-tailed error distributions such as log-Student-t, log-power-exponential, log-hyperbolic, log-slash, log-contaminated-normal, and the extensions of the harmonic law, Birnbaum Saunders and Birnbaum Saunders-t distributions, among others. For all of these distributions the scale and power parameters may be interpreted as the median and the skewness (or the relative dispersion), respectively. Shape parameters appear in some log-symmetric distributions as, for instance, the degrees of freedom parameter in the log-Student-t distribution. Similarly to the log-normal case, multiplicative nonlinear models are proposed in which the errors follow a log-symmetric distribution. However, the estimation, inference and model checking are performed under the symmetric error model obtained after a logarithm transformation to the response variable of the multiplicative model with log-symmetric errors.

This document is organized as follows. In Chapter 1, the log-symmetric class is derived and some of its statistical properties as well as some of its members, are presented. Quantile-based measures for the location, dispersion, relative dispersion, skewness and kurtosis are derived for the log-symmetric class. Parameter estimation and inference under the classic and the Bayesian approaches are discussed. Two applications to real datasets are provided. A multiplicative nonlinear model with log-symmetric errors is defined in Chapter 2, where the scale and power parameters, which are interpreted as the median and the skewness (or the relative dispersion) of the response variable distribution, may be modeled in a semi-parametric manner, so that, the distribution of the response variable and a set of covariates are related by a sum of unspecified functions whose functional forms are estimated from the data. Two possible modeling forms are proposed with the respective joint iterative processes being based on the Fisher scoring and backfitting algorithms. Asymptotic properties of the maximum penalized likelihood estimators are studied analytically and by using simulation experiments. Discussions on effective degrees of freedom estimation and simultaneous confidence intervals are given and some diagnostic procedures, such as deviance-type residuals and local influence measures, are developed.

In Chapter 3, a very flexible accelerated failure time model is proposed, which is an extension to handle non-informative left- and right-censored observations of the methods discussed in Chapter 2. This model provides high versatility in its random and systematic components, but it retains the well known appealing features of accelerated failure time models or log-location-scale models. Similarly to the uncensored case, the median and the skewness (or the relative dispersion) of the failure time distribution are explicitly modeled by semi-parametric functions. The parameters are estimated by using nonlinear optimization algorithms such as Gauss-Seidel, ECM and backfitting. Asymptotic properties of the maximum penalized likelihood estimators are studied analytically and by using simulation experiments. Deviance-type residuals and local influence measures are also derived.

Chapter 4 describes the capabilities and features of the new package **ssym**, which is an implementation of the log-symmetric models that is available from the Comprehensive R Archive Network (CRAN) at `http://CRAN.R-project.org/package=ssym`. This package allows to fit models under the presence of uncensored, left- and right-censored observations, where the median and the skewness (or the relative dispersion) are explicitly modeled by semi-parametric functions of covariates, whose nonparametric components are approximated by using natural cubic splines or P-splines. This package enables performing residual analysis, as well as sensitivity studies through local influence under usual perturbation schemes. Finally, in Chapter 5, the statistical tools presented in the previous chapters are illustrated by analyzing some real datasets using the R package **ssym**. Further examples can be found at `http://cran.r-project.org/web/packages/ssym/ssym.pdf`.

CHAPTER 2

LOG-SYMMETRIC DISTRIBUTIONS

Data whose interest variable is continuous, strictly positive, and asymmetric with possible outlying observations are commonly employed in various fields of practice. In fact, there is an extensive body of literature about distributions whose support is the interval $(0, \infty)$. Some of the more flexible distributions include the generalized modified Weibull (Carrasco *et al.*, 2008), generalized Inverse Gaussian (see, e.g., Jørgensen, 1982), and generalized Gamma (Stacy, 1962) distributions. However, according to Limpert *et al.* (2001), the log-normal distribution has been successfully applied in an enormous range of applications. Thus, to describe the behavior of strictly positive data, the log-symmetric distribution class is considered, because it is a generalization of the log-normal distribution that is flexible enough to include bimodal distributions as special cases and distributions that have heavier/lighter tails than those of the log-normal distribution.

Furthermore, the log-symmetric distributions are endowed with two interesting properties, closure under change of scale and closure under reciprocals, which, according to Puig (2008), are very desirable properties for distributions that are used to describe data with ratios of positive magnitudes. The log-symmetric class also generalizes and makes more flexible the distributions that have been developed to describe lifetimes under the assumption of accumulated damage (e.g., the Birnbaum-Saunders (Birnbaum and Saunders , 1969)), Birnbaum-Saunders-t (see, e.g., Barros *et al.*, 2008; Paula *et al.*, 2012) and generalized Birnbaum-Saunders (Díaz-García and Leiva , 2005) distributions) by introducing therein an additional parameter that may be used to control the shape of the hazard function and to regulate the skewness and the relative dispersion. Furthermore, as illustrated in this work, the log-symmetric class has several desirable statistical properties that may make it preferable to alternative distributions. For instance, the two parameters of the log-symmetric class are orthogonal and they may be interpreted directly as median and skewness, which are, in the context of asymmetric distributions, the most meaningful measures of location and shape, respectively. In addition, the extension of the log-symmetric class to the multivariate case is straightforward (Marchenko and Genton , 2010).

In this chapter, the log-symmetric class is characterized, and some of its main statistical properties are studied. In particular, quantile-based measures of location, dispersion, skewness, relative dispersion and kurtosis are derived, which are appropriate in the context of asymmetric and heavy-tailed distributions. Inference under the classic and the Bayesian approaches are also addressed. The practical use of the log-symmetric class of distributions is illustrated by describing data on per capita gross domestic product, and body fat percentage, in which the performance of the log-symmetric class is compared with that of some competitive and very flexible distributions such as the generalized modified Weibull, generalized Inverse Gaussian, generalized Gamma, log-skew-t (see, e.g., Azzalini *et al.*, 2003) and Box-cox-t (see, e.g., Rigby and Stasinopoulos, 2006) distributions.

2.1 Definition

Let Y be a continuous and symmetric random variable whose distribution belongs to the symmetric class (see, e.g., Fang *et al.*, 1990) that has location parameter $-\infty < \mu < \infty$, scale parameter $\phi > 0$ and density generator $g(\cdot)$. It is denoted as $Y \sim \mathcal{S}(\mu, \phi, g(\cdot))$. Its probability density function is given by $f_Y(y) = g[(y-\mu)^2/\phi]/\sqrt{\phi}$ for $-\infty < y < \infty$ provided that $g(u) > 0$ for $u > 0$ and $\int_0^\infty u^{-\frac{1}{2}} g(u) \partial u = 1$. Then, by setting $T = \exp(Y)$, a new class of distributions, the so-called log-symmetric class, is obtained. It is denoted this class, whose support is the interval $(0, \infty)$, as $T \sim \mathcal{LS}(\eta, \phi, g(\cdot))$, where $\eta = \exp(\mu)$ and ϕ are its scale and power parameters, respectively (see, for instance, Marshall and Olkin, 2007, chapter 7). The probability density function of T reduces to

$$f_T(t) = \frac{g(\tilde{t}^{\,2})}{t\sqrt{\phi}}, \qquad t > 0, \tag{2.1}$$

where $\tilde{t} = \log\Big[(t/\eta)^{\frac{1}{\sqrt{\phi}}}\Big]$. The density generator $g(\cdot)$ may involve an extra parameter (or an extra parameter vector), which is denoted here as ζ. Members of the class of distributions characterized by (2.1) include the log-normal, log-Student-t, log-power-exponential, log-logistic type I and II, log-hyperbolic, log-slash, log-contaminated-normal, harmonic law (see Puig, 2008, and references there in), Birnbaum-Saunders (Birnbaum and Saunders, 1969), Birnbaum-Saunders-t (see, e.g., Barros *et al.*, 2008; Paula *et al.*, 2012) and generalized Birnbaum-Saunders (see, e.g. Díaz-García and Leiva, 2005; Leiva *et al.*, 2008) distributions. Note that the Birnbaum-Saunders, Birnbaum-Saunders-t and generalized Birnbaum-Saunders distributions cited above are special cases (in which $\phi = 4$) of the (extended) homonymous distributions that will be considered here. Similarly, the harmonic law cited above is a special case (in which $\phi = 1$) of the (extended) homonymous distribution that will be considered here.

2.2 Some statistical properties

If $T \sim \mathcal{LS}(\eta, \phi, g(\cdot))$ then, one can verify that:

(P1) The cumulative distribution function (cdf) of T may be written as $F_T(t) = F_Z(\tilde{t})$, where $F_Z(\cdot)$ is the cdf of $Z = (Y-\mu)/\sqrt{\phi} \sim \mathcal{S}(0, 1, g(\cdot))$.

(P2) $T^* = (T/\eta)^{\frac{1}{\sqrt{\phi}}} \sim \mathcal{LS}(1, 1, g(\cdot))$, i.e., T^* follows standard log-symmetric distribution.

(P3) $cT \sim \mathcal{LS}(c\eta, \phi, g(\cdot))$ for all constant $c > 0$.

(P4) $T^c \sim \mathcal{LS}(\eta^c, c^2\phi, g(\cdot))$ for all constant $c \neq 0$.

(P5) (T/η) and (η/T) are random variables that are identically distributed.

(P6) If $\mathrm{E}(T^r)$ and $\mathrm{E}(Z)$ exist, then $\mathrm{E}(T^r) \geq \eta^r$.

(P7) If $\mathrm{M}_Y(r)$ exists, then $\mathrm{E}(T^r) = \mathrm{M}_Y(r)$, where $\mathrm{M}_Y(r)$ is the moment generating function of $Y = \log(T)$.

(P8) The quantile function of T is given by $\vartheta(q) = \eta \exp(\sqrt{\phi}\, Z_\zeta^{(q)})$, where $Z_\zeta^{(q)}$ is the $100(q)\%$ quantile of $Z = (Y-\mu)/\sqrt{\phi} \sim \mathcal{S}(0, 1, g(\cdot))$.

(P9) The Shannon entropy of T, which is denoted as $\mathrm{ET}(T)$, may be expressed as $\mathrm{ET}(T) = \log\big[\eta\sqrt{\phi}\big] + \mathrm{ET}(Z)$, provided that $\mathrm{ET}(Z)$ and $\mathrm{E}(Z)$ exist.

(P10) If the $\mathrm{W}(\cdot)$ function is such that $\log[\mathrm{W}(x)] = h[\log(x)]$, where $h\colon \mathbb{R} \to \mathbb{R}$ is an injective and differentiable odd function, then the distribution of $\overline{T} = \mathrm{W}(T^*)$ is $\mathcal{LS}(1, 1, \overline{g}(\cdot))$, where $\overline{g}(u) = g\big\{[h^{-1}(\sqrt{u})]^2\big\}/h'[h^{-1}(\sqrt{u})]$.

Let $\mathbf{t} = (t_1, t_2, \ldots, t_n)^\top$ be a random sample of size n from $T \sim \mathcal{LS}(\eta, \phi, g(\cdot))$, in which ζ is assumed to be known or fixed. Then, the maximum likelihood estimates (MLEs) of η and ϕ (which are denoted here as $\hat{\eta}$ and $\hat{\phi}$, respectively) may be written as

$$\hat{\eta} = \left\{\prod_{k=1}^{n} t_k^{\mathrm{v}\left(\hat{\tilde{t}}_k\right)}\right\}^{1/\sum_{k=1}^{n}\mathrm{v}\left(\hat{\tilde{t}}_k\right)} \quad \text{and} \quad \hat{\phi} \propto \frac{\sum_{k=1}^{n}\mathrm{v}\left(\hat{\tilde{t}}_k\right)[\log(t_k/\hat{\eta})]^2}{\sum_{k=1}^{n}\mathrm{v}\left(\hat{\tilde{t}}_k\right)},$$

where $\hat{\tilde{t}}_k = \log\left[(t_k/\hat{\eta})^{\frac{1}{\sqrt{\hat{\phi}}}}\right]$ and $\mathrm{v}(t) = -2g'(t^2)/g(t^2)$ is a weight function induced by $g(\cdot)$, and $\mathrm{v}\left(\hat{\tilde{t}}_1\right), \ldots, \mathrm{v}\left(\hat{\tilde{t}}_n\right)$ is a set of positive weights if the $g(u)$ function is monotonically decreasing for $u > 0$, with $g'(u) = \partial g(u)/\partial u$. Therefore, when $\mathrm{v}\left(\hat{\tilde{t}}_k\right) > 0$ for $k = 1, \ldots, n$, the MLE of η may be interpreted as a weighted geometric mean of $t_1, t_2, \ldots, t_n$, whereas the MLE of ϕ is proportional to a weighted arithmetic mean, for which $\mathrm{v}\left(\hat{\tilde{t}}_1\right), \ldots, \mathrm{v}\left(\hat{\tilde{t}}_n\right)$ are the individual-specific weights. Thus, the choice of $g(\cdot)$ may induce a $\mathrm{v}(\cdot)$ function that enables one to estimate the parameters using the maximum likelihood method in a manner that is robust to extreme or outlying observations (i.e., the $g(\cdot)$ function may induce a $\mathrm{v}(t)$ function whose value decreases as the t value departs from the "centre" of the T distribution).

2.3 Members of the log-symmetric class

- Log-normal(η, ϕ):

$$g(u) \propto \exp\left[-\frac{1}{2}u\right] \qquad \text{and} \qquad \mathrm{v}(t) = 1.$$

- Log-Student-$t(\eta, \phi, \zeta)$:

$$g(u) \propto \left[1 + \frac{u}{\zeta}\right]^{-\frac{\zeta+1}{2}}, \ \zeta > 0, \qquad \text{and} \qquad \mathrm{v}(t) = \frac{\zeta + 1}{\zeta + t^2}.$$

If $\phi > (\zeta + 1)^2/4\zeta$, then the function $f_T(t)$ is monotonically decreasing.

- Log-power-exponential(η, ϕ, ζ):

$$g(u) \propto \exp\left[-\frac{1}{2}u^{\frac{1}{1+\zeta}}\right], \ -1 < \zeta \leq 1, \quad \text{and} \quad \mathrm{v}(t) = \frac{|t|^{-\frac{2\zeta}{\zeta+1}}}{1+\zeta}.$$

The log-normal ($\zeta = 0$) and the log-Laplace ($\zeta = 1$) distributions are special cases. If $\zeta = 1$ and $\phi > 1/4$, then the function $f_T(t)$ is monotonically decreasing. The distribution of $Y = \log(T)$ is power exponential (Box and Tiao, 1973).

- Log-hyperbolic(η, ϕ, ζ):

$$g(u) \propto \exp\left[-\zeta\sqrt{1+u}\right], \ \zeta > 0, \quad \text{and} \quad \mathrm{v}(t) = \frac{\zeta}{\sqrt{1+t^2}}.$$

The log-normal(η, σ^2) distribution is a limiting case when $\phi \to \infty$ and $\phi/\zeta \to \sigma^2$. Similarly, the log-Laplace(η, σ^2) distribution is a limiting case when $\phi \to 0$ and $\phi/\zeta^2 \to 4\sigma^2$ (Fonseca *et al.*, 2012). If $\phi \geq \zeta^2$, then the function $f_T(t)$ is monotonically decreasing. The moments of T are given by

$$\mathrm{E}(T^r) = \eta^r \frac{\mathrm{K}_1\left(\sqrt{\zeta^2 - \phi r^2}\right)}{\mathrm{K}_1(\zeta)} \frac{\zeta}{\sqrt{\zeta^2 - \phi r^2}}, \qquad |r| < \zeta/\sqrt{\phi},$$

in which $\mathrm{K}_r(\zeta)=\frac{1}{2}\int_0^{\infty}x^{r-1}\exp\Big[-\frac{\zeta}{2}(x+\frac{1}{x})\Big]\partial x$ is the modified Bessel function of third-order and index r. The distribution of $Y=\log(T)$ is symmetric hyperbolic (Barndoff-Nielsen, 1977).

- Log-slash(η,ϕ,ζ):

$$g(u)\propto\mathrm{IGF}\Big(\zeta+\frac{1}{2},\frac{u}{2}\Big)\ \text{ and }\ \mathrm{v}(t)=\mathrm{IGF}\Big(\zeta+\frac{3}{2},\frac{t^2}{2}\Big)/\mathrm{IGF}\Big(\zeta+\frac{1}{2},\frac{t^2}{2}\Big),$$

where $\zeta>0$, $\mathrm{IGF}(a,x)=\int\limits_0^1\exp(-tx)t^{a-1}\partial t$ is the incomplete gamma function for $a>0$ and $x\geq 0$. The distribution of $Y=\log(T)$ is slash (Rogers and Tukey, 1972).

- Log-contaminated-normal$(\eta,\phi,\zeta=(\zeta_1,\zeta_2)^\top)$:

$$g(u)\propto\sqrt{\zeta_2}\exp\Big[-\frac{1}{2}\zeta_2 u\Big]+\frac{(1-\zeta_1)}{\zeta_1}\exp\Big[-\frac{1}{2}u\Big],\quad \zeta_1,\zeta_2\in(0,1);$$

and

$$\mathrm{v}(t)=\frac{\zeta_2{}^{\frac{3}{2}}\zeta_1\exp\Big[(1-\zeta_2)\frac{t^2}{2}\Big]+(1-\zeta_1)}{\zeta_2{}^{\frac{1}{2}}\zeta_1\exp\Big[(1-\zeta_2)\frac{t^2}{2}\Big]+(1-\zeta_1)}.$$

The moments of T are given by

$$\mathrm{E}(T^r)=\eta^r\exp\Big[\frac{1}{2}\phi r^2\Big]\Big\{\zeta_1\exp\Big[\frac{1-\zeta_2}{2\zeta_2}\phi r^2\Big]+(1-\zeta_1)\Big\}.$$

The mode of T is within the interval $(\eta\exp(-\phi/\zeta_2),\ \ \eta\exp(-\phi))$. The distribution of $Y=\log(T)$ is contaminated normal.

- (extended) Birnbaum-Saunders(η,ϕ,ζ):

$$g(u)\propto\cosh(u^{\frac{1}{2}})\exp\Big[-\frac{2}{\zeta^2}\sinh^2(u^{\frac{1}{2}})\Big],\quad \zeta>0,$$

and

$$\mathrm{v}(t)=\frac{\sinh(t)}{t}\left[\frac{4\cosh(t)}{\zeta^2}-\frac{1}{\cosh(t)}\right],$$

where $\sinh(\cdot)$ and $\cosh(\cdot)$ represent the hyperbolic sine and cosine functions, respectively. The weighting $\mathrm{v}(t)$ of the Birnbaum-Saunders distribution increases as $|t|$ also increases. In addition, the Birnbaum-Saunders distribution is bimodal if $\zeta>2$ and $\phi<\varrho(\zeta)$, where

$$\varrho(\zeta)=\Big(\sqrt{1+2\zeta^2}-3\Big)\left[\frac{1}{\sqrt{1+\sqrt{1+2\zeta^2}}}-\frac{\sqrt{1+\sqrt{1+2\zeta^2}}}{\zeta^2}\right]^2.$$

Note that $\varrho(\zeta)<1$ for all $\zeta>2$. The moments of T are given by (Rieck, 1999)

$$\mathrm{E}(T^r)=\eta^r\frac{\exp(1/\zeta^2)}{\zeta\sqrt{2\pi}}\left[\mathrm{K}_{r_1^*}(1/\zeta^2)+\mathrm{K}_{r_2^*}(1/\zeta^2)\right],$$

where $r_1^*=(r\sqrt{\phi}+1)/2$ and $r_2^*=(r\sqrt{\phi}-1)/2$. The distribution of $Y=\log(T)$ is sinh-normal (Rieck and Nedelman, 1991). From the properties (P3) and (P4), $\overline{T} = \eta(T/\eta)^{\frac{2}{\sqrt{\phi}}} \sim$ Birnbaum-Saunders$(\eta,\ 4,\ \zeta)$, i.e., $\overline{T}$ exhibits the probability distribution proposed by Birnbaum and Saunders (1969).

- (extended) Birnbaum-Saunders-$t(\eta, \phi, \zeta = (\zeta_1, \zeta_2)^\top)$:

$$g(u) \propto \cosh(u^{\frac{1}{2}})\Big[\zeta_2\zeta_1^2 + 4\sinh^2(u^{\frac{1}{2}})\Big]^{-\frac{\zeta_2+1}{2}}, \quad \zeta_1 > 0,\ \ \zeta_2 > 0;$$

and

$$\mathrm{v}(t) = \frac{\sinh(t)}{t}\left[\frac{4(\zeta_2+1)\cosh(t)}{\zeta_2\zeta_1^2+4\sinh^2(t)} - \frac{1}{\cosh(t)}\right].$$

The Birnbaum-Saunders-t distribution is bimodal if $\zeta_1 > 2\sqrt{1+1/\zeta_2}$ and $\phi < [\mathrm{v}(t_1)t_1]^2$, where $t_1 = \log\left[\sqrt{t_0^2-1}+t_0\right]$ and t_0 is given by

$$t_0 = \frac{1}{2}\left[(\zeta_2+1) + \frac{2\zeta_2}{\zeta_1^2\zeta_2 - 4}\right]^{-\frac{1}{2}} \times \left[(\zeta_2+3)+\sqrt{(\zeta_2+3)^2 + 2(\zeta_2+1)(\zeta_1^2\zeta_2-4) + 4\zeta_2}\right]^{\frac{1}{2}}.$$

The distribution of $Y = \log(T)$ is sinh-t (see, e.g., Barros *et al.* , 2008; Paula *et al.* , 2012). From the properties (P3) and (P4), $\overline{T} = \eta(T/\eta)^{\frac{2}{\sqrt{\phi}}} \sim$ Birnbaum-Saunders-$t(\eta, 4, \zeta = (\zeta_1, \zeta_2)^\top)$, i.e., $\overline{T}$ exhibits the probability distribution studied by Barros *et al.* (2008) and Paula *et al.* (2012).

- (extended) Generalized Birnbaum-Saunders$(\eta, \phi, \zeta = (\zeta_1, \zeta_2)^\top)$:

$$g(u) \propto \cosh(u^{\frac{1}{2}}) \times h_{\zeta_2}\left[\frac{4}{\zeta_1^2}\sinh^2(u^{\frac{1}{2}})\right], \quad \zeta_1 > 0,$$

where $h_{\zeta_2}(\cdot)$ represents the kernel of the symmetric distribution (indexed by the extra parameter ζ_2) that describes the *cumulative damage*, where $h_{\zeta_2}(u) > 0$ for $u > 0$ and $\int_0^\infty u^{-\frac{1}{2}} h_{\zeta_2}(u)\partial u = 1$. The distribution of $Z^* = (2/\zeta_1)\sinh(\tilde{t})$ is given by $f_{Z^*}(z) = h_{\zeta_2}(z^2)$ (Leiva *et al.*, 2008). The Birnbaum-Saunders, Birnbaum-Saunders-t and slash-Birnbaum-Saunders distributions are special cases (Balakrishnan *et al.*, 2009).

- (extended) Harmonic-law(η, ϕ, ζ):

$$g(u) \propto \exp\Big[-\zeta\cosh(u^{\frac{1}{2}})\Big],\ \zeta > 0, \quad \text{and} \quad \mathrm{v}(t) = \zeta\frac{\sinh(t)}{t}.$$

The moments of T are given by

$$\mathrm{E}(T^r) = \eta^r \frac{\mathrm{K}_{r^*}(\zeta)}{\mathrm{K}_0(\zeta)},$$

where $r^* = r\sqrt{\phi}$. The mode of T is $\eta\Big[\zeta^{-1}\Big(\sqrt{\phi+\zeta^2} - \sqrt{\phi}\Big)\Big]^{\sqrt{\phi}}$. From the properties (P3) and (P4), $\overline{T} = \eta(T/\eta)^{\frac{1}{\sqrt{\phi}}} \sim$ Harmonic-law$(\eta, 1, \zeta)$, i.e., $\overline{T}$ exhibits the probability distribution studied by Puig (2008).

The values of the individual-specific weights $\mathrm{v}(t)$ are strictly positive for the log-normal, log-Student-t, log-power-exponential, log-slash, log-hyperbolic, log-contaminated-normal, har-

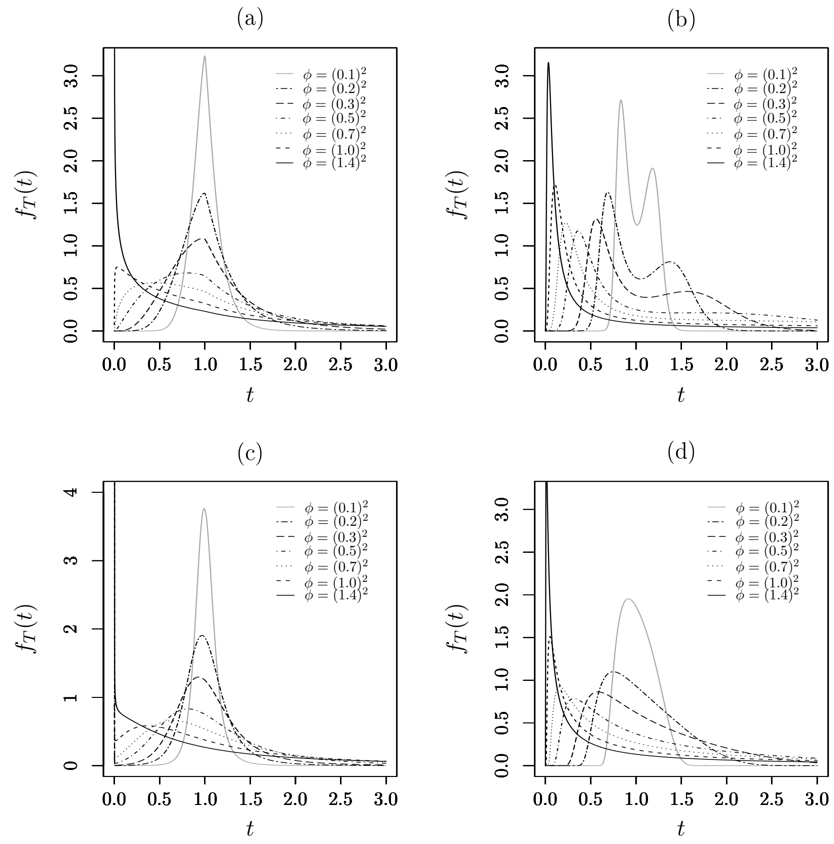

Figure 2.1: *Graphs of the density function of the log-power-exponential* ($\eta = 1, \phi, \zeta = 0.5$) *(a), Birnbaum-Saunders-t* ($\eta = 1, \phi, \zeta = (6, 4)^\top$) *(b), log-Student-t* ($\eta = 1, \phi, \zeta = 4$) *(c), and harmonic-law* ($\eta = 1, \phi, \zeta = 0.1$) *(d) distributions.*

monic law, Birnbaum-Saunders (for $\zeta \leq 2$) and Birnbaum-Saunders-t (for $\zeta_1 \leq 2\sqrt{1 + 1/\zeta_2}$) distributions. Furthermore, for distributions with tails that are heavier than those of the log-normal distribution (e.g., the log-Student-t, log-power-exponential (for $0 < \zeta \leq 1$), log-slash, log-hyperbolic (for small ζ) and log-contaminated-normal distributions) the individual-specific weights tend to be smaller as t departs from the "centre" of the T distribution. Therefore, for distributions that have heavier tails, the MLEs of η and ϕ are less sensitive to extreme or outlying observations than for the log-normal distribution. Similar results hold for the distributions that were developed to describe lifetimes under the assumption of cumulative damage. In fact, one can verify that the weights of extreme or outlying observations for the Birnbaum-Saunders-t distribution are (relatively) smaller than those for the Birnbaum-Saunders distribution. Figure 2.1 shows the probability density functions of some log-symmetric distributions. This figure illustrates the flexibility of the log-symmetric class.

2.4 Summary of the shape

The measures that are most frequently used for assessing the location, dispersion, relative dispersion, skewness and kurtosis are based on moments. However, because of their derivation, such measures may be not appropriate in the context of asymmetric distributions. Furthermore, sometimes the moments are not finite or are quite difficult to calculate. Therefore, in this section, quantile-based measures of the location, dispersion, relative dispersion, skewness and kurtosis for the log-symmetric class are derived; these measures, exist even for distributions for which no moments exist, are appropriate in the context of asymmetric distributions, are easier to calculate and/or interprete than those based on moments, and are invariant under changes in the extreme tails of the distribution. Furthermore, some of these measures (i.e., those used to measure the relative dispersion, skewness and kurtosis) are invariant under location-scale transformations (see, e.g., Groeneveld and Meeden, 1984).

2.4.1 Location

The median of $T \sim \mathcal{LS}(\eta, \phi, g(\cdot))$ is η, and the mode(s) of T may be written as $M_T = \eta \exp\big(t_T \sqrt{\phi}\big)$ provided that $f_T'(t)$ is continuous in M_T, in which t_T is(are) the solution(s) of

$$-\mathrm{v}(t)t = \sqrt{\phi} \quad \text{restricted to} \quad \mathrm{v}'(t)t^2 \mathrm{sign}(t) > \sqrt{\phi}\,\mathrm{sign}(t),$$

where $\mathrm{v}'(t) = \partial \mathrm{v}(t)/\partial t$. In addition, it is possible to verify that $M_T < \eta$ if both $g(u)$ is monotonically decreasing for $u > 0$ and $f_T'(t)$ is continuous in M_T.

2.4.2 Dispersion

The interquartile range of $T \sim \mathcal{LS}(\eta, \phi, g(\cdot))$ may be expressed as

$$\varsigma = \vartheta(0.75) - \vartheta(0.25) = 2\eta \sinh\Big(\sqrt{\phi}\, Z_\zeta^{(0.75)}\Big), \quad \varsigma \in (0, \infty).$$

2.4.3 Relative dispersion

The coefficient of quartile variation (Zwillinger and Kokoska, 2000, page 17) is given by

$$\varpi = \frac{\vartheta(0.75) - \vartheta(0.25)}{\vartheta(0.75) + \vartheta(0.25)} = \tanh\Big(\sqrt{\phi} Z_\zeta^{(0.75)}\Big), \quad \varpi \in (0, 1),$$

where $\tanh(\cdot)$ is the hyperbolic tangent function. Note that ϖ is a monotonically increasing function of ϕ for fixed ζ. According to Bonett (2006), ϖ may be preferable to the coefficient of variation for describing the relative dispersion in asymmetric distributions.

2.4.4 Skewness

A quantile-based measure of skewness (see, e.g., Groeneveld and Meeden, 1984; Hinkley, 1975) is given by

$$\begin{aligned} \varkappa(q) &= \frac{\vartheta(q) + \vartheta(1-q) - 2\vartheta(1/2)}{\vartheta(1-q) - \vartheta(q)} \\ &= \operatorname{cosech}\Big(\sqrt{\phi} Z_\zeta^{(q)}\Big) - \operatorname{cotanh}\Big(\sqrt{\phi} Z_\zeta^{(q)}\Big), \end{aligned}$$

where $\varkappa(q) \in (0, 1)$, $q \in (0, \frac{1}{2})$, and $\operatorname{cotanh}(\cdot)$ and $\operatorname{cosech}(\cdot)$ represent the hyperbolic cotangent and cosecant functions, respectively. A simple derivative reveals that for all $q \in (0, \frac{1}{2})$ the measure of skewness $\varkappa(q)$ is a monotonically increasing function of ϕ for fixed ζ. Therefore, ϕ may be interpreted as the skewness of T for fixed ζ.

2.4.5 Kurtosis

The kurtosis proposed by Moors (1988) reduces to

$$\begin{aligned} \bar{\varsigma} &= \frac{\vartheta(7/8) - \vartheta(5/8) + \vartheta(3/8) - \vartheta(1/8)}{\vartheta(6/8) - \vartheta(2/8)} \\ &= \frac{\sinh\Big(\sqrt{\phi} Z_\zeta^{(7/8)}\Big) - \sinh\Big(\sqrt{\phi} Z_\zeta^{(5/8)}\Big)}{\sinh\Big(\sqrt{\phi} Z_\zeta^{(6/8)}\Big)}, \end{aligned}$$

where $\bar{\varsigma} \in [0, \infty)$.

The main conclusion of this section is that, irrespective of the value of ϕ, η is the median of the T distribution. Similarly, irrespective of the value of η, ϕ is the skewness (or the relative dispersion) of T for fixed ζ.

2.5 Maximum likelihood estimation

The log-likelihood function of the interest parameters can be written as

$$\mathsf{L}(\boldsymbol{\theta}) = -\frac{n}{2}\log(\phi) - \sum_{k=1}^{n}\log(t_k) + \sum_{k=1}^{n}\log[g(\tilde{t}_k^2)].$$

To calculate the maximum likelihood estimate of $\boldsymbol{\theta} = (\eta, \phi)^\top$, denoted as $\hat{\boldsymbol{\theta}}$, the system of equations given by $\Big(\mathrm{U}_\eta(\hat{\boldsymbol{\theta}}), \mathrm{U}_\phi(\hat{\boldsymbol{\theta}})\Big) = \Big(\partial\mathsf{L}(\hat{\boldsymbol{\theta}})/\partial\hat{\eta}, \partial\mathsf{L}(\hat{\boldsymbol{\theta}})/\partial\hat{\phi}\Big) = (0,0)$ is solved using the Fisher scoring algorithm, where

$$\mathrm{U}_\eta(\boldsymbol{\theta}) = \frac{1}{\eta\phi}\log\left[\prod_{k=1}^{n}(t_k/\eta)^{\mathrm{v}(\tilde{t}_k)}\right]$$

and

$$\mathrm{U}_\phi(\boldsymbol{\theta}) = -\frac{n}{2\phi} + \frac{1}{2\phi}\sum_{k=1}^{n}\mathrm{v}(\tilde{t}_k)\tilde{t}_k^2.$$

The (expected) Fisher information matrix, which is denoted as $\mathbf{K}(\boldsymbol{\theta})$, is given by

$$n^{-1}\mathbf{K}(\boldsymbol{\theta}) = -n^{-1}\mathrm{E}[\partial^2\ell(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top] = \begin{bmatrix} d_g(\zeta)/\phi\eta^2 & 0 \\ 0 & [f_g(\zeta)-1]/4\phi^2 \end{bmatrix},$$

where $d_g(\zeta) = \mathrm{E}[\mathrm{v}^2(Z)Z^2]$ and $f_g(\zeta) = \mathrm{E}[\mathrm{v}^2(Z)Z^4]$ for $Z \sim \mathcal{S}(0,1,g(\cdot))$. For instance, the quantity $d_g(\zeta)$ is equal to 1, $(\zeta+1)/(\zeta+3)$, $\{2^{1-\zeta}\Gamma[(3-\zeta)/2]\}/\{(1+\zeta)^2\Gamma[(1+\zeta)/2]\}$ and $2+\frac{4}{\zeta^2}-\frac{\sqrt{2\pi}}{\zeta}\left\{1-\mathrm{erf}\left(\frac{\sqrt{2}}{\zeta}\right)\right\}\exp\!\left(\frac{2}{\zeta^2}\right)$ when T is assumed to exhibit the log-normal, log-Student-t, log-power-exponential and Birnbaum-Saunders distributions, respectively, where $\Gamma(\cdot)$ represents the gamma function and $\mathrm{erf}(x) = (2/\sqrt{\pi})\int_0^x e^{-t^2}dt$. Similarly, the quantity $f_g(\zeta)$ is equal to 3, $3(\zeta+1)/(\zeta+3)$ and $(\zeta+3)/(\zeta+1)$ when T is assumed to exhibit the log-normal, log-Student-t and log-power-exponential distributions, respectively (see, e.g., Cordeiro *et al.* (2000); Villegas *et al.* (2013)). Then, the Fisher scoring algorithm becomes

Algorithm 1.1

Step 1. Initialize the counter as $l = 0$ and set the initial value to $\boldsymbol{\theta}^{(0)}$.

Step 2. Based on $\boldsymbol{\theta}^{(l)}$ calculate the following expressions:

$$\eta^{(l+1)} = \eta^{(l)}\left\{\prod_{k=1}^{n}\left[t_k/\eta^{(l)}\right]^{\rho\left(\tilde{t}_k^{(l)}\right)}\right\}^{\frac{1}{n}} \quad \text{and}$$

$$\log[\phi^{(l+1)}] = \log[\phi^{(l)}] + \frac{2}{f_g(\zeta)-1}\left\{n^{-1}\sum_{k=1}^{n}\mathrm{v}\left(\tilde{t}_k^{(l)}\right)\left[\tilde{t}_k^{(l)}\right]^2 - 1\right\},$$

where $\tilde{t}_k^{(l)} = \log\left[\left(t_k/\eta^{(l)}\right)^{1/\sqrt{\phi^{(l)}}}\right]$ and $\rho(t) = \mathrm{v}(t)/d_g(\zeta)$.

Step 3. Update $l = (l+1)$ and $\boldsymbol{\theta}^{(l)}$.

Step 4. Repeat steps 3 and 4 until convergence of $\boldsymbol{\theta}^{(l)}$.

Because the MLEs of η and ϕ for the log-normal distribution have closed forms, they can be used as initial values for the iterative procedure for other log-symmetric distributions. Because some distributions, such as the log-Student-t, log-power-exponential (for $0 \leq \zeta \leq 1$), log-slash, log-hyperbolic and log-contaminated-normal distributions, may be obtained as a power mixture of log-normal distributions (see, e.g., Andrews and Mallows, 1974; Barndoff-Nielsen, 1977; West , 1987), the EM algorithm (Dempster *et al.*, 1977) can be used in those cases to develop an more efficient iterative process for parameter estimation. Then, for these distributions the *Step 2* of the *Algorithm 1.1* reduces to

$$\eta^{(l+1)} = \left\{\prod_{k=1}^{n} t_k^{\mathrm{v}\left(\tilde{t}_k^{(l)}\right)}\right\}^{1/\sum_{k=1}^{n}\mathrm{v}\left(\tilde{t}_k^{(l)}\right)} \quad \text{and}$$

$$\phi^{(l+1)} = n^{-1}\sum_{k=1}^{n}\mathrm{v}\left(\tilde{t}_k^{(l)}\right)\Big[\log(t_k/\eta^{(l)})\Big]^2.$$

Moreover, according to Balakrishnan *et al.* (2009), some distributions of the generalized Birnbaum-Saunders class (e.g., the Birnbaum-Saunders-t and slash-Birnbaum-Saunders distributions) can be obtained as scale mixtures of Birnbaum-Saunders distributions. Thus, the EM algorithm can also be used in those cases to develop a more efficient iterative procedure of parameter estimation. For example, under the Birnbaum-Saunders-t distribution the joint iterative process for $\hat{\eta}$ and $\hat{\phi}$ becomes

Algorithm 1.2

Step 1. Initialize the counter as $m = 0$ and set the initial value to $\boldsymbol{\theta}^{(0)}$.

Step 2. Calculate $\mathbf{u}^{(m)} = (u_1^{(m)}, \dots, u_n^{(m)})^\top$ based on $\boldsymbol{\theta}^{(m)}$ as follows:

$$u_k^{(m)} = \frac{\zeta_1^2(\zeta_2+1)}{\zeta_1^2\zeta_2 + \Big[2\sinh(\tilde{t}_k^{(m)})\Big]^2} \qquad \text{for} \quad k = 1, \dots, n.$$

Step 3. Calculate $d_g^*\Big(\zeta_1/[u_1^{(m)}]^{\frac{1}{2}}\Big), \dots, d_g^*\Big(\zeta_1/[u_n^{(m)}]^{\frac{1}{2}}\Big)$, where

$$d_g^*(\zeta) = 2 + \frac{4}{\zeta^2} - \frac{\sqrt{2\pi}}{\zeta}\left\{1 - \frac{2}{\sqrt{\pi}}\int_0^{\frac{\sqrt{2}}{\zeta}}\exp(-t^2)\partial t\right\}\exp\left(\frac{2}{\zeta^2}\right).$$

Step 4. Calculate $f_g^*\Big(\zeta_1/[u_1^{(m)}]^{\frac{1}{2}}\Big), \dots, f_g^*\Big(\zeta_1/[u_n^{(m)}]^{\frac{1}{2}}\Big)$, where

$$f_g^*(\zeta) = \mathrm{E}\Big[\big(4\sinh(z)\cosh(z)z/\zeta^2 - \tanh(z)z\big)^2\Big], \quad \text{and} \quad \exp(z) \sim \text{Birnbaum-Saunders}(1,1,\zeta).$$

Step 5. Initialize the counter as $l = 0$ and set the initial value to $\boldsymbol{\theta}_*^{(0)} = \boldsymbol{\theta}^{(m)}$.

Step 6. Perform the following algorithm based on $\boldsymbol{\theta}_*^{(l)}$:

(A) Compute the following expressions:

$$\eta^{(l+1)} = \eta^{(l)}\left\{\prod_{k=1}^{n}\Big[t_k/\eta^{(l)}\Big]^{\rho^*\left(\tilde{t}_k^{(l)}, u_k^{(m)}\right)}\right\} \quad \text{and}$$

$$\log[\phi^{(l+1)}] = \log[\phi^{(l)}] + \frac{2\sum_{k=1}^{n}\left\{\mathrm{v}^*\left(\tilde{t}_k^{(l)}, u_k^{(m)}\right)\Big[\tilde{t}_k^{(l)}\Big]^2 - 1\right\}}{\sum_{k=1}^{n}\left\{f_g^*\Big(\zeta_1/[u_k^{(m)}]^{\frac{1}{2}}\Big) - 1\right\}},$$

where

$$\rho^*\left(\tilde{t}_k^{(l)}, u_k^{(m)}\right) = \mathrm{v}^*\left(\tilde{t}_k^{(l)}, u_k^{(m)}\right) / \sum_{i=1}^{n} d_g^*\left(\zeta_1/[u_i^{(m)}]^{\frac{1}{2}}\right), \quad k = 1, \dots, n,$$

and

$$\mathrm{v}^*(\tilde{t}, u) = \frac{\sinh(\tilde{t})}{\tilde{t}} \left[\frac{4\cosh(\tilde{t})u}{\zeta_1^2} - \frac{1}{\cosh(\tilde{t})}\right].$$

(B) Update $l = (l + 1)$ and $\boldsymbol{\theta}_*^{(l)}$.

(C) Repeat steps (A) and (B) until convergence of $\boldsymbol{\theta}_*^{(l)}$.

Step 7. Update $m = (m + 1)$ and $\boldsymbol{\theta}^{(m)} = \boldsymbol{\theta}_*^{(l)}$.

Step 8. Repeat steps 2, 3, 4, 5, 6 and 7 until convergence of $\boldsymbol{\theta}^{(m)}$.

The usual regularity conditions of large sample theory are fulfilled by all of the log-symmetric distributions listed above except for the log-Laplace distribution (i.e., the log-power-exponential distribution for $\zeta = 1$) (see Cordeiro *et al.* (2000)). Thus, the asymptotic distribution of the maximum likelihood estimator of $\boldsymbol{\theta}$ is the following:

$$\sqrt{n}\begin{pmatrix}\hat{\eta} - \eta \\ \hat{\phi} - \phi\end{pmatrix} \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}_2\left(\mathbf{0}; \begin{bmatrix}\phi\eta^2/d_g(\zeta) & 0 \\ 0 & 4\phi^2/[f_g(\zeta) - 1]\end{bmatrix}\right).$$

Hence, $\hat{\eta}$ and $\hat{\phi}$ are asymptotically independent.

2.5.1 Measuring goodness-of-fit

The goodness-of-fit is quantified using the following statistic, which is quite intuitive and has the advantage of graphical representation:

$$\Upsilon = n^{-1}\sum_{k=1}^{n}\left|\Phi^{-1}[F_T(\hat{\tilde{t}}^{(k)})] - \upsilon^{(k)}\right|,$$

where $F_T(\cdot)$ is the cumulative distribution function of T, $\hat{\tilde{t}}^{(k)}$ is the k-th order statistic of $\hat{\tilde{t}}$, $\upsilon^{(k)}$ is the expectation of the k-th order statistic for a random sample of size n of a standard normal distribution and $\Phi(\cdot)$ is the cumulative distribution function of a standard normal distribution. Note that $\Phi^{-1}[F_T(\tilde{t}^{(1)})], \dots, \Phi^{-1}[F_T(\tilde{t}^{(n)})]$ represent an ordered random sample from a standard normal distribution and $\hat{\eta}$ and $\hat{\phi}$ are a consistent estimators. Then, smaller values of Υ indicate better goodness-of-fit. Graphically, the criterion Υ indicates that the smaller the difference between the normal Q-Q plot of $\Phi^{-1}[F_T(\hat{\tilde{t}}^{(1)})], \dots, \Phi^{-1}[F_T(\hat{\tilde{t}}^{(n)})]$ and a straight line (with zero intercept and unit slope), the better the goodness-of-fit. One advantage of this criterion is that it allows for graphically evaluating the appropriateness or agreement with the data of the tails (heavier or lighter) and/or the unimodality/bimodality of the distribution postulated for T.

2.5.2 Choosing the extra parameter value

The density generator $g(\cdot)$ considered in the T distribution involves the extra parameter ζ, which is assumed to be known or fixed in the estimation process described above. This assumption ensures easy calculation of confidence regions and hypothesis testing for η and ϕ using Wald- and Rao-type statistics because the Fisher information matrix is diagonal. The motivation for this assumption also comes from the paper by Lucas (1997), which demonstrated that the robustness against outlying observations of Student-t models remains only if the degrees of freedom are fixed instead of estimated using the maximum likelihood method. In addition, Kano *et al.* (1993) (and references therein) reported difficulties in calculating the extra parameter using the maximum likelihood method for the power exponential and contaminated normal distributions. Thus, to consider a unified approach for log-symmetric distributions, it is proposed choosing the extra parameter value by minimizing the Υ statistic. In fact, if the estimator $\hat{\zeta} = \text{argmin}\Upsilon$ is consistent; and $d_g(\cdot)$ and $f_g(\cdot)$ are continuous functions, then the multivariate Slutsky's theorem allows one to demonstrate that the ζ value may be replaced with the value obtained by minimizing the Υ statistic without changing the asymptotic distribution of $\hat{\eta}$ and $\hat{\phi}$.

To investigate the performance of the proposed criterion, a simulation study is performed. First, a sample of size n is generated from a standard log-symmetric distribution. The resulting sample is used to estimate η and ϕ using the maximum likelihood method and to choose the extra parameter value by minimizing the Υ statistic. This process is replicated $R = 5000$ times. To consider different simulation scenarios, different log-symmetric distributions are used (i.e., the log-Student-t, log-power-exponential, log-hyperbolic, log-slash and Birnbaum-Saunders distributions) and the sample size is modified by considering n = 50, 100, 200, 400 and 800. As summary measures, the median (M) and interquartile range (IR) of the R chosen extra parameter values are considered. The results are presented in Table 2.1. It can be observed that in all scenarios, the median of the extra parameter values yielded by the proposed method tends to the true value as the size of the sample increases. Similarly, the variability around the median decreases as the size of the sample increases. These results indicate that for large sample sizes, the difference between the extra parameter value yielded by the proposed method and the true parameter value is *small*. Therefore, for large sample sizes, the inference on η and ϕ could be based on the asymptotic distribution described above even when the extra parameter value is unknown but has been chosen using the proposed method.

Table 2.1: *The median (M) and the interquartile range (IR) of the $R = 5000$ chosen extra parameter values by minimizing the Υ statistic.*

Distribution	ζ	$n = 50$		$n = 100$		$n = 200$		$n = 400$		$n = 800$	
		M	IR	M	IR	M	IR	M	IR	M	IR
Log-Student-t	2	1.80	1.10	1.90	0.70	1.95	0.50	2.00	0.40	2.00	0.20
	4	3.30	3.10	3.60	2.30	3.80	1.80	3.90	1.23	3.95	0.90
	6	4.60	6.20	5.20	4.80	5.50	3.70	5.80	2.60	5.90	2.60
	8	5.40	9.50	6.40	7.90	7.00	6.10	7.60	4.50	7.85	3.20
Log-power-exp.	0.1	0.17	0.39	0.14	0.33	0.12	0.23	0.11	0.17	0.10	0.12
	0.2	0.26	0.46	0.23	0.35	0.22	0.25	0.21	0.17	0.20	0.13
	0.3	0.32	0.46	0.31	0.34	0.31	0.26	0.30	0.19	0.30	0.13
	0.4	0.39	0.44	0.39	0.38	0.40	0.26	0.40	0.19	0.40	0.14
Log-hyperbolic	1.0	0.95	2.00	1.00	1.60	1.00	1.10	1.00	0.80	1.00	0.50
	1.1	1.01	2.20	1.03	1.50	1.08	1.10	1.10	0.80	1.10	0.60
	1.2	1.05	2.20	1.10	1.60	1.13	1.20	1.18	0.80	1.20	0.60
	1.3	1.10	2.20	1.15	1.70	1.20	1.10	1.25	0.90	1.30	0.60
Log-slash	0.8	0.75	0.30	0.78	0.21	0.79	0.16	0.81	0.12	0.80	0.08
	0.9	0.80	0.34	0.85	0.26	0.89	0.20	0.89	0.14	0.90	0.10
	1.0	0.89	0.39	0.92	0.33	0.98	0.24	0.99	0.16	1.00	0.13
	1.1	0.97	0.41	1.03	0.36	1.06	0.25	1.08	0.20	1.10	0.15
Birnbaum-Saunders	1.5	1.20	1.80	1.30	0.90	1.40	0.60	1.40	0.50	1.50	0.30
	2.0	1.70	1.70	1.80	1.10	1.90	0.80	2.00	0.50	2.00	0.30
	2.5	2.10	1.90	2.30	1.30	2.40	0.80	2.40	0.50	2.50	0.40
	3.0	2.75	2.20	2.85	1.40	2.95	1.00	2.95	0.60	3.00	0.40

2.6 Bayesian Inference

The method for inference about the interest parameters using the classical approach is based on the asymptotic properties of the maximum likelihood estimator. Therefore, for small sample sizes, such inference may be inadequate. Furthermore, some distributions, such as log-Laplace distribution, do not satisfy the usual regularity conditions; consequently, for such distributions, parameter inference cannot be performed using the standard method. Thus, this section considers inference using a Bayesian approach based on Markov chain Monte Carlo (MCMC) methods. One of the main advantages of Bayesian inference is that it is exact and available for any parametric model. For simplicity, it is supposed that η and ϕ are independent and have the following prior distributions:

$$\eta \sim \text{log-normal}(a_\eta, b_\eta) \quad \text{and} \quad \phi^{-1} \sim \text{Gamma}(c_\phi, d_\phi),$$

where $a_\eta > 0$, $b_\eta > 0$, $c_\phi > 0$ and $d_\phi > 0$ are assumed to be known. Next, we describe how samples are drawn from the posterior distribution of $\boldsymbol{\theta}$.

2.6.1 Log-normal, log-Student-t, log-slash, log-contaminated-normal, log-hyperbolic and log-Laplace distributions

One can sample from a joint posterior distribution of η and ϕ using Gibbs sampling (see, e.g., Gelfand and Smith , 1990), which involves successive sampling from the complete conditional densities. Because some distributions, such as the log-Student-t, log-slash, log-contaminated-normal, log-hyperbolic and log-Laplace distributions, may be obtained as a shape mixture of log-normal distributions, the algorithm can be easily developed using a data augmentation scheme, in which the complete conditional densities are known distributions. Next, a simple algorithm is described.

Algorithm 1.3

Step 1. Set the initial value of the parameter vector to $\boldsymbol{\theta}^{(0)}$.

Step 2. Based on $\boldsymbol{\theta}^{(l)}$ sample $\mathbf{u}^{(l+1)} = (u_1^{(l+1)}, \ldots, u_n^{(l+1)})$ independent as follows:

(A) Log-normal: $\mathrm{P}[u_k^{(l+1)} = 1] = 1$.

(B) Log-Student-t:

$$u_k|\eta^{(l)}, \phi^{(l)} \sim \text{Gamma}\left(\frac{\zeta+1}{2}, \frac{1}{2}\left(\left[\tilde{t}_k^{(l)}\right]^2 + \zeta\right)\right),$$

where $u \sim \text{Gamma}(a, b)$ represents a random variable with probability density function given by $f(u) \propto u^{a-1}\exp(-bu)$.

(D) Log-slash:

$$u_k|\eta^{(l)}, \phi^{(l)} \sim \text{TGamma}\left(\frac{2\zeta+1}{2}, \frac{1}{2}\left[\tilde{t}_k^{(l)}\right]^2; (0,1)\right),$$

where $\text{TGamma}(\cdot,\cdot;(0,1))$ represents a random variable with truncated gamma distribution within the interval $(0,1)$ (see, e.g., Nadarajah and Kotz, 2006).

(E) Log-contaminated-normal:

$$u_k|\eta^{(l)}, \phi^{(l)} \sim \begin{cases} \zeta_2 \ \text{with probability} \ p \propto \zeta_2^{\frac{1}{2}}\zeta_1 \exp\left(-\dfrac{\zeta_2}{2}\left[\tilde{t}_k^{(l)}\right]^2\right) \\ 1 \ \text{with probability} \ q \propto (1-\zeta_1)\exp\left(-\dfrac{1}{2}\left[\tilde{t}_k^{(l)}\right]^2\right) \end{cases}$$

(F) Log-Laplace:

$$u_k|\eta^{(l)}, \phi^{(l)} \sim \text{GIG}\left(\frac{1}{2}, \left[\tilde{t}_k^{(l)}\right]^2, \frac{1}{4}\right),$$

where $u \sim \text{GIG}(a, b, c)$ is a random variable with generalized Inverse Gaussian distribution (see, e.g., Hörmann and Leydold, 2013) and density function given by

$$f(u) \propto u^{a-1}\exp\left(-\frac{1}{2}[b/u + c\,u]\right).$$

(G) Log-hyperbolic:

$$u_k|\eta^{(l)}, \phi^{(l)} \sim \text{GIG}\left(\frac{1}{2}, \left[\tilde{t}_k^{(l)}\right]^2 + 1, \zeta^2\right).$$

Step 3. Based on $\mathbf{u}^{(l+1)}$ and $\phi^{(l)}$, sample $\eta^{(l+1)}$ as follows:

$$\eta|\mathbf{u}^{(l+1)}, \phi^{(l)} \sim \text{log-normal}\left(a^{(l+1)}, b^{(l+1)}\right), \qquad \text{where}$$

(A) Log-normal, log-Student-t, log-slash and log-contaminated-normal:

$$b^{(l+1)} = \left(\sum_{k=1}^{n} \frac{u_k^{(l+1)}}{\phi^{(l)}} + \frac{1}{b_\eta}\right)^{-1} \text{ and } a^{(l+1)} = \left[a_\eta^{\frac{\phi^{(l)}}{b_\eta}} \left(\prod_{k=1}^{n} t_k^{u_k^{(l+1)}}\right)\right]^{\frac{b^{(l+1)}}{\phi^{(l)}}}.$$

(B) Log-Laplace and log-hyperbolic:

$$b^{(l+1)} = \left(\sum_{k=1}^{n} \frac{1/u_k^{(l+1)}}{\phi^{(l)}} + \frac{1}{b_\eta}\right)^{-1} \text{ and } a^{(l+1)} = \left[a_\eta^{\frac{\phi^{(l)}}{b_\eta}} \left(\prod_{k=1}^{n} t_k^{1/u_k^{(l+1)}}\right)\right]^{\frac{b^{(l+1)}}{\phi^{(l)}}}.$$

Step 4. Based on $\mathbf{u}^{(l+1)}$ and $\eta^{(l+1)}$, sample $\phi^{(l+1)}$ as follows:

$$\phi^{-1}|\mathbf{u}^{(l+1)}, \eta^{(l+1)} \sim \text{Gamma}\left(\frac{n}{2} + c_\phi, d^{(l+1)}\right), \qquad \text{where}$$

(A) Log-normal, log-Student-t, log-slash and log-contaminated-normal:

$$d^{(l+1)} = \frac{1}{2} \sum_{k=1}^{n} \left[\log\left(\frac{t_k}{\eta^{(l+1)}}\right)\right]^2 u_k^{(l+1)} + d_\phi.$$

(B) Log-Laplace and log-hyperbolic:

$$d^{(l+1)} = \frac{1}{2} \sum_{k=1}^{n} \frac{1}{u_k^{(l+1)}} \left[\log\left(\frac{t_k}{\eta^{(l+1)}}\right)\right]^2 + d_\phi.$$

Step 5. Repeat steps 2, 3 and 4 until convergence is reached.

2.6.2 Harmonic law, Log-power-exponential ($-1 < \zeta < 1$), Birnbaum-Saunders and Birnbaum-Saunders-t distributions

For these distributions, the posterior conditional densities of η given ϕ and ϕ given η are unknown. Therefore, samples from the complete conditional densities are drawn using the Metropolis-Hastings method (see, e.g., Chib and Greenberg, 1995). Next, a simple algorithm is presented.

Algorithm 1.4

Step 1. Set the initial value of the parameter vector to $\boldsymbol{\theta}^{(0)}$.

Step 2. Based on $\boldsymbol{\theta}^{(l)}$ sample η^* from the log-normal($a^{(l+1)}, b^{(l+1)}$) distribution, where

$$b^{(l+1)} = \left(\frac{n}{\lambda\phi^{(l)}} + \frac{1}{b_\eta}\right)^{-1} \quad \text{and} \quad a^{(l+1)} = \left[a_\eta^{\frac{\phi^{(l)}}{b_\eta}} \left(\prod_{k=1}^{n} t_k\right)\right]^{\frac{b^{(l+1)}}{\lambda\phi^{(l)}}},$$

where $\lambda > 0$ is a tunning parameter. Then, a new value $\eta^{(l+1)} = \eta^*$ is accepted with probability

$$\min\left\{1, \frac{f_\eta(\eta^*|\phi^{(l)})}{f_\eta(\eta^{(l)}|\phi^{(l)})}\right\}, \qquad \text{where}$$

$$f_\eta(\eta|\phi) \propto \eta^{-1}\left[\prod_{k=1}^{n} g(\tilde{t}_k^2)\right]\exp\left\{-\frac{1}{2b_\eta}\left[\log\left(\frac{\eta}{a_\eta}\right)\right]^2\right\}.$$

Step 3. Based on $\eta^{(l+1)}$ and $\phi^{(l)}$ sample $1/\phi^*$ from the Gamma$(c^{(l+1)}, d^{(l+1)})$ distribution, where

$$c^{(l+1)} = \frac{n}{2} + c_\phi \quad \text{and} \quad d^{(l+1)} = \frac{1}{2\lambda}\sum_{k=1}^{n}\left[\log\left(\frac{t_k}{\eta^{(l+1)}}\right)\right]^2 + d_\phi.$$

Then, a new value $\phi^{(l+1)} = \phi^*$ is accepted with probability

$$\min\left\{1, \frac{f_\phi(\phi^*|\eta^{(l+1)})}{f_\phi(\phi^{(l)}|\eta^{(l+1)})}\right\}, \qquad \text{where}$$

$$f_\phi(\phi|\eta) \propto \left[\prod_{k=1}^{n} g(\tilde{t}_k^2)\right](1/\phi)^{\frac{n}{2}+c_\phi-1}\exp\left(-\frac{d_\phi}{\phi}\right).$$

Step 4. Repeat steps 2 and 3 until convergence is reached.

The value of the tuning parameter in the *Algorithm 1.4* may be set to $\lambda = \mathrm{Var}(\log(T^*))$, where T^* exhibits a standard log-symmetric distribution.

2.6.3 Unknown extra parameter

When the extra parameter ζ is unknown, an additional *step* may be included in the algorithms described above to draw samples from the posterior conditional distribution of ζ given $\boldsymbol{\theta}$; it is assumed that ζ and $\boldsymbol{\theta}$ have independent prior distributions. For instance,

(A) Log-hyperbolic distribution:

ζ is assumed to exhibit the log-normal(c_ζ, d_ζ) distribution. Thus, ζ^* is sampled from the log-normal$(\zeta^{(l)}, \lambda)$ distribution, where $\lambda > 0$ is a tunning parameter. A new value $\zeta^{(l+1)} = \zeta^*$ is accepted with probability given by

$$\min\left\{1, \frac{f_\zeta(\zeta^*|\mathbf{u}^{(l+1)})}{f_\zeta(\zeta^{(l)}|\mathbf{u}^{(l+1)})}\right\}, \qquad \text{where}$$

$$f_\zeta(\zeta|\mathbf{u}) \propto [\mathrm{K}_1(\sqrt{\zeta})]^{-n}\exp\left(-\zeta\left[\frac{1}{2}\sum_{k=1}^{n} u_k\right] - \frac{1}{2d_\zeta}\left[\log\left(\frac{\zeta}{c_\zeta}\right)\right]^2\right).$$

(B) Log-slash distribution:

ζ is assumed to exhibit the Gamma(c_ζ, d_ζ) distribution. Then, based on $\mathbf{u}^{(l+1)}$, $\zeta^{(l+1)}$ is sampled from the Gamma$(c_\zeta^{(l+1)}, d_\zeta^{(l+1)})$ distribution, where $c_\zeta^{(l+1)} = c_\zeta + n$ and $d_\zeta^{(l+1)} = -\sum_{k=1}^{n}\log(u_k^{(l+1)}) + d_\zeta$.

(C) Log-contaminated-normal distribution:

$\zeta_1 \sim \mathrm{Beta}(c_{\zeta_1}, d_{\zeta_1})$ and $\zeta_2 \sim \mathrm{TGamma}(c_{\zeta_2}, d_{\zeta_2};(0,1))$ are independent. Then, based on $\mathbf{u}^{(l+1)}$, $\eta^{(l+1)}$ and $\phi^{(l+1)}$, $\zeta_2^{(l+1)}$ and $\zeta_1^{(l+1)}$ are sampled from the TGamma$(c_{\zeta_2}^{(l+1)}, d_{\zeta_2}^{(l+1)};(0,1))$ and Beta$(c_{\zeta_1}^{(l+1)}, d_{\zeta_1}^{(l+1)})$ distributions, where $c_{\zeta_1}^{(l+1)} = c_{\zeta_1} + \sum_{k=1}^{n}\mathrm{I}(u_k^{(l+1)} = \zeta_2)$, $d_{\zeta_1}^{(l+1)} = d_{\zeta_1} + \sum_{k=1}^{n}\mathrm{I}(u_k^{(l+1)} = 1)$, $c_{\zeta_2}^{(l+1)} = c_{\zeta_2} + \frac{1}{2}\sum_{k=1}^{n}\mathrm{I}(u_k^{(l+1)} = \zeta_2)$ and $d_{\zeta_2}^{(l+1)} = d_{\zeta_2} + \frac{1}{2}\sum_{k=1}^{n}[\tilde{t}_k^{(l+1)}]^2\mathrm{I}(u_k^{(l+1)} = \zeta_2)$.

(D) Birnbaum-Saunders distribution:

$1/\zeta^2$ is assumed to exhibit the Gamma(c_ζ, d_ζ) distribution. Based on $\eta^{(l+1)}$ and $\phi^{(l+1)}$, $1/[\zeta^{(l+1)}]^2$ is sampled from the Gamma$(c_\zeta^{(l+1)}, d_\zeta^{(l+1)})$ distribution, where $c_\zeta^{(l+1)} = c_\zeta + \frac{n}{2}$ and $d_\zeta^{(l+1)} = 2\sum_{k=1}^{n} \sinh^2(\tilde{t}_k^{(l+1)}) + d_\zeta$.

Maximum likelihood estimates for the same family of distributions may be used as initial values for *Algorithm 1.3* and *Algorithm 1.4*. Inferences about the parameters or functions of them are available from the approximate posterior marginal density. For example, it is possible summarize the simulated posterior distribution of η and ϕ by computing the summary statistics (i.e., the posterior means, medians, and standard deviations) and credible intervals. In the case of non-informative priors, comparisons with the maximum likelihood approach may be performed.

2.7 Applications

2.7.1 Gross domestic product

Gross domestic product (GDP) divided by midyear population is known as the per capita GDP. The per capita GDP is most likely the best measure of a country's overall well being. The GDP is the sum of the gross value added by all resident producers in the economy and any product taxes and minus any subsidies that are not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources. The data set considered here corresponds to the per capita GDP (current US$) of 190 countries during 2010, and it was downloaded from the World Bank's DataBank website (`http://databank.worldbank.org/data/`).

Maximum likelihood estimation

Table 2.2: *Values of $-2\mathsf{L}(\hat{\boldsymbol{\theta}})$, AIC and BIC for the fitted distributions to the GDP data.*

	Log normal	Birnbaum Saunders	Log skew-t	Box Cox-t	General. Modified Weibull	General. Gamma	General. Inverse Gaussian
$-2\mathsf{L}(\hat{\boldsymbol{\theta}})$	3919.61	**3902.82**	3919.58	3919.58	3920.06	3921.62	3905.74
AIC	3923.60	**3906.82**	3927.58	3927.58	3928.06	3927.62	3911.75
BIC	3930.10	**3913.31**	3940.57	3940.57	3941.05	3937.36	3921.49

Table 2.2 lists the goodness-of-fit statistics $-2\mathsf{L}(\hat{\boldsymbol{\theta}})$, AIC (Akaike, 1973) and BIC (Schwarz, 1978) for the log-normal, Birnbaum-Saunders($\zeta = 2.2$), log-skew-t, Box-cox-t (see e.g. Rigby and Stasinopoulos , 2006), generalized modified Weibull, generalized Gamma and generalized Inverse Gaussian distributions fitted to the GDP data. The extra parameter ζ of the Birnbaum-Saunders distribution was chosen by minimizing the criterion Υ, as illustrated in Figure 2.2(a).

The Birnbaum-Saunders($\zeta = 2.2$) distribution has the lowest $-2\mathsf{L}(\hat{\boldsymbol{\theta}})$, AIC and BIC values among all the fitted models; thus, it could be considered to be the best model in the sense of these information measures. Figure 2.2(b) shows a plot of $\Phi^{-1}[F_T(\hat{\tilde{t}}^{(k)})]$ versus $\upsilon^{(k)}$ for the fitted Birnbaum-Saunders distribution; this plot indicates that the distribution describes the data adequately. A plot of the Birnbaum-Saunders($\zeta = 2.2$) density distribution is shown in Figure 2.3(a) (together with the data histogram). Similarly, Figure 2.3(b) presents the empirical cumulative distribution function of the per capita GDP and the cumulative distribution function of the Birnbaum-Saunders ($\zeta = 2.2$) model. The maximum likelihood estimates (and the corresponding approximate standard errors, which are given in parentheses) of the model parameters of the fitted Birnbaum-Saunders($\zeta = 2.2$) distribution are

$$\hat{\eta} = 4891.135(427.07) \quad \text{and} \quad \hat{\phi} = 3.187(0.21).$$

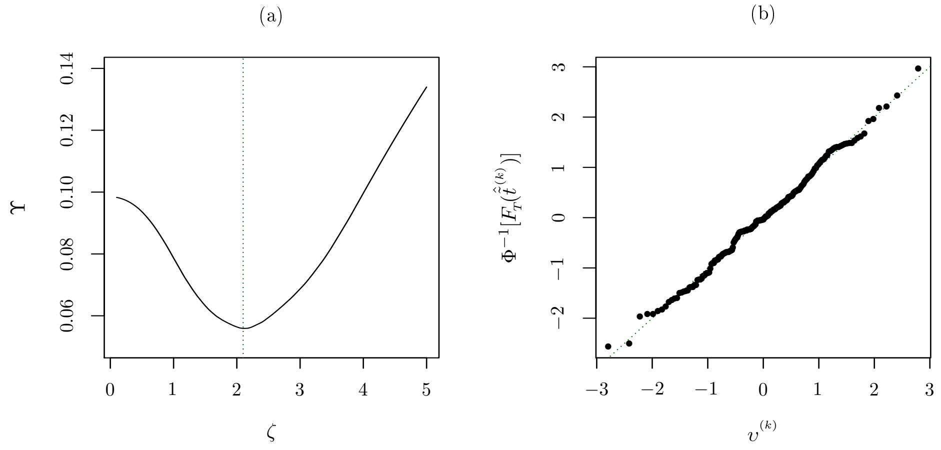

Figure 2.2: (a) *Graph of* Υ *under the Birnbaum-Saunders distribution;* (b) *plot of* $\Phi^{-1}[F_T(\hat{\hat{t}}^{(k)})]$ *versus* $v^{(k)}$ *under the Birnbaum-Saunders(*$\zeta = 2.2$*) distribution fitted to the GDP data.*

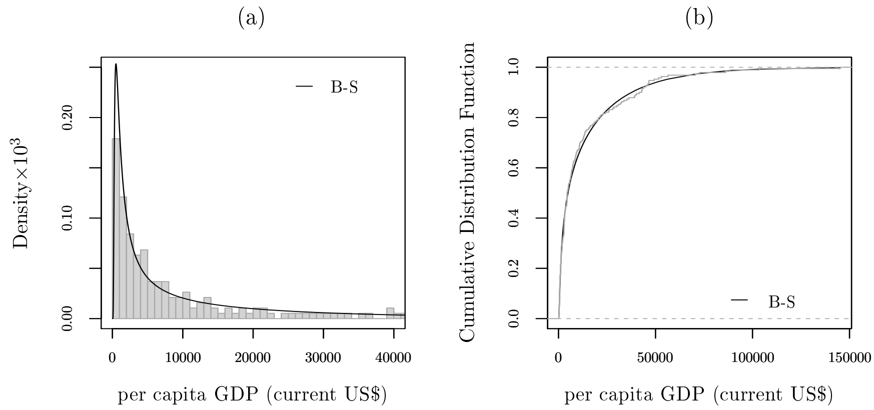

Figure 2.3: (a) *Histogram and* (b) *empirical cumulative distribution function of per capita GDP (current US$) of 190 countries during 2010.*

Because the Birnbaum-Saunders($\zeta = 2.2$) distribution was identified as the best model, and from the properties (P3) and (P4) described in Section 2.2, one can conclude that the probability distribution of any macroeconomic indicator that can be expressed as $c_1 T^{c_2}$ also belongs to the log-symmetric class, where T represents the per capita GDP during 2010 and $c_1 > 0$ and $c_2 \neq 0$ are known constants. The Birnbaum-Saunders($\zeta = 2.2$) distribution was also fitted to the per capita GDP for 2009; $\hat{\eta} = 4823.88(424.96)$ and $\hat{\phi} = 3.313(0.22)$ were obtained. Then, ignoring the variability associated with the point estimates of η and ϕ it may be concluded that the median of the per capita GDP distribution increased in 2010, whereas in the same year, the skewness and relative dispersion of the per capita GDP distribution decreased. Similarly, the modes of the per capita GDP distributions were US$ 466.056 and US$ 500.174 in 2009 and 2010, respectively.

2.7.2 Bayesian inference

It is consider the prior distributions described in Section 2.6 with hyperparameters fixed as follows: $a_\eta = 1$, $b_\eta = 10000$, $c_\phi = 0.0001$, $d_\phi = 0.0001$, $c_\zeta = 0.0001$, and $d_\zeta = 0.0001$. This setup allows for comparisons with the maximum likelihood approach. One chain of size 110000 for each parameter was simulated, and the first 10000 iterations were discarded to eliminate the effect of initial values. To avoid correlation, a spacing of size 10 was used, thereby obtaining an effective sample of size 10000. Table 2.3 lists the summary statistics of the posterior distribution and the 95% credible interval for the parameters of the log-normal and Birnbaum-Saunders models fitted to the GDP data. The statistic DIC (see, e.g., Gelman *et al.*, 2004) presented in Table 2.3 indicates that the Birnbaum-Saunders model describes the data better than the log-normal model. The inferential results are very similar to the results obtained using the maximum likelihood approach. Figure 2.6 displays the history of the chains and the approximate posterior marginal densities of the parameters η and ϕ for the Birnbaum-Saunders model.

Table 2.3: *Posterior mean, median, standard deviation (SD) and 95% credible interval for parameters of the log-normal and Birnbaum-Saunders distributions fitted to GDP data.*

	Log-normal					
	Mean	Median	SD	2.5%	97.5%	DIC
η	4843.55	4808.20	541.03	3870.75	5978.52	3923.61
ϕ	2.35	2.33	0.25	1.92	2.88	
	Birnbaum-Saunders					
	Mean	Median	SD	2.5%	97.5%	DIC
η	4837.86	4821.61	362.48	4181.02	5573.93	3910.80
ϕ	3.25	3.23	0.30	2.70	3.90	
ζ	2.20	2.19	0.19	1.84	2.60	

2.7.3 Body fat percentage

This data set, collected from the Australian Institute of Sport, consists of 13 variables measured on 102 male and 100 female athletes, and it was downloaded from the website `http://www.statsci.or` The main objective of this analysis is to describe the distribution of the body fat percentage of athletes. This distribution is expected to be bimodal because the data set comprises of male and female athletes.

Table 2.4 lists the goodness-of-fit statistics $-2\mathsf{L}(\hat{\boldsymbol{\theta}})$, AIC and BIC for the log-normal, Birnbaum-Saunders-$t(\zeta = (4.5, 4)^\top)$, log-skew-t, Box-cox-t, generalized modified Weibull, generalized Gamma and generalized Inverse Gaussian distributions fitted to the body fat percentage of athletes. The extra parameter ζ of the Birnbaum-Saunders-t distribution was chosen by minimizing the criterion Υ, as illustrated in Figure 2.5(a). The Birnbaum-Saunders-$t(\zeta = (4.5, 4)^\top)$ distribution has the lowest $-2\mathsf{L}(\hat{\boldsymbol{\theta}})$, AIC and BIC values among all the fitted models; thus, it could be considered to be the best model. Figure 2.5(b) shows a plot of $\Phi^{-1}[F_T(\hat{\tilde{t}}^{(k)})]$ versus $\upsilon^{(k)}$ for the fitted Birnbaum-Saunders-t distribution; this plot indicates that the distribution describes the data adequately. A plot of the Birnbaum-Saunders-$t(\zeta = (4.5, 4)^\top)$ density distribution is shown in Figure 2.6(a) (together with the data histogram). Similarly, Figure 2.6(b) presents the empirical cumulative distribution function of the body fat percentage and the cumulative distribution function of the Birnbaum-Saunders-$t(\zeta = (4.5, 4)^\top)$ model. The maximum likelihood estimates (and the corresponding approximate standard errors, which are given in parentheses) of the model parameters of the fitted Birnbaum-Saunders-$t(\zeta = (4.5, 4)^\top)$ distribution are

$$\hat{\eta} = 12.467(0.278) \quad \text{and} \quad \hat{\phi} = 0.09116(0.0058).$$

Therefore, the median and the modes of the distribution of body fat percentage are 12.46%, 7.63% and 17.31%, respectively.

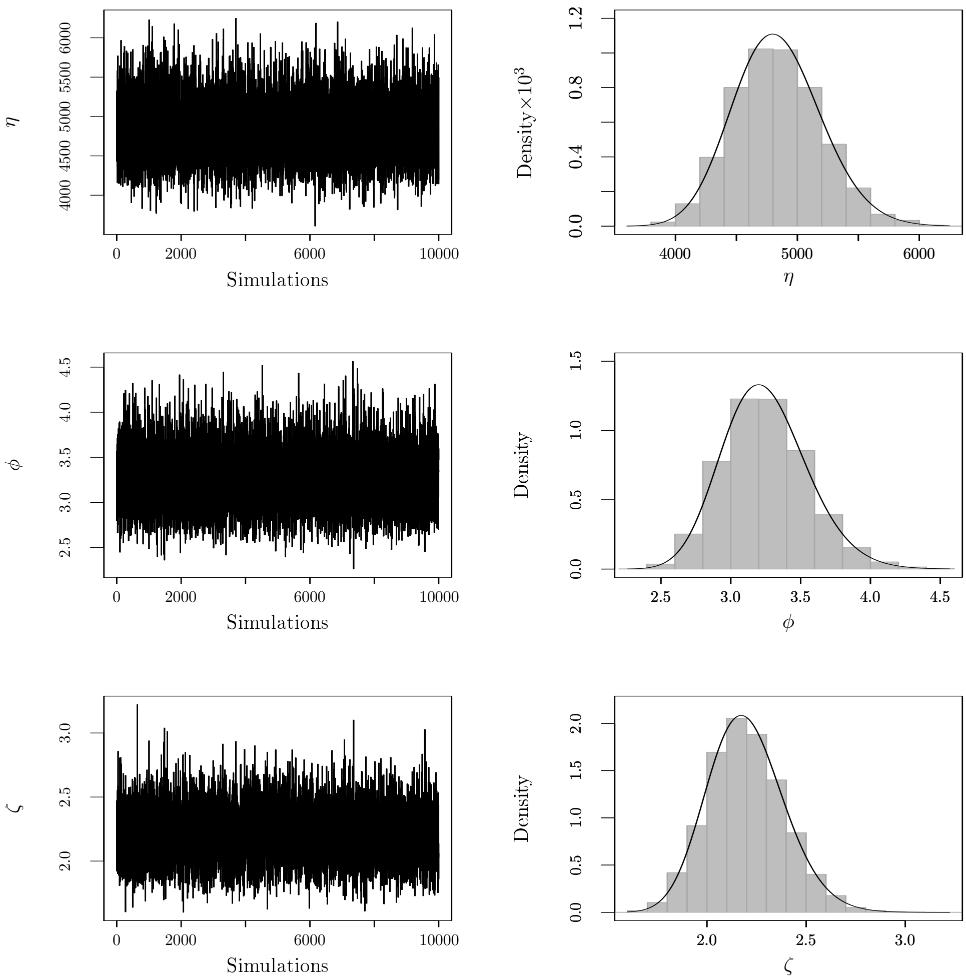

Figure 2.4: *History of the chains and the approximate posterior marginal densities of η, ϕ and ζ for the Birnbaum-Saunders distribution fitted to GDP data.*

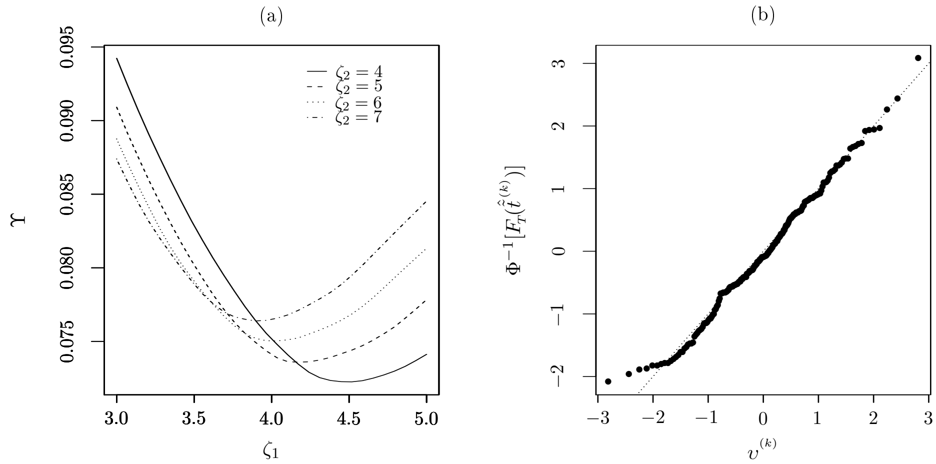

Figure 2.5: *(a) Graph of Υ under the Birnbaum-Saunders-t distribution; (b) plot of $\Phi^{-1}[F_T(\hat{\hat{t}}^{(k)})]$ versus $\upsilon^{(k)}$ under the Birnbaum-Saunders-t($\zeta = (4.5, 4)^\top$) distribution fitted to the data of body fat.*

Table 2.4: *Values of $-2\mathsf{L}(\hat{\boldsymbol{\theta}})$, AIC and BIC for the fitted distributions to the body fat percentage of athletes.*

	Log normal	Birnbaum Saunders-t	Log skew-t	Box Cox-t	General. Modified Weibull	General. Gamma	General. Inverse Gaussian
$-2\mathsf{L}(\hat{\boldsymbol{\theta}})$	1262.24	**1224.38**	1255.15	1261.21	1260.68	1264.32	1269.02
AIC	1266.24	**1228.38**	1263.15	1269.21	1268.68	1270.32	1275.02
BIC	1272.86	**1235.00**	1276.39	1282.44	1281.92	1280.24	1284.94

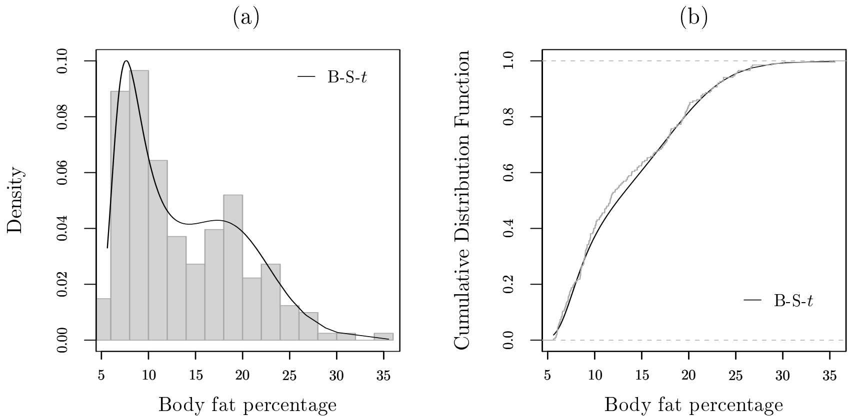

Figure 2.6: *(a) Histogram and (b) empirical cumulative distribution function of the body fat percentage of athletes.*

CHAPTER 3

LOG-SYMMETRIC REGRESSION MODELS

Linear and nonlinear regression models are commonly applied in areas such as Biology, Chemistry, Medicine, Economics and Engineering. The analysis based on models under normal errors and constant variance is the most popular when the variable of interest is continuous due to desirable statistical properties and a comprehensively developed theory. Nevertheless, the application of such models may be inadequate in certain scenarios commonly found in practice. For instance, as shown by Vanegas and Paula (2015a), ignoring the skewness of the response variable distribution may introduce biases into the parameter estimates and/or the estimation of the associated variability measures. To address this problem, certain proposals have been made in the literature to replace the normality assumption by more flexible classes of distributions. For example, in the context of asymmetric and heavy-tailed responses, Lin *et al.* (2009) derived diagnostic methods in nonlinear skew-t-normal regression models; Cancho *et al.* (2010) studied nonlinear skew-normal regression models using classical and Bayesian approaches; Lachos *et al.* (2011) introduced heteroscedastic nonlinear regression models based on scale mixtures of skew-normal distributions; and Labra *et al.* (2012) derived diagnostic methods for the class of regression models introduced previously by Lachos *et al.* (2011).

Although the models studied in these papers are attractive, they have certain limitations, for instance, modeling the mean instead of the median, assuming that the skewness parameter is constant across the observations and do not admit the presence of nonparametric effects. Therefore, this work provides a unified theoretical framework for semi-parametric regression analysis based on log-normal, log-Student-t, Birnbaum-Saunders, Birnbaum-Saunders-t, harmonic law and other right-skewed and strictly positive distributions, in which both, the median and the skewness (or the relative dispersion) of the response variable distribution are explicitly modeled. In this setup, here termed log-symmetric regression models, both the median and the skewness (or the relative dispersion) are described using semi-parametric functions of explanatory variables, in which their nonparametric components are approximated by natural cubic splines (see, e.g., Green and Silverman, 1994; Lancaster and Salkauskas, 1986) or P-splines (Eilers and Marx , 1996). The flexibility provided by the systematic component under this model lies in its capacity to relate the distribution of the response variable with a set of covariates using a sum of arbitrary functions, whose functional forms are estimated from the data. Under this approach, the parameter interpretation is based on the multiplicative effects of the covariates acting on the median and the skewness (or the relative dispersion) of the distribution of the response variable. However, if the skewness (or the relative dispersion) is specified to be constant, the multiplicative effect of the covariates directly affects the quantiles (of any order) of the distribution of the response variable, which turns the parameter interpretation simpler. Moreover, some of the log-symmetric distributions exhibit heavier tails than those of the log-normal one, which allows to estimate the model parameters in a robust manner under the presence of extreme or outlying observations.

In the context of nonparametric and semi-parametric models, it is possible to cite some of the most important contributions. For example, Hastie and Tibshirani (1990) introduced the class of generalized additive models, and Rigby and Stasinopoulos (2005) introduced the generalized additive models for location, scale and shape (GAMLSS), which address the semi-parametric mixed joint modeling of all parameters in a general class of distributions. Rigby and Stasinopoulos (2006) illustrated the use of semi-parametric models based on Box-cox-t distribution. Wood (2006) studied generalized additive models using the R statistical package to illustrate their applications. Recently, Ibacache-Pulgar *et al.* (2013) derived diagnostic tools in symmetric homoscedastic semi-parametric models; Wu and Yu (2014) studied quantile regression based on both, partially linear single-index models and partially linear additive models.

In this chapter, the log-symmetric regression models are introduced. A detailed description of the parameter estimation process is also provided, which combines Fisher scoring, backfitting and expectation-maximization (EM) algorithms. Discussions on effective degrees of freedom estimation and simultaneous confidence intervals for nonparametric components are given, and some diagnostic procedures, such as deviance-type residuals and local influence measures, are derived. The asymptotic behaviour of the maximum penalized likelihood estimator is studied under a fixed-knot assumption. A simulation study, which concerns with the behavior of the maximum penalized likelihood estimates in log-symmetric regression models, is also presented.

3.1 Formulation of the model

Let $t_1, \ldots, t_n$ be measurements of a quantitative interest characteristic performed on n subjects or experimental units, which are assumed to be realizations of n independent random variables $T_1, \ldots, T_n$ whose distribution is continuous, strictly positive, right-skewed and heavy/light-tailed. Thus, T_k is assumed to be obtained as

$$T_k = \eta_k \, \xi_k^{\sqrt{\phi_k}}, \qquad k = 1, \ldots, n, \tag{3.1}$$

where $\eta_k > 0$ and $\phi_k > 0$ represent the median and the skewness (or the relative dispersion), respectively, of the T_k distribution, whereas $\xi_1, \ldots, \xi_n$ is a set of independent and multiplicative random errors that exhibit a standard log-symmetric distribution with the extra parameter (or extra parameter vector) ζ, whose probability density function is given by

$$f_{\xi_k}(\xi; g(\cdot)) = \frac{1}{\xi}\, g\left\{[\log(\xi)]^2\right\}, \qquad \xi > 0, \tag{3.2}$$

for some function $g(u)$, where $g(u) > 0$ for $u > 0$ and $\int_0^\infty u^{-\frac{1}{2}} g(u) \partial u = 1$. Furthermore, the quantile of order $0 < w < 1$ of T_k can be written as

$$\vartheta_{T^*}(w) = \eta_k \, [\vartheta_\xi(w)]^{\sqrt{\phi_k}},$$

where $\vartheta_\xi(w)$ is the quantile of order w of ξ_k. The class of distributions for the model error includes the standardized versions of the log-normal, log-Student-$t(\zeta)$, log-power-exponential (ζ), log-contaminated-normal$(\zeta = (\zeta_1, \zeta_2)^\top)$, log-hyperbolic$(\zeta)$, log-slash$(\zeta)$, Birnbaum-Saunders$(\zeta)$, Birnbaum-Saunders-$t(\zeta = (\zeta_1, \zeta_2)^\top)$ and harmonic law(ζ) distributions. From the statistical properties of the class of distributions described in (4.2), one may conclude that the distribution of T belongs to the same family compared with the model error. The distribution of T is flexible enough to include bimodal distributions as special cases (e.g., Birnbaum-Saunders and Birnbaum-Saunders-t distributions for some combinations of ϕ and ζ) and distributions that have heavier (e.g., log-Student-t, log-slash, log-contaminated-normal, log-power-exponential for $0 < \zeta \le 1$ and log-hyperbolic for small ζ) and lighter (e.g., log-power-exponential for $-1 < \zeta < 0$ and log-hyperbolic for large ζ) tails than those of the log-normal distribution.

Otherwise, it is assumed that η_k and ϕ_k are related to explanatory variables by either of the following setups:

$$(I) \quad \begin{cases} \eta_k = \eta(\mathbf{x}_k, \boldsymbol{\beta}) \quad \text{and} \\ \log(\phi_k) = \mathbf{w}_k^\top \boldsymbol{\gamma} + \sum_{r=1}^{q'} \mathrm{f}_{\phi_r}(b_{k,r}); \end{cases} \qquad (II) \quad \begin{cases} \log(\eta_k) = \mathbf{x}_k^\top \boldsymbol{\beta} + \sum_{j=1}^{p'} \mathrm{f}_{\eta_j}(a_{k,j}) \quad \text{and} \\ \log(\phi_k) = \mathbf{w}_k^\top \boldsymbol{\gamma} + \sum_{r=1}^{q'} \mathrm{f}_{\phi_r}(b_{k,r}), \end{cases} \tag{3.3}$$

where $\mathbf{x}_k^* = (\mathbf{x}_k^\top, a_{k,1}, \ldots, a_{k,p'})^\top$ and $\mathbf{w}_k^* = (\mathbf{w}_k^\top, b_{k,1}, \ldots, b_{k,q'})^\top$ are vectors of explanatory variables for η_k and ϕ_k, respectively; $\log[\eta(\mathbf{x}_k, \cdot)]$ is a continuous and twice differentiable function for all $\boldsymbol{\beta} \in \Omega_\beta$; $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_q)^\top$ are vectors of unknown parameters; $\mathrm{f}_{\eta_j}(a)$ $(j = 1, \ldots, p')$ and $\mathrm{f}_{\phi_r}(b)$ $(r = 1, \ldots, q')$ are continuous, smooth and nonparametric functions of the quantitative explanatory variables a and b, respectively; and Ω_β is a compact set with interior points. The matrices $\mathbf{X} = (\mathbf{x}_1^*, \ldots, \mathbf{x}_n^*)^\top$, $\mathbf{W} = (\mathbf{w}_1^*, \ldots, \mathbf{w}_n^*)^\top$ and $\mathbf{D}_\beta = \partial\boldsymbol{\eta}/\partial\boldsymbol{\beta}^\top$ are assumed to be of full column rank for all $\boldsymbol{\beta} \in \Omega_\beta$, where $\boldsymbol{\eta} = (\eta(\mathbf{x}_1, \boldsymbol{\beta}), \ldots, \eta(\mathbf{x}_n, \boldsymbol{\beta}))^\top$. The functions $\mathrm{f}_{\eta_j}(a)$ $(j = 1, \ldots, p')$ and $\mathrm{f}_{\phi_r}(b)$ $(r = 1, \ldots, q')$ are approximated by using natural cubic splines (see, e.g., Green and Silverman, 1994; Lancaster and Salkauskas, 1986, sections 4.6 and 4.7) or P-splines (Eilers and Marx, 1996).

3.1.1 Natural cubic splines

Let $\min(\mathbf{a}_j)$, $\max(\mathbf{a}_j)$, $\min(\mathbf{b}_r)$ and $\max(\mathbf{b}_r)$ be the minimum and the maximum values of $(a_{1,j}, \ldots, a_{n,j})$ and $(b_{1,r}, \ldots, b_{n,r})$, respectively. From two sets of pre-selected knots such that $\min(\mathbf{a}_j) = a_{j(1)} < a_{j(2)} < \ldots < a_{j(p_j')} = \max(\mathbf{a}_j)$ and $\min(\mathbf{b}_r) = b_{r(1)} < b_{r(2)} < \ldots < b_{r(q_r')} = \max(\mathbf{b}_r)$, the values of $\mathrm{f}_{\eta_j}(a_{1,j}), \ldots, \mathrm{f}_{\eta_j}(a_{n,j})$ and $\mathrm{f}_{\phi_r}(b_{1,r}), \ldots, \mathrm{f}_{\phi_r}(b_{n,r})$ can be approximated by

$$(\mathrm{f}_{\eta_j}(a_{1,j}), \ldots, \mathrm{f}_{\eta_j}(a_{n,j}))^\top = \dot{\mathbf{N}}_{\eta_j} \dot{\boldsymbol{\tau}}_{\eta_j} \quad \text{and} \quad (\mathrm{f}_{\phi_r}(b_{1,r}), \ldots, \mathrm{f}_{\phi_r}(b_{n,r}))^\top = \dot{\mathbf{N}}_{\phi_r} \dot{\boldsymbol{\tau}}_{\phi_r},$$

where $\dot{\mathbf{N}}_{\eta_j}$ and $\dot{\mathbf{N}}_{\phi_r}$ are basis matrices; $\dot{\boldsymbol{\tau}}_{\eta_j} = (\dot{\tau}_{j,1}^{\eta}, \ldots, \dot{\tau}_{j,p_j'}^{\eta})^\top$ and $\dot{\boldsymbol{\tau}}_{\phi_r} = (\dot{\tau}_{r,1}^{\phi}, \ldots, \dot{\tau}_{r,q_r'}^{\phi})^\top$ are vectors of unknown parameters; p_j' and q_r' are the sizes of the knot vectors, which satisfy $3 \le p_j' \le \bar{p}_j'$ and $3 \le q_r' \le \bar{q}_r'$, where $\bar{p}_j'$ and $\bar{q}_r'$ are the number of different values in $(a_{1,j}, \ldots, a_{n,j})$ and $(b_{1,r}, \ldots, b_{n,r})$, respectively. For $i = 1, \ldots, p_j' - 1$ define $h_i = a_{j(i+1)} - a_{j(i)}$. Then, the approximation of $\mathrm{f}_{\eta_j}(a)$ via natural cubic splines can be written as

$$\begin{aligned} \mathrm{f}_{\eta_j}(a) = {} & \frac{(a - a_{j(i)})}{h_i} \mathrm{f}_{\eta_j}[a_{j(i+1)}] + \frac{(a_{j(i+1)} - a)}{h_i} \mathrm{f}_{\eta_j}[a_{j(i)}] \\ & - \frac{1}{6}(a - a_{j(i)})(a_{j(i+1)} - a) \left\{ \left(1 + \frac{a - a_{j(i)}}{h_i}\right) \mathrm{f}''_{\eta_j}[a_{j(i+1)}] + \left(1 + \frac{a_{j(i+1)} - a}{h_i}\right) \mathrm{f}''_{\eta_j}[a_{j(i)}] \right\}, \end{aligned}$$

for $a_{j(i)} \le a \le a_{j(i+1)}$, $i = 1, \ldots, p_j' - 1$. In addition, $\mathrm{f}_{\eta_j}(a)$ is a twice-differentiable function on $[a_{j(1)}, a_{j(p_j')}]$, $\mathrm{f}''_{\eta_j}[a_{j(1)}] = \mathrm{f}''_{\eta_j}[a_{j(p_j')}] = 0$ and $\dot{\mathbf{N}}_{\eta_j} \mathbf{1}_{p_j'} = \mathbf{1}_n$. To avoid overfitting, the approximations of $\mathrm{f}_{\eta_j}(\cdot)$ and $\mathrm{f}_{\phi_r}(\cdot)$ via natural cubic splines consider the following penalty terms:

$$-\frac{\lambda_{\eta_j}}{2} \int \left[\mathrm{f}''_{\eta_j}(t)\right]^2 \partial t = -\frac{\lambda_{\eta_j}}{2} \dot{\boldsymbol{\tau}}_{\eta_j}^\top \dot{\mathbf{M}}_{\eta} \dot{\boldsymbol{\tau}}_{\eta_j} \quad \text{and} \quad -\frac{\lambda_{\phi_r}}{2} \int \left[\mathrm{f}''_{\phi_r}(t)\right]^2 \partial t = -\frac{\lambda_{\phi_r}}{2} \dot{\boldsymbol{\tau}}_{\phi_r}^\top \dot{\mathbf{M}}_{\phi_r} \dot{\boldsymbol{\tau}}_{\phi_r},$$

respectively, where $\lambda_{\eta_j} > 0$ and $\lambda_{\phi_r} > 0$ are smoothing parameters and $\dot{\mathbf{M}}_{\eta_j}$ and $\dot{\mathbf{M}}_{\phi_r}$ are symmetric and nonnegative definite matrices. The structures of $\dot{\mathbf{N}}_{\eta_j}$, $\dot{\mathbf{N}}_{\phi_r}$, $\dot{\mathbf{M}}_{\eta_j}$ and $\dot{\mathbf{M}}_{\phi_r}$ do not depend on $\dot{\boldsymbol{\tau}}_{\eta_j}$ or $\dot{\boldsymbol{\tau}}_{\phi_r}$, and their explicit forms may be found in Appendix A. If the set of knots coincide with the set of different values of the explanatory variable, then this approach becomes the topic studied by Green and Silverman (1994).

3.1.2 P-splines

Let $\tilde{p}_j'$ and $\tilde{q}_r'$ be the sizes of the internal knot vectors for $\mathrm{f}_{\eta_j}(\cdot)$ and $\mathrm{f}_{\phi_r}(\cdot)$, respectively. From two sets of pre-selected knots, $a_{j(1)} < \ldots a_{j(1+d_j^\eta)} < \ldots a_{j(\tilde{p}_j'+d_j^\eta)} < \ldots a_{j(\tilde{p}_j'+2d_j^\eta)}$ and $b_{r(1)} < \ldots b_{r(1+d_r^\phi)} < \ldots b_{j(\tilde{q}_r'+d_r^\phi)} < \ldots b_{r(\tilde{q}_r'+2d_r^\phi)}$, the values of $\mathrm{f}_{\eta_j}(a_{1j}), \ldots, \mathrm{f}_{\eta_j}(a_{nj})$ and $\mathrm{f}_{\phi_r}(b_{1r}), \ldots, \mathrm{f}_{\phi_r}(b_{nr})$ can be approximated by

$$(\mathrm{f}_{\eta_j}(a_{1,j}), \ldots, \mathrm{f}_{\eta_j}(a_{n,j}))^\top = \dot{\mathbf{N}}_{\eta_j}\dot{\boldsymbol{\tau}}_{\eta_j} \quad \text{and} \quad (\mathrm{f}_{\phi_r}(b_{1,r}), \ldots, \mathrm{f}_{\phi_r}(b_{n,r}))^\top = \dot{\mathbf{N}}_{\phi_r}\dot{\boldsymbol{\tau}}_{\phi_r},$$

where $a_{j(1+d_j^\eta)} = \min(\mathbf{a}_j)$, $a_{j(\tilde{p}_j'+d_j^\eta)} = \max(\mathbf{a}_j)$, $b_{r(1+d_r^\phi)} = \min(\mathbf{b}_r)$, and $b_{r(\tilde{q}_r'+d_r^\phi)} = \max(\mathbf{b}_r)$; $\dot{\mathbf{N}}_{\eta_j}$ and $\dot{\mathbf{N}}_{\phi_r}$ are B-splines basis matrices of degree $d_j^\eta \leq 3$ and $d_r^\phi \leq 3$, respectively (see Boor , 1978); $\dot{\boldsymbol{\tau}}_{\eta_j} = (\dot{\tau}_{j,1}^\eta, \ldots, \dot{\tau}_{j,p_j'}^\eta)^\top$ and $\dot{\boldsymbol{\tau}}_{\phi_r} = (\dot{\tau}_{r,1}^\phi, \ldots, \dot{\tau}_{r,q_r'}^\phi)^\top$ are vectors of unknown parameters of dimension $p_j' = \tilde{p}_j' + d_j^\eta - 1$ and $q_r' = \tilde{q}_r' + d_r^\phi - 1$, respectively. Then, the approximation of $\mathrm{f}_{\eta_j}(a)$ via P-splines can be written as

$$\mathrm{f}_{\eta_j}(a) = \sum_{i=1}^{p_j'} \mathrm{B}_i(a; d_j^\eta - 1)\tau_{ji}^\eta,$$

where

$$\mathrm{B}_i(a; d_j^\eta - 1) = \frac{a - a_{j(i)}}{a_{j(i+d_j^\eta)} - a_{j(i)}}\mathrm{B}_i(a; d_j^\eta - 2) + \frac{a_{j(i+1+d_j^\eta)} - a}{a_{j(i+1+d_j^\eta)} - a_{j(i+1)}}\mathrm{B}_{i+1}(a; d_j^\eta - 2),$$

in which $\mathrm{B}_i(a; -1) = 1$ if $a_{j(i)} \leq a < a_{j(i+1)}$ and $\mathrm{B}_i(a; -1) = 0$ otherwise (see, e.g., Wood (2006, page 148); Lancaster and Salkauskas (1986, page 90)). Thus, the k-th row of $\dot{\mathbf{N}}_{\eta_j}$ is given by $[\mathrm{B}_1(a_{k,j}; d_j^\eta - 1), \ldots, \mathrm{B}_{p_j'}(a_{k,j}; d_j^\eta - 1)]$ and $\dot{\mathbf{N}}_{\eta_j}\mathbf{1}_{p_j'} = \mathbf{1}_n$. To avoid overfitting, the approximations of $\mathrm{f}_{\eta_j}(\cdot)$ and $\mathrm{f}_{\phi_r}(\cdot)$ via P-splines consider the following *difference penalty terms* of order $\varsigma_j^\eta > 0$ and $\varsigma_r^\phi > 0$, respectively:

$$-\frac{\lambda_{\eta_j}}{2}\sum_{i=\varsigma_j^\eta+1}^{p_j'}\left[\Delta^{\varsigma_j^\eta}\dot{\tau}_{j,i}^\eta\right]^2 = -\frac{\lambda_{\eta_j}}{2}\dot{\boldsymbol{\tau}}_{\eta_j}^\top\dot{\mathbf{M}}_{\eta_j}\dot{\boldsymbol{\tau}}_{\eta_j} \quad \text{and}$$

$$-\frac{\lambda_{\phi_r}}{2}\sum_{i=\varsigma_r^\phi+1}^{q_r'}\left[\Delta^{\varsigma_r^\phi}\dot{\tau}_{r,i}^\phi\right]^2 = -\frac{\lambda_{\phi_r}}{2}\dot{\boldsymbol{\tau}}_{\phi_r}^\top\dot{\mathbf{M}}_{\phi_r}\dot{\boldsymbol{\tau}}_{\phi_r},$$

where $\Delta\dot{\tau}_{j,i} = \dot{\tau}_{j,i}^\eta - \dot{\tau}_{j,i-1}^\eta$, $\dot{\mathbf{M}}_{\eta_j} = \mathbf{D}_{\varsigma_j^\eta}^\top\mathbf{D}_{\varsigma_j^\eta}$, $\dot{\mathbf{M}}_{\phi_r} = \mathbf{D}_{\varsigma_r^\phi}^\top\mathbf{D}_{\varsigma_r^\phi}$, with $\mathbf{D}_{\varsigma_j^\eta}$ and $\mathbf{D}_{\varsigma_r^\phi}$ as the matrix representations of the difference operators $\Delta^{\varsigma_j^\eta}$ and $\Delta^{\varsigma_r^\phi}$, respectively.

One of the advantages of this model is the high flexibility of their random and systematic components, because it considers a rich class of distributions with many desirable properties for the model error/response, and it has the capacity to assume parametric, semi-parametric or nonparametric systematic components. The second main advantage of this model is the the easy and straightforward interpretation of the results because the parameters η and ϕ, which may be directly interpreted as the median and the skewness (or the relative dispersion) of the T distribution, are explicitly modeled as semi-parametric functions of explanatory variables. The third main advantage is the simplicity of calculation of confidence regions and hypothesis testing based on the Wald- and Rao-type statistics due to the orthogonality between the regression parameters associated with η and ϕ (provided that the extra parameter ζ is assumed to be known or fixed).

3.2 Parameter estimation

In matrix form, (3.3) may be expressed as

$$\underset{(I)}{\begin{cases} \boldsymbol{\eta} = \boldsymbol{\eta}(\mathbf{X}, \boldsymbol{\beta}) \quad \text{and} \\ \log(\boldsymbol{\phi}) = \mathbf{W}\boldsymbol{\gamma} + \dot{\mathbf{N}}_{\phi_1}\dot{\boldsymbol{\tau}}_{\phi_1} + \ldots + \dot{\mathbf{N}}_{\phi_{q'}}\dot{\boldsymbol{\tau}}_{\phi_{q'}}; \end{cases}} \qquad \underset{(II)}{\begin{cases} \log(\boldsymbol{\eta}) = \mathbf{X}\boldsymbol{\beta} + \dot{\mathbf{N}}_{\eta_1}\dot{\boldsymbol{\tau}}_{\eta_1} + \ldots + \dot{\mathbf{N}}_{\eta_{p'}}\dot{\boldsymbol{\tau}}_{\eta_{p'}} \quad \text{and} \\ \log(\boldsymbol{\phi}) = \mathbf{W}\boldsymbol{\gamma} + \dot{\mathbf{N}}_{\phi_1}\dot{\boldsymbol{\tau}}_{\phi_1} + \ldots + \dot{\mathbf{N}}_{\phi_{q'}}\dot{\boldsymbol{\tau}}_{\phi_{q'}}, \end{cases}}$$

where $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)^\top$, $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_n)^\top$, $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^\top$ and $\mathbf{W} = (\mathbf{w}_1, \ldots, \mathbf{w}_n)^\top$. This model exhibits identification problems (see, e.g., Fahrmeir *et al.* (2013, pages 50 and 536); Wood (2006, page 163)). For instance, if a constant $c \neq 0$ is added to $\dot{\boldsymbol{\tau}}_{\eta_j}$ and at the same time c is subtracted from $\dot{\boldsymbol{\tau}}_{\eta_{j'}}$, the sum $\dot{\mathbf{N}}_{\eta_j}[\dot{\boldsymbol{\tau}}_{\eta_j} + c\mathbf{1}_{p'_j}] + \dot{\mathbf{N}}_{\eta_{j'}}[\dot{\boldsymbol{\tau}}_{\eta_{j'}} - c\mathbf{1}_{p'_{j'}}] = \dot{\mathbf{N}}_{\eta_j}\dot{\boldsymbol{\tau}}_{\eta_j} + c\mathbf{1}_n + \dot{\mathbf{N}}_{\eta_{j'}}\dot{\boldsymbol{\tau}}_{\eta_{j'}} - c\mathbf{1}_n$ remains the same, i.e., $\log(\boldsymbol{\eta})$ does not change if $\dot{\boldsymbol{\tau}}_{\eta_j}$ changes to $\dot{\boldsymbol{\tau}}_{\eta_j} + c\mathbf{1}_{p'_j}$ and $\dot{\boldsymbol{\tau}}_{\eta_{j'}}$ changes to $\dot{\boldsymbol{\tau}}_{\eta_{j'}} - c\mathbf{1}_{p'_{j'}}$. Then, to avoid identification problems, $\dot{\boldsymbol{\tau}}_{\eta_j}$ $(j = 1, \ldots, p')$ and $\dot{\boldsymbol{\tau}}_{\phi_r}$ $(r = 1, \ldots, q')$ are restricted to satisfy $\mathbf{1}_{p'_j}^\top\dot{\boldsymbol{\tau}}_{\eta_j} = \mathbf{1}_{q'_r}^\top\dot{\boldsymbol{\tau}}_{\phi_r} = 0$. These linear constraints may be introduced by writing the model in terms of $\mathbf{N}_{\eta_j}$, $\boldsymbol{\tau}_{\eta_j}$, $\mathbf{M}_{\eta_j}$ $(j = 1, \ldots, p')$ and $\mathbf{N}_{\phi_r}$, $\boldsymbol{\tau}_{\phi_r}$, $\mathbf{M}_{\phi_r}$ $(r = 1, \ldots, q')$, in which

$$\mathbf{N}_{\eta_j} = \dot{\mathbf{N}}_{\eta_j}\mathbf{C}^{\eta}_{j}, \qquad \dot{\boldsymbol{\tau}}_{\eta_j} = \mathbf{C}^{\eta}_{j}\boldsymbol{\tau}_{\eta_j}, \qquad \mathbf{M}_{\eta_j} = [\mathbf{C}^{\eta}_{j}]^\top\dot{\mathbf{M}}_{\eta_j}\mathbf{C}^{\eta}_{j},$$

$$\mathbf{N}_{\phi_r} = \dot{\mathbf{N}}_{\phi_r}\mathbf{C}^{\phi}_{r}, \qquad \dot{\boldsymbol{\tau}}_{\phi_r} = \mathbf{C}^{\phi}_{r}\boldsymbol{\tau}_{\phi_r}, \quad \text{and} \quad \mathbf{M}_{\phi_r} = [\mathbf{C}^{\phi}_{r}]^\top\dot{\mathbf{M}}_{\phi_r}\mathbf{C}^{\phi}_{r},$$

where $\mathbf{C}^{\eta}_{j}$ (of dimension $p'_j \times (p'_j - 1)$) and $\mathbf{C}^{\phi}_{r}$ (of dimension $q'_r \times (q'_r - 1)$) are obtained via the QR decomposition of $\mathbf{1}_{p'_j}$ and $\mathbf{1}_{q'_r}$, respectively (see, e.g., Wood (2006, page 45)), in which they are such that $[\mathbf{C}^{\eta}_{j}]^\top\mathbf{C}^{\eta}_{j} = \mathbf{I}_{p'_j - 1}$ and $[\mathbf{C}^{\phi}_{r}]^\top\mathbf{C}^{\phi}_{r} = \mathbf{I}_{q'_r - 1}$, with $\mathbf{I}_t$ as the identity matrix of order t. Then, the parameter vector is given by $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top, \boldsymbol{\tau}_{\phi}^\top)^\top$ and $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\tau}_{\eta}^\top, \boldsymbol{\gamma}^\top, \boldsymbol{\tau}_{\phi}^\top)^\top$ under the setups (I) and (II), respectively, where $\boldsymbol{\tau}_{\eta} = (\boldsymbol{\tau}_{\eta_1}, \ldots, \boldsymbol{\tau}_{\eta_{p'}})^\top$ and $\boldsymbol{\tau}_{\phi} = (\boldsymbol{\tau}_{\phi_1}, \ldots, \boldsymbol{\tau}_{\phi_{q'}})^\top$.

Similar to Cordeiro and Andrade (2011), the estimation of $\boldsymbol{\theta}$ is performed by fitting a symmetric heteroscedastic semi-parametric model to the transformed response variable (i.e., $Y = \log(T)$), in which the systematic component of the location parameter is given by $\mu_k = \log(\eta_k)$, $k = 1, \ldots, n$, and the systematic component of the dispersion parameter ϕ_k is a semi-parametric function with logarithmic link, where the nonparametric functions are approximated by natural cubic splines or P-splines, and the extra parameter ζ is assumed to be known or fixed. For a known ζ, this approach generalizes the random and systematic components of the linear log-Birnbaum-Saunders, the nonlinear log-Birnbaum-Saunders and the linear log-Birnbaum-Saunders-t models proposed by Rieck and Nedelman (1991), Lemonte and Cordeiro (2009) and Paula *et al.* (2012), respectively. Then, the parameter estimates are the values that maximize the penalized log-likelihood function of $\boldsymbol{\theta}$, which is denoted here as $\mathsf{PL}(\boldsymbol{\theta}) = \mathsf{L}(\boldsymbol{\theta}) + \mathsf{P}(\boldsymbol{\theta})$, where $\mathsf{L}(\boldsymbol{\theta})$ and $\mathsf{P}(\boldsymbol{\theta})$ represent the log-likelihood function (which is based on the joint distribution of $Y_1, \ldots, Y_n$, whose observed values are denoted here as $\mathbf{y} = (y_1, \ldots, y_n)^\top$) and the penalty term of $\boldsymbol{\theta}$, respectively. The penalty term of $\boldsymbol{\theta}$ may be written as

$$\underset{(I)}{\mathsf{P}(\boldsymbol{\theta}) = -\sum_{r=1}^{q'}\frac{\lambda_{\phi_r}}{2}\boldsymbol{\tau}_{\phi_r}^\top\mathbf{M}_{\phi_r}\boldsymbol{\tau}_{\phi_r}} \quad \text{and} \quad \underset{(II)}{\mathsf{P}(\boldsymbol{\theta}) = -\sum_{j=1}^{p'}\frac{\lambda_{\eta_j}}{2}\boldsymbol{\tau}_{\eta_j}^\top\mathbf{M}_{\eta_j}\boldsymbol{\tau}_{\eta_j} - \sum_{r=1}^{q'}\frac{\lambda_{\phi_r}}{2}\boldsymbol{\tau}_{\phi_r}^\top\mathbf{M}_{\phi_r}\boldsymbol{\tau}_{\phi_r}}$$

under the setups described by I and II, respectively, in the expression (3.3). The log-likelihood function of $\boldsymbol{\theta}$ is given by

$$\mathsf{L}(\boldsymbol{\theta}) = \sum_{k=1}^{n} \mathsf{L}_k(\mu_k, \phi_k),$$

in which $\mathsf{L}_k(\mu_k, \phi_k) = \log[g(z_k^2)] - \frac{1}{2}\log(\phi_k)$ is the contribution of the k-th individual to the function $\mathsf{L}(\boldsymbol{\theta})$, where $z_k = (y_k - \mu_k)/\sqrt{\phi_k}$ for $k = 1, \ldots, n$. In the next section, explicit forms of the score function and the Fisher information matrix are provided.

3.2.1 Score function and Fisher information matrix

The score function or estimating equation of $\boldsymbol{\theta}$ is given by $\partial \mathsf{PL}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}^\top = \mathbf{U}(\boldsymbol{\theta}) - \mathbf{M}\boldsymbol{\theta}$, in which

$$\mathbf{U}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{U}_{\eta}(\boldsymbol{\theta}) \\ \mathbf{U}_{\phi}(\boldsymbol{\theta}) \end{bmatrix} = \begin{bmatrix} \mathbf{D}_{\beta}^\top \boldsymbol{\Omega}^{-1} \mathbf{D}_{(\mathrm{v})}(\mathbf{y} - \boldsymbol{\mu}) \\ \frac{1}{2}\mathbf{W}^\top(\mathbf{s} - \mathbf{1}_n) \\ \frac{1}{2}\mathbf{N}_{\phi}^\top(\mathbf{s} - \mathbf{1}_n) \end{bmatrix} \quad \text{and} \quad \mathbf{M} = \mathrm{diag}\{\tilde{\mathbf{0}}_p, \overline{\mathbf{M}}_{\phi}\}$$

under the setup (I), whereas

$$\mathbf{U}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{U}_{\eta}(\boldsymbol{\theta}) \\ \mathbf{U}_{\phi}(\boldsymbol{\theta}) \end{bmatrix} = \begin{bmatrix} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{D}_{(\mathrm{v})}(\mathbf{y} - \boldsymbol{\mu}) \\ \mathbf{N}_{\eta}^\top \boldsymbol{\Omega}^{-1} \mathbf{D}_{(\mathrm{v})}(\mathbf{y} - \boldsymbol{\mu}) \\ \frac{1}{2}\mathbf{W}^\top(\mathbf{s} - \mathbf{1}_n) \\ \frac{1}{2}\mathbf{N}_{\phi}^\top(\mathbf{s} - \mathbf{1}_n) \end{bmatrix} \quad \text{and} \quad \mathbf{M} = \mathrm{diag}\{\overline{\mathbf{M}}_{\eta}, \overline{\mathbf{M}}_{\phi}\}$$

under the setup (II), where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^\top$, $\boldsymbol{\Omega} = \mathrm{diag}\{\phi_1, \ldots, \phi_n\}$, $\mathbf{D}_{(\mathrm{v})} = \mathrm{diag}\{\mathrm{v}_1, \ldots, \mathrm{v}_n\}$, $\mathbf{s} = (\mathrm{v}_1 z_1^2, \ldots, \mathrm{v}_n z_n^2)^\top$, $\mathbf{N}_{\eta} = [\mathbf{N}_{\eta_1}, \ldots, \mathbf{N}_{\eta_{p'}}]$, $\overline{\mathbf{M}}_{\eta} = \mathrm{diag}\{\tilde{\mathbf{0}}_p, \mathbf{M}_{\eta}\}$, $\mathbf{M}_{\eta} = \mathrm{diag}\{\lambda_{\eta_1}\mathbf{M}_{\eta_1}, \ldots, \lambda_{\eta_{p'}}\mathbf{M}_{\eta_{p'}}\}$, $\mathbf{N}_{\phi} = [\mathbf{N}_{\phi_1}, \ldots, \mathbf{N}_{\phi_{q'}}]$, $\overline{\mathbf{M}}_{\phi} = \mathrm{diag}\{\tilde{\mathbf{0}}_q, \mathbf{M}_{\phi}\}$, $\mathbf{M}_{\phi} = \mathrm{diag}\{\lambda_{\phi_1}\mathbf{M}_{\phi_1}, \ldots, \lambda_{\phi_{q'}}\mathbf{M}_{\phi_{q'}}\}$, $\tilde{\mathbf{0}}_t$ is a $t \times t$ zero matrix, and $\mathrm{v}_k \equiv \mathrm{v}(z_k) = -2g'(z_k^2)/g(z_k^2)$ for $k = 1, \ldots, n$. Note that the estimate of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}$, is the solution of the equation $\mathbf{U}(\hat{\boldsymbol{\theta}}) = \mathbf{M}\hat{\boldsymbol{\theta}}$.

Assuming that $\boldsymbol{\mu}$ and $\boldsymbol{\Omega}$ are the true parameter values, the Fisher information matrix may be calculated as $-\mathrm{E}\big[\partial \mathsf{PL}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top\big] = \mathbf{K}(\boldsymbol{\theta}) + \mathbf{M}$, in which $\mathbf{K}(\boldsymbol{\theta}) = \mathrm{diag}\{\mathbf{K}_{\eta}(\boldsymbol{\theta}), \mathbf{K}_{\phi}(\boldsymbol{\theta})\}$, where

$$\mathbf{K}_{\eta}(\boldsymbol{\theta}) = d_g(\zeta)\big(\mathbf{D}_{\beta}^\top \boldsymbol{\Omega}^{-1} \mathbf{D}_{\beta}\big) \quad \text{and} \quad \mathbf{K}_{\phi}(\boldsymbol{\theta}) = \frac{f_g(\zeta) - 1}{4}\big(\overline{\mathbf{W}}^\top \overline{\mathbf{W}}\big)$$

under the setup (I), whereas

$$\mathbf{K}_{\eta}(\boldsymbol{\theta}) = d_g(\zeta)\big(\overline{\mathbf{X}}^\top \boldsymbol{\Omega}^{-1} \overline{\mathbf{X}}\big) \quad \text{and} \quad \mathbf{K}_{\phi}(\boldsymbol{\theta}) = \frac{f_g(\zeta) - 1}{4}\big(\overline{\mathbf{W}}^\top \overline{\mathbf{W}}\big)$$

under the setup (II), in which $\overline{\mathbf{X}} = \big[\mathbf{X}, \mathbf{N}_{\eta}\big]$ and $\overline{\mathbf{W}} = \big[\mathbf{W}, \mathbf{N}_{\phi}\big]$.

3.2.2 Iterative process

The iterative process of parameter estimation considers the backfitting algorithm (see, e.g., Hastie and Tibshirani, 1990; Wood, 2006) to accomplish each step of the Fisher scoring algorithm. Define $\mathbf{N}_{\eta_j} = \mathbf{X}$ and $\mathbf{N}_{\phi_r} = \mathbf{W}$ for $j = r = 0$. Then, for a fixed value of the smoothing parameter $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_{\eta}^\top, \boldsymbol{\lambda}_{\phi}^\top)^\top$, the iterative process to estimate $\boldsymbol{\theta}$ reduces to:

Algorithm 2.1

Step 1. Initialize the counter as $l = 0$ and set the initial value to $\boldsymbol{\theta}^{(0)}$.

Step 2. Based on $\boldsymbol{\theta}^{(l)}$ do the following:

(A.1) Under the setup (I) described in (3.3) compute

$$\boldsymbol{\beta}^{(l+1)} = \Big\{[\mathbf{D}_\beta^{(l)}]^\top \boldsymbol{\Omega}_{(l)}^{-1}\mathbf{D}_\beta^{(l)}\Big\}^{-1}[\mathbf{D}_\beta^{(l)}]^\top \boldsymbol{\Omega}_{(l)}^{-1}\Big\{\mathbf{D}_\beta^{(l)}\boldsymbol{\beta}^{(l)} + (1/d_g(\zeta))\mathbf{D}_{(\mathrm{v})}^{(l)}(\mathbf{y}-\boldsymbol{\mu}^{(l)})\Big\}.$$

(A.2) Under the setup (II) described in (3.3), solve the following equations for $\boldsymbol{\beta}^{(l+1)}$ and $\boldsymbol{\tau}_{\eta_j}^{(l+1)}$ $(j = 1, \ldots, p')$ via the backfitting algorithm:

$$\begin{aligned}
\boldsymbol{\beta}^{(l+1)} &= \Big[\mathbf{X}^\top\boldsymbol{\Omega}_{(l)}^{-1}\mathbf{X}\Big]^{-1}\mathbf{X}^\top\boldsymbol{\Omega}_{(l)}^{-1}\Big[\tilde{\mathbf{y}}^{(l)} - \sum_{j\neq 0}\mathbf{N}_{\eta_j}\boldsymbol{\tau}_{\eta_j}^{(l+1)}\Big],\\
\boldsymbol{\tau}_{\eta_1}^{(l+1)} &= \Big[\mathbf{N}_{\eta_1}^\top\boldsymbol{\Omega}_{(l)}^{-1}\mathbf{N}_{\eta_1} + \lambda_{\eta_1}^*\mathbf{M}_{\eta_1}\Big]^{-1}\mathbf{N}_{\eta_1}^\top\boldsymbol{\Omega}_{(l)}^{-1}\Big[\tilde{\mathbf{y}}^{(l)} - \sum_{j\neq 1}\mathbf{N}_{\eta_j}\boldsymbol{\tau}_{\eta_j}^{(l+1)}\Big],\\
&\vdots\\
\boldsymbol{\tau}_{\eta_{p'}}^{(l+1)} &= \Big[\mathbf{N}_{\eta_{p'}}^\top\boldsymbol{\Omega}_{(l)}^{-1}\mathbf{N}_{\eta_{p'}} + \lambda_{\eta_{p'}}^*\mathbf{M}_{\eta_{p'}}\Big]^{-1}\mathbf{N}_{\eta_{p'}}^\top\boldsymbol{\Omega}_{(l)}^{-1}\Big[\tilde{\mathbf{y}}^{(l)} - \sum_{j\neq p'}\mathbf{N}_{\eta_j}\boldsymbol{\tau}_{\eta_j}^{(l+1)}\Big],
\end{aligned}$$

where $\lambda_{\eta_j}^* = \lambda_{\eta_j}/d_g(\zeta)$ and $\tilde{\mathbf{y}}^{(l)} = \boldsymbol{\mu}^{(l)} + (1/d_g(\zeta))\mathbf{D}_{(\mathrm{v})}^{(l)}(\mathbf{y}-\boldsymbol{\mu}^{(l)})$ is a working response variable.

Step 3. Solve the following equations for $\boldsymbol{\gamma}^{(l+1)}$ and $\boldsymbol{\tau}_{\phi_r}^{(l+1)}$ $(r = 1, \ldots, q')$ via the backfitting algorithm:

$$\begin{aligned}
\boldsymbol{\gamma}^{(l+1)} &= \Big(\mathbf{W}^\top\mathbf{W}\Big)^{-1}\mathbf{W}^\top\Big\{\tilde{\mathbf{z}}^{(l)} - \sum_{r\neq 0}\mathbf{N}_{\phi_r}\boldsymbol{\tau}_{\phi_r}^{(l+1)}\Big\},\\
\boldsymbol{\tau}_{\phi_1}^{(l+1)} &= \Big[\mathbf{N}_{\phi_1}^\top\mathbf{N}_{\phi_1} + \lambda_{\phi_1}^*\mathbf{M}_{\phi_1}\Big]^{-1}\mathbf{N}_{\phi_1}^\top\Big\{\tilde{\mathbf{z}}^{(l)} - \sum_{r\neq 1}\mathbf{N}_{\phi_r}\boldsymbol{\tau}_{\phi_r}^{(l+1)}\Big\},\\
&\vdots\\
\boldsymbol{\tau}_{\phi_{q'}}^{(l+1)} &= \Big[\mathbf{N}_{\phi_{q'}}^\top\mathbf{N}_{\phi_{q'}} + \lambda_{\phi_{q'}}^*\mathbf{M}_{\phi_{q'}}\Big]^{-1}\mathbf{N}_{\phi_{q'}}^\top\Big\{\tilde{\mathbf{z}}^{(l)} - \sum_{r\neq q'}\mathbf{N}_{\phi_r}\boldsymbol{\tau}_{\phi_r}^{(l+1)}\Big\},
\end{aligned}$$

where $\lambda_{\phi_r}^* = 4\lambda_{\phi_r}/(f_g(\zeta)-1)$ and $\tilde{\mathbf{z}}^{(l)} = \log[\boldsymbol{\phi}^{(l)}] + \frac{2}{f_g(\zeta)-1}(\mathbf{s}^{(l)} - \mathbf{1}_n)$ is a working response variable.

Step 4. Update $l = (l+1)$ and $\boldsymbol{\theta}^{(l)}$.

Step 5. Repeat steps 2, 3 and 4 until convergence of $\boldsymbol{\theta}^{(l)}$.

Because distributions such as the log-Student-t, log-power-exponential (for $0 \leq \zeta \leq 1$), log-contaminated-normal, log-slash and log-hyperbolic can be obtained as a power mixture of the log-normal distribution (see Andrews and Mallows, 1974; Barndoff-Nielsen, 1977; West, 1987) the EM algorithm (Dempster *et al.*, 1977) can be applied to obtain an more efficient process. In these cases, $\xi|U = u \sim$ log-normal$(1, h(u))$ and the density generator of the ξ distribution may be written as $g(t^2) = \int_{\mathbb{R}+}\Big\{\phi\Big[t/\sqrt{h(u)}\Big]/\sqrt{h(u)}\Big\}f_U(u)\partial u$, where $\phi(\cdot)$ is the density function of the standard normal distribution, $h(u) > 0$ for $u > 0$ and $f_U(\cdot)$ is the probability density function (or probability mass function) of the random variable U. The E step of the EM algorithm is accomplished by calculating $\mathrm{v}(z_k^{(m)}) = \mathrm{E}_U\Big[1/h(U)\,|\,Y_k = y_k; \boldsymbol{\theta}^{(m)}\Big]$ for $k = 1, \ldots, n$. The M step is accomplished by maximizing $\mathsf{PL}^{(m)}(\boldsymbol{\theta}) = \mathsf{L}^{(m)}(\boldsymbol{\theta}) + \mathsf{P}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, where

$$\mathsf{L}^{(m)}(\boldsymbol{\theta}) = \sum_{k=1}^n \log\left\{\phi\Big[z_k\sqrt{\mathrm{v}(z_k^{(m)})}\Big]/\sqrt{\phi_k}\right\}.$$

Then, the iterative process of parameter estimation for a fixed value of the smoothing parameter reduces to the following form:

Algorithm 2.2

Step 1. Initialize the counter as $m = 0$ and set the initial value of the parameter vector to $\boldsymbol{\theta}^{(0)}$.

Step 2. Calculate $\mathbf{D}_{(\mathrm{v})}^{(m)}=\mathrm{diag}\{\mathrm{v}(z_1^{(m)}),\ldots,\mathrm{v}(z_n^{(m)})\}$ from $\boldsymbol{\theta}^{(m)}$.

Step 3. Initialize the counter as $l = 0$ and set the initial value as $\boldsymbol{\theta}_*^{(0)} = \boldsymbol{\theta}^{(m)}$.

Step 4. Based on $\boldsymbol{\theta}_*^{(l)}$ and $\boldsymbol{\theta}^{(m)}$ do the following:

(A.1) Under the setup (I) described in (3.3) compute

$$\boldsymbol{\beta}^{(l+1)} = \left\{[\mathbf{D}_\beta^{(l)}]^\top \boldsymbol{\Omega}_{(l)}^{-1}\mathbf{D}_{(\mathrm{v})}^{(m)}\mathbf{D}_\beta^{(l)}\right\}^{-1}[\mathbf{D}_\beta^{(l)}]^\top \boldsymbol{\Omega}_{(l)}^{-1}\mathbf{D}_{(\mathrm{v})}^{(m)}\left\{\mathbf{D}_\beta^{(l)}\boldsymbol{\beta}^{(l)} + (\mathbf{y} - \boldsymbol{\mu}^{(l)})\right\}.$$

(A.2) Under the setup (II) described in (3.3), solve the following equations for $\boldsymbol{\beta}^{(l+1)}$ and $\boldsymbol{\tau}_{\eta_j}^{(l+1)}$ $(j = 1,\ldots,p')$ via the backfitting algorithm:

$$\begin{aligned}
\boldsymbol{\beta}^{(l+1)} &= \left[\mathbf{X}^\top \boldsymbol{\Omega}_{(l)}^{-1}\mathbf{D}_{(\mathrm{v})}^{(m)}\mathbf{X}\right]^{-1}\mathbf{X}^\top \boldsymbol{\Omega}_{(l)}^{-1}\mathbf{D}_{(\mathrm{v})}^{(m)}\left[\mathbf{y} - \sum_{j\neq 0}\mathbf{N}_{\eta_j}\boldsymbol{\tau}_{\eta_j}^{(l+1)}\right],\\
\boldsymbol{\tau}_{\eta_1}^{(l+1)} &= \left[\mathbf{N}_{\eta_1}^\top \boldsymbol{\Omega}_{(l)}^{-1}\mathbf{D}_{(\mathrm{v})}^{(m)}\mathbf{N}_{\eta_1} + \lambda_{\eta_1}\mathbf{M}_{\eta_1}\right]^{-1}\mathbf{N}_{\eta_1}^\top \boldsymbol{\Omega}_{(l)}^{-1}\mathbf{D}_{(\mathrm{v})}^{(m)}\left[\mathbf{y} - \sum_{j\neq 1}\mathbf{N}_{\eta_j}\boldsymbol{\tau}_{\eta_j}^{(l+1)}\right],\\
&\ \vdots\\
\boldsymbol{\tau}_{\eta_{p'}}^{(l+1)} &= \left[\mathbf{N}_{\eta_{p'}}^\top \boldsymbol{\Omega}_{(l)}^{-1}\mathbf{D}_{(\mathrm{v})}^{(m)}\mathbf{N}_{\eta_{p'}} + \lambda_{\eta_{p'}}\mathbf{M}_{\eta_{p'}}\right]^{-1}\mathbf{N}_{\eta_{p'}}^\top \boldsymbol{\Omega}_{(l)}^{-1}\mathbf{D}_{(\mathrm{v})}^{(m)}\left[\mathbf{y} - \sum_{j\neq p'}\mathbf{N}_{\eta_j}\boldsymbol{\tau}_{\eta_j}^{(l+1)}\right].
\end{aligned}$$

(B) Solve the following equations for $\boldsymbol{\gamma}^{(l+1)}$ and $\boldsymbol{\tau}_{\phi_r}^{(l+1)}$ $(r = 1,\ldots,q')$ via the backfitting algorithm:

$$\begin{aligned}
\boldsymbol{\gamma}^{(l+1)} &= \left(\mathbf{W}^\top\mathbf{W}\right)^{-1}\mathbf{W}^\top\left\{\left[\log\left(\boldsymbol{\phi}^{(l)}\right) + \mathbf{s}^{(l)}\right] - \sum_{r\neq 0}\mathbf{N}_{\phi_r}\boldsymbol{\tau}_{\phi_r}^{(l+1)}\right\},\\
\boldsymbol{\tau}_{\phi_1}^{(l+1)} &= \left[\mathbf{N}_{\phi_1}^\top\mathbf{N}_{\phi_1} + 2\lambda_{\phi_1}\mathbf{M}_{\phi_1}\right]^{-1}\mathbf{N}_{\phi_1}^\top\left\{\left[\log\left(\boldsymbol{\phi}^{(l)}\right) + \mathbf{s}^{(l)}\right] - \sum_{r\neq 1}\mathbf{N}_{\phi_r}\boldsymbol{\tau}_{\phi_r}^{(l+1)}\right\},\\
&\ \vdots\\
\boldsymbol{\tau}_{\phi_{q'}}^{(l+1)} &= \left[\mathbf{N}_{\phi_{q'}}^\top\mathbf{N}_{\phi_{q'}} + 2\lambda_{\phi_{q'}}\mathbf{M}_{\phi_{q'}}\right]^{-1}\mathbf{N}_{\phi_{q'}}^\top\left\{\left[\log\left(\boldsymbol{\phi}^{(l)}\right) + \mathbf{s}^{(l)}\right] - \sum_{r\neq q'}\mathbf{N}_{\phi_r}\boldsymbol{\tau}_{\phi_r}^{(l+1)}\right\},
\end{aligned}$$

where $\mathbf{s}^{(l)} = \left(\mathrm{v}(z_1^{(m)})[z_1^{(l)}]^2 - 1,\ldots,\mathrm{v}(z_n^{(m)})[z_n^{(l)}]^2 - 1\right)^\top$.

(C) Update $l = (l + 1)$ and $\boldsymbol{\theta}_*^{(l)}$.

(D) Repeat steps (A), (B) and (C) until convergence of $\boldsymbol{\theta}_*^{(l)}$.

Step 5. Update $m = (m + 1)$ and $\boldsymbol{\theta}^{(m)} = \boldsymbol{\theta}_*^{(l)}$.

Step 6. Repeat steps 2, 3, 4 and 5 until convergence of $\boldsymbol{\theta}^{(m)}$.

The EM algorithm also is applied when the distribution of the model error is Birnbaum-Saunders-$t(\zeta = (\zeta_1,\zeta_2)^\top)$ because it may be obtained as a shape mixture of the Birnbaum-Saunders distribution (Balakrishnan *et al.*, 2009). In this case, $\xi|U = u \sim$ Birnbaum-Saunders$(1, 1, \zeta_1/\sqrt{u})$

and the density generator of the ξ distribution is $g(t^2) = \int_{\mathbb{R}+}\{\overline{\phi}[t, \zeta_1/\sqrt{u}]\}f_U(u)\partial u$, where $\overline{\phi}(\cdot, \zeta)$ is the probability density function of a standard sinh-normal distribution with extra parameter ζ, and $f_U(u) \propto u^{\frac{\zeta_2+1}{2}}\exp(-u\zeta_2/2)$. The E step of the EM algorithm is accomplished by calculating $u_k^{(m)} = \mathrm{E}_U\Big[U \,|\, y_k; \boldsymbol{\theta}^{(m)}\Big]$ for $k = 1, \ldots, n$. The M step is accomplished by maximizing $\mathsf{PL}^{(m)}(\boldsymbol{\theta}) = \mathsf{L}^{(m)}(\boldsymbol{\theta}) + \mathsf{P}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, where

$$\mathsf{L}^{(m)}(\boldsymbol{\theta}) = \sum_{k=1}^{n} \log\left\{\overline{\phi}\Big[z_k, \zeta_1/\sqrt{u_k^{(m)}}\Big]/\sqrt{\phi_k}\right\}.$$

Next, the resulting iterative process for a fixed value of the smoothing parameter is described.

Algorithm 2.3

Step 1. Initialize the counter as $m = 0$ and set the initial value of the parameter vector to $\boldsymbol{\theta}^{(0)}$.

Step 2. Calculate $\mathbf{u}^{(m)} = (u_1^{(m)}, \ldots, u_n^{(m)})^\top$ from $\boldsymbol{\theta}^{(m)}$ in the following manner:

$$u_k^{(m)} = \frac{\zeta_1^2(\zeta_2+1)}{\zeta_1^2\zeta_2 + \Big[2\sinh(z_k^{(m)})\Big]^2} \qquad \text{for} \quad k = 1, \ldots, n.$$

Step 3. Calculate $\mathbf{D}_{(d_g)}^{(m)} = \mathrm{diag}\Big\{d_g^*\Big(\zeta_1/[u_1^{(m)}]^{\frac{1}{2}}\Big), \ldots, d_g^*\Big(\zeta_1/[u_n^{(m)}]^{\frac{1}{2}}\Big)\Big\}$, where

$$d_g^*(\zeta) = 2 + \frac{4}{\zeta^2} - \frac{\sqrt{2\pi}}{\zeta}\left\{1 - \frac{2}{\sqrt{\pi}}\int_0^{\frac{\sqrt{2}}{\zeta}} \exp\left(-t^2\right)\partial t\right\}\exp\left(\frac{2}{\zeta^2}\right).$$

Step 4. Calculate $\mathbf{D}_{(f_g)}^{(m)} = \frac{1}{4}\mathrm{diag}\left\{f_g^*\Big(\zeta_1/[u_1^{(m)}]^{\frac{1}{2}}\Big), \ldots, f_g^*\Big(\zeta_1/[u_n^{(m)}]^{\frac{1}{2}}\Big)\right\}$, where

$$f_g^*(\zeta) = \mathrm{E}\Big[\big(4\sinh(Z)\cosh(Z)Z/\zeta^2 - \tanh(Z)Z\big)^2\Big] - 1,$$

where Z exhibit a standard sinh-normal distribution with extra parameter ζ.

Step 5. Initialize the counter as $l = 0$ and set the initial value as $\boldsymbol{\theta}_*^{(0)} = \boldsymbol{\theta}^{(m)}$.

Step 6. Based on $\boldsymbol{\theta}_*^{(l)}$ and $\boldsymbol{\theta}^{(m)}$ do the following:

(A.1) Under the setup (I) described in (3.3) compute

$$\boldsymbol{\beta}^{(l+1)} = \left\{[\mathbf{D}_\beta^{(l)}]^\top \boldsymbol{\Omega}_{(l)}^{-1}\mathbf{D}_{(d_g)}^{(m)}\mathbf{D}_\beta^{(l)}\right\}^{-1}[\mathbf{D}_\beta^{(l)}]^\top \boldsymbol{\Omega}_{(l)}^{-1}\mathbf{D}_{(d_g)}^{(m)}\Big\{\mathbf{D}_\beta^{(l)}\boldsymbol{\beta}^{(l)} + \mathbf{D}_{(\tilde{\rho})}^{(l)}(\mathbf{y} - \boldsymbol{\mu}^{(l)})\Big\}, \quad \text{where}$$

$$\mathbf{D}_{(\tilde{\rho})}^{(l)} = \mathrm{diag}\Big\{\tilde{\rho}(z_1^{(l)}, u_1^{(m)}), \ldots, \tilde{\rho}(z_n^{(l)}, u_n^{(m)})\Big\}, \tilde{\rho}(z, u) = \frac{4\sinh(z)\cosh(z)u}{z\,\zeta_1^2\, d_g^*(\zeta_1/u^{\frac{1}{2}})} - \frac{\tanh(z)}{z\, d_g^*(\zeta_1/u^{\frac{1}{2}})}.$$

(A.2) Under the setup (II) described in (3.3), solve the following equations for $\boldsymbol{\beta}^{(l+1)}$ and

$\boldsymbol{\tau}_{\eta_j}^{(l+1)}$ $(j = 1, \ldots, p')$ via the backfitting algorithm:

$$\boldsymbol{\beta}^{(l+1)} = \Big[\mathbf{X}^\top \boldsymbol{\Omega}_{(l)}^{-1}\mathbf{D}_{(d_g)}^{(m)}\mathbf{X}\Big]^{-1}\mathbf{X}^\top \boldsymbol{\Omega}_{(l)}^{-1}\mathbf{D}_{(d_g)}^{(m)}\Big[\tilde{\mathbf{y}}^{(l)} - \sum_{j\neq 0}\mathbf{N}_{\eta_j}\boldsymbol{\tau}_{\eta_j}^{(l+1)}\Big],$$

$$\boldsymbol{\tau}_{\eta_1}^{(l+1)} = \Big[\mathbf{N}_{\eta_1}^\top \boldsymbol{\Omega}_{(l)}^{-1}\mathbf{D}_{(d_g)}^{(m)}\mathbf{N}_{\eta_1} + \lambda_{\eta_1}\mathbf{M}_{\eta_1}\Big]^{-1}\mathbf{N}_{\eta_1}^\top \boldsymbol{\Omega}_{(l)}^{-1}\mathbf{D}_{(d_g)}^{(m)}\Big[\tilde{\mathbf{y}}^{(l)} - \sum_{j\neq 1}\mathbf{N}_{\eta_j}\boldsymbol{\tau}_{\eta_j}^{(l+1)}\Big],$$

$$\vdots$$

$$\boldsymbol{\tau}_{\eta_{p'}}^{(l+1)} = \Big[\mathbf{N}_{\eta_{p'}}^\top \boldsymbol{\Omega}_{(l)}^{-1}\mathbf{D}_{(d_g)}^{(m)}\mathbf{N}_{\eta_{p'}} + \lambda_{\eta_{p'}}\mathbf{M}_{\eta_{p'}}\Big]^{-1}\mathbf{N}_{\eta_{p'}}^\top \boldsymbol{\Omega}_{(l)}^{-1}\mathbf{D}_{(d_g)}^{(m)}\Big[\tilde{\mathbf{y}}^{(l)} - \sum_{j\neq p'}\mathbf{N}_{\eta_j}\boldsymbol{\tau}_{\eta_j}^{(l+1)}\Big],$$

where $\tilde{\mathbf{y}}^{(l)} = \boldsymbol{\mu}^{(l)} + \mathbf{D}_{(\tilde{\rho})}^{(l)}(\mathbf{y} - \boldsymbol{\mu}^{(l)})$ is a working response variable.

(B) Solve the following equations for $\boldsymbol{\gamma}^{(l+1)}$ and $\boldsymbol{\tau}_{\phi_r}^{(l+1)}$ $(r = 1, \ldots, q')$ via the backfitting algorithm:

$$\boldsymbol{\gamma}^{(l+1)} = \Big(\mathbf{W}^\top \mathbf{D}_{(f_g)}^{(m)}\mathbf{W}\Big)^{-1}\mathbf{W}^\top \mathbf{D}_{(f_g)}^{(m)}\Big\{\tilde{\mathbf{z}}^{(l)} - \sum_{r\neq 0}\mathbf{N}_{\phi_r}\boldsymbol{\tau}_{\phi_r}^{(l+1)}\Big\},$$

$$\boldsymbol{\tau}_{\phi_1}^{(l+1)} = \Big[\mathbf{N}_{\phi_1}^\top \mathbf{D}_{(f_g)}^{(m)}\mathbf{N}_{\phi_1} + \lambda_{\phi_1}\mathbf{M}_{\phi_1}\Big]^{-1}\mathbf{N}_{\phi_1}^\top \mathbf{D}_{(f_g)}^{(m)}\Big[\tilde{\mathbf{z}}^{(l)} - \sum_{r\neq 1}\mathbf{N}_{\phi_r}\boldsymbol{\tau}_{\phi_r}^{(l+1)}\Big\},$$

$$\vdots$$

$$\boldsymbol{\tau}_{\phi_{q'}}^{(l+1)} = \Big[\mathbf{N}_{\phi_{q'}}^\top \mathbf{D}_{(f_g)}^{(m)}\mathbf{N}_{\phi_{q'}} + \lambda_{\phi_{q'}}\mathbf{M}_{\phi_{q'}}\Big]^{-1}\mathbf{N}_{\phi_{q'}}^\top \mathbf{D}_{(f_g)}^{(m)}\Big[\tilde{\mathbf{z}}^{(l)} - \sum_{r\neq q'}\mathbf{N}_{\phi_r}\boldsymbol{\tau}_{\phi_r}^{(l+1)}\Big\},$$

where $\tilde{\mathbf{z}}^{(l)} = \log[\boldsymbol{\phi}^{(l)}] + \frac{1}{2}[\mathbf{D}_{(f_g)}^{(m)}]^{-1}\Big(\mathbf{D}_{(d_g)}\mathbf{D}_{(\tilde{\rho})}\bar{\mathbf{s}}^{(l)} - \mathbf{1}_n\Big)$, $\bar{\mathbf{s}}^{(l)} = \big([z_1^{(l)}]^2, \ldots, [z_n^{(l)}]^2\big)^\top$.

(C) Update $l = (l+1)$ and $\boldsymbol{\theta}_*^{(l)}$.

(D) Repeat steps (A), (B) and (C) until convergence of $\boldsymbol{\theta}_*^{(l)}$.

Step 7. Update $m = (m+1)$ and $\boldsymbol{\theta}^{(m)} = \boldsymbol{\theta}_*^{(l)}$.

Step 8. Repeat steps 2, 3, 4, 5, 6 and 7 until convergence of $\boldsymbol{\theta}^{(m)}$.

Note that the iterative estimation process consider a set of individual-specific weights $(\mathrm{v}(z_k^{(l)}),\ k = 1, \ldots, n)$, which are related to the relative importance of each individual in the iterative estimation process and are dependent on the standardized difference between the observed response and the fitted value. For further details on the function $\mathrm{v}(\cdot)$ see Section 1.3. Hereinafter, $\hat{\eta}_k$, $\hat{\mu}_k = \log(\hat{\eta}_k)$ and $\hat{\phi}_k$ represent the fitted values of η_k, $\mu_k = \log(\eta_k)$ and ϕ_k, respectively, for $k = 1, \ldots, n$.

3.3 Asymptotic theory

Let $\hat{\boldsymbol{\theta}}$, $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\gamma}}$, $\hat{\boldsymbol{\tau}}_\eta$ and $\hat{\boldsymbol{\tau}}_\phi$ represent the maximum penalized likelihood estimators of $\boldsymbol{\theta}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, $\boldsymbol{\tau}_\eta$ and $\boldsymbol{\tau}_\phi$, respectively. Under the absence of nonparametric effects in the systematic component (3.3), the model setup coincides with the topic addressed by Cysneiros *et al.* (2010), which described the asymptotic properties of $\hat{\boldsymbol{\theta}}$, such as consistency, efficiency and normality. The asymptotic behaviour of $\hat{\boldsymbol{\theta}}$ under the general case of the systematic component (3.3) is studied here by using a framework of fixed-knot (see Wu and Yu, 2014; Yu and Ruppert, 2002), which implies that the penalty matrices and the size of $\boldsymbol{\tau}_\eta$ and $\boldsymbol{\tau}_\phi$ are not dependent on the sample size n. Next, the required conditions for some asymptotic results be valid, are listed.

Conditions:

1. There is $\boldsymbol{\tau}^{[0]}_{\eta_j} \in \mathbb{R}^{p'_j-1}$ such that $\mathrm{f}_{\eta_j}(a) = \mathbf{N}^{\top}_{\eta_j}(a)\boldsymbol{\tau}^{[0]}_{\eta_j}$ for all $a \in (\underline{a}_j, \overline{a}_j)$, $j = 1, \ldots, p'$. Then, $\boldsymbol{\tau}^{[0]}_{\eta} = \left(\left[\boldsymbol{\tau}^{[0]}_{\eta_1}\right]^{\top}, \ldots, \left[\boldsymbol{\tau}^{[0]}_{\eta_{p'}}\right]^{\top}\right)^{\top}$ is the true parameter vector for $\boldsymbol{\tau}_{\eta}$.

2. There is $\boldsymbol{\tau}^{[0]}_{\phi_r} \in \mathbb{R}^{q'_r-1}$ such that $\mathrm{f}_{\phi_r}(b) = \mathbf{N}^{\top}_{\phi_r}(b)\boldsymbol{\tau}^{[0]}_{\phi_r}$ for all $b \in (\underline{b}_r, \overline{b}_r)$, $r = 1, \ldots, q'$. Then, $\boldsymbol{\tau}^{[0]}_{\phi} = \left(\left[\boldsymbol{\tau}^{[0]}_{\phi_1}\right]^{\top}, \ldots, \left[\boldsymbol{\tau}^{[0]}_{\phi_{q'}}\right]^{\top}\right)^{\top}$ is the true parameter vector for $\boldsymbol{\tau}_{\phi}$.

3. The usual regularity conditions of large sample theory are fulfilled (see Cox and Hinkley, 1974, chapter 9). This conditions ensure that

   (a) $n^{-1}\mathbf{K}(\boldsymbol{\theta}^{[0]}) \underset{n\to\infty}{\longrightarrow} \boldsymbol{\Sigma}(\boldsymbol{\theta}^{[0]})$, where $\mathbf{K}(\boldsymbol{\theta}^{[0]})$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta}^{[0]})$ are positive definite matrices, and $\boldsymbol{\theta}^{[0]} = \left(\left[\boldsymbol{\beta}^{[0]}\right]^{\top}, \left[\boldsymbol{\tau}^{[0]}_{\eta}\right]^{\top}, \left[\boldsymbol{\gamma}^{[0]}\right]^{\top}, \left[\boldsymbol{\tau}^{[0]}_{\phi}\right]^{\top}\right)^{\top}$ is the true parameter vector for $\boldsymbol{\theta}$.

   (b) $\mathbf{K}^{-\frac{1}{2}}(\boldsymbol{\theta}^{[0]})\mathbf{U}(\boldsymbol{\theta}^{[0]}) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(\mathbf{0}, \mathbf{I})$.

   (c) $n^{-1}\mathbf{U}(\boldsymbol{\theta}^{[0]}) \xrightarrow[n\to\infty]{\mathcal{P}} \mathbf{0}$.

   (d) $n^{-1}\left[\dfrac{\partial^2 \mathsf{L}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\top}}(\boldsymbol{\theta}^{[0]}) + \mathbf{K}(\boldsymbol{\theta}^{[0]})\right] \xrightarrow[n\to\infty]{\mathcal{P}} \mathbf{0}$.

   (e) $n^{-1}\left[\dfrac{\partial^2 \mathsf{L}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\top}}(\boldsymbol{\theta}^{*}) - \dfrac{\partial^2 \mathsf{L}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\top}}(\boldsymbol{\theta}^{[0]})\right] \xrightarrow[n\to\infty]{\mathcal{P}} \mathbf{0}$ for all $\boldsymbol{\theta}^{*}$ in line segment joining $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}^{[0]}$.

4. The smoothing parameter may be dependent on the sample size. Let $\boldsymbol{\lambda}^{(n)}$ be the value of the smoothing parameter under a sample of size n. Then, $\|\boldsymbol{\lambda}^{(n)}\| \underset{n\to\infty}{\longrightarrow} \overline{\lambda} < \infty$.

Theorem 1. *Under (1)-(4) it follows that $\hat{\boldsymbol{\theta}}$ is a consistent estimator of $\boldsymbol{\theta}^{[0]}$, and*

$$\mathbf{K}^{-\frac{1}{2}}(\boldsymbol{\theta}^{[0]})\left[\mathbf{K}(\boldsymbol{\theta}^{[0]}) + \mathbf{M}\right](\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{[0]}) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(\mathbf{0}, \mathbf{I}).$$

Under (1)-(4) and for large sample sizes, $\hat{\boldsymbol{\theta}}$ is an unbiased estimator of $\boldsymbol{\theta}^{[0]}$ whose variance-covariance matrix may be written as

$$\begin{aligned}\mathrm{Var}\big[\hat{\boldsymbol{\theta}}\big] &= \left[\mathbf{K}(\boldsymbol{\theta}^{[0]}) + \mathbf{M}\right]^{-1}\mathbf{K}(\boldsymbol{\theta}^{[0]})\left[\mathbf{K}(\boldsymbol{\theta}^{[0]}) + \mathbf{M}\right]^{-1} \\ &= \begin{bmatrix}\mathrm{Var}\big[\hat{\bar{\boldsymbol{\beta}}}\big] & \mathbf{0} \\ \mathbf{0} & \mathrm{Var}\big[\hat{\bar{\gamma}}\big]\end{bmatrix},\end{aligned}$$

in which $\bar{\boldsymbol{\beta}} = (\boldsymbol{\beta}^{\top}, \boldsymbol{\tau}_{\eta}^{\top})^{\top}$, $\hat{\bar{\boldsymbol{\beta}}} = (\hat{\boldsymbol{\beta}}^{\top}, \hat{\boldsymbol{\tau}}_{\eta}^{\top})^{\top}$, $\bar{\gamma} = (\gamma^{\top}, \boldsymbol{\tau}_{\phi}^{\top})^{\top}$ and $\hat{\bar{\gamma}} = (\hat{\gamma}^{\top}, \hat{\boldsymbol{\tau}}_{\phi}^{\top})^{\top}$. The variance-covariance matrices $\mathrm{Var}\big[\hat{\bar{\boldsymbol{\beta}}}\big]$ and $\mathrm{Var}\big[\hat{\bar{\gamma}}\big]$ are

$$\begin{aligned}\mathrm{Var}\big[\hat{\bar{\boldsymbol{\beta}}}\big] &= \frac{1}{d_g(\zeta)}\left(\mathbf{D}_{\beta}^{\top}\boldsymbol{\Omega}^{-1}\mathbf{D}_{\beta}\right)^{-1} \qquad \text{and} \\ \mathrm{Var}\big[\hat{\bar{\gamma}}\big] &= \frac{4}{f_g(\zeta) - 1}\left[\overline{\mathbf{W}}^{\top}\overline{\mathbf{W}} + \overline{\mathbf{M}}^{*}_{\phi}\right]^{-1}\left(\overline{\mathbf{W}}^{\top}\overline{\mathbf{W}}\right)\left[\overline{\mathbf{W}}^{\top}\overline{\mathbf{W}} + \overline{\mathbf{M}}^{*}_{\phi}\right]^{-1}\end{aligned}$$

under the setup (I) described in (3.3), where $\overline{\mathbf{M}}^{*}_{\phi} = (4/(f_g(\zeta) - 1))\overline{\mathbf{M}}_{\phi}$. Similarly, under the setup (II) described in (3.3), $\mathrm{Var}\big[\hat{\bar{\boldsymbol{\beta}}}\big]$ and $\mathrm{Var}\big[\hat{\bar{\gamma}}\big]$ are given by

$$\mathrm{Var}\big[\hat{\bar{\boldsymbol{\beta}}}\big] = \frac{1}{d_g(\zeta)}\left[\overline{\mathbf{X}}^{\top}\boldsymbol{\Omega}^{-1}\overline{\mathbf{X}} + \overline{\mathbf{M}}^{*}_{\eta}\right]^{-1}\left(\overline{\mathbf{X}}^{\top}\boldsymbol{\Omega}^{-1}\overline{\mathbf{X}}\right)\left[\overline{\mathbf{X}}^{\top}\boldsymbol{\Omega}^{-1}\overline{\mathbf{X}} + \overline{\mathbf{M}}^{*}_{\eta}\right]^{-1}$$

and

$$\mathrm{Var}\big[\hat{\boldsymbol{\gamma}}\big] = \frac{4}{f_g(\zeta)-1}\Big[\overline{\mathbf{W}}^{\top}\overline{\mathbf{W}} + \overline{\mathbf{M}}_{\phi}^{*}\Big]^{-1}\Big(\overline{\mathbf{W}}^{\top}\overline{\mathbf{W}}\Big)\Big[\overline{\mathbf{W}}^{\top}\overline{\mathbf{W}} + \overline{\mathbf{M}}_{\phi}^{*}\Big]^{-1},$$

where $\overline{\mathbf{M}}_{\eta}^{*} = (1/d_g(\zeta))\overline{\mathbf{M}}_{\eta}$. An intuitive estimator of the asymptotic variance-covariance matrix of $\hat{\boldsymbol{\theta}}$ reduces to $\mathrm{Var}\big[\hat{\boldsymbol{\theta}}\big]$ evaluated at the $\boldsymbol{\theta}$ estimate.

3.4 Testing no effect of covariates

In this section, testing no effect of an explanatory variable whose effect is nonparametrically modeled, is discussed. If the interest effect was specified as $\mathrm{f}_{\eta_j}(a)$ for some $j = 1,\ldots,p'$, testing no effect of the explanatory variable a on $\boldsymbol{\eta}$, means assessing if indeed $(\mathrm{f}_{\eta_j}(a_{1,j}),\ldots,\mathrm{f}_{\eta_j}(a_{n,j}))^{\top} = \dot{\mathbf{N}}_{\eta_j}\dot{\boldsymbol{\tau}}_{\eta_j} = c\mathbf{1}_n$ for some constant c. Because $\dot{\mathbf{N}}_{\eta_j}\mathbf{1}_{p_j'} = \mathbf{1}_n$, testing no effect of a implies assessing $\mathrm{H}_0 : \dot{\boldsymbol{\tau}}_{\eta_j} = c\mathbf{1}_{p_j'}$ versus $\mathrm{H}_1 : \dot{\boldsymbol{\tau}}_{\eta_j} \neq c\mathbf{1}_{p_j'}$. Under the new parametrization, which was introduced in previous sections to avoid identification problems, testing no effect of a on $\boldsymbol{\eta}$ means assessing the following hypothesis system

$$\mathrm{H}_0 : \mathbf{C}_j^{\eta}\boldsymbol{\tau}_{\eta_j} = c\mathbf{1}_{p_j'} \quad \text{versus} \quad \mathrm{H}_1 : \mathbf{C}_j^{\eta}\boldsymbol{\tau}_{\eta_j} \neq c\mathbf{1}_{p_j'}. \tag{3.4}$$

Let $\mathbf{L}_{t+1} = [\mathbf{1}_t, -\mathbf{I}_t]$ be a contrast matrix. Thus, (3.4) may be written as $\mathrm{H}_0 : \mathbf{L}_{p_j'}\mathbf{C}_j^{\eta}\boldsymbol{\tau}_{\eta_j} = \mathbf{0}_{p_j'-1}$ versus $\mathrm{H}_1 : \mathbf{L}_{p_j'}\mathbf{C}_j^{\eta}\boldsymbol{\tau}_{\eta_j} \neq \mathbf{0}_{p_j'-1}$. However, because $\mathrm{rank}(\mathbf{L}_{p_j'}) = \mathrm{rank}(\mathbf{C}_j^{\eta}) = p_j' - 1$, (3.4) also may be written as

$$\mathrm{H}_0 : \boldsymbol{\tau}_{\eta_j} = \mathbf{0}_{p_j'-1} \quad \text{versus} \quad \mathrm{H}_1 : \boldsymbol{\tau}_{\eta_j} \neq \mathbf{0}_{p_j'-1}.$$

Therefore, the effect of a on $\boldsymbol{\eta}$ may be considered as being "null" for "small" values of the following statistic

$$\mathsf{F}_{\eta_j} = \hat{\boldsymbol{\tau}}_{\eta_j}^{\top}\mathrm{Var}^{-1}[\hat{\boldsymbol{\tau}}_{\eta_j}]\hat{\boldsymbol{\tau}}_{\eta_j},$$

which, asymptotically and under H_0, exhibits a chi-square distribution with $p_j' - 1$ degrees of freedom. Similar results hold for testing no effect of an explanatory variable b on $\boldsymbol{\phi}$, whose effect was specified by $\mathrm{f}_{\phi_r}(b)$ for some $r = 1,\ldots,q'$.

3.5 Simultaneous confidence intervals

Let $100(1-\alpha)\%$ be the desired simultaneous confidence level of $\mathcal{CI}_{1,j}^{(\alpha^*)},\ldots,\mathcal{CI}_{n,j}^{(\alpha^*)}$, where $\mathcal{CI}_{k,j}^{(\alpha^*)} = \Big\{\mathrm{f}\in\mathbb{R} : |\hat{\mathrm{f}}_{\eta_j}(a_{k,j}) - \mathrm{f}| \leq \Phi^{-1}(1-\frac{\alpha^*}{2})[\hat{\mathrm{V}}\mathrm{ar}(\hat{\mathrm{f}}_{\eta_j}(a_{k,j}))]^{\frac{1}{2}}\Big\}$ is the normality-based $100(1-\alpha^*)\%$ confidence interval of $\mathrm{f}_{\eta_j}(a_{k,j})$, with $\hat{\mathrm{V}}\mathrm{ar}(\hat{\mathrm{f}}_{\eta_j}(a_{k,j}))$ being the estimate of $\mathrm{Var}(\hat{\mathrm{f}}_{\eta_j}(a_{k,j}))$. The simultaneous confidence level of $\mathcal{CI}_{1,j}^{(\alpha^*)},\ldots,\mathcal{CI}_{n,j}^{(\alpha^*)}$ can be calculated as

$$\begin{aligned}\mathrm{P}\left\{\bigcap_{k=1}^{n}\Big[\mathrm{f}_{\eta_j}(a_{k,j}) \in \mathcal{CI}_{k,j}^{(\alpha^*)}\Big]\right\} &= 1 - \mathrm{P}\left\{\bigcup_{k=1}^{n}\Big[\mathrm{f}_{\eta_j}(a_{k,j}) \notin \mathcal{CI}_{k,j}^{(\alpha^*)}\Big]\right\}\\ &\geq 1 - \sum \mathrm{P}\Big[\mathrm{f}_{\eta_j}(a_{k,j}) \notin \mathcal{CI}_{k,j}^{(\alpha^*)}\Big]\\ &\geq 1 - \bar{n}_{\eta_j}\alpha^*,\end{aligned}$$

where $\bar{n}_{\eta_j}$ is the number of different values in $(a_{1,j},\ldots,a_{n,j})$. Thus, according to the Bonferroni method, α^* is set to $\alpha^* = \alpha/\bar{n}_{\eta_j}$ so that the simultaneous confidence level of $\mathcal{CI}_{1,j}^{(\alpha^*)},\ldots,\mathcal{CI}_{n,j}^{(\alpha^*)}$ is at least $100(1-\alpha)\%$. Similar results hold for the simultaneous confidence intervals of $\exp[\mathrm{f}_{\eta_j}(a_{k,j})]$, $\mathrm{f}_{\phi_r}(b_{k,r})$ and $\exp[\mathrm{f}_{\phi_r}(b_{k,r})]$. The simultaneous confidence intervals may be used to informally testing no effect of an explanatory variable whose effect is nonparametrically modeled. In fact, a straight line of zero slope that may be located within the simultaneous confidence intervals is an informal evidence of null effect of the interest covariate.

3.6 Degrees of freedom

This section deals with the approximate calculation of the degrees of freedom used in the estimation of $\boldsymbol{\eta}$ and $\boldsymbol{\phi}$. For this purpose, an analogy with the parametric linear case is established. Indeed, it is can be seen that, at the convergence of the Algorithm 2.1, the estimates of $\bar{\boldsymbol{\beta}}$ and $\bar{\gamma}$ under the setup (II) can be written as

$$\begin{aligned}\hat{\bar{\boldsymbol{\beta}}} &= [\mathbf{K}_{\eta}(\hat{\boldsymbol{\theta}}) + \overline{\mathbf{M}}_{\eta}]^{-1}[\mathbf{U}_{\eta}(\hat{\boldsymbol{\theta}}) + \mathbf{K}_{\eta}(\hat{\boldsymbol{\theta}})\hat{\bar{\boldsymbol{\beta}}}] \\ &= \left(\overline{\mathbf{X}}^{\top}\hat{\boldsymbol{\Omega}}^{-1}\overline{\mathbf{X}} + \overline{\mathbf{M}}_{\eta}^{*}\right)^{-1}\overline{\mathbf{X}}^{\top}\hat{\boldsymbol{\Omega}}^{-1}[\overline{\mathbf{X}}\hat{\bar{\boldsymbol{\beta}}} + (1/d_g(\zeta))\hat{\mathbf{D}}_{(\mathrm{v})}(\mathbf{y} - \hat{\boldsymbol{\mu}})]\end{aligned}$$

and

$$\begin{aligned}\hat{\bar{\gamma}} &= [\mathbf{K}_{\phi}(\hat{\boldsymbol{\theta}}) + \overline{\mathbf{M}}_{\phi}]^{-1}[\mathbf{U}_{\phi}(\hat{\boldsymbol{\theta}}) + \mathbf{K}_{\phi}(\hat{\boldsymbol{\theta}})\hat{\bar{\gamma}}] \\ &= \left(\overline{\mathbf{W}}^{\top}\overline{\mathbf{W}} + \overline{\mathbf{M}}_{\phi}^{*}\right)^{-1}\overline{\mathbf{W}}^{\top}[\overline{\mathbf{W}}\hat{\bar{\gamma}} + (2/(f_g(\zeta) - 1))(\hat{\mathbf{s}} - \mathbf{1}_n)],\end{aligned}$$

where $\hat{\boldsymbol{\Omega}}$, $\hat{\mathbf{D}}_{(\mathrm{v})}$, $\hat{\boldsymbol{\mu}} = \log(\hat{\boldsymbol{\eta}})$, and $\hat{\mathbf{s}}$ represent $\boldsymbol{\Omega}$, $\mathbf{D}_{(\mathrm{v})}$, $\boldsymbol{\mu} = \log(\boldsymbol{\eta})$, and $\mathbf{s}$ avaliated at the estimate of $\boldsymbol{\theta}$. By analogy with the parametric case, the degrees of freedom used in the median submodel are given by $\mathsf{df}(\hat{\boldsymbol{\eta}}) = \mathsf{tr}(\hat{\mathbf{H}}_{\eta})$, where $\mathsf{tr}(\mathbf{H})$ represents the trace of $\mathbf{H}$, and $\hat{\mathbf{H}}_{\eta}$ is a matrix such that $\log(\hat{\boldsymbol{\eta}}) = \overline{\mathbf{X}}\hat{\bar{\boldsymbol{\beta}}} = \hat{\mathbf{H}}_{\eta}[\mathbf{y}_{\eta}(\hat{\boldsymbol{\theta}})]$, in which $\mathbf{y}_{\eta}(\hat{\boldsymbol{\theta}}) = \overline{\mathbf{X}}\hat{\bar{\boldsymbol{\beta}}} + (1/d_g(\zeta))\hat{\mathbf{D}}_{(\mathrm{v})}(\mathbf{y} - \hat{\boldsymbol{\mu}})$ is a local response variable. According to Eilers and Marx (1996), $\mathsf{df}(\hat{\boldsymbol{\eta}})$ can be written as

$$\begin{aligned}\mathsf{df}(\hat{\boldsymbol{\eta}}) &= \mathsf{tr}\left[\overline{\mathbf{X}}\left(\overline{\mathbf{X}}^{\top}\hat{\boldsymbol{\Omega}}^{-1}\overline{\mathbf{X}} + \overline{\mathbf{M}}_{\eta}^{*}\right)^{-1}\overline{\mathbf{X}}^{\top}\hat{\boldsymbol{\Omega}}^{-1}\right] \\ &= \mathsf{tr}\left\{\left(\overline{\mathbf{X}}^{\top}\hat{\boldsymbol{\Omega}}^{-1}\overline{\mathbf{X}} + \overline{\mathbf{M}}_{\eta}^{*}\right)^{-1}\overline{\mathbf{X}}^{\top}\hat{\boldsymbol{\Omega}}^{-1}\overline{\mathbf{X}}\right\} \\ &= \mathsf{tr}\left\{\left[\mathbf{I} + \mathbf{Q}^{-\frac{1}{2}}\overline{\mathbf{M}}_{\eta}^{*}\mathbf{Q}^{-\frac{1}{2}}\right]^{-1}\right\} \\ &= \sum_{i=1}^{\mathsf{dim}(\overline{\mathbf{M}}_{\eta}^{*})}\frac{1}{1 + \alpha_i^{(\eta)}},\end{aligned}$$

where $\alpha_i^{(\eta)} \geq 0$ are the eigenvalues of the nonnegative definite matrix $\mathbf{Q}^{-\frac{1}{2}}\overline{\mathbf{M}}_{\eta}^{*}\mathbf{Q}^{-\frac{1}{2}}$, $\mathbf{Q}^{\frac{1}{2}}$ is a positive definite matrix such that $\overline{\mathbf{X}}^{\top}\hat{\boldsymbol{\Omega}}^{-1}\overline{\mathbf{X}} = \mathbf{Q}^{\frac{1}{2}}\mathbf{Q}^{\frac{1}{2}}$, and $\mathsf{dim}(\overline{\mathbf{M}}_{\eta}^{*})$ is the number of rows (or columns) of $\overline{\mathbf{M}}_{\eta}^{*}$. Note that the first eigenvalues of $\mathbf{Q}^{-\frac{1}{2}}\overline{\mathbf{M}}_{\eta}^{*}\mathbf{Q}^{-\frac{1}{2}}$ and $\overline{\mathbf{M}}_{\eta}^{*}$ are zero. Therefore,

$$\mathsf{df}(\hat{\boldsymbol{\eta}}) = p + \sum_{i=p+1}^{\mathsf{dim}(\overline{\mathbf{M}}_{\eta}^{*})}\frac{1}{1 + \alpha_i^{(\eta)}}.$$

Moreover, because the number of eigenvalues equal to zero of $\mathbf{Q}^{-\frac{1}{2}}\overline{\mathbf{M}}_{\eta}^{*}\mathbf{Q}^{-\frac{1}{2}}$ and $\overline{\mathbf{M}}_{\eta}^{*}$ coincide, it follows that $\mathsf{dim}(\overline{\mathbf{M}}_{\eta}^{*}) > \mathsf{df}(\hat{\boldsymbol{\eta}}) > p + \sum_{j=1}^{p'}\alpha_{\eta j}^{(0)}$, where $\alpha_{\eta j}^{(0)}$ is the number of eigenvalues of $\mathbf{M}_{\eta_j}$ equal to zero. Finally, also by analogy with the parametric case, the number of degrees of freedom associated with the i-th covariate corresponds to the i-th element of the main diagonal of $(\overline{\mathbf{X}}^{\top}\hat{\boldsymbol{\Omega}}^{-1}\overline{\mathbf{X}} + \overline{\mathbf{M}}_{\eta}^{*})^{-1}\overline{\mathbf{X}}^{\top}\hat{\boldsymbol{\Omega}}^{-1}\overline{\mathbf{X}}$.

Similarly, because $\log(\hat{\boldsymbol{\phi}}) = \overline{\mathbf{W}}\hat{\bar{\gamma}} = \hat{\mathbf{H}}_{\phi}[\mathbf{y}_{\phi}(\hat{\boldsymbol{\theta}})]$, where $\hat{\mathbf{H}}_{\phi} = \overline{\mathbf{W}}\left(\overline{\mathbf{W}}^{\top}\overline{\mathbf{W}} + \overline{\mathbf{M}}_{\phi}^{*}\right)^{-1}\overline{\mathbf{W}}^{\top}$, the degrees of freedom used in the skewness submodel are given by

$$\begin{aligned}\mathsf{df}(\hat{\boldsymbol{\phi}}) &= \mathsf{tr}\Big\{\big(\overline{\mathbf{W}}^\top\overline{\mathbf{W}} + \overline{\mathbf{M}}_\phi^*\big)^{-1}\overline{\mathbf{W}}^\top\overline{\mathbf{W}}\Big\}\\ &= \mathsf{tr}\Big\{\big[\mathbf{I} + \mathbf{Q}^{-\frac{1}{2}}\overline{\mathbf{M}}_\phi^*\mathbf{Q}^{-\frac{1}{2}}\big]^{-1}\Big\}\\ &= q + \sum_{i=q+1}^{\mathsf{dim}(\overline{\mathbf{M}}_\phi^*)}\frac{1}{1+\alpha_i^{(\phi)}},\end{aligned}$$

where $\alpha_i^{(\phi)} \geq 0$ are the eigenvalues of the nonnegative definite matrix $\mathbf{Q}^{-\frac{1}{2}}\overline{\mathbf{M}}_\phi^*\mathbf{Q}^{-\frac{1}{2}}$, and $\mathbf{Q}^{\frac{1}{2}}$ is a positive definite matrix such that $\overline{\mathbf{W}}^\top\overline{\mathbf{W}} = \mathbf{Q}^{\frac{1}{2}}\mathbf{Q}^{\frac{1}{2}}$. Because the number of eigenvalues equal to zero of $\mathbf{Q}^{-\frac{1}{2}}\overline{\mathbf{M}}_\phi^*\mathbf{Q}^{-\frac{1}{2}}$ and $\overline{\mathbf{M}}_\phi^*$ coincide, it follows that $\mathsf{dim}(\overline{\mathbf{M}}_\phi^*) > \mathsf{df}(\hat{\boldsymbol{\phi}}) > q + \sum_{r=1}^{q'}\alpha_{\phi_r}^{(0)}$, where $\alpha_{\phi_r}^{(0)}$ is the number of eigenvalues of $\mathbf{M}_{\phi_r}$ equal to zero. Moreover, the number of degrees of freedom associated with the i-th covariate corresponds to the i-th element of the main diagonal of $(\overline{\mathbf{W}}^\top\overline{\mathbf{W}} + \overline{\mathbf{M}}_\phi^*)^{-1}\overline{\mathbf{W}}^\top\overline{\mathbf{W}}$.

3.7 Choosing the smoothing parameter

Choosing the smoothing parameter by using a criterion that ensures a compromise between "low" model complexity and "high" goodness-of-fit was considered by Hastie and Tibshirani (1990), Rigby and Stasinopoulos (2005), Wood (2006) and Wu and Yu (2014). Then, in this work, the value of the smoothing parameter $\boldsymbol{\lambda}$ is chosen by minimizing the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC) through an outer iteration. The AIC and BIC criteria are given by

$$\mathsf{AIC}(\hat{\boldsymbol{\theta}}|\boldsymbol{\lambda}) = -2\mathsf{L}(\hat{\boldsymbol{\theta}}|\boldsymbol{\lambda}) + 2[\mathsf{df}(\hat{\boldsymbol{\eta}}|\boldsymbol{\lambda}) + \mathsf{df}(\hat{\boldsymbol{\phi}}|\boldsymbol{\lambda})]$$

and

$$\mathsf{BIC}(\hat{\boldsymbol{\theta}}|\boldsymbol{\lambda}) = -2\mathsf{L}(\hat{\boldsymbol{\theta}}|\boldsymbol{\lambda}) + \log(n)[\mathsf{df}(\hat{\boldsymbol{\eta}}|\boldsymbol{\lambda}) + \mathsf{df}(\hat{\boldsymbol{\phi}}|\boldsymbol{\lambda})],$$

where $\hat{\boldsymbol{\theta}}|\boldsymbol{\lambda}$, $\hat{\boldsymbol{\eta}}|\boldsymbol{\lambda}$ and $\hat{\boldsymbol{\phi}}|\boldsymbol{\lambda}$ are the estimates of $\boldsymbol{\theta}$, $\boldsymbol{\eta}$ and $\boldsymbol{\phi}$ given a particular value of the smoothing parameter $\boldsymbol{\lambda}$, respectively.

3.8 Simulation Results I

This section presents a simulation study to assess the statistical properties of the maximum penalized likelihood estimates in log-symmetric regression models under two scenarios, denoted by A and B. For this purpose, a data set of size n is simulated in each scenario, in which the response variable is generated from a log-symmetric distribution, where its median (η) and its skewness (or relative dispersion) (ϕ) can be written as

$$\overset{(A)}{\begin{cases}\eta = \exp(\beta_1 x_1 + \beta_3 a^{\beta_2}),\\ \log(\phi) = \gamma_1 x_1 + \mathrm{f}_\phi(a),\end{cases}} \qquad \overset{(B)}{\begin{cases}\log(\eta) = \beta_1 x_1 + \beta_2 x_2 + \mathrm{f}_\eta(a),\\ \log(\phi) = \gamma_1 x_1 + \mathrm{f}_\phi(a),\end{cases}} \tag{3.5}$$

under the scenarios A and B, respectively, in which $x_1 \sim \text{Bernoulli}(0.5)$, $x_2 \sim \text{log-normal}(1,1)$, $\mathrm{f}_\eta(a) = 5a + \sin(2\pi a)$, $\mathrm{f}_\phi(a) = 1.2[1.166 - \sin(\pi a)]$, and a is a sequence of seventy values in the interval $[0.05, 0.95]$, which is replicated several times until the sample size is reached. The values assigned to the parameters are $\beta_1 = 2$, $\beta_2 = 0.5$, $\beta_3 = 10$ and $\gamma_1 = -0.2$. To describe the random component, several log-symmetric distributions (e.g., log-normal, log-Student-t, log-slash, log-hyperbolic, log-power-exponential, log-contaminated-normal, Birnbaum-Saunders and Birnbaum-Saunders-t) and several values of their extra parameters are considered. The most of these distributions exhibit heavier tails than those of the log-normal distribution. The generated sample is used to fit a log-symmetric model with the systematic component described in (3.5); however, the functional forms of $\mathrm{f}_\eta(\cdot)$ and $\mathrm{f}_\phi(\cdot)$ are assumed to be unknown. Then, under the scenario A described in (3.5), the function $\mathrm{f}_\phi(\cdot)$ is approximated by a natural cubic spline with eight knots (given by $q(a,0/7), q(a,1/7), q(a,2/7), \ldots, q(a,7/7)$, in which $q(a,t)$ is the quantile of order t of a). Similarly, under the scenario B described in (3.5), the functions $\mathrm{f}_\eta(\cdot)$ and $\mathrm{f}_\phi(\cdot)$ are approximated by cubic P-splines with eight internal knots (given by $q(a,0/7), q(a,1/7), q(a,2/7), \ldots, q(a,7/7)$), and a *difference penalty* term of order 2. The smoothing parameters are chosen by minimizing the AIC criterion. This process is replicated $R = 5000$ times, keeping the values of x_1, x_2 and a fixed. For the R estimates of β_1, β_2, β_3 and γ_1, the following summary measures are calculated: i) empirical expected value, i.e., $\bar{\hat{\theta}} = R^{-1}\sum_{i=1}^{R}\hat{\theta}^{(i)}$, where $\hat{\theta}^{(i)}$ is the estimate of $\hat{\theta}$ in the i-th replication; ii) coverage rate of the normality-based 95% confidence interval, i.e., $\mathrm{CR}(\hat{\theta}) = 100\times R^{-1}\sum_{i=1}^{R}\mathrm{I}\Big[|\hat{\theta}^{(i)} - \theta\mid/[\hat{\mathrm{Var}}^{(i)}(\hat{\theta})]^{\frac{1}{2}}, [0, 1.96]\Big]$, where $\hat{\mathrm{Var}}^{(i)}(\hat{\theta})$ is the estimate of $\mathrm{Var}(\hat{\theta})$ in the i-th replication, and $\mathrm{I}[x,\Theta] = 1$ if $x \in \Theta$ and $\mathrm{I}[x,\Theta] = 0$ in other cases; and iii) p-value of the one-sample Kolmogorov-Smirnov test (see Conover, 1971) to judge the normality of the sample $\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(R)}$. Additionally, as a summary measure of the R estimates of the nonparametric functions $\mathrm{f}_\eta(\cdot)$ and $\mathrm{f}_\phi(\cdot)$, the coverage rate of the simultaneous normality-based 95% confidence intervals is used, i.e., $\mathrm{CR}(\hat{\mathrm{f}}) = 100 \times R^{-1}\sum_{i=1}^{R}\prod_{k=1}^{n}\mathrm{I}\Big[|\hat{\mathrm{f}}^{(i)}(a_k) - \mathrm{f}(a_k)|/[\hat{\mathrm{Var}}^{(i)}(\hat{\mathrm{f}}(a_k))]^{\frac{1}{2}}, \Theta\Big]$, where $\Theta = [0, \Phi^{-1}(1 - 0.05/2\bar{n})]$, $\bar{n} = 70$ is the number of different values of a in the sample, and $\hat{\mathrm{f}}^{(r)}(a_k)$ and $\hat{\mathrm{Var}}^{(r)}(\hat{\mathrm{f}}(a_k))$ are the estimates of $\mathrm{f}(a_k)$ and $\mathrm{Var}(\hat{\mathrm{f}}(a_k))$ in the i-th replication, respectively. The results are presented in Tables 3.1 and 3.2 under the scenarios A and B, respectively, described in (3.5).

The results indicate that the empirical expected values are close to the true values of the parameters. The coverage rates of the 95% confidence intervals for β_1, β_2, β_3 and γ_1 are close to the nominal values, mainly under scenario A. In addition, the Kolmogorov-Smirnov test indicates that the empirical distributions of the estimates of β_1, β_2, β_3 and γ_1 are fairly close to the normal distribution. Also, the coverage rates of the simultaneous 95% confidence intervals for $\mathrm{f}_\eta(\cdot)$ and $\mathrm{f}_\phi(\cdot)$ are higher than 90%, except under the Birnbaum-Saunders model in the scenario B. In conclusion, the maximum penalized likelihood estimators of β_1, β_2, β_3 and γ_1 seem to be approximately unbiased, and their distributions seem to be approximately normal, even though: i) the size of the sample is not very large; ii) the systematic component of the fitted models is slightly complex; iii) in most cases, the distribution of the random error exhibits heavier tails than those of the log-normal distribution; and iv) the value of the smoothing parameter is unknown but estimated by minimizing the AIC criterion. However, the variance of these estimators seem to be slightly underestimated, especially under the Birnbaum-Saunders model. In addition, the interval estimates of $\mathrm{f}_\eta(\cdot)$ and $\mathrm{f}_\phi(\cdot)$ exhibit good behavior because in most cases, their coverage rates are close to the nominal values, mainly under scenario A.

3.9 Diagnostic methods

In this section, some diagnostic methods such as deviance-type residuals for the median and the skewness (or relative dispersion) submodels, overall goodness-of-fit criterion, and local influence measures under log-symmetric regression models, are addressed.

3.9.1 Individual goodness-of-fit

To evaluate the goodness-of-fit of the median and skewness (or relative dispersion) submodels, the measure known in the statistical literature as *deviance* is considered. Then, the deviance for $\hat{\boldsymbol{\eta}}$ given $\hat{\boldsymbol{\phi}}$, denoted as $\mathsf{D}(\hat{\boldsymbol{\eta}}|\hat{\boldsymbol{\phi}})$, and the deviance for $\hat{\boldsymbol{\phi}}$ given $\hat{\boldsymbol{\eta}}$, denoted as $\mathsf{D}(\hat{\boldsymbol{\phi}}|\hat{\boldsymbol{\eta}})$, are defined. The deviance value is always non-negative, and the lower is its value, the better is the goodness-of-fit of the assessed submodel. Thus, the deviance-type residuals (see, e.g., Davison and Gigli, 1989;

Table 3.1: *Results of the simulation study I under the scenario A and* $n = 140$.

Error distribution	$\hat{\theta}$				$\mathrm{CR}(\hat{\theta})$				p-value of K-S test				$\mathrm{CR}(\hat{\mathrm{f}})$
	β_1	β_2	β_3	γ_1	β_1	β_2	β_3	γ_1	β_1	β_2	β_3	γ_1	f_ϕ
log-normal	2.00	0.50	10.00	-0.20	93.42	92.28	92.10	93.64	0.81	0.62	0.79	0.50	89.24
log-Student-t(4)	2.00	0.50	10.00	-0.21	93.10	92.50	92.12	93.58	0.92	0.21	0.94	0.51	93.74
log-Student-t(6)	2.00	0.50	10.00	-0.20	92.78	92.64	92.98	93.76	0.99	0.31	0.89	0.97	92.20
log-Student-t(8)	2.00	0.50	10.00	-0.20	93.60	92.90	92.26	94.14	0.91	0.04	0.78	0.99	91.78
log-Student-t(10)	2.00	0.50	10.01	-0.20	93.56	92.44	92.28	94.32	0.31	0.96	0.90	0.96	92.34
log-power-exp.(0.2)	2.00	0.50	10.00	-0.20	93.36	92.98	92.58	93.50	0.90	0.50	0.34	1.00	90.72
log-power-exp.(0.3)	2.00	0.50	10.01	-0.20	93.18	92.54	92.56	93.62	0.84	0.30	0.84	0.95	90.58
log-power-exp.(0.4)	2.00	0.50	10.01	-0.20	93.40	92.74	92.36	94.12	0.69	0.14	0.75	0.73	91.36
log-power-exp.(0.5)	2.00	0.50	10.01	-0.21	93.00	92.22	91.66	93.48	0.44	0.10	0.91	0.68	91.26
log-hyperbolic(1.0)	1.99	0.50	10.01	-0.20	93.00	92.82	92.70	94.04	0.53	0.62	0.22	0.92	92.46
log-hyperbolic(0.9)	2.00	0.50	10.00	-0.20	93.46	92.52	92.48	94.36	0.69	0.30	0.34	0.67	91.70
log-hyperbolic(0.8)	2.00	0.50	10.01	-0.20	93.62	93.10	92.88	94.50	0.53	0.29	0.13	0.47	92.82
log-hyperbolic(0.7)	2.00	0.50	10.00	-0.20	94.42	92.50	92.44	94.02	0.62	0.04	0.76	0.56	92.60
log-slash(1.5)	2.00	0.50	10.01	-0.20	93.42	92.16	93.14	94.36	1.00	0.14	0.53	1.00	92.84
log-slash(1.4)	2.00	0.50	10.01	-0.20	93.92	92.16	92.60	94.54	0.56	0.48	0.99	0.54	92.64
log-slash(1.3)	2.01	0.50	10.00	-0.21	93.56	92.64	92.32	94.46	0.73	0.43	0.70	0.64	93.26
log-slash(1.2)	2.00	0.50	10.00	-0.21	93.63	92.30	91.96	94.06	0.95	0.19	0.95	0.79	92.92
log-cont-nor(0.3,0.3)	2.00	0.50	10.00	-0.20	93.12	92.30	92.34	94.60	1.00	0.08	0.58	0.80	91.64
log-cont-nor(0.5,0.3)	2.00	0.50	10.00	-0.20	93.02	91.92	92.08	94.78	0.99	0.14	0.87	0.77	91.00
log-cont-nor(0.7,0.3)	2.00	0.50	10.01	-0.20	93.18	91.92	92.06	94.34	0.87	0.02	0.90	0.40	90.64
log-cont-nor(0.9,0.3)	2.00	0.50	10.01	-0.20	93.26	92.30	92.56	94.26	0.94	0.03	0.76	0.80	90.10
B-S(0.1)	2.00	0.50	10.00	-0.20	93.48	92.48	91.92	93.64	0.83	1.00	0.78	0.52	89.24
B-S(0.3)	2.00	0.50	10.00	-0.20	93.38	92.34	91.92	93.66	0.78	0.96	0.68	0.36	88.90
B-S(0.5)	2.00	0.50	10.00	-0.20	93.56	92.34	91.68	93.50	0.88	0.98	0.85	0.72	88.28
B-S(0.7)	2.00	0.50	10.00	-0.20	93.68	92.08	92.00	93.42	0.43	0.97	0.58	0.89	87.80
B-S-t(0.1,4)	2.00	0.50	10.00	-0.20	93.58	92.62	92.38	93.46	0.68	0.81	0.89	0.95	93.34
B-S-t(0.3,4)	2.00	0.50	10.00	-0.20	93.52	92.64	92.48	93.36	0.79	0.93	0.97	0.99	92.96
B-S-t(0.5,4)	2.00	0.50	10.00	-0.21	93.63	92.70	92.59	93.85	0.74	0.97	0.85	0.97	92.70
B-S-t(0.7,4)	2.00	0.50	10.00	-0.21	93.93	92.72	93.13	93.89	0.80	0.94	0.96	0.89	90.71

Pierce and Shafer, 1986) for the median and the skewness (or relative dispersion) submodels are defined as the signed square root of the contribution to the deviance of each individual. The residuals may be employed to identify observations marginally discrepant and to assess the appropriateness of the proposed submodel.

Goodness-of-fit of the median submodel

The deviance for $\hat{\boldsymbol{\eta}}$ given $\hat{\boldsymbol{\phi}}$ reduces to

$$\mathsf{D}(\hat{\boldsymbol{\eta}}|\hat{\boldsymbol{\phi}}) = 2\sum_{k=1}^{n}\Big[\mathsf{L}_k(\tilde{\mu}_k, \hat{\phi}_k) - \mathsf{L}_k(\hat{\mu}_k, \hat{\phi}_k)\Big]\,,$$

where $\tilde{\mu}_k$ is the value of $\mu \in \mathbb{R}$ that maximizes the function $\mathsf{L}_k(\mu, \hat{\phi}_k)$. Hence, $\mathsf{D}(\hat{\boldsymbol{\eta}}|\hat{\boldsymbol{\phi}})$ becomes

$$\mathsf{D}(\hat{\boldsymbol{\eta}}|\hat{\boldsymbol{\phi}}) = \sum_{k=1}^{n}\mathsf{d}_k(\hat{\boldsymbol{\eta}}|\hat{\boldsymbol{\phi}}),$$

Table 3.2: *Results of the simulation study I under the scenario B and n = 210*

Error distribution	$\bar{\hat{\theta}}$			CR($\hat{\theta}$)			p-value of K-S test			CR($\hat{\mathrm{f}}$)	
	β_1	β_2	γ_1	β_1	β_2	γ_1	β_1	β_2	γ_1	f_η	f_ϕ
log-normal	1.98	0.50	-0.21	93.56	92.80	91.66	0.98	0.55	0.75	93.12	90.00
log-Student-t(4)	1.97	0.50	-0.21	92.86	93.10	92.66	0.84	0.48	0.46	90.94	91.82
log-Student-t(6)	1.98	0.50	-0.20	93.04	93.00	93.92	0.87	0.88	0.80	91.74	91.26
log-Student-t(8)	1.98	0.50	-0.21	93.18	93.18	92.64	0.96	0.94	0.83	92.14	91.22
log-Student-t(10)	1.98	0.50	-0.20	93.28	93.28	92.94	0.81	0.94	0.79	92.72	90.88
log-power-exp.(0.2)	1.98	0.50	-0.21	92.44	92.54	92.78	0.56	0.84	0.99	91.46	90.18
log-power-exp.(0.3)	1.97	0.50	-0.21	93.82	92.60	93.08	0.96	0.61	0.93	90.26	90.42
log-power-exp.(0.4)	1.98	0.50	-0.21	92.84	92.60	92.40	0.90	0.33	0.99	90.28	91.18
log-power-exp.(0.5)	1.97	0.50	-0.21	92.06	92.14	92.80	1.00	0.93	0.94	90.06	91.60
log-hyperbolic(1.0)	1.98	0.50	-0.21	93.20	92.61	92.96	0.87	0.99	0.98	90.43	90.92
log-hyperbolic(0.9)	1.98	0.50	-0.20	93.26	92.84	92.88	0.80	0.94	0.97	90.20	91.82
log-hyperbolic(0.8)	1.99	0.50	-0.21	92.60	91.98	93.10	0.87	0.20	0.95	91.10	91.28
log-hyperbolic(0.7)	1.99	0.50	-0.21	93.06	92.38	93.50	0.77	0.20	0.66	91.04	92.22
log-slash(1.5)	1.99	0.50	-0.21	93.60	92.66	93.36	0.70	0.74	0.84	90.68	91.38
log-slash(1.4)	1.99	0.50	-0.21	92.47	92.59	93.15	0.99	0.33	0.98	90.00	91.66
log-slash(1.3)	1.98	0.50	-0.20	92.63	92.29	93.12	0.97	0.92	0.86	90.05	90.85
log-slash(1.2)	1.97	0.50	-0.21	92.94	92.71	92.31	0.48	0.95	0.99	90.67	91.86
log-cont-nor(0.3,0.3)	1.99	0.50	-0.21	92.98	92.10	92.80	0.77	0.53	0.98	90.40	90.94
log-cont-nor(0.5,0.3)	1.98	0.50	-0.21	93.16	92.22	92.38	0.69	0.57	0.49	90.80	90.84
log-cont-nor(0.7,0.3)	1.99	0.50	-0.21	93.18	92.22	91.76	0.77	0.64	0.38	90.38	90.28
log-cont-nor(0.9,0.3)	1.99	0.50	-0.21	93.28	92.38	92.40	0.47	0.83	0.84	90.70	90.92
B-S(0.1)	2.00	0.50	-0.21	93.04	92.28	90.60	0.92	0.37	0.68	91.60	87.68
B-S(0.3)	2.00	0.50	-0.21	93.28	92.38	91.00	0.68	0.78	0.69	91.80	88.36
B-S(0.5)	2.00	0.50	-0.21	93.24	92.28	91.12	0.88	0.98	0.67	92.34	88.06
B-S(0.5)	1.99	0.50	-0.21	93.16	92.32	90.82	0.97	0.93	0.77	92.94	87.72
B-S-t(0.1,4)	2.00	0.50	-0.21	92.82	91.22	92.30	0.72	0.52	0.87	91.20	90.64
B-S-t(0.3,4)	2.00	0.50	-0.20	92.70	91.40	93.02	0.84	0.47	0.80	91.28	91.04
B-S-t(0.5,4)	2.00	0.50	-0.21	92.88	91.50	93.12	0.87	0.75	0.74	92.14	90.80
B-S-t(0.7,4)	1.99	0.50	-0.21	92.84	91.56	93.24	0.45	0.54	0.97	93.14	90.14

where $\mathsf{d}_k(\hat{\boldsymbol{\eta}}|\hat{\boldsymbol{\phi}})$ may be interpreted as the contribution of the k-th observation to the deviance of $\hat{\boldsymbol{\eta}}$ given $\hat{\boldsymbol{\phi}}$. Therefore, $\mathsf{d}_k(\hat{\boldsymbol{\eta}}|\hat{\boldsymbol{\phi}})$ may be used to define a residual associated with $\hat{\eta}_k$ (i.e., a measure of the individual goodness-of-fit in the median submodel) as follows:

$$t_\eta(\hat{z}_k) = \mathrm{sign}(\hat{z}_k)\Big[\mathsf{d}_k(\hat{\boldsymbol{\eta}}|\hat{\boldsymbol{\phi}})\Big]^{\frac{1}{2}}.$$

If the function $g(\cdot)$ is monotically decreasing for $u \geq 0$, then $\mathsf{d}_k(\hat{\boldsymbol{\eta}}|\hat{\boldsymbol{\phi}}) = 2\log[g(0)/g(\hat{z}_k^2)]$, where $\hat{z}_k = (y_k - \hat{\mu}_k)/[\hat{\phi}_k]^{\frac{1}{2}}$, $k = 1, \ldots, n$. The residual $t_\eta(\hat{z}_k)$ is an odds and twice differentiable function of $\hat{z}_k$. Hence, the distribution of $t_\eta(\hat{z}_k)$ is symmetric around zero if the distribution of $\hat{z}_k$ is symmetric around zero. Table 3.3 provides the expressions of $\mathsf{d}_k(\hat{\boldsymbol{\eta}}|\hat{\boldsymbol{\phi}})$ for some log-symmetric distributions.

Table 3.3: *Expressions of* $\mathsf{d}_k(\hat{\boldsymbol{\eta}}|\hat{\boldsymbol{\phi}})$ *for some log-symmetric distributions.*

Distribution	$\mathsf{d}_k(\hat{\boldsymbol{\eta}}|\hat{\boldsymbol{\phi}})$
log-normal	$\hat{z}_k^2$
log-Student-t	$(\zeta+1)\log\left(1+\dfrac{\hat{z}_k^2}{\zeta}\right)$
log-power-exponential	$|\hat{z}_k|^{2/(\zeta+1)}$
log-hyperbolic	$2\zeta\left[\sqrt{1+\hat{z}_k^2}-1\right]$
Binrbaum Saunders	$\dfrac{4}{\zeta^2}\sinh^2(\hat{z}_k)-\log\left[\cosh^2(\hat{z}_k)\right],\ \ \zeta<2$

Goodness-of-fit of the skewness (or the relative dispersion) submodel

The deviance for $\hat{\boldsymbol{\phi}}$ given $\hat{\boldsymbol{\eta}}$ reduces to

$$\mathsf{D}(\hat{\boldsymbol{\phi}}|\hat{\boldsymbol{\eta}}) = 2\sum_{k=1}^{n}\left[\mathsf{L}_k(\hat{\mu}_k,\tilde{\phi}_k)-\mathsf{L}_k(\hat{\mu}_k,\hat{\phi}_k)\right],$$

where $\tilde{\phi}_k$ is the value of $\phi\in\mathbb{R}^+$ that maximizes the function $\mathsf{L}_k(\hat{\mu}_k,\phi)$. Then, $\mathsf{D}(\hat{\boldsymbol{\phi}}|\hat{\boldsymbol{\eta}})$ becomes

$$\mathsf{D}(\hat{\boldsymbol{\phi}}|\hat{\boldsymbol{\eta}}) = \sum_{k=1}^{n}\mathsf{d}_k(\hat{\boldsymbol{\phi}}|\hat{\boldsymbol{\eta}}),$$

where $\mathsf{d}_k(\hat{\boldsymbol{\phi}}|\hat{\boldsymbol{\eta}}) = 2\log[g(\varrho^2)/g(\hat{z}_k^2)]-\log[\hat{z}_k^2/\varrho^2]$ is a non-negative and monotone increasing function of the difference between $\hat{z}_k^2$ and ϱ^2, which may be interpreted as the contribution of the k-th observation to the deviance of $\hat{\boldsymbol{\phi}}$ given $\hat{\boldsymbol{\eta}}$. If the function $\mathsf{L}_k(\hat{\mu}_k,\phi)$ has just one critic point, then ϱ is the solution of the equation $\mathrm{v}(\varrho)\varrho^2=1$. $\mathsf{d}_k(\hat{\boldsymbol{\phi}}|\hat{\boldsymbol{\eta}})$ may be used to define a residual associated with $\hat{\phi}_k$ (i.e., a measure of the individual goodness-of-fit in the skewness (or relative dispersion) submodel) as follows:

$$t_\phi(\hat{z}_k) = \mathrm{sign}(\hat{z}_k)[\mathsf{d}_k(\hat{\boldsymbol{\phi}}|\hat{\boldsymbol{\eta}})]^{\frac{1}{2}}.$$

The residual $t_\phi(\hat{z}_k)$ is an odds function of $\hat{z}_k$ such that $t_\phi(\pm|\varrho|)=0$. Thus, $t_\phi(\hat{z}_k)$ is symmetric around zero if $\hat{z}_k$ is symmetric around zero. Table 3.4 provides the expressions of $\mathsf{d}_k(\hat{\boldsymbol{\phi}}|\hat{\boldsymbol{\eta}})$ for some log-symmetric distributions.

Table 3.4: *Expressions of* $\mathsf{d}_k(\hat{\boldsymbol{\phi}}|\hat{\boldsymbol{\eta}})$ *and* ϱ^2 *for some log-symmetric distributions.*

Distribution	$\mathsf{d}_k(\hat{\boldsymbol{\phi}}|\hat{\boldsymbol{\eta}})$	ϱ^2
log-normal	$\hat{z}_k^2-1-\log(\hat{z}_k^2)$	1
log-Student-t	$(\zeta+1)\log\left(\dfrac{\zeta+\hat{z}_k^2}{\zeta+1}\right)-\log\left(\hat{z}_k^2\right)$	1
log-power-exponential	$|\hat{z}_k|^{2/(\zeta+1)}-(1+\zeta)-\log\left[\hat{z}_k^2(1+\zeta)^{-(1+\zeta)}\right]$	$(1+\zeta)^{(1+\zeta)}$
log-hyperbolic	$2\zeta\left[\sqrt{1+\hat{z}_k^2}-\sqrt{1+\varrho^2}\right]-\log(\hat{z}_k^2/\varrho^2)$	$\frac{1+\sqrt{1+4\zeta^2}}{2\zeta^2}$

3.9.2 Overall goodness-of-fit

The overall goodness-of-fit is measured through the following statistic, which has the advantage of graphical representation as it is based on the quantile-quantile plot (see, e.g., Waller and Turnbull , 1992):

$$\Upsilon = n^{-1}\sum_{k=1}^{n}\left|\Phi^{-1}[F_{\xi^*}(\hat{z}^{(k)})] - \upsilon^{(k)}\right|,$$

where $F_{\xi^*}(\cdot)$ is the cumulative distribution function of $\log(\xi)$; $\hat{z}^{(k)}$ is the k-th order statistic of $\hat{z}_1, \ldots, \hat{z}_n$; and $\upsilon^{(k)}$ is the expectation of the k-th order statistic in a sample of size n of the standard normal distribution. The quantile residuals (see, e.g., Cysneiros and Vanegas , 2008; Dunn and Smith , 1996) are given by $\Phi^{-1}[F_{\xi^*}(\hat{z}_k)]$, $k = 1, \ldots, n$, and are used here as overall residuals (i.e., a measure of the individual goodness-of-fit). Therefore, if the estimates coincide with the true values for $\boldsymbol{\eta}$, $\boldsymbol{\phi}$ and ζ, the order statistics of the overall residuals, given by $\left\{\Phi^{-1}[F_{\xi^*}(\hat{z}^{(1)})], \ldots, \Phi^{-1}[F_{\xi^*}(\hat{z}^{(n)})]\right\}$, represent an ordered random sample from the standard normal distribution. The smaller is the value of Υ, the better is the goodness-of-fit. Graphically, the criterion Υ indicates that the smaller is the difference between the normal Q-Q plot of the overall residuals and a straight line (with zero intercept and unit slope), the better is the goodness-of-fit. In addition, if ζ is unknown, its value may be selected by minimizing the Υ statistic as illustrated by Vanegas and Paula (2014b).

3.9.3 Influence or sensitivity analysis

The general idea of diagnostic methods is to study changes in the model estimates under perturbations in the model/data. Individual cases or clusters that, when deleted, lead to substantial changes in the model estimates, particularly inferential changes, are classified as influential. The perturbed log-likelihood function is the usual way to study the influence of perturbations in the model/data on the parameter estimates. A natural extension to semi-parametric models is to consider the penalized log-likelihood function as follows

$$\mathsf{PL}^*(\boldsymbol{\theta}|\boldsymbol{\omega}) = \mathsf{L}(\boldsymbol{\theta}|\boldsymbol{\omega}) + \mathsf{P}(\boldsymbol{\theta}),$$

where $\mathsf{PL}^*(\boldsymbol{\theta}|\boldsymbol{\omega})$ denotes the perturbed log-likelihood function and $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^\top$ is the perturbation vector. A general procedure that does not require the elimination of observations, proposed by Cook (1986) and called *local influence*, consists of studying the influence of small perturbations in the model/data on the parameter estimates. The idea is to study the behavior of the penalized likelihood displacement $\mathsf{LD}_{\boldsymbol{\omega}} = 2\{\mathsf{PL}(\hat{\boldsymbol{\theta}}) - \mathsf{PL}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}})\}$ around the no perturbation vector $\boldsymbol{\omega}_0$, such that $\mathsf{PL}(\boldsymbol{\theta}) = \mathsf{PL}^*(\boldsymbol{\theta}|\boldsymbol{\omega}_0)$, where $\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$ denotes the maximum penalized likelihood estimate under the perturbed model. The suggestion of Cook (1986) is to consider the normal curvature in the direction $\mathbf{d}$, such that $\|\mathbf{d}\| = 1$, defined as

$$\mathsf{C}_d(\hat{\boldsymbol{\theta}}) = 2|\mathbf{d}^\top \Delta^\top(\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}_0)[\mathbf{J}(\hat{\boldsymbol{\theta}})]^{-1}\Delta(\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}_0)\,\mathbf{d}|, \tag{3.6}$$

where $\Delta(\boldsymbol{\theta}, \boldsymbol{\omega}) = \partial^2\mathsf{PL}(\boldsymbol{\theta}|\boldsymbol{\omega})/\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^\top$ and $\mathbf{J}(\boldsymbol{\theta}) = \partial^2\mathsf{PL}^*(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$. To have a curvature invariant under uniform change of scale, Poon and Poon (1999) proposed the conformal normal curvature defined as $\mathsf{C}^*_d(\hat{\boldsymbol{\theta}}) = \mathsf{C}_d(\hat{\boldsymbol{\theta}})/2\sqrt{\mathsf{tr}(\mathbf{V}^\top\mathbf{V})}$, where $\mathbf{V} = \Delta^\top(\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}_0)[\mathbf{J}(\hat{\boldsymbol{\theta}})]^{-1}\Delta(\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}_0)$. This curvature allows that $0 \leq \mathsf{C}^*_d(\hat{\boldsymbol{\theta}}) \leq 1$ for any unitary direction $\mathbf{d}$. A maximum curvature, denoted by $\mathsf{C}^*_{d_{\max}}$, is obtained in the direction $\mathbf{d}^*_{\max}$, where $\mathsf{C}^*_{d_{\max}}$ is the largest eigenvalue of $\mathbf{V}/\sqrt{\mathsf{tr}(\mathbf{V}^\top\mathbf{V})}$, and $\mathbf{d}^*_{\max}$ is its corresponding eigenvalue. The local influence measure can be used to identify observations that may jointly influence the fitted model, and it is calculated from the eigenvector that corresponds to the highest eigenvalue of the matrix of conformal normal curvature. Similarly, the total local influence measure can be used to identify observations that may individually exert influence on the fitted model, and it is calculated from the main diagonal of the matrix of conformal normal curvature.

Next, the expressions of $\mathbf{J}(\boldsymbol{\theta})$ and $\Delta(\hat{\boldsymbol{\theta}},\boldsymbol{\omega}_0)$ under two usual perturbation schemes are provided.

Second derivative matrix

The second derivative matrix $\partial^2\mathsf{L}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$ is given by

$$\mathbf{J}(\boldsymbol{\theta}) = \begin{bmatrix} -\mathbf{D}_\beta^\top\boldsymbol{\Omega}^{-1}\mathbf{D}_{(\mathrm{c})}\mathbf{D}_\beta + \sum\limits_{k=1}^{n}\frac{z_k}{\sqrt{\phi_k}}\mathrm{v}(z_k)\mathbf{D}_{\beta\beta}^{(k)} & -\mathbf{D}_\beta^\top\boldsymbol{\Omega}^{-\frac{1}{2}}\mathbf{D}_{(\bar{\mathrm{c}})}\overline{\mathbf{W}} \\ -\overline{\mathbf{W}}^\top\boldsymbol{\Omega}^{-\frac{1}{2}}\mathbf{D}_{(\bar{\mathrm{c}})}\mathbf{D}_\beta & -\overline{\mathbf{W}}^\top\mathbf{D}_{(\underline{\mathrm{c}})}\overline{\mathbf{W}} \end{bmatrix},$$

under the setup (I) described in (3.3), whereas

$$\mathbf{J}(\boldsymbol{\theta}) = \begin{bmatrix} -\overline{\mathbf{X}}^\top\boldsymbol{\Omega}^{-1}\mathbf{D}_{(\mathrm{c})}\overline{\mathbf{X}} & -\overline{\mathbf{X}}^\top\boldsymbol{\Omega}^{-\frac{1}{2}}\mathbf{D}_{(\bar{\mathrm{c}})}\overline{\mathbf{W}} \\ -\overline{\mathbf{W}}^\top\boldsymbol{\Omega}^{-\frac{1}{2}}\mathbf{D}_{(\bar{\mathrm{c}})}\overline{\mathbf{X}} & -\overline{\mathbf{W}}^\top\mathbf{D}_{(\underline{\mathrm{c}})}\overline{\mathbf{W}} \end{bmatrix}$$

under the setup (II) described in (3.3), where $\mathbf{D}_{(\mathrm{c})} = \mathrm{diag}\{\mathrm{c}_1,\ldots,\mathrm{c}_n\}$, $\mathbf{D}_{(\bar{\mathrm{c}})} = \mathrm{diag}\{\bar{\mathrm{c}}_1,\ldots,\bar{\mathrm{c}}_n\}$, $\mathbf{D}_{\underline{\mathrm{c}}} = \mathrm{diag}\{\underline{\mathrm{c}}_1,\ldots,\underline{\mathrm{c}}_n\}$, $\mathbf{D}_{\beta\beta}^{(k)}=\left[\partial^2\mu_k/\partial\beta_i\partial\beta_{i'}\right]$ for $i,i' = 1,\ldots,p$, $\mathrm{c}_k = \mathrm{v}(z_k) + \mathrm{v}'(z_k)z_k$, $\bar{\mathrm{c}}_k = \mathrm{v}(z_k)z_k + z_k^2\mathrm{v}'(z_k)/2$, and $\underline{\mathrm{c}}_k = \bar{\mathrm{c}}_k z_k/2$.

Case-weight perturbation scheme

Under this perturbation scheme $\boldsymbol{\omega}_0 = (1,\ldots,1)^\top$ and $\Delta(\hat{\boldsymbol{\theta}},\boldsymbol{\omega}_0)$ is given by

$$\Delta(\hat{\boldsymbol{\theta}},\boldsymbol{\omega}_0) = \begin{bmatrix} \hat{\mathbf{D}}_\beta^\top\hat{\mathbf{D}}_{(\mathrm{v})}\hat{\boldsymbol{\Omega}}^{-\frac{1}{2}}\hat{\mathbf{D}}_{(z)} \\ \frac{1}{2}\overline{\mathbf{W}}^\top\left(\hat{\mathbf{D}}_{(\mathrm{v})}\hat{\mathbf{D}}_{(z)}^2 - \mathbf{I}\right) \end{bmatrix}$$

under the setup (I) described in (3.3), whereas

$$\Delta(\hat{\boldsymbol{\theta}},\boldsymbol{\omega}_0) = \begin{bmatrix} \overline{\mathbf{X}}^\top\hat{\mathbf{D}}_{(\mathrm{v})}\hat{\boldsymbol{\Omega}}^{-\frac{1}{2}}\hat{\mathbf{D}}_{(z)} \\ \frac{1}{2}\overline{\mathbf{W}}^\top\left(\hat{\mathbf{D}}_{(\mathrm{v})}\hat{\mathbf{D}}_{(z)}^2 - \mathbf{I}\right) \end{bmatrix}$$

under the setup (II) described in (3.3), where $\hat{\mathbf{D}}_{(z)}$ correspond to $\mathbf{D}_{(z)} = \mathrm{diag}\{z_1,\ldots,z_n\}$ evaluated at the $\boldsymbol{\theta}$ estimate.

Response perturbation scheme

Under this perturbation scheme y_k is replaced by $y_k^{(\omega)} = y_k + \omega_k$ and $\boldsymbol{\omega}_0 = (0,\ldots,0)^\top$. Then,

$$\Delta(\hat{\boldsymbol{\theta}},\boldsymbol{\omega}_0) = \begin{bmatrix} \hat{\mathbf{D}}_\beta^\top\hat{\boldsymbol{\Omega}}^{-1}\hat{\mathbf{D}}_{(\mathrm{c})} \\ \overline{\mathbf{W}}^\top\hat{\boldsymbol{\Omega}}^{-\frac{1}{2}}\hat{\mathbf{D}}_{(\bar{\mathrm{c}})} \end{bmatrix}$$

under the setup (I) described in (3.3), whereas

$$\Delta(\hat{\boldsymbol{\theta}},\boldsymbol{\omega}_0) = \begin{bmatrix} \overline{\mathbf{X}}^\top\hat{\boldsymbol{\Omega}}^{-1}\hat{\mathbf{D}}_{(\mathrm{c})} \\ \overline{\mathbf{W}}^\top\hat{\boldsymbol{\Omega}}^{-\frac{1}{2}}\hat{\mathbf{D}}_{(\bar{\mathrm{c}})} \end{bmatrix}$$

under the setup (II) described in (3.3), where $\hat{\mathbf{D}}_{(\mathrm{c})}$ and $\hat{\mathbf{D}}_{(\bar{\mathrm{c}})}$ correspond to $\mathbf{D}_{(\mathrm{c})} = \mathrm{diag}\{\mathrm{c}_1,\ldots,\mathrm{c}_n\}$ and $\mathbf{D}_{(\bar{\mathrm{c}})} = \mathrm{diag}\{\bar{\mathrm{c}}_1,\ldots,\bar{\mathrm{c}}_n\}$ evaluated at the $\boldsymbol{\theta}$ estimate, respectively.

3.10 Simulation Results II

The probability distributions of the deviance-type residuals are unknown. Therefore, in this section, some of their statistical properties are studied via Monte carlo simulation. For this purpose, the simulation scenarios A and B that were described in section 3.8, are considered once again. To describe the random component, several log-symmetric distributions and several values of their extra parameters are considered. The simulated samples from each scenario are used to estimate the parameters of the models in (3.5) as described in section 3.8. Afterwards, the deviance-type residuals of the median and skewness (or relative dispersion) submodels are calculated for each individual in the sample. This process is replicated $R = 5000$ times, keeping the values of the explanatory variables fixed. Then, the R values of the residual of the k-th individual (i.e., $t(\hat{z}_k^{(1)}), \ldots, t(\hat{z}_k^{(R)})$, where $t(\hat{z}_k^{(r)})$ is the deviance-type residual of the individual k in the r-th replication) are used to calculate the empirical values of the mean, coefficient of skewness and quantiles of order 97.5% and 99.5% of $t(\hat{z}_k)$. Finally, as summary measures, the means of the n values of these four statistics are calculated. Tables 3.5 and 3.6 present the results for $t_\eta(\hat{z})$ and $t_\phi(\hat{z})$ under the simulation scenarios A and B, respectively.

It can be seen that in all cases the values of the mean and coefficient of skewness are quite close to zero, which indicates that the deviance-type residuals have approximately zero mean, and they exhibit probability distributions approximately symmetric around zero. Additionally, it can be seen that, the quantiles of order 97.5% and 99.5% of the distribution of the deviance-type residuals are dependent on the error distribution. However, the simulation results also show that in most cases, the individuals/observations whose deviance-type residuals (i.e., $t_\eta(\hat{z})$) are outside the interval (-3,3) may be considered to be marginally discrepant in the median submodel. Similarly, the simulation results suggest that, the individuals whose deviance-type residuals (i.e., $t_\phi(\hat{z})$) are outside the interval (-3,3) may be considered to be marginally discrepant in the skewness (or relative dispersion) submodel. Nonetheless, because the joint probability distribution of $t(\hat{z}_1), \ldots, t(\hat{z}_n)$ is unknown, the normal probability plots with simulated envelopes are recommended to identify discrepant/outlying observations in both submodels.

Table 3.5: *Results of the simulation study II under the scenario A and $n = 140$.*

Error distribution	$t_\eta(\hat{z})$				$t_\phi(\hat{z})$			
	Mean	Skewness	Q. 97.5%	Q. 99.5%	Mean	Skewness	Q. 97.5%	Q. 99.5%
log-normal	0.000	0.002	1.939	2.474	0.001	-0.003	2.151	2.771
log-Student-t(10)	0.000	0.002	2.282	2.980	0.002	0.001	2.177	2.816
log-Student-t(8)	0.000	0.002	2.169	2.813	0.000	0.004	2.168	2.787
log-Student-t(6)	0.000	0.002	2.112	2.727	0.000	0.002	2.157	2.790
log-Student-t(4)	-0.001	-0.002	2.072	2.680	-0.001	0.001	2.152	2.785
log-power-exp.(0.2)	0.000	-0.002	2.055	2.589	0.000	-0.002	2.135	2.780
log-power-exp.(0.3)	-0.001	-0.001	2.110	2.651	0.001	0.002	2.136	2.800
log-power-exp.(0.4)	-0.001	0.001	2.161	2.702	0.001	0.002	2.148	2.833
log-power-exp.(0.5)	-0.001	0.003	2.211	2.756	0.002	0.000	2.149	2.888
log-hyperbolic(1.0)	0.000	-0.001	2.226	2.797	-0.002	-0.001	2.122	2.747
log-hyperbolic(0.9)	0.000	0.001	2.244	2.815	0.001	-0.001	2.118	2.750
log-hyperbolic(0.8)	0.000	0.001	2.252	2.820	0.001	-0.002	2.112	2.740
log-hyperbolic(0.7)	0.000	0.002	2.268	2.833	0.001	-0.001	2.115	2.740
log-slash(1.5)	0.001	0.005	2.222	3.022	0.002	-0.004	2.211	2.834
log-slash(1.4)	-0.001	-0.004	2.252	3.054	-0.001	-0.002	2.216	2.853
log-slash(1.3)	0.000	0.004	2.303	3.111	0.001	0.000	2.226	2.865
log-slash(1.2)	-0.014	0.067	2.494	3.367	-0.028	0.070	2.244	3.044
log-cont-nor(0.3,0.3)	0.000	0.003	2.162	2.751	0.000	-0.003	2.153	2.759
log-cont-nor(0.5,0.3)	0.000	0.005	2.153	2.670	0.002	0.002	2.127	2.753
log-cont-nor(0.7,0.3)	0.000	0.003	2.082	2.594	0.000	0.002	2.119	2.751
log-cont-nor(0.9,0.3)	0.000	0.003	1.993	2.511	-0.001	0.000	2.131	2.763
B-S(0.1)	0.000	0.001	1.937	2.470	0.001	0.000	2.151	2.774
B-S(0.5)	0.000	0.001	1.919	2.443	0.002	0.001	2.150	2.772
B-S(1.0)	0.000	0.002	1.887	2.394	0.002	0.006	2.149	2.784
B-S(1.5)	0.000	0.000	1.842	2.331	0.002	0.000	2.146	2.775
B-S-t(0.1,4)	-0.001	0.001	2.173	2.793	-0.001	0.001	2.173	2.793
B-S-t(0.5,4)	0.001	0.003	2.255	2.901	0.000	0.001	2.162	2.777
B-S-t(1.0,4)	0.026	-0.049	2.205	2.694	0.014	-0.058	2.076	2.665
B-S-t(1.5,4)	0.039	-0.027	2.159	2.624	0.030	-0.004	2.079	2.668

Table 3.6: *Results of the simulation study II under the scenario B and n = 210.*

Error distribution	$t_\eta(\hat{z})$				$t_\phi(\hat{z})$			
	Mean	Skewness	Q. 97.5%	Q. 99.5%	Mean	Skewness	Q. 97.5%	Q. 99.5%
log-normal	0.000	-0.001	1.940	2.492	0.001	-0.001	2.158	2.778
log-Student-t(4)	0.001	0.003	2.298	3.007	0.002	0.003	2.188	2.821
log-Student-t(6)	0.000	0.001	2.182	2.840	0.002	0.001	2.174	2.802
log-Student-t(8)	0.000	-0.003	2.116	2.760	0.001	-0.006	2.166	2.787
log-Student-t(10)	0.000	0.001	2.082	2.700	0.001	0.000	2.163	2.783
log-power-exp.(0.2)	0.000	0.002	2.061	2.619	0.000	0.000	2.154	2.799
log-power-exp.(0.3)	0.000	0.001	2.116	2.675	0.002	0.004	2.170	2.854
log-power-exp.(0.4)	0.001	0.001	2.172	2.725	-0.001	0.001	2.177	2.903
log-power-exp.(0.5)	0.000	-0.003	2.219	2.768	-0.001	0.001	2.197	3.001
log-hyperbolic(1.0)	0.000	0.005	1.814	2.513	0.001	0.013	1.698	2.428
log-hyperbolic(0.9)	0.000	-0.001	2.247	2.834	0.000	-0.002	2.129	2.755
log-hyperbolic(0.8)	0.000	0.000	2.264	2.849	0.000	-0.001	2.129	2.756
log-hyperbolic(0.7)	0.001	0.000	2.276	2.856	0.001	0.001	2.128	2.757
log-slash(1.5)	0.006	-0.037	2.081	2.767	0.019	-0.039	2.152	2.798
log-slash(1.4)	-0.037	0.057	2.123	2.967	-0.041	0.022	2.142	2.647
log-slash(1.3)	-0.075	0.016	2.245	3.282	-0.040	0.096	2.174	2.937
log-slash(1.2)	-0.032	0.015	2.255	3.248	-0.029	-0.009	2.175	2.847
log-cont-nor(0.3,0.3)	0.000	0.002	2.172	2.764	0.000	-0.002	2.159	2.773
log-cont-nor(0.5,0.3)	0.000	0.001	2.156	2.690	0.001	0.003	2.136	2.763
log-cont-nor(0.7,0.3)	0.000	0.000	2.084	2.611	0.001	-0.001	2.129	2.750
log-cont-nor(0.9,0.3)	0.000	0.000	1.991	2.533	0.000	-0.001	2.143	2.772
B-S(0.1)	0.000	-0.001	1.935	2.486	0.000	-0.004	2.152	2.775
B-S(0.3)	0.000	0.001	1.919	2.460	0.001	0.002	2.154	2.781
B-S(0.5)	0.000	0.001	1.885	2.411	0.002	0.001	2.153	2.773
B-S(0.7)	0.000	0.001	1.839	2.348	0.001	0.000	2.148	2.773
B-S-t(0.1,4)	0.000	0.001	2.303	2.995	-0.001	-0.002	2.186	2.813
B-S-t(0.3,4)	0.000	0.003	2.271	2.937	0.000	-0.001	2.176	2.798
B-S-t(0.50,4)	0.000	0.001	2.216	2.844	0.001	0.000	2.164	2.787
B-S-t(0.7,4)	0.000	0.000	2.151	2.750	0.001	0.001	2.156	2.779

CHAPTER 4

CENSORED LOG-SYMMETRIC REGRESSION MODELS

Accelerated failure time models or log-location-scale models with a specified error distribution and non-informative right-censored observations (see, e.g., Bagdonavičius and Nikulin, 2001, Chapter 5) have received a lot of attention in recent years. This type of models and the relative risk or Cox models (see, e.g., Kalbfleisch and Prentice, 2002, Chapter 4) are competitive. The accelerated failure time model is appealing as it allows to specify a multiplicative effect of covariates acting on the quantiles (of any order) of the failure time distribution, which enables a straightforward parameter interpretation. Although these regression models are very interesting, they have limitations, for instance, usually just one parameter of the failure time distribution is modeled, and they do not admit the presence of nonparametric effects in their systematic component. Therefore, in this work, a very flexible accelerated failure time model is proposed, where the location and scale parameters of the log-lifetime distribution are modeled by using semi-parametric functions of explanatory variables, and whose nonparametric components are approximated by natural cubic splines or P-splines. The flexibility provided by the systematic component under this model lies in its capacity to relate the distribution of the lifetime (or failure time) with a set of covariates using a sum of arbitrary functions, whose functional forms are estimated from the data. Obviously, if the log-scale parameter is specified to be constant, this approach retains the direct link between the multiplicative effect of covariates and the quantiles (of any order) of the failure time distribution. In addition, if the location and scale parameters of the log-lifetime distribution are specified to be affected by covariates, the regression parameters can be interpreted by taking into account their multiplicative effect acting on the median and the skewness (or the relative dispersion) of the failure time distribution. Furthermore, the random component of the model is described by a very flexible class of probability distributions (i.e., the log-symmetric class), which in turn induces a wide range of shapes for the failure or hazard rate function (e.g., increasing, decreasing and upside-down bathtub shaped). Particular cases of this approach include models based on the Birnbaum-Saunders and Birnbaum-Saunders-t distributions, which have been extensively studied in the context of failure times under the fatigue or cumulated damage assumption (see, e.g., Barros *et al.*, 2008; Paula *et al.*, 2012). In addition, some of the log-symmetric distributions exhibit heavier tails than those of the log-normal one, which allows to estimate the model parameters in a robust manner under the presence of extreme or outlying observations. Finally, due to the properties of the log-symmetric class, the statistical methodology addressed in this chapter can also be used to analyze strictly positive data under the presence of non-informative left-censored observations.

The remainder of this chapter concerns with the formulation of the model setup, termed here as censored log-symmetric regression model. Iterative processes of parameter estimation based on Gauss-Seidel, backfitting and expectation-constrained/maximization (ECM) algorithms are presented. An approximate method to calculate the degrees of freedom used in the estimation process, is also discussed. Asymptotic behaviour of the maximum penalized likelihood estimator under a fixed-knot assumption is studied, analytically and by using simulation experiments. Diagnostic methods including deviance-type residuals, overall goodness-of-fit criterion based on a quantile-quantile plot, and local influence measures, are also derived.

4.1 Formulation of the model

Let $T_1^*, \ldots, T_n^*$ be the lifetime of n objects or individuals. These lifetimes are described by independent, strictly positive and right-skewed random variables belonging to the log-symmetric class, which is flexible enough so that distributions with lighter and heavier tails than the log-normal ones, as well as distributions with bimodality, are particular cases. Thus, T_k^* is assumed to be obtained as

$$T_k^* = \eta_k\, \xi_k^{\sqrt{\phi_k}}, \qquad k = 1, \ldots, n, \tag{4.1}$$

where $\eta_k > 0$ and $\phi_k > 0$ represent the median and the skewness (or the relative dispersion), respectively, of the T_k^* distribution, whereas $\xi_1, \ldots, \xi_n$ is a set of independent and multiplicative random errors exhibiting a standard log-symmetric distribution with the extra parameter (or extra parameter vector) ζ, whose probability density function is given by

$$f_{\xi_k}(\xi; g(\cdot)) = \frac{1}{\xi}\, g\big\{[\log(\xi)]^2\big\}, \qquad \xi > 0, \tag{4.2}$$

for some function $g(\cdot)$, where $g(u) > 0$ for $u > 0$ and $\int_0^\infty u^{-\frac{1}{2}} g(u) \partial u = 1$. Furthermore, the survival function and the quantile of order $0 < w < 1$ of T_k^* can be written as

$$S_{T^*}(t) = S_\xi\Big[(t/\eta_k)^{\frac{1}{\sqrt{\phi_k}}}\Big] \qquad \text{and} \qquad \vartheta_{T^*}(w) = \eta_k\, [\vartheta_\xi(w)]^{\sqrt{\phi_k}},$$

respectively, where $S_\xi(\cdot)$ and $\vartheta_\xi(w)$ represent the survival function and the quantile of order w of ξ_k, respectively. The behaviour of T^* is frequently characterized by its failure or hazard rate function, denoted by $R_{T^*}(t) = f_{T^*}(t)/S_{T^*}(t)$. Figure 4.1 illustrates the flexibility of $R_{T^*}(t)$ under some log-symmetric distributions.

Moreover, it is assumed that the median (η_k) and the skewness (or the relative dispersion) (ϕ_k) of the distribution of T_k^* are given by

$$\begin{cases} \log(\eta_k) = \mathbf{x}_k^\top \boldsymbol{\beta} + \mathrm{f}_{\eta_1}(a_{k,1}) + \ldots + \mathrm{f}_{\eta_{p'}}(a_{k,p'}) & \text{and} \\ \log(\phi_k) = \mathbf{w}_k^\top \boldsymbol{\gamma} + \mathrm{f}_{\phi_1}(b_{k,1}) + \ldots + \mathrm{f}_{\phi_{q'}}(b_{k,q'}), \end{cases} \tag{4.3}$$

where $\mathbf{x}_k^* = (\mathbf{x}_k^\top, a_{k,1}, \ldots, a_{k,p'})^\top$ and $\mathbf{w}_k^* = (\mathbf{w}_k^\top, b_{k,1}, \ldots, b_{k,q'})^\top$ are explanatory variables values for η_k and ϕ_k, respectively; $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_q)^\top$ are vectors of unknown parameters; $\mathrm{f}_{\eta_j}(a)$ $(j = 1, \ldots, p')$ and $\mathrm{f}_{\phi_r}(b)$ $(r = 1, \ldots, q')$ are continuous, smooth and nonparametric functions of the quantitative explanatory variables a and b, respectively, which are approximated by using natural cubic splines (see, e.g., Green and Silverman, 1994; Lancaster and Salkauskas, 1986, sections 4.6 and 4.7) or P-splines (Eilers and Marx, 1996). The matrices $\mathbf{X} = (\mathbf{x}_1^*, \ldots, \mathbf{x}_n^*)^\top$ and $\mathbf{W} = (\mathbf{w}_1^*, \ldots, \mathbf{w}_n^*)^\top$ are assumed to be of full column rank. In addition, there are n independent and positive random variables $C_1, \ldots, C_n$ that represent censoring times and are independent of $T_1^*, \ldots, T_n^*$. Then, the observed bivariate data, $(t_1, \delta_1), \ldots, (t_n, \delta_n)$, are assumed to be realizations of $(T_1, \bar{\delta}_1), \ldots, (T_n, \bar{\delta}_n)$, where $T_k = \min(T_k^*, C_k)$ and $\bar{\delta}_k = \mathsf{I}(T_k^* > C_k)$, in which $\mathsf{I}(\cdot)$ is the indicator function. Due to the properties of the log-symmetric class, the formulated model is a log-location-scale model or an accelerated failure time model, where the log-lifetime distribution belongs to the symmetric class, that is,

$$Y_k^* = \log(T_k^*) = \log(\eta_k) + \sqrt{\phi_k}\, \xi_k^*, \qquad k = 1, \ldots, n,$$

in which $\xi_1^* = \log(\xi_1), \ldots, \xi_n^* = \log(\xi_n)$ are independent and identically distributed errors exhibiting standard symmetric distribution (i.e., $f_{\xi_k^*}(\xi; g(\cdot)) = g(\xi^2)$). Particular cases of the formulated

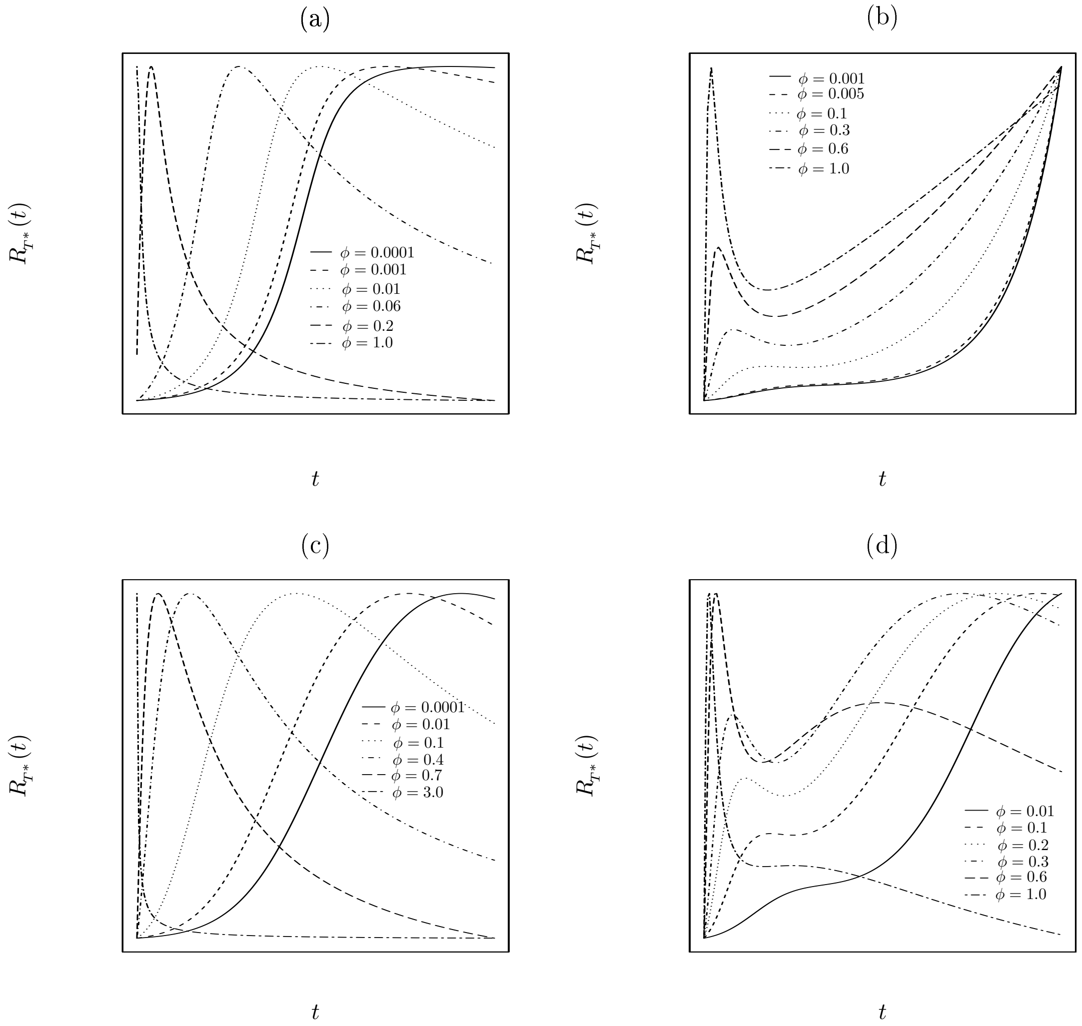

Figure 4.1: *Graph of the hazard rate function under log-hyperbolic* $(\eta = 1, \phi, \zeta = 1)$ (a)*, Birnbaum-Saunders* $(\eta = 1, \phi, \zeta = 3)$ (b)*, log-Student-t* $(\eta = 1, \phi, \zeta = 10)$ (c)*, and Birnbaum-Saunders-t* $(\eta = 1, \phi, \zeta = 3)$ (d) *distributions.*

model include log-normal, Birnbaum-Saunders and Birnbaum-Saunders-t models.

4.2 Parameter estimation

Similar to the uncensored case, to allow the model identification, the systematic component (4.3) of the formulated model is written as

$$\begin{cases} \log(\boldsymbol{\eta}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{N}_{\eta_1}\boldsymbol{\tau}_{\eta_1} + \ldots + \mathbf{N}_{\eta_{p'}}\boldsymbol{\tau}_{\eta_{p'}} \quad \text{and} \\ \log(\boldsymbol{\phi}) = \mathbf{W}\boldsymbol{\gamma} + \mathbf{N}_{\phi_1}\boldsymbol{\tau}_{\phi_1} + \ldots + \mathbf{N}_{\phi_{q'}}\boldsymbol{\tau}_{\phi_{q'}}, \end{cases} \tag{4.4}$$

where $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)^\top$, $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_n)^\top$, $\mathbf{N}_{\eta_j}(j = 1, \ldots, p')$ and $\mathbf{N}_{\phi_r}(r = 1, \ldots, q')$ are basis matrices of dimension $n \times (p'_j - 1)$ and $n \times (q'_r - 1)$, respectively; $\boldsymbol{\tau}_{\eta_j}(j = 1, \ldots, p')$ and $\boldsymbol{\tau}_{\phi_r}(r = 1 \ldots, q')$ are unknown parameter vectors to be estimated, of dimension $(p'_j - 1)$ and $(q'_r - 1)$, respectively. In addition, to avoid overfitting, a quadratic penalty term is introduced, in which $\mathbf{M}_{\eta_j}(j = 1, \ldots, p')$ and $\mathbf{M}_{\phi_r}(r = 1, \ldots, q')$ are square penalty matrices of dimension $(p'_j - 1)$ and $(q'_r - 1)$, respectively. Note that the structure of $\mathbf{N}$ and $\mathbf{M}$ is dependent on the type of spline (natural cubic spline or P-spline) which will be used to approximate each one of the functions $\mathrm{f}_{\eta_j}(\cdot)(j = 1, \ldots, p')$ and $\mathrm{f}_{\phi_r}(\cdot)(r = 1, \ldots, q')$.

The estimation of $\boldsymbol{\theta}$ is performed by fitting a symmetric heteroscedastic semi-parametric model to the transformed lifetime (or censoring time) (i.e., $Y = \log(T)$), in which the systematic component of the location parameter is given by $\mu_k = \log(\eta_k)$, $k = 1, \ldots, n$, and the systematic component of the dispersion parameter ϕ_k is a semi-parametric function with logarithmic link, where the nonparametric functions are approximated by natural cubic splines or P-splines, and the extra parameter ζ is assumed to be known or fixed. For a known ζ, this approach generalizes the random and systematic components of the models discussed by Barros *et al.* (2008) and Li *et al.* (2012).

Let $y_1, \ldots, y_n$ be the observed values of $Y_1 = \log(T_1), \ldots, Y_n = \log(T_n)$. Thus, the penalized log-likelihood function of $\boldsymbol{\theta}$ is given by $\mathsf{PL}(\boldsymbol{\theta}) = \mathsf{L}(\boldsymbol{\theta}) + \mathsf{P}(\boldsymbol{\theta})$, where the penalty term of $\boldsymbol{\theta}$ becomes

$$\mathsf{P}(\boldsymbol{\theta}) = \sum_{j=1}^{p'} \frac{\lambda_{\eta_j}}{2} \boldsymbol{\tau}_{\eta_j}^\top \mathbf{M}_{\eta_j} \boldsymbol{\tau}_{\eta_j} + \sum_{r=1}^{q'} \frac{\lambda_{\phi_r}}{2} \boldsymbol{\tau}_{\phi_r}^\top \mathbf{M}_{\phi_r} \boldsymbol{\tau}_{\phi_r},$$

and the log-likelihood function of $\boldsymbol{\theta}$ is given by

$$\mathsf{L}(\boldsymbol{\theta}) = \sum_{k=1}^{n} \Big\{ \delta_k \log[1 - F_{\xi^*}(z_k)] + (1 - \delta_k) \mathsf{L}_k(\mu_k, \phi_k) \Big\},$$

in which $\log[1 - F_{\xi^*}(\tilde{t}_k)]$ and $\mathsf{L}_k(\mu_k, \phi_k) = \log[g(z_k^2)] - \frac{1}{2}\log(\phi_k)$ are the contributions of a censored and an uncensored observations to the log-likelihood function of $\boldsymbol{\theta}$, respectively, $z_k = (y_k - \mu_k)/\sqrt{\phi_k}$, $\mu_k = \log(\eta_k)$, and $F_{\xi^*}(\cdot)$ is the cumulative distribution function of ξ^*.

4.2.1 Score function and Hessian matrix

The score function or estimating equation of $\boldsymbol{\theta}$ is given by $\partial \mathsf{PL}(\boldsymbol{\theta})/\partial \boldsymbol{\theta} = \mathbf{U}(\boldsymbol{\theta}) - \mathbf{M}\boldsymbol{\theta}$, where

$$\mathbf{U}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{U}_\eta(\boldsymbol{\theta}) \\ \mathbf{U}_\phi(\boldsymbol{\theta}) \end{bmatrix} = \begin{bmatrix} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{D}_{(\tilde{\mathrm{v}})} (\mathbf{y} - \boldsymbol{\mu}) \\ \mathbf{N}_\eta^\top \boldsymbol{\Omega}^{-1} \mathbf{D}_{(\tilde{\mathrm{v}})} (\mathbf{y} - \boldsymbol{\mu}) \\ \frac{1}{2} \mathbf{W}^\top (\tilde{\mathbf{s}} - \mathbf{1}_n) \\ \frac{1}{2} \mathbf{N}_\phi^\top (\tilde{\mathbf{s}} - \mathbf{1}_n) \end{bmatrix} \quad \text{and} \quad \mathbf{M} = \mathrm{diag}\left\{ \overline{\mathbf{M}}_\eta, \overline{\mathbf{M}}_\phi \right\},$$

where $\mathbf{y} = (y_1, \ldots, y_n)^\top$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^\top$, $\boldsymbol{\Omega} = \mathrm{diag}\{\phi_1, \ldots, \phi_n\}$, $\mathbf{D}_{(\tilde{\mathrm{v}})} = \mathrm{diag}\{\tilde{\mathrm{v}}_1, \ldots, \tilde{\mathrm{v}}_n\}$, $\tilde{\mathbf{s}} = (\tilde{s}_1, \ldots, \tilde{s}_n)^\top$, $\mathbf{N}_\eta = [\mathbf{N}_{\eta_1}, \ldots, \mathbf{N}_{\eta_{p'}}]$, $\overline{\mathbf{M}}_\eta = \mathrm{diag}\left\{\tilde{\mathbf{0}}_p, \mathbf{M}_\eta\right\}$, $\mathbf{M}_\eta = \mathrm{diag}\{\lambda_{\eta_1}\mathbf{M}_{\eta_1}, \ldots, \lambda_{\eta_{p'}}\mathbf{M}_{\eta_{p'}}\}$, $\mathbf{N}_\phi = [\mathbf{N}_{\phi_1}, \ldots, \mathbf{N}_{\phi_{q'}}]$, $\overline{\mathbf{M}}_\phi = \mathrm{diag}\left\{\tilde{\mathbf{0}}_q, \mathbf{M}_\phi\right\}$, $\mathbf{M}_\phi = \mathrm{diag}\{\lambda_{\phi_1}\mathbf{M}_{\phi_1}, \ldots, \lambda_{\phi_{q'}}\mathbf{M}_{\phi_{q'}}\}$, $\tilde{\mathrm{v}}_k = (1 - \delta_k)\mathrm{v}(z_k) + \delta_k\, R_{\xi^*}(z_k)/z_k]$, $\tilde{s}_k = (1 - \delta_k)\mathrm{v}(z_k) z_k^2 + \delta_k[R_{\xi^*}(z_k) z_k + 1]$, and $R_{\xi^*}(z_k) = g(z_k^2)/[1 - F_{\xi^*}(z_k)]$. Obviously, the maximum penalized likelihood estimate of $\boldsymbol{\theta}$, denoted as $\hat{\boldsymbol{\theta}}$, is the solution of $\mathbf{U}(\hat{\boldsymbol{\theta}}) = \mathbf{M}\hat{\boldsymbol{\theta}}$.

Moreover, the hessian matrix of $\boldsymbol{\theta}$ is given by $\partial^2 \mathsf{PL}(\boldsymbol{\theta})/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top = \mathbf{J}(\boldsymbol{\theta}) - \mathbf{M}$, where

$$\mathbf{J}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{J}_{\eta\eta}(\boldsymbol{\theta}) & \mathbf{J}_{\eta\phi}(\boldsymbol{\theta}) \\ \mathbf{J}_{\phi\eta}(\boldsymbol{\theta}) & \mathbf{J}_{\phi\phi}(\boldsymbol{\theta}) \end{bmatrix} = \begin{bmatrix} -\overline{\mathbf{X}}^\top \boldsymbol{\Omega}^{-1} \mathbf{D}_{(\mathrm{c})} \overline{\mathbf{X}} & -\overline{\mathbf{X}}^\top \boldsymbol{\Omega}^{-\frac{1}{2}} \mathbf{D}_{(\bar{\mathrm{c}})} \overline{\mathbf{W}} \\ -\overline{\mathbf{W}}^\top \boldsymbol{\Omega}^{-\frac{1}{2}} \mathbf{D}_{(\bar{\mathrm{c}})} \overline{\mathbf{X}} & -\overline{\mathbf{W}}^\top \mathbf{D}_{(\underline{\mathrm{c}})} \overline{\mathbf{W}} \end{bmatrix},$$

in which $\mathbf{D}_{(\mathrm{c})} = \mathrm{diag}\{\mathrm{c}_1, \ldots, \mathrm{c}_n\}$, $\mathbf{D}_{(\bar{\mathrm{c}})} = \mathrm{diag}\{\bar{\mathrm{c}}_1, \ldots, \bar{\mathrm{c}}_n\}$, $\mathbf{D}_{(\underline{\mathrm{c}})} = \mathrm{diag}\{\underline{\mathrm{c}}_1, \ldots, \underline{\mathrm{c}}_n\}$, $c_k = \delta_k R'_{\xi^*}(z_k) + (1 - \delta_k)[\mathrm{v}(z_k) + \mathrm{v}'(z_k) z_k]$, $2\bar{c}_k = \delta_k[R_{\xi^*}(z_k) + R'_{\xi^*}(z_k) z_k] + (1 - \delta_k)[2\mathrm{v}(z_k) z_k + z_k^2 \mathrm{v}'(z_k)]$, $\underline{c}_k = \bar{c}_k z_k/2$, $\overline{\mathbf{X}} = [\mathbf{X}, \mathbf{N}_\eta]$, and $\overline{\mathbf{W}} = [\mathbf{W}, \mathbf{N}_\phi]$.

4.2.2 Iterative processes

To solve the equation $\mathbf{U}(\hat{\boldsymbol{\theta}}) = \mathbf{M}\hat{\boldsymbol{\theta}}$, the Newton-Raphson method can be used. However, in this work, the nonlinear Gauss-Seidel algorithm (see, e.g., Ortega, 1970, page 219) is used instead, as it allows to employ in each iteration a *diagonalized* version of the hessian matrix

$\mathbf{J}(\boldsymbol{\theta})$, which in turn facilitates the introduction of the backfitting algorithm. Let $\bar{\boldsymbol{\beta}} = (\boldsymbol{\beta}^\top, \boldsymbol{\tau}_\eta^\top)^\top$ and $\bar{\boldsymbol{\gamma}} = (\boldsymbol{\gamma}^\top, \boldsymbol{\tau}_\phi^\top)^\top$ be the parameter vectors associated to the median and the skewness (or the relative dispersion) submodels, respectively. Then, for fixed value of the smoothing parameter, the resulting algorithm is as follows

Algorithm 4.1

Step 1. Initialize the counter and the parameter vector as $l = 0$ and $\boldsymbol{\theta}^{(0)} = (\bar{\boldsymbol{\beta}}^{(0)}, \bar{\boldsymbol{\gamma}}^{(0)})$.

Step 2. Calculate $\bar{\boldsymbol{\beta}}^{(l+1)}$ as the vector that maximizes $\mathsf{PL}(\bar{\boldsymbol{\beta}}, \bar{\boldsymbol{\gamma}}^{(l)})$ with respect to $\bar{\boldsymbol{\beta}}$ (i.e., the solution of $\mathbf{U}_\eta(\bar{\boldsymbol{\beta}}, \bar{\boldsymbol{\gamma}}^{(l)}) = \overline{\mathbf{M}}_\eta \bar{\boldsymbol{\beta}}$ with respect to $\bar{\boldsymbol{\beta}}$). To do this, the Newton-Raphson algorithm is used, in which each one of its stages is accomplished by using the backfitting algorithm.

Step 3. Calculate $\bar{\boldsymbol{\gamma}}^{(l+1)}$ as the vector that maximizes $\mathsf{PL}(\bar{\boldsymbol{\beta}}^{(l+1)}, \bar{\boldsymbol{\gamma}})$ with respect to $\bar{\boldsymbol{\gamma}}$ (i.e., the solution of $\mathbf{U}_\phi(\bar{\boldsymbol{\beta}}^{(l+1)}, \bar{\boldsymbol{\gamma}}) = \overline{\mathbf{M}}_\phi \bar{\boldsymbol{\gamma}}$ with respect to $\bar{\boldsymbol{\gamma}}$). . To do this, the Newton-Raphson algorithm is used, in which each one of its stages is accomplished by using the backfitting algorithm.

Step 4. Update $l = (l + 1)$ and $\boldsymbol{\theta}^{(l+1)} = (\bar{\boldsymbol{\beta}}^{(l+1)}, \bar{\boldsymbol{\gamma}}^{(l+1)})$.

Step 5. Repeat Steps 2, 3, and 4 until converge of $\boldsymbol{\theta}^{(l)}$.

The efficiency of the proposed algorithm can be improved by introducing the Expectation/-Conditional Maximization (ECM) algorithm (Meng and Rubin , 1993) when the error distribution is a power mixture of the log-normal distribution (see Andrews and Mallows , 1974; Barndoff-Nielsen, 1977; West, 1987). In these cases, $Y_k|[\delta_k = 0, U_k = u_k] \sim \text{Normal}(\mu_k, \phi_k h(u_k))$, $Y_k^*|[Y_k^* > y_k, \delta_k = 1, U_k = u_k] \sim \text{TNormal}(\mu_k, \phi_k h(u_k), (y_k, \infty))$ and the density generator of the ξ^* distribution is $g(z^2) = \int_{\mathbb{R}+}\Big\{\phi\Big(z/\sqrt{h(u)}\Big)/\sqrt{h(u)}\Big\} f_U(u)\partial u$, where $\text{TNormal}(\mu, \phi, A)$ represents a truncated normal distribution that lies within the interval A, exhibiting location parameter μ and scale parameter ϕ, $\phi(\cdot)$ is the density function of the standard normal distribution, $h(u) > 0$ for $u > 0$, and $f_U(\cdot)$ is the probability density function (or probability mass function) of U. Thus, the E step of the ECM algorithm is accomplished by calculating $\bar{\mathrm{v}}(z_k^{(l)}) = \mathrm{E}_U\Big[1/h(U)\,|\,Y_k^* > y_k; \boldsymbol{\theta}^{(l)}\Big]$ and $\check{m}(z_k^{(l)}) = \mathrm{E}_U\Big[\phi\Big(z_k/\sqrt{h(U)}\Big)/\Big\{\sqrt{h(U)}\Big[1 - \Phi\Big(z_k/\sqrt{h(U)}\Big)\Big]\Big\}\,|\,Y_k^* > y_k; \boldsymbol{\theta}^{(l)}\Big]$ for censored observations; and $\bar{\mathrm{v}}(z_k^{(l)}) = \mathrm{E}_U\Big[1/h(U)\,|\,Y_k^* = y_k; \boldsymbol{\theta}^{(l)}\Big]$ for uncensored observations, with $\Phi(\cdot)$ being the cumulative distribution function of the standard normal distribution. The two CM steps are accomplished by calculating $\bar{\boldsymbol{\beta}}^{(l+1)}$ as the argument that maximizes $\mathsf{PL}^{(l)}(\bar{\boldsymbol{\beta}}, \bar{\boldsymbol{\gamma}}^{(l)})$ with respect to $\bar{\boldsymbol{\beta}}$, and, then, by calculating $\bar{\boldsymbol{\gamma}}^{(l+1)}$ as the argument that maximizes $\mathsf{PL}^{(l)}(\bar{\boldsymbol{\beta}}^{(l+1)}, \bar{\boldsymbol{\gamma}})$ with respect to $\bar{\boldsymbol{\gamma}}$, where $\mathsf{PL}^{(l)}(\boldsymbol{\theta}) = \mathsf{L}^{(l)}(\boldsymbol{\theta}) + \mathsf{P}(\boldsymbol{\theta})$, and

$$\mathsf{L}^{(l)}(\boldsymbol{\theta}) = -\frac{1}{2}\sum_{k=1}^{n}\left\{\delta_k \bar{\mathrm{v}}(z_k^{(l)})\frac{(\bar{m}_k^{(l)} - \mu_k)^2 + \tilde{m}_k^{(l)} - [\bar{m}_k^{(l)}]^2}{\phi_k} + (1 - \delta_k)\bar{\mathrm{v}}(z_k^{(l)})\frac{(y_k - \mu_k)^2}{\phi_k}\right\},$$

in which $\bar{m}_k^{(l)}$ and $\tilde{m}_k^{(l)}$ represent $\bar{m}_k = \mu_k + \sqrt{\phi_k}\,\check{m}(z_k)/\bar{\mathrm{v}}(z_k)$ and $\tilde{m}_k = \mu_k^2 + \sqrt{\phi_k}\,[\check{m}(z_k)(y_k + \mu_k) + \sqrt{\phi_k}]/\bar{\mathrm{v}}(z_k)$ avaliated at $\boldsymbol{\theta}^{(l)}$, respectively. Then, for fixed smoothing parameter, the resulting algorithm is described as follows

Algorithm 4.2

Step 1. Initialize the counter and the parameter vector as $l = 0$ and $\boldsymbol{\theta}^{(0)} = (\bar{\boldsymbol{\beta}}^{(0)}, \bar{\boldsymbol{\gamma}}^{(0)})$.

Step 2. Calculate $\bar{\mathrm{v}}(z_1^{(l)}),\ldots,\bar{\mathrm{v}}(z_n^{(l)})$ from $\boldsymbol{\theta}^{(l)}$, where $\bar{\mathrm{v}}(z_k)=\mathrm{v}(z_k)=-2g'(z_k^2)/g(z_k^2)$ for uncensored observations, and $\bar{\mathrm{v}}(z_k)$ is described in Table 4.1 for censored observations.

Step 3. Calculate $\bar{m}_k^{(l)}$ and $\tilde{m}_k^{(l)}$ from $\boldsymbol{\theta}^{(l)}$ for the censored observations, where

$$\bar{m}_k=\mu_k+\frac{\sqrt{\phi_k}}{\bar{\mathrm{v}}(z_k)}\frac{g(z_k^2)}{1-F_{\xi^*}(z_k)} \qquad \text{and}$$
$$\tilde{m}_k=\mu_k^2+\frac{\sqrt{\phi_k}}{\bar{\mathrm{v}}(z_k)}\left[\frac{g(z_k^2)}{1-F_{\xi^*}(z_k)}(y_k+\mu_k)+\sqrt{\phi_k}\right].$$

Step 4. Calculate $\bar{\boldsymbol{\beta}}^{(l+1)}$ as the vector that maximizes $\mathsf{PL}^{(l)}(\bar{\boldsymbol{\beta}},\bar{\boldsymbol{\gamma}}^{(l)})$ with respect to $\bar{\boldsymbol{\beta}}$. To do this, the Newton-Raphson algorithm is used, in which each one of its stages is accomplished by using the backfitting algorithm.

Step 5. Calculate $\bar{\boldsymbol{\gamma}}^{(l+1)}$ as the vector that maximizes $\mathsf{PL}^{(l)}(\bar{\boldsymbol{\beta}}^{(l+1)},\bar{\boldsymbol{\gamma}})$ with respect to $\bar{\boldsymbol{\gamma}}$. To do this, the Newton-Raphson algorithm is used, in which each one of its stages is accomplished by using the backfitting algorithm.

Step 6. Update $l=(l+1)$ and $\boldsymbol{\theta}^{(l)}$.

Step 7. Repeat Steps 2, 3, 4, 5 and 6 until convergence of $\boldsymbol{\theta}^{(l)}$.

Hereinafter, $\hat{\eta}_k$, $\hat{\mu}_k=\log(\hat{\eta}_k)$ and $\hat{\phi}_k$ represent the fitted values of η_k, $\mu_k=\log(\eta_k)$ and ϕ_k, respectively, for $k=1,\ldots,n$.

4.3 Degrees of freedom

The numer of degrees of freedom used to estimate $\boldsymbol{\eta}$ and $\boldsymbol{\phi}$ are frequently employed to quantify the model complexity. In this section, a method to calculate it is discussed. To do this, it can be seen that, at the convergence of the Algorithm 3.1 (in particular, the convergence of the Newton-Raphson processes described by the Steps 2 and 3), the estimates of $\bar{\boldsymbol{\beta}}$ and $\bar{\boldsymbol{\gamma}}$ can be written as

$$\begin{aligned}\hat{\bar{\boldsymbol{\beta}}}&=\left[-\mathbf{J}_{\eta\eta}(\hat{\boldsymbol{\theta}})+\overline{\mathbf{M}}_{\eta}\right]^{-1}\left[\mathbf{U}_{\eta}(\hat{\boldsymbol{\theta}})-\mathbf{J}_{\eta\eta}(\hat{\boldsymbol{\theta}})\,\hat{\bar{\boldsymbol{\beta}}}\right]\\&=\left[\overline{\mathbf{X}}^{T}\hat{\boldsymbol{\Omega}}^{-1}\hat{\mathbf{D}}_{(\mathrm{c})}\overline{\mathbf{X}}+\overline{\mathbf{M}}_{\eta}\right]^{-1}\overline{\mathbf{X}}^{\top}\hat{\boldsymbol{\Omega}}^{-1}\hat{\mathbf{D}}_{(\mathrm{c})}\left[\hat{\mathbf{D}}_{(\mathrm{c})}^{-1}\hat{\mathbf{D}}_{(\bar{\mathrm{v}})}(\mathbf{y}-\hat{\boldsymbol{\mu}})+\overline{\mathbf{X}}\hat{\bar{\boldsymbol{\beta}}}\right],\qquad\text{and}\\\hat{\bar{\boldsymbol{\gamma}}}&=\left[-\mathbf{J}_{\phi\phi}(\hat{\boldsymbol{\theta}})+\overline{\mathbf{M}}_{\eta}\right]^{-1}\left[\mathbf{U}_{\phi}(\hat{\boldsymbol{\theta}})-\mathbf{J}_{\phi\phi}(\hat{\boldsymbol{\theta}})\,\hat{\bar{\boldsymbol{\gamma}}}\right]\\&=\left[\overline{\mathbf{W}}^{T}\hat{\mathbf{D}}_{(\underline{\mathrm{c}})}\overline{\mathbf{W}}+\overline{\mathbf{M}}_{\phi}\right]^{-1}\overline{\mathbf{W}}^{\top}\hat{\mathbf{D}}_{(\underline{\mathrm{c}})}\left[\frac{1}{2}\hat{\mathbf{D}}_{(\underline{\mathrm{c}})}^{-1}(\hat{\tilde{\mathbf{s}}}-\mathbf{1}_{n})+\overline{\mathbf{W}}\hat{\bar{\boldsymbol{\gamma}}}\right],\end{aligned}$$

where $\hat{\boldsymbol{\Omega}}$, $\hat{\mathbf{D}}_{(\mathrm{c})}$, $\hat{\boldsymbol{\mu}}$, $\hat{\bar{\boldsymbol{\beta}}}$, $\hat{\bar{\boldsymbol{\gamma}}}$, $\hat{\mathbf{D}}_{(\underline{\mathrm{c}})}$, $\hat{\mathbf{D}}_{(\bar{\mathrm{v}})}$ and $\hat{\tilde{\mathbf{s}}}$ represent $\boldsymbol{\Omega}$, $\mathbf{D}_{(\mathrm{c})}$, $\boldsymbol{\mu}$, $\bar{\boldsymbol{\beta}}$, $\bar{\boldsymbol{\gamma}}$, $\mathbf{D}_{(\underline{\mathrm{c}})}$, $\mathbf{D}_{(\bar{\mathrm{v}})}$ and $\tilde{\mathbf{s}}$ avaliated at the estimate of $\boldsymbol{\theta}$. By analogy with the parametric case, the degrees of freedom associated to the estimate of $\boldsymbol{\eta}$ can be estimated by $\mathsf{df}(\hat{\boldsymbol{\eta}})=\mathsf{tr}(\hat{\mathbf{H}}_{\eta})$, where $\hat{\mathbf{H}}_{\eta}=\overline{\mathbf{X}}\left[\overline{\mathbf{X}}^{\top}\hat{\boldsymbol{\Omega}}^{-1}\hat{\mathbf{D}}_{(\mathrm{c})}\overline{\mathbf{X}}+\overline{\mathbf{M}}_{\eta}\right]^{-1}\overline{\mathbf{X}}^{\top}\hat{\boldsymbol{\Omega}}^{-1}\hat{\mathbf{D}}_{(\mathrm{c})}$ is such that $\log(\hat{\boldsymbol{\eta}})=\hat{\mathbf{H}}_{\eta}[\mathbf{y}_{\eta}(\hat{\boldsymbol{\theta}})]$, in which $\mathbf{y}_{\eta}(\boldsymbol{\theta})=\overline{\mathbf{X}}\bar{\boldsymbol{\beta}}+\mathbf{D}_{(\mathrm{c})}^{-1}\mathbf{D}_{(\bar{\mathrm{v}})}(\mathbf{y}-\boldsymbol{\mu})$ is a local response variable. Then, according to Eilers and Marx (1996), it is possible to write

$$\begin{aligned}\mathsf{df}(\hat{\boldsymbol{\eta}})&=\mathsf{tr}\left\{\left[\overline{\mathbf{X}}^{\top}\hat{\boldsymbol{\Omega}}^{-1}\hat{\mathbf{D}}_{(\mathrm{c})}\overline{\mathbf{X}}+\overline{\mathbf{M}}_{\eta}\right]^{-1}\overline{\mathbf{X}}^{\top}\hat{\boldsymbol{\Omega}}^{-1}\hat{\mathbf{D}}_{(\mathrm{c})}\overline{\mathbf{X}}\right\}\\&=\mathsf{tr}\left\{\left[\mathbf{I}+\mathbf{Q}^{-\frac{1}{2}}\overline{\mathbf{M}}_{\eta}\mathbf{Q}^{-\frac{1}{2}}\right]^{-1}\right\}\\&=p+\sum_{i=p+1}^{\mathsf{dim}(\overline{\mathbf{M}}_{\eta})}\frac{1}{1+\alpha_i^{(\eta)}},\end{aligned}$$

where $[\mathbf{I}+\mathbf{Q}^{-\frac{1}{2}}\overline{\mathbf{M}}_{\eta}\mathbf{Q}^{-\frac{1}{2}}]$ is a positive definite matrix, $\alpha_i^{(\eta)} \geq 0$ are the eigenvalues of $\mathbf{Q}^{-\frac{1}{2}}\overline{\mathbf{M}}_{\eta}\mathbf{Q}^{-\frac{1}{2}}$ and $\mathsf{dim}(\mathbf{M})$ is the number of rows (or columns) of $\mathbf{M}$, in which $\mathbf{Q}^{\frac{1}{2}}$ is a positive definite matrix such that $\overline{\mathbf{X}}^{\top}\hat{\boldsymbol{\Omega}}^{-1}\hat{\mathbf{D}}_{(\mathrm{c})}\overline{\mathbf{X}} = \mathbf{Q}^{\frac{1}{2}}\mathbf{Q}^{\frac{1}{2}}$. Because the number of eigenvalues equal to zero of $\mathbf{Q}^{-\frac{1}{2}}\overline{\mathbf{M}}_{\eta}\mathbf{Q}^{-\frac{1}{2}}$ and $\overline{\mathbf{M}}_{\eta}$ coincide, it follows that $\mathsf{dim}(\overline{\mathbf{M}}_{\eta}) > \mathsf{df}(\hat{\boldsymbol{\eta}}) > p + \sum_{j=1}^{p'} \alpha_{\eta_j}^{(0)}$, where $\alpha_{\eta_j}^{(0)}$ is the number of eigenvalues of $\mathbf{M}_{\eta_j}$ equal to zero.

The degrees of freedom associated to the estimate of $\boldsymbol{\phi}$ can be estimated as $\mathsf{df}(\hat{\boldsymbol{\phi}}) = \mathsf{tr}(\hat{\mathbf{H}}_{\phi})$, where $\hat{\mathbf{H}}_{\phi} = \overline{\mathbf{W}}[\overline{\mathbf{W}}^{\top}\hat{\mathbf{D}}_{(\underline{\mathrm{c}})}\overline{\mathbf{W}} + \overline{\mathbf{M}}_{\phi}]^{-1}\overline{\mathbf{W}}\hat{\mathbf{D}}_{(\underline{\mathrm{c}})}$ is such that $\log(\hat{\boldsymbol{\phi}}) = \hat{\mathbf{H}}_{\phi}[\mathbf{y}_{\phi}(\hat{\boldsymbol{\theta}})]$, in which $\mathbf{y}_{\phi}(\hat{\boldsymbol{\theta}}) = \overline{\mathbf{W}}\bar{\boldsymbol{\gamma}} + \frac{1}{2}\mathbf{D}_{(\underline{\mathrm{c}})}^{-1}(\tilde{\mathbf{s}} - \mathbf{1}_n)$ is a local response variable. Then, according to Eilers and Marx (1996), it is possible to write

$$
\begin{aligned}
\mathsf{df}(\hat{\boldsymbol{\phi}}) &= \mathsf{tr}\Big\{\big[\overline{\mathbf{W}}^{\top}\hat{\mathbf{D}}_{(\underline{\mathrm{c}})}\overline{\mathbf{W}} + \overline{\mathbf{M}}_{\phi}\big]^{-1}\overline{\mathbf{W}}^{\top}\hat{\mathbf{D}}_{(\underline{\mathrm{c}})}\overline{\mathbf{W}}\Big\} \\
&= \mathsf{tr}\Big\{\big[\mathbf{I} + \mathbf{Q}^{-\frac{1}{2}}\overline{\mathbf{M}}_{\phi}\mathbf{Q}^{-\frac{1}{2}}\big]^{-1}\Big\} \\
&= q + \sum_{r=q+1}^{\mathsf{dim}(\overline{\mathbf{M}}_{\phi})} \frac{1}{1+\alpha_r^{(\phi)}},
\end{aligned}
$$

where where $[\mathbf{I}+\mathbf{Q}^{-\frac{1}{2}}\overline{\mathbf{M}}_{\eta}\mathbf{Q}^{-\frac{1}{2}}]$ is a positive definite matrix, $\alpha_r^{(\phi)} \geq 0$ are the eigenvalues of $\mathbf{Q}^{-\frac{1}{2}}\overline{\mathbf{M}}_{\phi}\mathbf{Q}^{-\frac{1}{2}}$ and $\mathbf{Q}^{\frac{1}{2}}$ is a positive definite matrix such that $\overline{\mathbf{W}}^{\top}\hat{\mathbf{D}}_{(\underline{\mathrm{c}})}\overline{\mathbf{W}} = \mathbf{Q}^{\frac{1}{2}}\mathbf{Q}^{\frac{1}{2}}$. Because the number of eigenvalues equal to zero of $\mathbf{Q}^{-\frac{1}{2}}\overline{\mathbf{M}}_{\phi}\mathbf{Q}^{-\frac{1}{2}}$ and $\overline{\mathbf{M}}_{\phi}$ coincide, it follows that $\mathsf{dim}(\overline{\mathbf{M}}_{\phi}) > \mathsf{df}(\hat{\boldsymbol{\phi}}) > q + \sum_{r=1}^{q'} \alpha_{\phi_r}^{(0)}$, where $\alpha_{\phi_r}^{(0)}$ is the number of eigenvalues of $\mathbf{M}_{\phi_r}$ equal to zero.

4.4 Assymptotic theory

Under the absence of nonparametric effects in the systematic component (4.3), the model setup coincides with the topic addressed by Bagdonavičius and Nikulin (2001, chapter 4), which described the asymptotic properties of $\hat{\boldsymbol{\theta}}$. The asymptotic behaviour of $\hat{\boldsymbol{\theta}}$ under the general case of the systematic component (4.3) is studied here by using a framework of fixed-knot (see Wu and Yu, 2014; Yu and Ruppert, 2002), which implies that the penalty matrices and the size of $\boldsymbol{\tau}_{\eta}$ and $\boldsymbol{\tau}_{\phi}$ are not dependent on the sample size n. Next, the required conditions for some asymptotic results be valid, are listed.

<u>**Conditions:**</u>

1. There is $\boldsymbol{\tau}_{\eta_j}^{[0]} \in \mathbb{R}^{p_j'-1}$ such that $\mathrm{f}_{\eta_j}(a) = \mathbf{N}_{\eta_j}^{\top}(a)\boldsymbol{\tau}_{\eta_j}^{[0]}$ for all $a \in (\underline{a}_j, \overline{a}_j)$, $j = 1, \ldots, p'$. Then, $\boldsymbol{\tau}_{\eta}^{[0]} = \left(\left[\boldsymbol{\tau}_{\eta_1}^{[0]}\right]^{\top}, \ldots, \left[\boldsymbol{\tau}_{\eta_{p'}}^{[0]}\right]^{\top}\right)^{\top}$ is the true parameter vector for $\boldsymbol{\tau}_{\eta}$.

2. There is $\boldsymbol{\tau}_{\phi_r}^{[0]} \in \mathbb{R}^{q_r'-1}$ such that $\mathrm{f}_{\phi_r}(b) = \mathbf{N}_{\phi_r}^{\top}(b)\boldsymbol{\tau}_{\phi_r}^{[0]}$ for all $b \in (\underline{b}_r, \overline{b}_r)$, $r = 1, \ldots, q'$. Then, $\boldsymbol{\tau}_{\phi}^{[0]} = \left(\left[\boldsymbol{\tau}_{\phi_1}^{[0]}\right]^{\top}, \ldots, \left[\boldsymbol{\tau}_{\phi_{q'}}^{[0]}\right]^{\top}\right)^{\top}$ is the true parameter vector for $\boldsymbol{\tau}_{\phi}$.

3. The regularity conditions described by Borgan (1984, section 4) and Bagdonavičius and Nikulin (2001, chapter 4) are fulfilled. This conditions ensure that

   (a) $n^{-1}\mathbf{U}(\boldsymbol{\theta}^{[0]}) \xrightarrow[n\to\infty]{\mathcal{P}} \mathbf{0}$, where $\boldsymbol{\theta}^{[0]} = \left(\left[\boldsymbol{\beta}^{[0]}\right]^{\top}, \left[\boldsymbol{\tau}_{\eta}^{[0]}\right]^{\top}, \left[\boldsymbol{\gamma}^{[0]}\right]^{\top}, \left[\boldsymbol{\tau}_{\phi}^{[0]}\right]^{\top}\right)^{\top}$ is the true parameter vector for $\boldsymbol{\theta}$.

   (b) $n^{-1}\mathbf{J}(\boldsymbol{\theta}^{[0]}) \xrightarrow[n\to\infty]{} -\boldsymbol{\Sigma}(\boldsymbol{\theta}^{[0]})$, where $\boldsymbol{\Sigma}(\boldsymbol{\theta}^{[0]})$ is a positive definite matrix.

   (c) $n^{-1}\mathbf{J}(\boldsymbol{\theta}^{*}) \xrightarrow[n\to\infty]{\mathcal{P}} -\boldsymbol{\Sigma}(\boldsymbol{\theta}^{[0]})$ for all $\boldsymbol{\theta}^{*}$ in line segment joining $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}^{[0]}$.

Table 4.1: *Expressions for $\bar{\mathrm{v}}(z_k)$ under some log-symmetric distributions.*

Distribution	$[1-F_{\xi^*}(z_k)]\bar{\mathrm{v}}(z_k)$
log-Student-t	$1-F_\epsilon\Big(z_k\sqrt{(\zeta+2)/\zeta},\zeta+2\Big)^\dagger$
log-slash	$[\zeta/(\zeta+1)]\,[1-F_\epsilon(z_k,\zeta+1)]^\dagger$
log-contaminated-normal	$\zeta_1\zeta_2\big[1-\Phi(z_k\sqrt{\zeta_2})\big]+(1-\zeta_1)[1-\Phi(z_k)]$
log-hyperbolic	$\dfrac{\zeta}{2\,\mathrm{K}_1(\zeta)}\int_{\log\left(z_k+\sqrt{z_k^2+1}\right)}^{\infty}\exp[-\zeta\cosh(t)]\partial t$

† $F_\epsilon(\cdot,\zeta)$ is the cumulative distribution function of a random variable ϵ, such that $\exp(\epsilon)$ exhibits a standard log-symmetric distribution with extra parameter ζ.

(d) $-\mathbf{J}^{-\frac{1}{2}}(\boldsymbol{\theta}^{[0]})\mathbf{U}(\boldsymbol{\theta}^{[0]}) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(\mathbf{0},\mathbf{I})$.

(e) $P\Big[n^{-1}|\partial\mathbf{J}_{il}(\boldsymbol{\theta}^*)/\partial\theta_{i'}|<\omega\Big] \xrightarrow[n\to\infty]{} 1$ for all i,j,i' and for all $\boldsymbol{\theta}^*$ in line segment joining $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}^{[0]}$, where ω is a constant, and $\mathbf{J}_{il}(\boldsymbol{\theta})$ is the (i,l)-th element of $\mathbf{J}(\boldsymbol{\theta})$.

4. The smoothing parameter may be dependent on the sample size. Let $\boldsymbol{\lambda}^{(n)}$ be the value of the smoothing parameter under a sample of size n. Then, $\|\boldsymbol{\lambda}^{(n)}\| \xrightarrow[n\to\infty]{} \bar{\lambda}<\infty$.

Theorem 2. *Under (1)-(4) the maximun penalized likelihood estimator of $\boldsymbol{\theta}$ is consistent and*

$$\Big[-\mathbf{J}(\boldsymbol{\theta}^{[0]})\Big]^{-\frac{1}{2}}\Big[-\mathbf{J}(\boldsymbol{\theta}^{[0]})+\mathbf{M}\Big]\Big(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^{[0]}\Big) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(\mathbf{0},\mathbf{I}).$$

Under (1)-(4) and for large sample sizes, $\hat{\boldsymbol{\theta}}$ is an unbiased estimator of $\boldsymbol{\theta}^{[0]}$ whose variance-covariance matrix may be written as

$$\mathrm{Var}\big[\hat{\boldsymbol{\theta}}\big]=\Big[-\mathbf{J}(\boldsymbol{\theta}^{[0]})+\mathbf{M}\Big]^{-1}\Big[-\mathbf{J}(\boldsymbol{\theta}^{[0]})\Big]\Big[-\mathbf{J}(\boldsymbol{\theta}^{[0]})+\mathbf{M}\Big]^{-1}.$$

An intuitive estimator of the asymptotic variance-covariance matrix of $\hat{\boldsymbol{\theta}}$ reduces to $\mathrm{Var}\big[\hat{\boldsymbol{\theta}}\big]$ evaluated at the $\boldsymbol{\theta}$ estimate.

4.5 Choosing the smoothing parameter

Choosing the smoothing parameter by using a criterion that ensures a compromise between "low" model complexity and "high" goodness-of-fit was considered by Hastie and Tibshirani (1990), Rigby and Stasinopoulos (2005), Wood (2006) and Wu and Yu (2014). Then, in this work, the value of the smoothing parameter $\boldsymbol{\lambda}$ is chosen by minimizing the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC) through an outer iteration. The AIC and BIC criteria are given by

$$\mathsf{AIC}(\hat{\boldsymbol{\theta}}|\boldsymbol{\lambda})=-2\mathsf{L}(\hat{\boldsymbol{\theta}}|\boldsymbol{\lambda})+2[\mathsf{df}(\hat{\boldsymbol{\eta}}|\boldsymbol{\lambda})+\mathsf{df}(\hat{\boldsymbol{\phi}}|\boldsymbol{\lambda})]$$

and

$$\mathsf{BIC}(\hat{\boldsymbol{\theta}}|\boldsymbol{\lambda})=-2\mathsf{L}(\hat{\boldsymbol{\theta}}|\boldsymbol{\lambda})+\log(n)[\mathsf{df}(\hat{\boldsymbol{\eta}}|\boldsymbol{\lambda})+\mathsf{df}(\hat{\boldsymbol{\phi}}|\boldsymbol{\lambda})],$$

where $\hat{\boldsymbol{\theta}}|\boldsymbol{\lambda}$, $\hat{\boldsymbol{\eta}}|\boldsymbol{\lambda}$ and $\hat{\boldsymbol{\phi}}|\boldsymbol{\lambda}$ are the estimates of $\boldsymbol{\theta}$, $\boldsymbol{\eta}$ and $\boldsymbol{\phi}$ given a particular value of the smoothing parameter $\boldsymbol{\lambda}$, respectively.

4.6 Simulation Results

This section presents a simulation study to assess the statistical properties of the maximum penalized likelihood estimates under censored log-symmetric regression models. For this purpose, a data set of size $n = 280$ is simulated, in which the response variable is generated from a log-symmetric distribution, where its median (η) and its skewness (or relative dispersion) (ϕ) can be written as

$$\begin{cases} \log(\eta) = \beta_1 x_1 + \beta_2 x_2 + \mathrm{f}_\eta(a) & \text{and} \\ \log(\phi) = \gamma_1 x_1 + \mathrm{f}_\phi(a), \end{cases}$$

where $x_1 \sim$ Bernoulli(0.5), $x_2 \sim$ log-normal(1, 1), $\mathrm{f}_\eta(a) = 5a + \sin(2\pi a)$, $\mathrm{f}_\phi(a) = 1.2[1.166 - \sin(\pi a)]$, and a is a sequence of seventy values in the interval $[0.05, 0.95]$, which is replicated several times until the sample size is reached. The values assigned to the parameters are $\beta_1 = 2$, $\beta_2 = 0.5$ and $\gamma_1 = -0.2$. To describe the random component, several log-symmetric distributions (e.g., log-normal, log-Student-t, log-slash, log-hyperbolic, log-power-exponential, log-contaminated-normal, Birnbaum-Saunders and Birnbaum-Saunders-t) and several values of their extra parameters are considered. Furthermore, a set of independent and identically distributed censoring times is generated from the exponential distribution with mean 550. So that, to build the response variable and the censoring status, the simulated censoring times and the lifetimes are compared. The resulting sample is used to fit a censored log-symmetric model, where the nonparametric functions $\mathrm{f}_\eta(\cdot)$ and $\mathrm{f}_\phi(\cdot)$ are assumed to be unknown but approximated by cubic P-splines with eight internal knots (given by $q(a, 0/7), q(a, 1/7), q(a, 2/7), \ldots, q(a, 7/7)$), and a *difference penalty* term of order 2. The smoothing parameters are chosen by minimizing the AIC criterion. This process is replicated $R = 5000$ times, keeping the values of x_1, x_2 and a fixed. For the R estimates of β_1, β_2 and γ_1, the following summary measures are calculated: i) empirical expected value, i.e., $\bar{\hat{\theta}} = R^{-1}\sum_{i=1}^{R} \hat{\theta}^{(i)}$; ii) coverage rate of the normality-based 95% confidence interval, i.e., $\mathrm{CR}(\hat{\theta}) = 100{\times}R^{-1}\sum_{i=1}^{R} \mathrm{I}\Big[|\hat{\theta}^{(i)} - \theta\ | \,/[\widehat{\mathrm{Var}}^{(i)}(\hat{\theta})]^{\frac{1}{2}}, [0, 1.96]\Big]$; and iii) p-value of the one-sample Kolmogorov-Smirnov test (see Conover, 1971) to judge the normality of the sample $\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(R)}$. Additionally, as a summary measure of the R estimates of the nonparametric functions $\mathrm{f}_\eta(\cdot)$ and $\mathrm{f}_\phi(\cdot)$, the coverage rate of the simultaneous normality-based 95% confidence intervals is used, i.e., $\mathrm{CR}(\hat{\mathrm{f}}) = 100 \times R^{-1}\sum_{i=1}^{R}\prod_{k=1}^{n} \mathrm{I}\Big[|\hat{\mathrm{f}}^{(i)}(a_k) - \mathrm{f}(a_k)|/[\widehat{\mathrm{Var}}^{(i)}(\hat{\mathrm{f}}(a_k))]^{\frac{1}{2}}, \Theta\Big]$, where $\Theta = [0, \Phi^{-1}(1 - 0.05/2\bar{n})]$, $\bar{n} = 70$ is the number of different values of a in the sample. In each replication, the percentage of censored observations is also calculated . Then, the mean and the standard deviation of the percentage of censored observations is computed. The results are presented in Table 4.2.

The simulation results indicate a good behaviour of the estimates of β_1, β_2 and γ_1 because their means are close to the parameter values and their distributions seem to be very close to the normal one. Furthermore, the coverage rates of the confidence intervals for β_1, β_2 and γ_1 are close to 95%. The simultaneous confidence intervals for $\mathrm{f}_\eta(\cdot)$ and $\mathrm{f}_\phi(\cdot)$ present a good performance, specially in the case of $\mathrm{f}_\eta(\cdot)$. Moreover, the percentage of censored observations is within the interval $(25\%, 30\%)$.

4.7 Diagnostic methods

In this section, some diagnostic methods such as deviance-type residuals for the median and the skewness (or relative dispersion) submodels, overall goodness-of-fit criterion, and local influence measures under right-censored log-symmetric regression models, are addressed.

Table 4.2: *Results of the simulation study*

Error distribution	$\bar{\hat{\theta}}$			$\mathrm{CR}(\hat{\theta})$			p-value of K-S test			$\mathrm{CR}(\hat{\mathrm{f}})$		$\overline{\%\mathrm{Cens.}}$
	β_1	β_2	γ_1	β_1	β_2	γ_1	β_1	β_2	γ_1	f_η	f_ϕ	
log-normal	2.00	0.51	-0.21	93.30	94.07	92.23	0.78	0.58	0.85	93.00	88.17	25.24
log-Student-t(4)	1.99	0.51	-0.21	93.20	93.90	93.23	0.86	0.12	0.97	92.23	87.57	26.52
log-Student-t(6)	2.00	0.51	-0.21	93.27	93.93	92.97	0.89	0.18	0.88	92.03	87.90	26.23
log-Student-t(8)	2.00	0.51	-0.21	92.20	94.33	92.30	0.94	0.17	0.97	92.87	88.43	25.97
log-Student-t(10)	2.00	0.51	-0.21	92.27	94.10	91.43	0.75	0.17	0.97	92.90	87.77	25.86
log-power-exp.(0.2)	2.00	0.51	-0.21	93.87	94.60	92.23	0.76	0.39	0.86	92.87	88.27	26.41
log-power-exp.(0.3)	2.00	0.51	-0.21	92.97	94.70	92.97	0.85	0.19	0.89	91.43	87.47	27.17
log-power-exp.(0.4)	2.00	0.51	-0.20	92.73	94.70	91.87	0.39	0.26	0.98	91.47	87.17	27.82
log-power-exp.(0.5)	1.99	0.50	-0.21	93.30	95.50	92.83	0.76	0.28	0.91	91.10	87.53	29.16
log-hyperbolic(1.0)	1.99	0.50	-0.21	92.37	95.10	92.60	0.91	0.19	0.94	88.73	87.23	29.18
log-hyperbolic(0.9)	1.99	0.51	-0.20	93.97	94.60	92.63	0.88	0.18	0.98	87.77	86.53	29.84
log-hyperbolic(0.8)	2.00	0.51	-0.21	93.20	95.30	92.47	1.00	0.10	0.52	86.97	86.53	30.59
log-hyperbolic(0.7)	1.99	0.51	-0.21	92.30	95.17	93.07	0.97	0.18	0.98	87.70	87.43	31.43
log-slash(1.5)	1.99	0.50	-0.21	92.87	95.23	92.73	0.74	0.04	0.65	87.57	87.77	28.33
log-slash(1.4)	2.00	0.50	-0.21	92.30	95.40	93.00	0.62	0.03	0.30	87.70	86.73	28.53
log-slash(1.3)	1.99	0.51	-0.22	92.27	95.20	92.27	0.98	0.18	0.75	86.70	87.27	28.66
log-slash(1.2)	1.99	0.50	-0.21	91.90	94.83	92.87	0.86	0.02	0.66	87.77	87.80	28.78
log-cont-nor(0.3,0.3)	2.00	0.50	-0.21	93.57	94.70	91.47	0.98	0.26	0.77	91.70	87.37	26.86
log-cont-nor(0.5,0.3)	1.99	0.51	-0.21	93.43	94.50	91.63	0.82	0.27	1.00	89.60	87.27	27.98
log-cont-nor(0.7,0.3)	1.99	0.51	-0.21	93.13	94.20	91.80	0.93	0.39	0.79	88.97	87.07	29.10
log-cont-nor(0.9,0.3)	2.00	0.51	-0.20	93.37	94.47	92.33	0.82	0.83	0.78	88.87	86.77	30.25
B-S(0.1)	2.00	0.50	-0.21	93.73	94.60	93.23	0.99	1.00	0.85	92.10	88.53	20.92
B-S(0.3)	2.00	0.50	-0.21	93.57	93.97	93.80	0.79	0.98	0.81	92.73	88.23	21.05
B-S(0.5)	2.00	0.50	-0.21	93.67	94.07	93.63	0.92	0.65	0.62	93.03	88.53	21.27
B-S(0.5)	2.00	0.50	-0.21	93.27	94.17	93.53	0.92	0.96	0.88	93.53	88.17	21.52
B-S-t(0.1,4)	2.00	0.50	-0.21	93.77	93.83	93.50	0.91	0.51	0.97	92.97	92.37	20.98
B-S-t(0.3,4)	2.00	0.50	-0.21	93.37	94.00	94.00	0.98	0.61	0.49	93.33	90.97	21.22
B-S-t(0.5,4)	2.00	0.50	-0.21	93.90	93.63	93.60	0.98	0.72	0.74	94.00	90.93	21.51
B-S-t(0.7,4)	2.00	0.50	-0.21	93.87	94.20	93.77	0.86	0.83	0.85	94.70	90.50	21.88

4.7.1 Individual goodness-of-fit

To evaluate the goodness-of-fit of the median and the skewness (or relative dispersion) submodels, the deviance is calculated. Thus, the deviance-type residuals for the median and the skewness (or the relative dispersion) submodels are defined as the signed square root of the contribution to the deviance of each individual.

Goodness-of-fit of the median submodel

The deviance of $\hat{\boldsymbol{\eta}}$ given $\hat{\boldsymbol{\phi}}$, denoted by $\mathsf{D}(\hat{\boldsymbol{\eta}}|\hat{\boldsymbol{\phi}})$, is given by

$$\mathsf{D}(\hat{\boldsymbol{\eta}}|\hat{\boldsymbol{\phi}}) = 2\sum_{k=1}^{n}\delta_k \log\Big[\Big(1 - F_{\xi^*}\Big([y_k - \tilde{\mu}_k]/[\hat{\phi}_k]^{\frac{1}{2}}\Big)\Big)/\Big(1 - F_{\xi^*}\Big([y_k - \hat{\mu}_k]/[\hat{\phi}_k]^{\frac{1}{2}}\Big)\Big)\Big] \\ + (1-\delta_k)\Big[\mathsf{L}_k(\tilde{\mu}_k, \hat{\phi}_k) - \mathsf{L}_k(\hat{\mu}_k, \hat{\phi}_k)\Big]\,,$$

where $\tilde{\mu}_k$ is the value of $\mathbb{R}\cup\{\infty\}$ that maximizes the function $\delta_k \log\Big[1 - F_{\xi^*}\Big([y_k - \mu]/[\hat{\phi}_k]^{\frac{1}{2}}\Big)\Big] +$ $(1-\delta_k)\mathsf{L}_k(\mu, \hat{\phi}_k)$ with respect to μ. Hence, $\mathsf{D}(\hat{\boldsymbol{\eta}}|\hat{\boldsymbol{\phi}})$ becomes

$$\mathsf{D}(\hat{\boldsymbol{\eta}}|\hat{\boldsymbol{\phi}}) = \sum_{k=1}^{n} \mathsf{d}_k(\hat{\boldsymbol{\eta}}|\hat{\boldsymbol{\phi}}),$$

where $\mathsf{d}_k(\hat{\boldsymbol{\eta}}|\hat{\boldsymbol{\phi}})$ may be used to define a residual associated with $\hat{\eta}_k$ as follows

$$t_\eta(\hat{z}_k) = \operatorname{sign}(\hat{z}_k)\Big[\mathsf{d}_k(\hat{\boldsymbol{\eta}}|\hat{\boldsymbol{\phi}})\Big]^{\frac{1}{2}}.$$

If the function $g(\cdot)$ is monotically decreasing for $u \geq 0$, then,

$$\mathsf{d}_k(\hat{\boldsymbol{\eta}}|\hat{\boldsymbol{\phi}}) = 2(1-\delta_k)\log[g(0)/g(\hat{z}_k^2)] - 2\delta_k \log[1 - F_{\xi^*}(\hat{z}_k)],$$

in which $\hat{z}_k = (y_k - \hat{\mu}_k)/[\hat{\phi}_k]^{\frac{1}{2}}$.

Goodness-of-fit of the skewness (or the relative dispersion) submodel

The deviance of $\hat{\boldsymbol{\phi}}$ given $\hat{\boldsymbol{\eta}}$, denoted by $\mathsf{D}(\hat{\boldsymbol{\phi}}|\hat{\boldsymbol{\eta}})$, is given by

$$\begin{aligned}\mathsf{D}(\hat{\boldsymbol{\phi}}|\hat{\boldsymbol{\eta}}) = 2\sum_{k=1}^{n}&\delta_k \log\Big[\Big(1 - F_{\xi^*}\Big([y_k - \hat{\mu}_k]/[\tilde{\phi}_k]^{\frac{1}{2}}\Big)\Big)\Big/\Big(1 - F_{\xi^*}\Big([y_k - \hat{\mu}_k]/[\hat{\phi}_k]^{\frac{1}{2}}\Big)\Big)\Big] \\ &+ (1-\delta_k)\Big[\mathsf{L}_k(\hat{\mu}_k, \tilde{\phi}_k) - \mathsf{L}_k(\hat{\mu}_k, \hat{\phi}_k)\Big],\end{aligned}$$

where $\tilde{\phi}_k$ is the value of $\mathbb{R}^+\cup\{\infty\}$ that maximizes the function $\delta_k \log\Big[1 - F_{\xi^*}\Big([y_k - \hat{\mu}_k]/[\phi]^{\frac{1}{2}}\Big)\Big] +$ $(1-\delta_k)\mathsf{L}_k(\hat{\mu}_k, \phi)$ with respect to ϕ. Then, $\mathsf{D}(\hat{\boldsymbol{\phi}}|\hat{\boldsymbol{\eta}})$ becomes

$$\mathsf{D}(\hat{\boldsymbol{\phi}}|\hat{\boldsymbol{\eta}}) = \sum_{k=1}^{n} \mathsf{d}_k(\hat{\boldsymbol{\phi}}|\hat{\boldsymbol{\eta}}),$$

where

$$\mathsf{d}_k(\hat{\boldsymbol{\phi}}|\hat{\boldsymbol{\eta}}) = 2(1-\delta_k)\log[g(\varrho^2)/g(\hat{z}_k^2)] - \log[\hat{z}_k^2/\varrho^2] - 2\delta_k \log[1 - F_{\xi^*}(\hat{z}_k)] - \delta_k[1 + \operatorname{sign}(\hat{z}_k)]\log(2).$$

If the function $\mathsf{L}_k(\hat{\mu}_k, \phi)$ has just one critic point, then ϱ is the solution of the equation $\mathrm{v}(\varrho)\varrho^2 = 1$. $\mathsf{d}_k(\hat{\boldsymbol{\phi}}|\hat{\boldsymbol{\eta}})$ may be used to define a residual associated with $\hat{\phi}_k$ by $t_\phi(\hat{z}_k) = \operatorname{sign}(\hat{z}_k)[\mathsf{d}_k(\hat{\boldsymbol{\phi}}|\hat{\boldsymbol{\eta}})]^{\frac{1}{2}}$.

4.7.2 Overall goodness-of-fit

The overall goodness-of-fit is measured through the following statistic, which has the advantage of graphical representation as it is based on the quantile-quantile plot for right-censored observations (see, e.g., Waller and Turnbull, 1992):

$$\Upsilon = \sum_{k=1}^{n}(1-\delta_k)\mid \Phi^{-1}[F_{\xi^*}(\hat{z}_k)] - \Phi^{-1}[\bar{F}_{\xi^*}(\hat{z}_k)]\mid \Big/ \sum_{k=1}^{n}(1-\delta_k),$$

where $\bar{F}_{\xi^*}(\cdot)$ is the cumulative distribution function of ξ^* estimated from $(\hat{z}_1, \delta_1), \ldots, (\hat{z}_n, \delta_n)$ by using the nonparametric Kaplan-Meier estimator (Kaplan and Meier, 1958). If the estimates of $\boldsymbol{\eta}$, $\boldsymbol{\phi}$ and ζ coincide with the true parameter values, then $(\hat{z}_1, \delta_1), \ldots, (\hat{z}_n, \delta_n)$ represents a right-censored sample obtained from the distribution of ξ^*, where the censoring times are independent but non-identically distributed variables (see, e.g., Zhou, 1991). The smaller is the value of Υ, the better is the goodness-of-fit of the model. Graphically, the criterion Υ indicates that the smaller is the difference between the plot of $\Phi^{-1}[F_{\xi^*}(\hat{z}_k)]$ versus $\Phi^{-1}[\bar{F}_{\xi^*}(\hat{z}_k)]$ (for all k such that

$\delta_k = 0$) and a straight line (with zero intercept and unit slope), the better is the goodness-of-fit of the model. Furthermore, if ζ is unknown, its value may be selected by minimizing Υ.

4.7.3 Influence or sensitivity analysis

The normal curvature in the direction $\mathbf{d}$, such that $\|\mathbf{d}\| = 1$, is defined as (Cook, 1986)

$$\mathsf{C}_d(\hat{\boldsymbol{\theta}}) = 2|\mathbf{d}^\top \Delta^\top(\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}_0)[\mathbf{J}(\hat{\boldsymbol{\theta}})]^{-1}\Delta(\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}_0)\,\mathbf{d}|, \tag{4.5}$$

where $\Delta(\boldsymbol{\theta}, \boldsymbol{\omega}) = \partial^2 \mathsf{PL}(\boldsymbol{\theta}|\boldsymbol{\omega})/\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^\top$. $\mathsf{C}^*_d(\hat{\boldsymbol{\theta}}) = \mathsf{C}_d(\hat{\boldsymbol{\theta}})/2\sqrt{\mathsf{tr}(\mathbf{V}^\top\mathbf{V})}$ is the conformal normal curvature proposed by Poon and Poon (1999), which is invariant under uniform change of scale, where $\mathbf{V} = \Delta^\top(\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}_0)[\mathbf{J}(\hat{\boldsymbol{\theta}})]^{-1}\Delta(\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}_0)$. This curvature allows that $0 \leq \mathsf{C}^*_d(\hat{\boldsymbol{\theta}}) \leq 1$ for any unitary direction $\mathbf{d}$. A maximum curvature, denoted by $\mathsf{C}^*_{d_{\max}}$, is obtained in the direction $\mathbf{d}^*_{\max}$, where $\mathsf{C}^*_{d_{\max}}$ is the largest eigenvalue of $\mathbf{V}/\sqrt{\mathsf{tr}(\mathbf{V}^\top\mathbf{V})}$, and $\mathbf{d}^*_{\max}$ is its corresponding eigenvalue. The local influence measure can be used to identify observations that may jointly influence the fitted model, and it is calculated from the eigenvector that corresponds to the highest eigenvalue of the matrix of conformal normal curvature. Similarly, the total local influence measure can be used to identify observations that may individually exert influence on the fitted model, and it is calculated from the main diagonal of the matrix of conformal normal curvature.

Next, the expressions of $\Delta(\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}_0)$ under two usual perturbation schemes are provided.

Case-weight perturbation scheme

Under this perturbation scheme $\boldsymbol{\omega}_0 = (1, \ldots, 1)^\top$ and $\Delta(\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}_0)$ is given by

$$\Delta(\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}_0) = \begin{bmatrix} \overline{\mathbf{X}}^\top \hat{\mathbf{D}}_{(\tilde{\mathrm{v}})} \hat{\boldsymbol{\Omega}}^{-\frac{1}{2}} \hat{\mathbf{D}}_{(z)} \\ \frac{1}{2}\overline{\mathbf{W}}^\top \left(\hat{\mathbf{D}}_{(\tilde{s})} - \mathbf{I}\right) \end{bmatrix},$$

where $\hat{\mathbf{D}}_{(z)}$ and $\hat{\mathbf{D}}_{(\tilde{s})}$ correspond to $\mathbf{D}_{(z)} = \mathrm{diag}\{z_1, \ldots, z_n\}$ and $\mathbf{D}_{(\tilde{s})} = \mathrm{diag}\{\tilde{s}_1, \ldots, \tilde{s}_n\}$ evaluated at the $\boldsymbol{\theta}$ estimate, respectively.

Response perturbation scheme

Under this perturbation scheme y_k is replaced by $y_k^{(\omega)} = y_k + \omega_k$ and $\boldsymbol{\omega}_0 = (0, \ldots, 0)^\top$. Then,

$$\Delta(\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}_0) = \begin{bmatrix} \overline{\mathbf{X}}^\top \hat{\boldsymbol{\Omega}}^{-1} \hat{\mathbf{D}}_{(\mathrm{c})} \\ \overline{\mathbf{W}}^\top \hat{\boldsymbol{\Omega}}^{-\frac{1}{2}} \hat{\mathbf{D}}_{(\bar{\mathrm{c}})} \end{bmatrix},$$

where $\hat{\mathbf{D}}_{(\bar{\mathrm{c}})}$ correspond to $\mathbf{D}_{(\bar{\mathrm{c}})}$ evaluated at the $\boldsymbol{\theta}$ estimate.

CHAPTER 5

THE PACKAGE SSYM

There are few packages in the R statistical computing environment (R Core Team, 2014) that facilitate the analysis of data for which the response variable is continuous, strictly positive, and asymmetric with possible outlying observations, especially when several parameters of the distribution of the response variable are dependent on explanatory variables. One such package is **gamlss** (Rigby and Stasinopoulos, 2007, 2014), which is an implementation of the generalized additive models for location, scale and shape (GAMLSS) (Rigby and Stasinopoulos, 2005). Similarly, heteroscedastic nonlinear regression models are implemented in the package **nlsmsn** (Garay *et al.*, 2013), which were introduced by Lachos *et al.* (2011) based on a scale mixture of skew-normal distribution. Although **gamlss** and **nlsmsn** are practical and flexible packages, they do not completely address the diagnostic methods (i.e., residuals and local influence measures) for each involved submodel, which frequently hinders the validation of the estimated models. In addition, **nlsmsn** has the limitation of assuming that the skewness of the response variable distribution is constant across the observations. Moreover, some routines of the R packages **survival** (Therneau, 2014), **rms** (Harrell, 2015) and **eha** (Brostom, 2014) allow to fit parametric accelerated failure time models. However, none of them are flexible enough to enable the specification of nonlinear effects whose functional form is assumed to be unknown. Furthermore, the model checking provided by these routines is based just on residual analysis and the influence or sensitivity analysis is not supported. Finally, none of these packages provides an implementation in the context of semi-parametric regression for distributions such as Birnbaum-Saunders and Birnbaum-Saunders-t, which have applications in several fields (see, e.g., Leiva *et al.*, 2008, and references therein) because they have been developed to describe lifetimes under the assumption of cumulative damage.

This chapter describes the capabilities and features of the new package **ssym** (Vanegas and Paula, 2014), which is an implementation of semiparametric log-symmetric models under the presence of right-censored or uncensored observations. This package, available from the Comprehensive R Archive Network (CRAN) at `http://CRAN.R-project.org/package=ssym`, also provides some functions to perform the residual analysis and the sensitivity analysis, as well as some graphic tools to draw the estimated nonparametric effects jointly with their simultaneous/pointwise confidence intervals.

5.1 Overview

The package **ssym** fits and obtains diagnostic statistics (deviance-type residuals for each submodel, local influence measures and goodness-of-fit statistics) for semi-parametric symmetric models, in which the distribution of the additive random error may be normal, Student-t, power exponential, contaminated normal, slash, symmetric hyperbolic, sinh-normal, or sinh-t, and in which the location (μ) and the dispersion (ϕ) parameters may be described using semi-parametric functions of explanatory variables in either of the following setups:

$$
(I)\quad \begin{cases} \mu_k = \mu(\mathbf{x}_k, \boldsymbol{\beta}) \quad \text{and} \\ \log(\phi_k) = \mathbf{w}_k^\top \boldsymbol{\gamma} + \sum_{r=1}^{q'} \mathrm{f}_{\phi_r}(b_{k,r}); \end{cases}
\qquad
(II)\quad \begin{cases} \mu_k = \mathbf{x}_k^\top \boldsymbol{\beta} + \sum_{j=1}^{p'} \mathrm{f}_{\eta_j}(a_{k,j}) \quad \text{and} \\ \log(\phi_k) = \mathbf{w}_k^\top \boldsymbol{\gamma} + \sum_{r=1}^{q'} \mathrm{f}_{\phi_r}(b_{k,r}), \end{cases} \tag{5.1}
$$

where $\mathbf{x}_k^* = (\mathbf{x}_k^\top, a_{k,1}, \ldots, a_{k,p'})^\top$ and $\mathbf{w}_k^* = (\mathbf{w}_k^\top, b_{k,1}, \ldots, b_{k,q'})^\top$ are vectors of explanatory variables for μ_k and ϕ_k, respectively. In particular, the functions `ssym.nl()` and `ssym.l()` of the package **ssym** fit symmetric models under the setups described by I and II, respectively, in expression (5.1), in which the nonparametric functions are approximated by natural cubic splines or P-splines. Thus, according to Section 3.2, the functions `ssym.nl()` and `ssym.l()` can also be used to fit log-symmetric models under the setups described by I and II, respectively, in expression (3.3) and where the distribution of the multiplicative random error may be log-normal, log-Student-t, log-power-exponential, log-contaminated-normal, log-slash, log-hyperbolic, (extended) Birnbaum-Saunders or (extended) Birnbaum-Saunders-t. Furthermore, the routine `ssym.l2()` allows to fit a log-symmetric model under the presence of random right-censored observations as described in the previous chapter.

5.2 The model-fitting functions

The arguments of the model-fitting functions `ssym.nl()`, `ssym.l()` and `ssym.l2()` are

```
ssym.nl(formula, start, family, xi, data, local.influence = FALSE,
        subset, maxiter = 1000, epsilon = 1e-07),
```

```
ssym.l(formula, family, xi, data, local.influence = FALSE,  subset,
       maxiter = 1000, epsilon = 1e-07),
```

and

```
ssym.l2(formula, family, xi, data, local.influence = FALSE,  subset,
        maxiter = 1000, epsilon = 1e-07),
```

respectively. The systematic component of the model must be specified in the argument `formula`. The function `Formula()` of the package **Formula**, which was written by Zeileis and Croissant (2010), has been invoked because it enables the simultaneous specification of the two submodels involved in the log-symmetric model. Thus, the argument `formula` comprises three parts, namely, the response variable in logarithmic scale, the regressors of $\log(\eta)$, and the regressors of $\log(\phi)$; the first two are separated by the symbol "$\sim$" and the second and third parts are separated by the symbol "|". For instance, under the presence of an uncensored sample, a log-symmetric model with the observed response variable t and the systematic component given by $\log(\eta) = \beta_1 + \beta_2 x_1 + \mathrm{f}_{\eta_1}(x_2) + \mathrm{f}_{\eta_2}(x_3)$ and $\log(\phi) = \gamma_1 + \gamma_2 z_1 + \mathrm{f}_{\phi_1}(z_2) + \mathrm{f}_{\phi_2}(z_3)$, should be specified as follows:

```
ssym.l(log(t) ~ x1 + ncs(x2) + ncs(x3) | z1 + ncs(z2) + ncs(z3),...)
```
or
```
ssym.l(log(t) ~ x1 + psp(x2) + psp(x3) | z1 + psp(z2) + psp(z3),...)
```

when the nonparametric functions are approximated by natural cubic splines or P-splines, respectively. Similarly, a log-symmetric model with the systematic component given by $\eta = \exp(\beta_1 x_1^{\beta_2})$ and $\log(\phi) = \gamma_1 + \gamma_2 z_1 + \mathrm{f}_{\phi_1}(z_2) + \mathrm{f}_{\phi_2}(z_3)$, should be specified as follows:

```
ssym.nl(log(t) ~ b1*x1^b2 | z1 + ncs(z2) + ncs(z3),...)
```

or
```
ssym.nl(log(t) ~ b1*x1^b2 | z1 + psp(z2) + psp(z3),...)
```

when the nonparametric functions are approximated by natural cubic splines or a P-splines, respectively. Moreover, under the presence of a right-censored sample, a log-symmetric model with the observed lifetime (or censoring time) t, censoring status `event`, and the systematic component given by $\log(\eta) = \beta_1 + \beta_2 x_1 + \mathrm{f}_{\eta_1}(x_2) + \mathrm{f}_{\eta_2}(x_3)$ and $\log(\phi) = \gamma_1 + \gamma_2 z_1 + \mathrm{f}_{\phi_1}(z_2) + \mathrm{f}_{\phi_2}(z_3)$, should be specified as follows:

```
ssym.l2(Surv(log(t),event) ~ x1 + ncs(x2) + ncs(x3) | z1 + ncs(z2)
+ ncs(z3),...)
```
or
```
ssym.l2(Surv(log(t),event) ~ x1 + psp(x2) + psp(x3) | z1 + psp(z2)
+ psp(z3),...)
```

when the nonparametric functions are approximated by natural cubic splines or P-splines, respectively. Note that the absence of the third part of the argument `formula` indicates to **ssym** that the model must be fitted assuming ϕ constant across the observations. In the functions `ssym.nl()`, `ssym.l()` and `ssym.l2()` the distribution of $\log(\xi)$ and its extra parameter value (ζ) are specified in the arguments `family` and `xi`, respectively. In effect, `family = "Normal"`, `"Student"`, `"Powerexp"`, `"Hyperbolic"`, `"Slash"`, `"Contnormal"`, `"Sinh-normal"` and `"Sinh-t"` correspond to normal, Student-t, power exponential, symmetric hyperbolic, slash, contaminated normal, sinh-normal and sinh-t distributions, respectively, where the parametric space of ζ was described in Section 2.3 for all distributions supported by **ssym**. The the functions `ssym.nl()`, `ssym.l()` and `ssym.l2()` also enable manage the iterative process of parameter estimation via the options `maxiter` and `epsilon`, which control the maximum number of iterations and the convergence criterion of the algorithm, respectively. Additionally, if the option `local.influence` is `TRUE`, then the local influence measures are calculated for $\hat{\tilde{\boldsymbol{\beta}}}$ and $\hat{\boldsymbol{\theta}}$ under case-weight and response perturbation schemes. In addition, the option `subset` enables a specified subset of individuals to be employed in the fitting process. By default, the smoothing parameter of a natural cubic spline or a P-spline is estimated from the data by minimizing the AIC or BIC criteria.

5.3 Standard functions

A set of standard extractor functions for the fitted model objects is available for the objects of the class "ssym", including methods for the generic functions `print()`, `summary()`, `plot()`, `coef()`, `vcov()`, `logLik()`, `AIC()`, `BIC()`, `residuals()` and `fitted()`. Next, a description of these standard functions is provided.

5.3.1 Summary

`summary()` produces a complete summary of the model fit including parameter estimates, associated standard errors, deviance values, p-values and degrees of freedom associated with the nonparametric components, as well as the values of the log-likelihood function (i.e., $\mathsf{L}(\hat{\boldsymbol{\theta}})$), AIC , BIC and the overall goodness-of-fit statistic Υ. In addition, `summary()` displays the quantiles of the standardized individual-specific weights (i.e., quantiles of $\rho(\hat{z}_k) = \mathrm{v}(\hat{z}_k)/d_g(\zeta)$, $k = 1, \ldots, n$) and the percentage of censored observations under the presence of uncensored and censored observations, respectively.

5.3.2 Estimating equations

`estfun.ssym()` extracts the score function evaluated at observed data and estimated parameters, which enables one to verify that the estimates provided by `ssym.nl()`, `ssym.l()` or `ssym.l2()` satisfy the estimating equations.

5.3.3 Goodness-of-fit statistics

The `logLik()`, `AIC()` and `BIC()` functions calculate the value of the log-likelihood function, as well as the AIC and BIC values, respectively, in the scale of the transformed variable, Y. In addition, the commands `attr(logLik(),"log")`, `attr(AIC(),"log")` and `attr(BIC(),"log")` display the value of the log-likelihood function, as well as the AIC and BIC values, respectively, in the scale of the original variable, T.

5.3.4 Diagnostic graphs

Under the presence of uncensored observations, the function `plot()` produces a graph of the standardized individual-specific weights versus the ordinary residuals, and a graph of the overall goodness-of-fit statistic Υ, i.e., a graph of $\Phi^{-1}[F^{*}_{\xi}(\hat{z}_k)]$ versus $\upsilon^{(k)}$, $k = 1, \ldots, n$. This function also displays graphs of the deviance-type residuals versus the fitted values for the median and the skewness (or the relative dispersion) submodels. However, under the presence of right-censored observations, the function `plot()` produces a graph of the hazard rate function of the error distribution ξ instead of a graph of the standardized individual-specific weights.

5.3.5 Parameter estimates

The function `coef()` extracts the parameter estimates for both submodels. Similarly, the function `vcov()` extracts the approximate variance-covariance matrix associated to the parameter estimates.

5.3.6 Fitted values

`fitted()` extracts the values of $\hat{\mu}_k = \log(\hat{\eta}_k)$ and $\hat{\phi}_k$, $k = 1, \ldots, n$.

5.3.7 Residuals

`residuals()` calculates the deviance-type residuals for both submodels. However, under the presence of uncensored observations, the overall residuals (i.e., $\Phi^{-1}[F^{*}_{\xi}(\hat{z}_k)]$) and the ordinary residuals (i.e., $\hat{z}_k$) are also calculated.

5.3.8 Local influence measures

If the option `local.influence` is `TRUE` in the call to `ssym.nl()`, `ssym.l()` or `ssym.l2()`, then the function `influence()` can extract the local influence measures (local influence and total local influence based on the conformal normal curvature) for $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\theta}}$ under the case-weight and response perturbation schemes. Graphs of the local influence measures for $\hat{\boldsymbol{\theta}}$ are automatically displayed by the call of the function `influence()`.

5.4 Other useful functions

5.4.1 Basis and penalty matrices

The arguments of the routines `ncs()` and `psp()` are

```
ncs(x,lambda, nknots, all.knots=FALSE)
```

and

```
psp(x, lambda, b.order=3, nknots, diff=2),
```

respectively. These functions are used to construct the basis ($\dot{\mathbf{N}}$) and the penalty ($\dot{\mathbf{M}}$) matrices to approximate a smooth function of x by using a natural cubic spline or a P-spline. The smoothing parameter may be provided by the user through the option `lambda`. By default, the routines `ncs()` and `psp()` use $m = [n^{\frac{1}{3}}] + 3$ knots (or internal knots, in the case of P-splines), which are given by $q(a, 1/(m+1)), \ldots, q(a, m/(m+1))$, with $q(a, w)$ being the quantile of order $0 < w < 1$ of a. However, the number the knots can be introduced by the user using the argument `nknots`. In addition, under P-splines, the degree of the B-spline and the degree of the *difference penalty term* can be specified by using the arguments `b.order` and `diff`, respectively.

5.4.2 Graphs of the nonparametric effects

From an object of class "ssym", the function

```
np.graph(object, which, exp=FALSE, simul=TRUE, obs=FALSE, var,...)
```

can display a graph of the fitted nonparametric effects jointly with their simultaneous (or pointwise provided that `simul=FALSE`) normality-based 95% confidence intervals, which are either natural cubic splines or P-splines. If `obs=TRUE` the displayed graph include the observed data adjusted by the parametric effects in the corresponding submodel, if they exist. The interest submodel is selected using the argument `which`, where 1 indicates median submodel and 2 indicates skewness (or the relative dispersion) submodel. The argument `var` allows to choosing the nonparametric effect using the name of the associated explanatory variable.

5.4.3 Simulated envelopes

From an object of class "ssym", the routine

```
envelope(object, reps=25, conf=0.95)
```

calculates and displays graphs of the deviance-type residuals with simulated envelope for the median and the skewness (or the relative dispersion) submodels. The arguments `reps` and `conf` represent the number of iterations and the confidence level for the simulated envelopes, respectively. A progress bar is displayed while the envelopes are calculated.

5.4.4 Choosing the extra parameter

From an object of class "ssym", the function

```
extra.parameter(object, lower, upper)
```

calculates and displays graphs of the overall goodness-of-fit statistic Υ and $-2\mathsf{L}(\hat{\boldsymbol{\theta}})$ versus the extra parameter ζ in the interval/region defined by the arguments `lower` and `upper`. These graphs may be used to choosing the extra parameter value. A progress bar is displayed while the graphs are calculated.

5.4.5 Random generation

The function

```
rvgs(n, family, xi)
```

enables the random generation of variates from the (standard) normal, Student-t, power exponential, slash, symmetric hyperbolic, contaminated normal, sinh-normal and sinh-t distributions. Then, the function `exp(rvgs())` can be used for the random generation of log-normal, log-Student-t, log-power-exponential, log-slash, log-hyperbolic, log-contaminated normal, Birnbaum-Saunders and Birnbaum-Saunders-t distributions. In the case of the power exponential distribution `rvgs()` calls the function `rnormp()` of the R package **normalp**. Syntax of the arguments `family` and `xi` coincides with that of the homonymous arguments in the functions `ssym.nl()` and `ssym.l()`.

5.4.6 Datasets

The package **ssym** includes several data sets to illustrate the use of its main functions. For example, it contains the following data sets: Snacks (Paula, 2013), Biaxial (Rieck and Nedelman, 1991), Claims (de Jong and Heller, 2008, pag 14), European Rabbits (Dudzinski and Mykytowycz , 1961), Gross Domestic Product (Vanegas and Paula, 2014b), and Ovocytes (LeGal *et al.*, 1984).

CHAPTER 6

APPLICATIONS OF LOG-SYMMETRIC MODELS

In this chapter, the statistical and computational tools presented in the previous chapters are illustrated by analyzing six real data sets using the R package **ssym**. Further examples can be found at `http://cran.r-project.org/web/packages/ssym/ssym.pdf`.

6.1 Boston housing

This dataset, recently analyzed by Ibacache-Pulgar *et al.* (2013) and Wu and Yu (2014) using semi-parametric models, is available at the R package **MASS** and consists of 506 observations of the dependent variable `medv`, median value of owner occupied homes in towns of Boston in 1970's (in thousands of US dollars), and other 13 independent variables, which include `crim`: per capita crime rate; `rm`: average number of rooms per dwelling; `dis`: weighted distances to five Boston employed centers; `tax`: full-value property-tax rate per USD 10,000; `lstat`: percentage of lower status of the population. It is proposed to analyze the dataset using a log-symmetric model where the median (η) and the skewness of the response distribution are given by

$$\begin{cases} \log(\eta_k) = \beta_1 + \beta_2\texttt{crim}_k + \beta_3\texttt{rm}_k + \beta_3\texttt{tax}_k + \mathrm{f}_{\eta_1}(\texttt{lstat}_k) + \mathrm{f}_{\eta_2}(\texttt{dis}_k), \\ \log(\phi_k) = \gamma_1 + \mathrm{f}_{\phi}(\texttt{lstat}_k), \qquad\qquad k = 1, \ldots, 506, \end{cases}$$

where $\mathrm{f}_{\eta_1}(\cdot)$, $\mathrm{f}_{\eta_2}(\cdot)$ and $\mathrm{f}_{\phi}(\cdot)$ are nonparametric functions approximated by using P-splines. Table 6.1 presents the values of Υ, AIC and BIC for the fitted models under various distributions of the model error. In all cases, the extra parameter ζ was selected by using the criterion of the overall goodness-of-fit Υ. It can be seen that the error distributions with heavy tails outperform the goodness-of-fit provided by the log-normal distribution. The log-slash model was selected to describe the data because it yields the lowest values of Υ, AIC and BIC. The behaviour of the overall goodness-of-fit statistic Υ with respect to the extra parameter ζ under the log-slash model is illustrated by the Figure 6.1(a).

Table 6.1: *Goodness-of-fit statistics for the fitted models to the Boston Housing data.*

Error distribution	Υ	AIC	BIC
log-normal	0.0642	-328.03	-206.04
log-Student-t(5.4)	0.0310	-379.96	-283.45
log-power-exponential(0.56)	0.0397	-362.14	-270.99
log-hyperbolic(1.2)	0.0350	-369.18	-275.50
log-slash(1.56)	0.0269	-386.22	-288.54
log-contaminated-normal(0.1,0.15)	0.0309	-383.87	-287.38

In **ssym**, the selected model can be fitted via

```
> data("Boston", package="MASS")
> fit <- ssym.l(log(medv) ~ crim + rm + tax + psp(lstat) + psp(dis) | psp(lstat),
+        data=Boston, family="Slash", xi=1.56, local.influence=TRUE)
> summary(fit)

     Family:  Slash ( 1.56 )
Sample size:  506
 Quantile of the Weights
   0%  25%  50%  75% 100%
 0.06 1.20 1.30 1.33 1.34
 ************************** Median/Location submodel **************************
 ******** Parametric component
              Estimate Std.Err z-value  Pr(>|z|)
(Intercept)  1.72590440  0.0829 20.8204 < 2.2e-16 ***
crim        -0.01076789  0.0014 -7.7320 1.059e-14 ***
rm           0.24020970  0.0120 20.0857 < 2.2e-16 ***
tax         -0.00026966  0.0001 -4.5669 4.951e-06 ***
 ******** Nonparametric component
          Smooth.param Basis.dimen   d.f. Statis. p-value
psp(lstat)        62.08      11.000  6.554   323.9  <2e-16 ***
psp(dis)          10.13      11.000  8.865   271.5  <2e-16 ***
 **** Deviance:  652
 ************************* Skewness/Dispersion submodel ***********************
 ******** Parametric component
           Estimate  Std.Err  z-value  Pr(>|z|)
(Intercept)  -4.2342   0.0967 -43.7757 < 2.2e-16 ***
 ******** Nonparametric component
          Smooth.param Basis.dimen   d.f. Statis. p-value
psp(lstat)        33.25      11.000  2.692   112.8  <2e-16 ***
 **** Deviance:  645.68
 ******************************************************************************
 Overall goodness-of-fit statistic:  0.026859
                -2*log-likelihood:  -432.445
                              AIC:  -386.224
                              BIC:  -288.545
```

Figure 6.1(a) can be reproduced by using the command `extra.parameter(fit,1.0,2.3)`. Figures 6.1(b)-(d) present the estimates of the nonparametric functions with the 95% simultaneous confidence intervals. These graphs can be reproduced by using the instructions `np.graph(fit,which=1,e` `np.graph(fit,which=1,exp=TRUE,"dis")` and `np.graph(fit,which=2,exp=TRUE,"lstat")`, respectively. The p-values associated with the nonparametric effects indicate that the distribution of the response variable depend on the variables `lstat` and `dis`. Figures 6.2(a)-(b) present the deviance-type residuals with simulated envelope. These graphs do not reveal any discrepant individual and they indicate that the fitted log-slash model describes the data adequately. The graphs of the local influence measures (Figures 6.2(c)-(d)) allows to identify the groups of individuals {369,370,371,372,373}, {370}, {371}, {372}, {233}, {268} and {419} as potentially influential on $\hat{\boldsymbol{\theta}}$. The elimination of these individuals does not introduce inferential changes.

6.2 Textures of five different types of snacks

A data set from an experiment developed in the School of Public Health - University of São Paulo, in which four different forms of light snacks (denoted by B, C, D, and E) were compared with a traditional snack (denoted by A) for 20 weeks. For the light snacks, the hydrogenated vegetable fat (hvf) was replaced by canola oil using different proportions: B (0% hvf, 22% canola oil), C (17% hvf, 5% canola oil), D (11% hvf, 11% canola oil) and E (5% hvf, 17% canola oil); A (22% hvf, 0% canola oil) contained no canola oil. The experiment was conducted such that a random sample of 15 units of each snack type was analyzed in a laboratory in each even week to measure various variables. A total of 75 units was analyzed in each even week; with 750 units being analyzed during the experiment (Paula, 2013). Only the variable `texture` was considered and compared over time for the five snack types. This data set (`Snacks`) is available in the package **ssym**, and the objects `texture`, `type` and `week` represent homonyms variables.

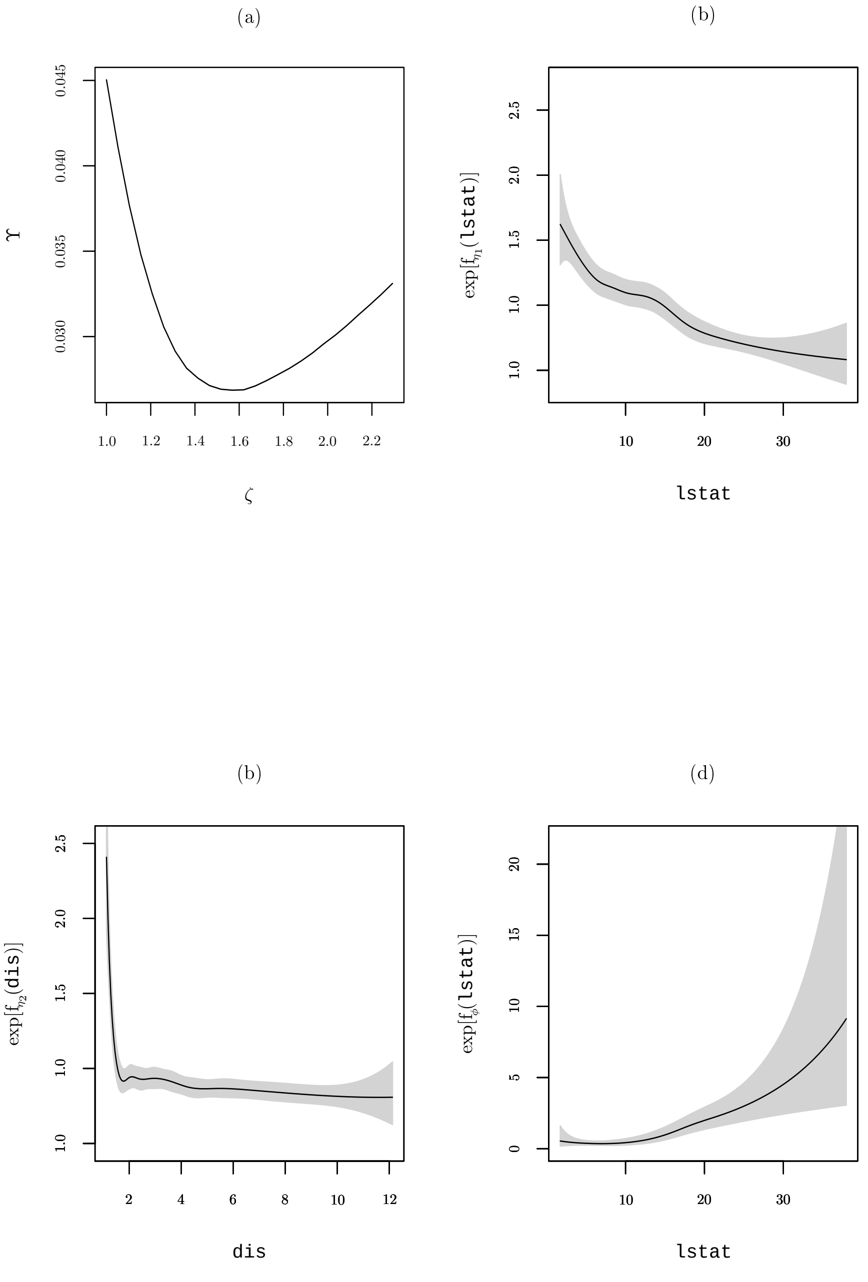

Figure 6.1: *Graphs of* Υ *versus* ζ *(a), simultaneous 95% confidence intervals for* $\exp[f_{\eta_1}(\texttt{lstat})]$ *(b),* $\exp[f_{\eta_2}(\texttt{dis})]$ *(c) and* $\exp[f_{\phi}(\texttt{lstat})]$ *(d) of the log-slash model fitted to the Boston Housing data.*

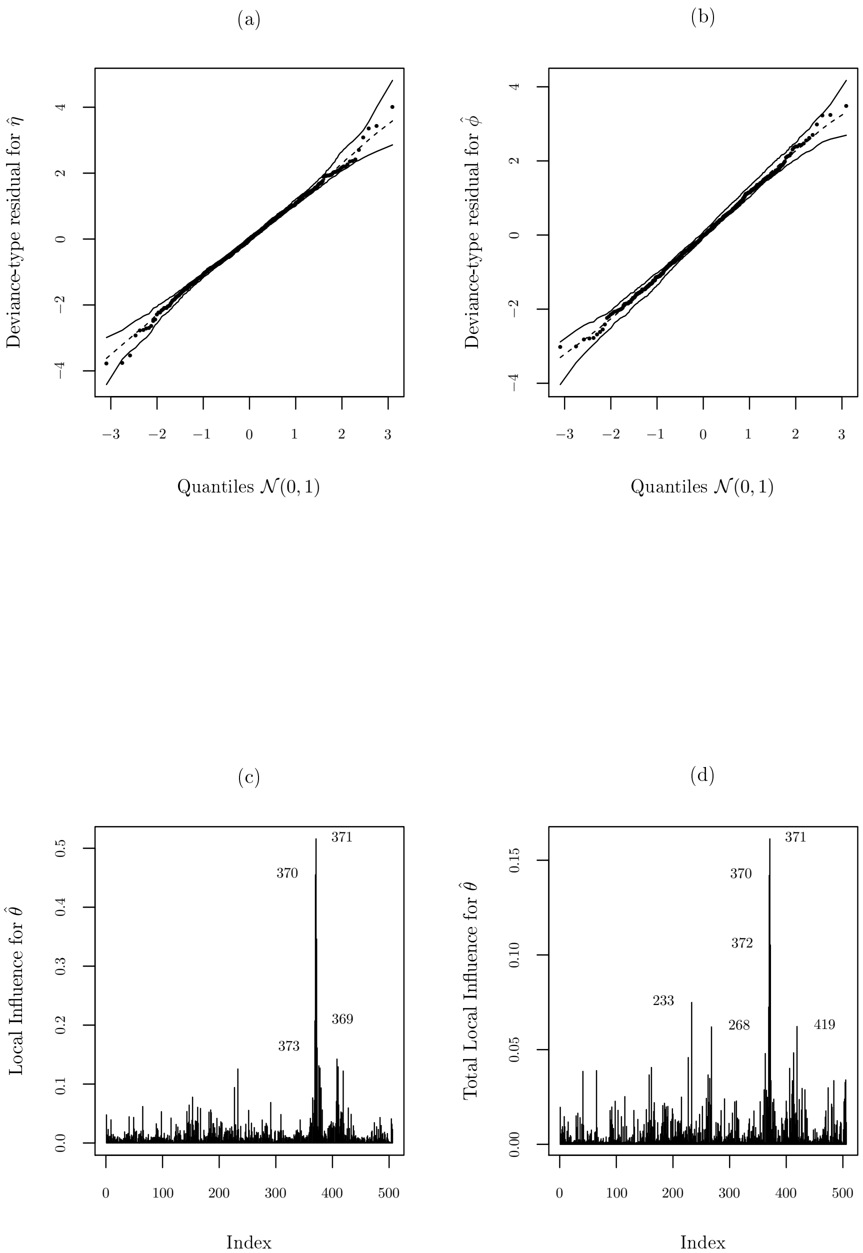

Figure 6.2: *Graphs of the deviance-type residuals with simulated envelope for $\hat{\eta}$ (a) and $\hat{\phi}$ (b), local influence for $\hat{\boldsymbol{\theta}}$ under the case-weight perturbation scheme (c), and total local influence for $\hat{\boldsymbol{\theta}}$ under the case-weight perturbation scheme (d) of the log-slash model fitted to the Boston Housing data.*

The preliminary analysis indicates that the conditional distribution of texture is right-skewed, its location is a nonlinear function of time (whose functional form is unknown) and the intensity of its skewness is dependent on the snack type. To describe the data, a model is proposed that assumes that the textures of the snack units are realizations of independent random variables with a log-symmetric distribution, in which its median (η) and its skewness (ϕ) are given by

$$\begin{cases} \log(\eta_k) = \beta_1 + \beta_2 x_{k,2} + \ldots + \beta_5 x_{k,5} + \mathrm{f}_\eta(\mathtt{week}_k), \\ \log(\phi_k) = \gamma_1 + \gamma_2 x_{k,2} + \ldots + \gamma_5 x_{k,5}, \qquad k = 1, \ldots, 750, \end{cases}$$

where $\mathrm{f}_\eta(\cdot)$ is a nonparametric function approximated by a natural cubic spline, and $x_{k,r}$ is a binary variable coded as $x_{k,r} = 1$ if the k-th snack unit belongs to the (recoded) snack type r (the levels 1-5 correspond to A-E snack types) and 0 otherwise. The basis matrix ($\dot{\mathbf{N}}_\eta$) and the penalty matrix ($\dot{\mathbf{M}}_\eta$) may be obtained using `attr(ncs(week),"N")` and `attr(ncs(week),"K")`, respectively. Table 6.2 presents the values of Υ, AIC and BIC for the fitted models under various distributions of the model error. In all cases, the extra parameter ζ was selected by using the criterion of the overall goodness-of-fit Υ. It can be seen that the error distributions with heavy tails outperform the goodness-of-fit provided by the log-normal distribution. The log-Student-t model was selected to describe the data because it yields the lowest values of AIC and BIC. The behaviour of the overall goodness-of-fit statistic Υ with respect to the extra parameter ζ under the log-Student-t model is illustrated by the Figure 6.3(a).

Table 6.2: *Goodness-of-fit statistics for the fitted models to Snacks data.*

Error distribution	Υ	AIC	BIC
log-normal	0.0377	-171.19	-85.27
log-Student-t(15)	0.0295	-176.42	-90.32
log-power-exponential(0.13)	0.0328	-175.25	-89.28
log-hyperbolic(7.3)	0.0301	-176.19	-90.13
log-slash(3)	0.0281	-176.06	-89.90
log-contaminated-normal(0.12,0.34)	0.0277	-176.40	-90.21

In **ssym**, the selected model can be fitted via

```
> data("Snacks", package="ssym")
> fit <- ssym.l(log(texture) ~ type + ncs(week) | type, data=Snacks,
+        family='Student', xi=15, local.influence=TRUE)
> summary(fit)

     Family:  Student ( 15 )
Sample size:  750
 Quantile of the Weights
   0%  25%  50%  75% 100%
 0.55 1.10 1.16 1.19  1.2
 ************************** Median/Location submodel **************************
 ******** Parametric component
             Estimate  Std.Err  z-value  Pr(>|z|)
(Intercept)  4.16019   0.0232 179.7028 < 2.2e-16 ***
type2       -0.17691   0.0283  -6.2459 4.214e-10 ***
type3       -0.08803   0.0320  -2.7493  0.005973 **
type4       -0.24908   0.0262  -9.5020 < 2.2e-16 ***
type5       -0.26823   0.0269  -9.9665 < 2.2e-16 ***
 ******** Nonparametric component
           Smooth.param Basis.dimen  d.f. Statis. p-value
ncs(week)        58.52       9.000 8.635   351.3  <2e-16 ***
 **** Deviance:  827.12
 ************************* Skewness/Dispersion submodel ***********************
 ******** Parametric component
```

```
             Estimate  Std.Err  z-value  Pr(>|z|)
(Intercept) -2.638636   0.1265 -20.8603 < 2.2e-16 ***
type2       -0.699481   0.1789  -3.9102 9.221e-05 ***
type3       -0.091042   0.1789  -0.5089    0.6108
type4       -1.265302   0.1789  -7.0733 1.513e-12 ***
type5       -1.045618   0.1789  -5.8452 5.060e-09 ***
 **** Deviance:  996.7
 ******************************************************************************
 Overall goodness-of-fit statistic:  0.029517
                -2*log-likelihood:  -213.687
                              AIC:  -176.417
                              BIC:  -90.321
```

Figure 6.3(a) can be reproduced by using the command `extra.parameter(fit,5,50)`. The AIC and BIC values (for the response in the original scale) are 5837.20 and 5946.40, respectively, which are lower than the AIC and BIC values obtained by Paula (2013), who described the data using a gamma model with varying dispersion. These values are produced by the commands `attr(AIC(fit),"log")` and `attr(BIC(fit),"log")`, respectively. The p-value associated with the nonparametric component indicates that the time significantly affects the median of the texture distribution. The graph of the fitted nonparametric function (Figure 6.3(b)), which is produced by the command `np.graph(fit,which=1,exp=TRUE)`, suggests that the median of the texture distribution achieves its highest level at approximately week 14. The summary of the fit also indicates that the snack type with the highest texture value is always the traditional snack (snack type A), which always exhibits the texture distribution with a more pronounced level of skewness. Note that the standardized individual-specific weights are within the interval $[0.55, 1.2]$ and the weight of a snack unit increases as its texture value approaches the median of its conditional distribution (Figure 6.3(c)). The plot of the overall goodness-of-fit statistic (Figure 6.3(d)) suggests that the fitted log-Student-t model satisfactorily describes the data. The plots of the deviance-type residuals versus the fitted values (omitted here) do not reveal any trend, pattern or evidence of a misspecified systematic component. The graphs of the deviance-type residuals with simulated envelope (Figure 6.4(a)-(b)) do not reveal any discrepant individual and they indicate that the log-Student-t model fits the data suitably. These graphs may be reproduced using the command `envelope(fit)`.

The instructions `ilm <- influence(fit)` extract the measures of local influence and construct their graphs versus the index of the observations (Figures 6.4(c)-(d)). For instance, under the case-weight perturbation scheme, the sets of snack units $\{601,661,675,676,691,720,749,750\}$, $\{91\}$, $\{691\}$ and $\{750\}$ were identified as potentially influential on $\hat{\boldsymbol{\theta}}$. The option `subset` of `ssym.l()` may be used to re-fit the log-Student-t model by eliminating the potentially influential snack units from the analysis and subsequently performing a comparison with the original fit. In this case, the elimination of the sets of potentially influential snack units does not introduce inferential changes. Under the response perturbation scheme, all snack units identified as potentially influential on $\hat{\boldsymbol{\theta}}$ belong to the snack type E, as illustrated by the graph (omitted here) produced by the following commands:

```
> plot(as.numeric(Snacks$type),ilm$theta$case.weight[,1],ylab="L.Influence",xaxt="n")
> axis(1,at=1:5,labels=c("A","B","C","D","E"))
```

6.3 Ultrasonic Calibration

This data set, which was previously analyzed by Lin *et al.* (2009), Lachos *et al.* (2011) and Labra *et al.* (2012) using parametric models with additive and asymmetric errors, consists of 214 observations generated in an ultrasonic calibration study, in which the interest variable is the ultrasonic response, and the explanatory variable is the distance to the metal. This data set (`Chwirut1`) is available in the `R` package **NISTnls**. The initial analysis indicates that the ultrasonic response decreases as the distance from the metal increases, and the skewness of the response distribution is dependent on the distance to the metal. Therefore, the ultrasonic

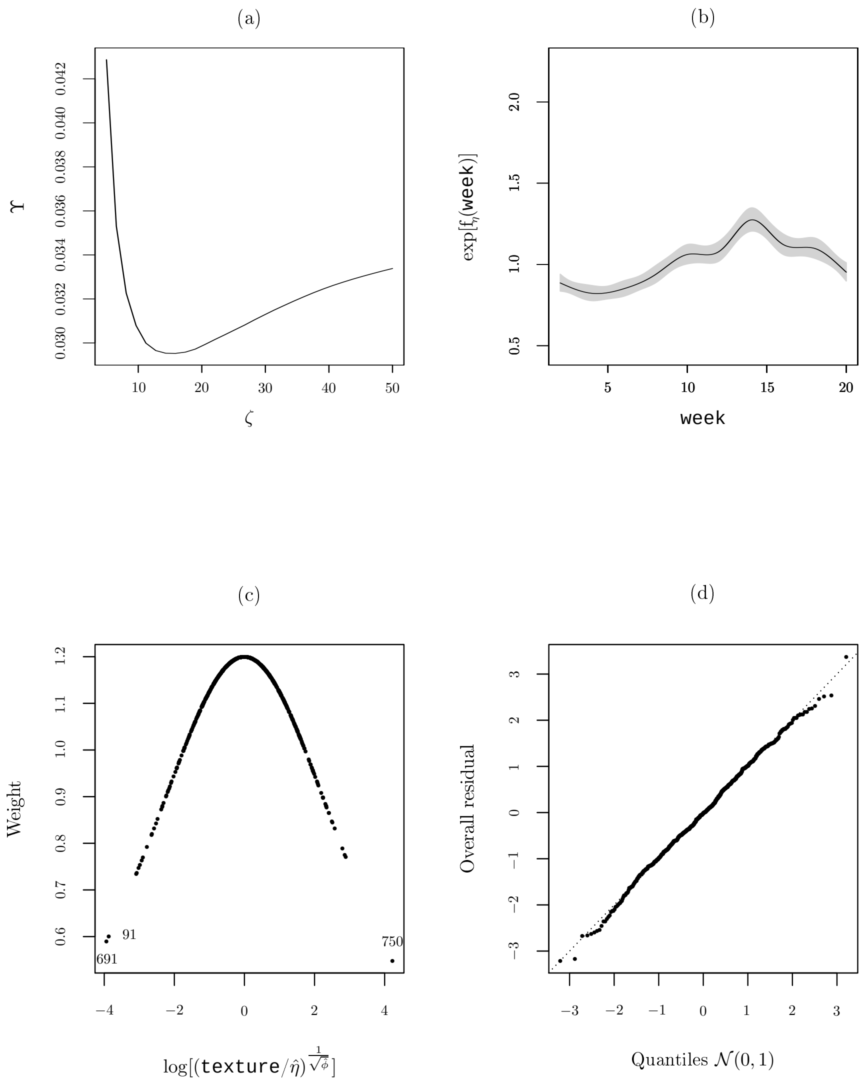

Figure 6.3: *Graphs of Υ versus ζ (a), $\exp[\mathrm{f}_\eta(\texttt{week})]$ and its simultaneous 95% confidence intervals (b), standardized individual-specific weights (c), and overall goodness-of-fit statistic (d) of the log-Student-t model fitted to Snacks data.*

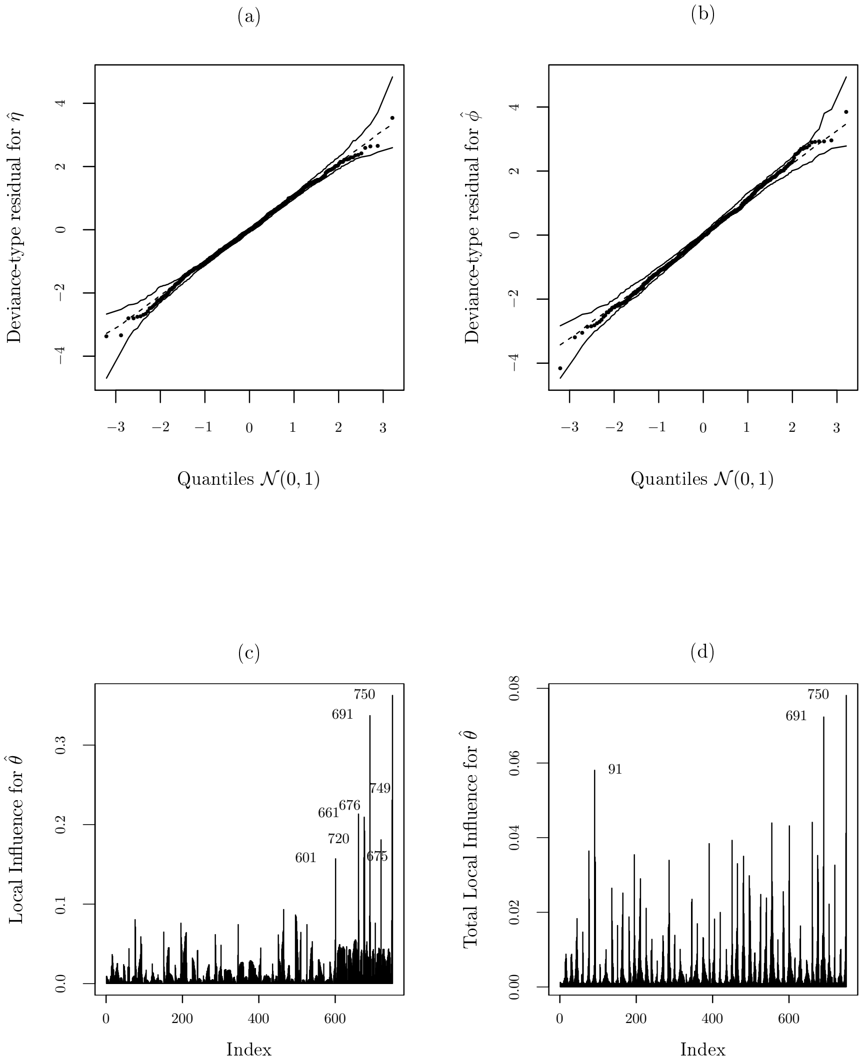

Figure 6.4: *Graphs of the deviance-type residuals with simulated envelope for $\hat{\eta}$ (a) and $\hat{\phi}$ (b), local influence for $\hat{\boldsymbol{\theta}}$ under the case-weight perturbation scheme (c), and total local influence for $\hat{\boldsymbol{\theta}}$ under the case-weight perturbation scheme (d) of the log-Student-t model fitted to Snacks data.*

response distribution may be described using a model with multiplicative and asymmetric errors, in which its median (η) and its skewness (ϕ) are described by

$$\begin{cases} \eta_k = \dfrac{\exp(-\beta_1 x_k)}{\beta_2 + \beta_3 x_k}, \\ \log(\phi_k) = \gamma_1 + \gamma_2 x_k, \qquad k = 1, \ldots, 214, \end{cases}$$

where x_k represents the distance to the metal of the individual k. Table 6.3 presents the values of Υ, AIC and BIC for the fitted models under various distributions of the model error. In all cases, the extra parameter ζ was selected by using the criterion of the overall goodness-of-fit Υ. It can be seen that the error distributions with heavy tails outperform the goodness-of-fit provided by the log-normal distribution. The log-contaminated-normal model was selected to describe the data because it yields the lowest values of AIC and BIC.

Table 6.3: *Goodness-of-fit statistics for the fitted models to Ultrasonic Calibration data.*

Error distribution	Υ	AIC	BIC
log-normal	0.1090	-270.94	-254.11
log-Student-t(3)	0.0612	-277.94	-261.11
log-power-exponential(0.7)	0.0505	-285.35	-268.52
log-hyperbolic(0.1)	0.0483	-284.46	-267.64
log-slash(1.1)	0.0688	-273.86	-257.03
log-contaminated-normal(0.68,0.1)	0.0499	-288.19	-271.36

The fitted log-contaminated-normal model is obtained using `ssym.nl()` in the following manner:

```
> data("Chwirut1", package="NISTnls")
> fit <- ssym.nl(log(y) ~ -b1*x-log(b2 + b3*x) | x, start=c(b1=0.15,b2=0.005,b3=0.012),
+        data=Chwirut1, family='Contnormal', xi=c(0.68,0.1), local.influence=TRUE)
summary(fit)
```

```
     Family:  Contnormal ( 0.68 , 0.1 )
Sample size:  214
 Quantile of the Weights
   0%  25%  50%  75% 100%
 0.47 0.61 2.13 2.83    3
 ************************** Median/Location submodel *************************
 ******** Parametric component
    Estimate Std.Err z-value  Pr(>|z|)
b1 0.1565732  0.0120 13.0543 < 2.2e-16 ***
b2 0.0055168  0.0003 17.5695 < 2.2e-16 ***
b3 0.0120526  0.0006 21.0148 < 2.2e-16 ***
 **** Deviance:  351.17
 ************************* Skewness/Dispersion submodel **********************
 ******** Parametric component
            Estimate  Std.Err  z-value  Pr(>|z|)
(Intercept) -7.55053   0.2149 -35.1400 < 2.2e-16 ***
x            0.56099   0.0710   7.9060 2.659e-15 ***
 **** Deviance:  191.77
 ****************************************************************************
 Overall goodness-of-fit statistic:  0.049915
                -2*log-likelihood:  -298.192
                              AIC:  -288.192
                              BIC:  -271.362
```

Figure 6.5(a) can be reproduced by using the command `extra.parameter(fit,c(0.4,0.08),c(0.9,` The graph of the fitted median submodel is presented in Figure 6.5(b). The AIC and BIC values (for the response in the original scale) are 1033.05 and 1049.88, respectively, which are

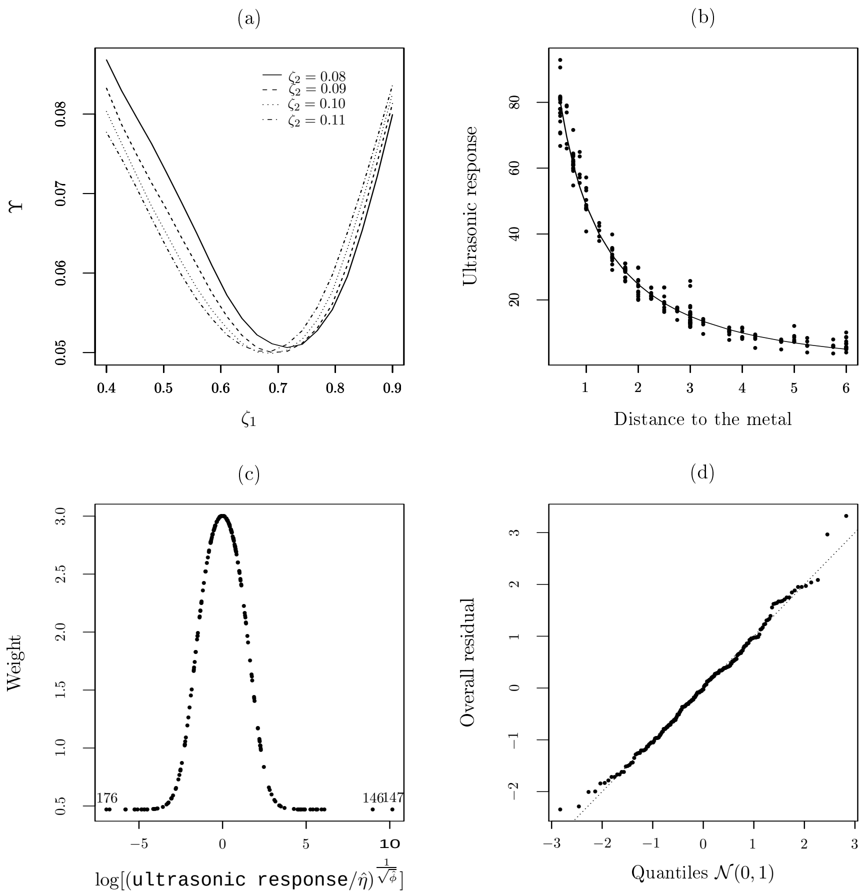

Figure 6.5: *Graphs of* Υ *versus* (ζ_1, ζ_2) *(a), fitted median submodel (b), standardized individual-specific weights (c), and overall goodness-of-fit statistic (d) of the log-contaminated-normal model fitted to Ultrasonic Calibration data.*

lower than the AIC and BIC values obtained by Lin *et al.* (2009), Lachos *et al.* (2011) and Labra *et al.* (2012). These values are obtained by the commands `attr(AIC(fit),"log")` and `attr(BIC(fit),"log")`, respectively. The approximate standard errors associated with $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$ are also lower than the ones obtained by these authors. The summary of the fitted model indicates that the skewness of the distribution of the ultrasonic calibration increases as the distance to the metal also increases. Note that the standardized individual-specific weights are within the interval $[0.47, 3]$ and the individuals with the higher ordinary residuals (i.e., individuals 146 and 147) have the lower weights (Figure 6.5(c)). The plot of the overall goodness-of-fit statistic (Figure 6.5(d)) suggests that the fitted log-contaminated-normal model suitably describes the data.

The graphs of the deviance-type residuals with simulated envelope (Figures 6.6(a)-(b)) do not reveal any discrepant observation and they indicate that the log-contaminated-normal fits the data adequately. The groups of individuals {142,146,147}, {142}, {146} and {147} are identified as potentially influential on $\hat{\boldsymbol{\theta}}$ (under the case-weight perturbation scheme, as shown in Figures6.6(c)-(d)) according to the graphs displayed by `ilm <- influence(fit)`. The following commands assess the impact on the fitted median submodel by eliminating the potentially influential individuals from the analysis:

```
> fit2 <- ssym.nl(log(y) ~ -b1*x - log(b2+b3*x) | x,
+         start=c(b1=0.15,b2=0.005,b3=0.012),
+         data=Chwirut1, family='Contnormal',
+         xi=c(0.68,0.1), subset=-c(142,146,147))
>
```

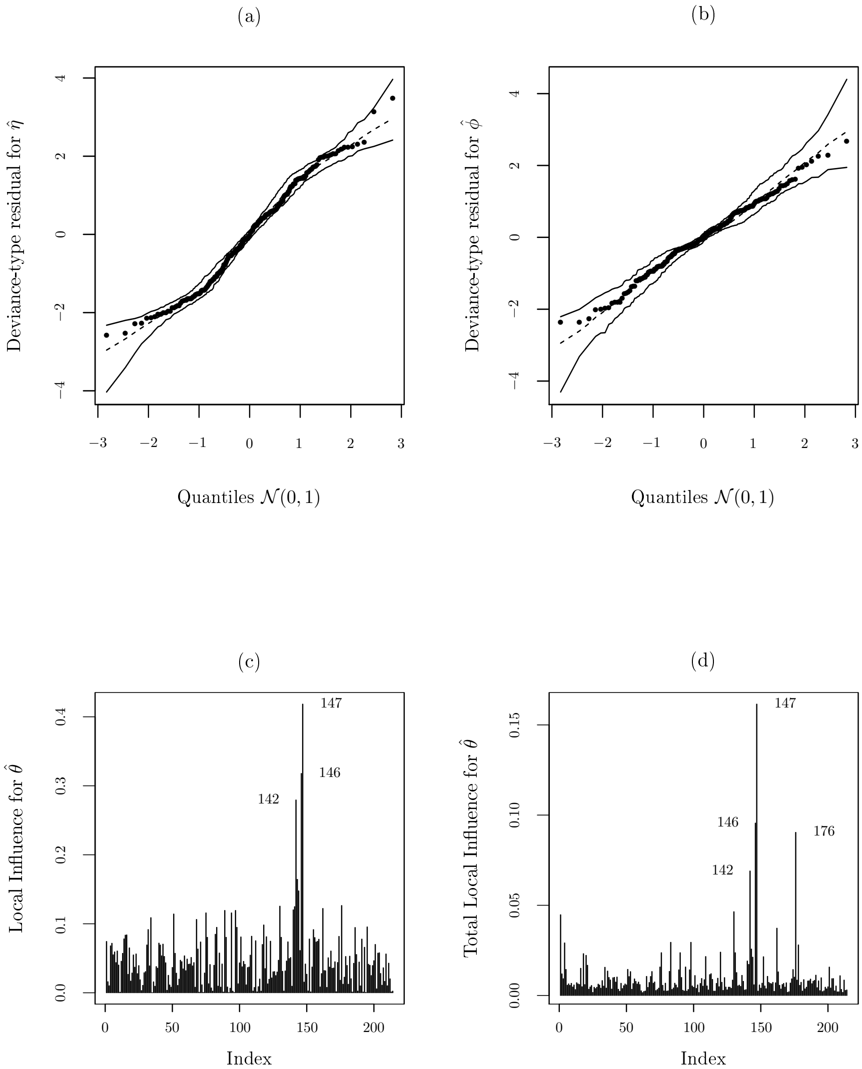

Figure 6.6: *Graphs of the deviance-type residuals with simulated envelope for $\hat{\eta}$ (a) and $\hat{\phi}$ (b), local influence for $\hat{\boldsymbol{\theta}}$ under the case-weight perturbation scheme (c), and total local influence for $\hat{\boldsymbol{\theta}}$ under the case-weight perturbation scheme (d) of the log-contaminated-normal model fitted to Ultrasonic Calibration data.*

```
> 100*(coef(fit2)$mu-coef(fit)$mu)/abs(coef(fit)$mu)
[1] -1.898305 -1.608060  1.419694
```

The percentage changes in the parameter estimates are less than 2%. Further, the elimination of these individuals does not introduce inferential changes. Regarding the local influence measures under the response perturbation scheme (omitted here), the estimate of $\boldsymbol{\theta}$ seems particularly sensitive to small perturbations on the response variable for individuals with the lowest value on the explanatory variable (i.e., $x = 0.5$), as illustrated by the graph (omitted here) produced by the following commands:

```
> plot(Chwirut1$x, ilm$theta$response[,1], ylab="L.Influence")
```

6.4 Goat Ovocytes

This data set, which was discussed by LeGal *et al.* (1984) and Huet *et al.* (1996), addresses an experiment comparing the responses of immature and mature goat ovocytes exposed to propanediol, a permeable compound. The fraction of cell volume during osmotic equilibration is recorded at each time for both type of ovocytes: immature and mature. This data set (`Ovocytes`) is available in the package **ssym**, and the objects `fraction`, `type` and `time` represent homonyms variables. The descriptive analysis indicates that the distribution of the response variable (i.e., fraction of cell volume) is asymmetric, especially for small values of the explanatory variable `time`. To describe the data, a log-symmetric model is proposed that assumes that the median and the skewness of the response variable distribution are given by

$$\begin{cases} \log(\eta_k) = \beta_1 + \beta_2\,\texttt{type}_k + \mathrm{f}_\eta(\texttt{time}_k), \\ \log(\phi_k) = \gamma_1 + \gamma_2\,\texttt{type}_k + \mathrm{f}_\phi(\texttt{time}_k), \end{cases} \qquad k = 1, \ldots, 161,$$

where the nonparametric functions $\mathrm{f}_\eta(\cdot)$ and $\mathrm{f}_\phi(\cdot)$ are approximated by using P-splines. The analysis under the log-normal distribution suggests that this data set should be described using an error distribution that exhibits lighter tails. Thus, a model with log-power-exponential errors is fitted, in which its extra parameter $\zeta = -0.55$ was selected by using the overall goodness-of-fit statistic Υ. In **ssym**, the log-power-exponential model can be fitted via

```
> data("Ovocytes", package="ssym")
> fit <- ssym.l(log(fraction) ~ type + psp(time) | type + psp(time),
+               data=Ovocytes, family='Powerexp', xi=-0.55, local.influence=TRUE)
summary(fit)

     Family:  Powerexp ( -0.55 )
Sample size:  161
 Quantile of the Weights
   0%  25%  50%  75% 100%
 0.00 0.03 0.20 0.54  1.6
 ************************** Median/Location submodel **************************
 ******** Parametric component
            Estimate  Std.Err  z-value  Pr(>|z|)
(Intercept) -0.35112   0.0116 -30.2597 < 2.2e-16 ***
typeMature   0.10167   0.0116   8.7297 < 2.2e-16 ***
 ******** Nonparametric component
          Smooth.param Basis.dimen d.f. Statis. p-value
psp(time)        77.62        9.00 6.07    1549  <2e-16 ***
 **** Deviance:  72.45

 ************************* Skewness/Dispersion submodel **********************
 ******** Parametric component
            Estimate  Std.Err  z-value  Pr(>|z|)
(Intercept)  -3.3820   0.1220 -27.7323 < 2.2e-16 ***
typeMature   -1.6898   0.1511 -11.1825 < 2.2e-16 ***
 ******** Nonparametric component
```

```
         Smooth.param Basis.dimen  d.f. Statis. p-value
psp(time)       0.4149       9.000 5.828   125.1  <2e-16 ***
 **** Deviance:  186.32
 ******************************************************************************
 Overall goodness-of-fit statistic:  0.100859
                 -2*log-likelihood:  -360.516
                               AIC:  -328.72
                               BIC:  -279.732
```

The p-values and the graphs (Figures 6.7(a)-(b)) associated with the nonparametric components indicate that the median and the skewness of the distribution of the fraction of cell volume are dependent on the time. Note that $\hat{\mathrm{f}}_{\eta}(\cdot)$ and $\hat{\mathrm{f}}_{\phi}(\cdot)$ are not monotone functions of time. These graphs, which include the simultaneous 95% confidence intervals, can be reproduced by the commands `np.graph(fit,which=1,exp=TRUE)` and `np.graph(fit,which=2,exp=TRUE)`, respectively. The summary of the fitted model also reveals that the distribution of the fraction of cell volume is dependent on the type of ovocyte. The AIC and BIC values (for the response in the original scale) are -429.265 and -394.239, respectively. These values are obtained by the commands `attr(AIC(fit),"log")` and `attr(BIC(fit),"log")`, respectively.

Figure 6.7(c) reveals that the highest standardized individual-specific weights correspond to the ovocytes with the highest differences between the observed and the fitted values. The graphs of the deviance-type residuals with simulated envelope (Figure 6.8(a)-(b)) indicates that the log-power-exponential model satisfactorily describes the data set. These graphs can be reproduced using the command `envelope(fit)`. Figures 6.8(c)-(d) present the local influence measures for $\hat{\boldsymbol{\theta}}$ under the case-weight perturbation scheme. The groups of individuals $\{113, 120\ \{6\}$, $\{120\}$, $\{147\}$ and $\{155\}$ are identified as potentially influential.

6.5 Personal Injury Insurance

A sample of the data set reported by de Jong and Heller (2008), which contains information about settled personal injury insurance claims from an Australian insurance company, was considered. The 540 claims in the sample had legal representation and were obtained for accidents that occurred from January 1998 to June 1999. This data set (`Claims`) is available in the package **ssym** and contains the variables `total`, `accmonth` and `op_time`, which correspond to the amount of money paid by an insurance policy (in thousands of Australian dollars), the month of occurrence of the accident (coded 103 (January 1998) to 120 (June 1999)) and the operational time (expressed as a percentage), respectively. Similar to Paula *et al.* (2012), the data are described using a model that assumes the Birnbaum-Saunders-t distribution for the response variable `total`; however, unlike Paula *et al.* (2012), varying skewness is assumed. Then, the median (η) and skewness (ϕ) of the response variable distribution are described by

$$\log(\eta_k) = \beta_1 + \beta_2\,\texttt{op\_time}_k, \quad \log(\phi_k) = \gamma_1 + \gamma_2\,\texttt{op\_time}_k, \quad k = 1, \ldots, 540.$$

By using the goodness-of-fit measure Υ, the extra parameter vector of the Birnbaum-Saunders-t distribution was selected to be $\zeta = (0.1, 4)^{\top}$. The summary of the proposed model is obtained using the following instructions:

```
> data("Claims", package="ssym")
> fit <- ssym.l(log(total) ~ op_time | op_time, data=Claims,
+        family='Sinh-t', xi=c(0.1,4), local.influence=TRUE)
summary(fit)

     Family:  Sinh-t ( 0.1 , 4 )
Sample size:  540
 Quantile of the Weights
   0%   25%  50%  75% 100%
 0.07 1.20 1.54 1.70 1.75
 *********** Median/Location submodel ***********
 **** Parametric component
```

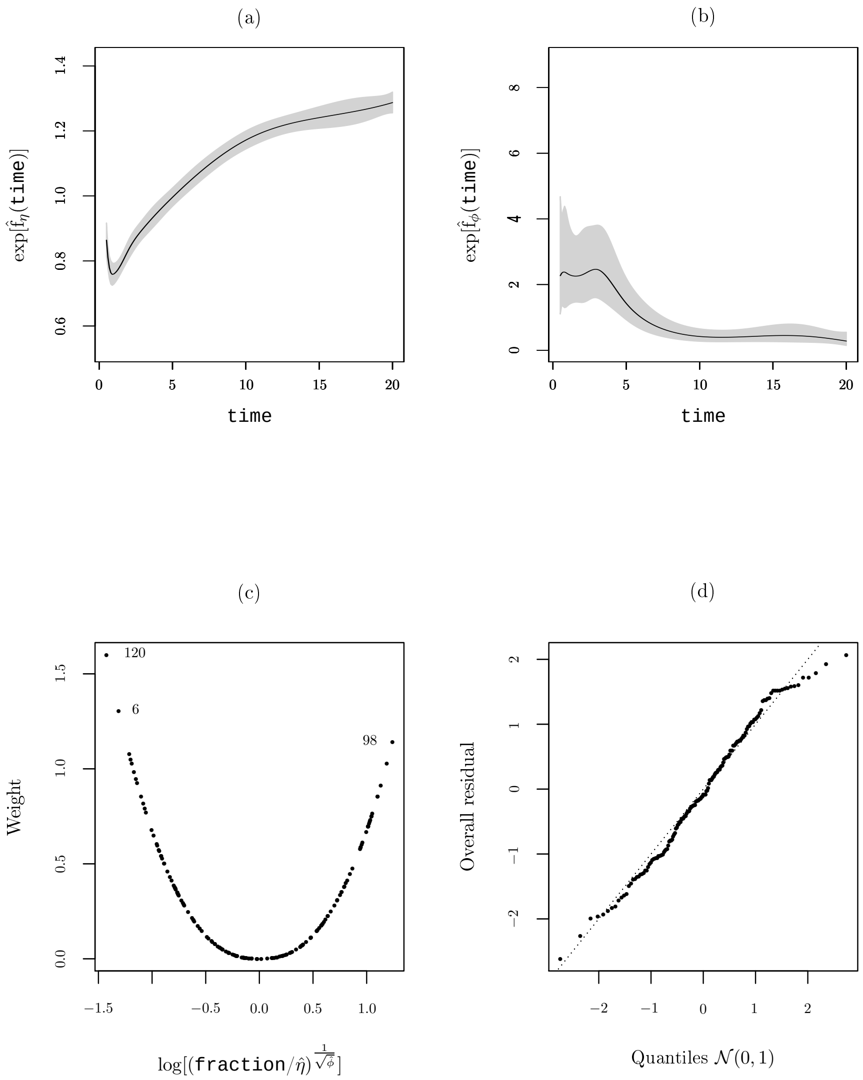

Figure 6.7: *Graphs of* $\exp[\hat{f}_\eta(\texttt{time})]$ *(a)* $\exp[\hat{f}_\phi(\texttt{time})]$ *(b) and their simultaneous 95% confidence intervals, standardized individual-specific weights (c), and overall goodness-of-fit statistic (d) of the log-power-exponential model fitted to Goat Ovocytes data.*

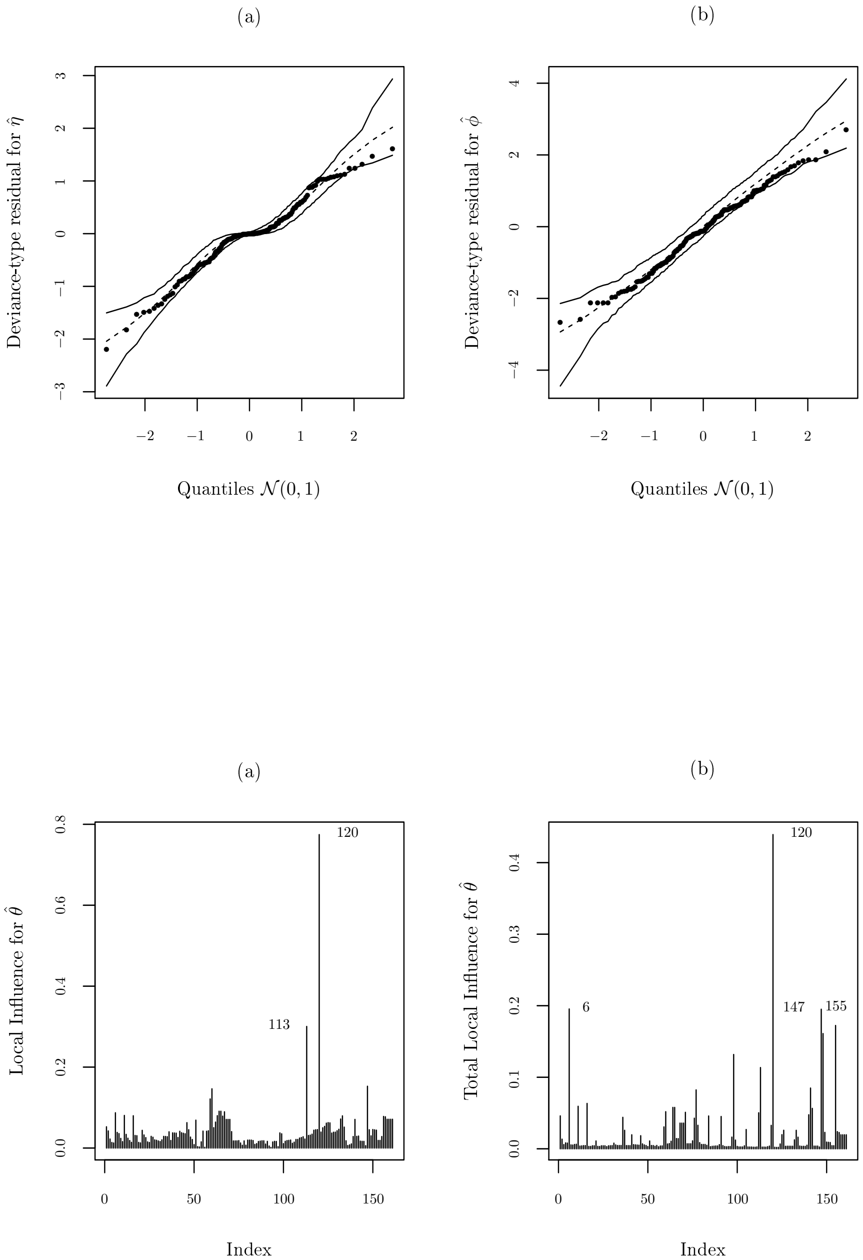

Figure 6.8: *Graphs of deviance-type residuals with simulated envelope for $\hat{\eta}$ (a) and $\hat{\phi}$ (b), local influence for $\hat{\boldsymbol{\theta}}$ under the case-weight perturbation scheme (c), and total local influence for $\hat{\boldsymbol{\theta}}$ under the case-weight perturbation scheme (d) of the log-power-exponential model fitted to Goat Ovocytes data.*

```
   Estimate Std.Err z-value Pr(>|z|)
            Estimate Std.Err z-value Pr(>|z|)
(Intercept)  1.74986 0.03981  43.955  < 2e-16 ***
op_time      0.01614 0.00336   4.804 1.56e-06 ***
 **** Deviance:  756.91
 ********* Skewness/Dispersion submodel *********
**** Parametric component
            Estimate Std.Err z-value Pr(>|z|)
(Intercept)  4.14050 0.13752  30.108  < 2e-16 ***
op_time      0.04064 0.00946   4.296 1.74e-05 ***
 **** Deviance:  763.19
 *************************************************
 Overall goodness-of-fit statistic:  0.03417
                 -2*log-likelihood:  1075.805
                               AIC:  1083.805
                               BIC:  1100.971
```

The median and the skewness of the distribution of `total` increase as the `op_time` also increases. For each increase of ten percentage points in the operational time, the median and skewness of `total` increase at approximately $100\times[\exp(\hat{\beta}_2\times 10)-1] = 17.5\%$ and $100\times[\exp(\hat{\gamma}_2\times 10)-1] = 50\%$, respectively. The AIC and BIC values are 1083.81 and 1100.97, respectively, which are lower than the AIC and BIC values obtained by Paula *et al.* (2012), who described the data using the homogeneous Birnbaum-Saunders-t model (i.e., a model in which $\phi_k = \phi = 4$ for all k). These values are obtained by the commands `AIC(fit)` and `BIC(fit)`, respectively. The plot of Υ (Figure 6.9(b)) suggests that the proposed Birnbaum-Saunders-t model satisfactorily fits the data.

Figures 6.9(c)-(d) show the residual plots versus the fitted values for both submodels. The plot of the deviance-type residuals for the median submodel reveals the claims 10 and 28 as marginally discrepant. These claims also present the lowest weights according to Figure 6.9(a). The plot of the deviance-type residuals for the skewness submodel reveals the claims 28, 75, 416, 476 and 533 as marginally discrepant. These plots do not present any pattern or tendency. Graphs of deviance-type residuals with simulated envelope for median and skewness submodels (Figure 6.10(a)-(b)) confirm that the Birnbaum-Saunders-t model satisfactorily describes the data. These graphs may be obtained via `envelope(fit,50)`. In the local influence measures under the case-weight perturbation scheme (Figures 6.10(c)-(d)), the sets of claims {2,3,10,28}, {2}, {3}, {10}, {28} and {537} were identified as potentially influential on $\hat{\boldsymbol{\theta}}$. The following commands can be used to assess the impact on the fitted median submodel by eliminating the potentially influential individuals from the analysis:

```
> fit2 <- ssym.l(log(total) ~ op_time | op_time, data=Claims,
+                  family='Sinh-t', xi=c(0.1,4), subset=-c(2,3,10,28))
>
> 100*(coef(fit2)$mu-coef(fit)$mu)/abs(coef(fit)$mu)
[1]    0.08012387 -0.48482172
```

Note that the percentage changes in the parameter estimates are less than 1%. In addition, the elimation of these individuals does not introduce inferential changes.

6.6 Biaxial Fatigue

This data set, available in the object `Biaxial` of the package **ssym**, describes the life of a metal piece subjected to cyclic stretching and compressing, where `Life`, the number of cycles to failure of the metal specimen, is the response variable, T, and `Work`, the work per cycle, is the explanatory variable, x. This data set was analyzed by Rieck and Nedelman (1991) and Lemonte and Patriota (2011) using the linear and nonlinear Birnbaum-Saunders models, respectively, in which the skewness parameter was fixed to be $\phi = 4$. Similar to Lemonte and Patriota (2011), the data are described here using a nonlinear model with Birnbaum-Saunders errors and where ϕ is assumed to be constant; however, unlike Rieck and Nedelman (1991) and Lemonte and Patriota

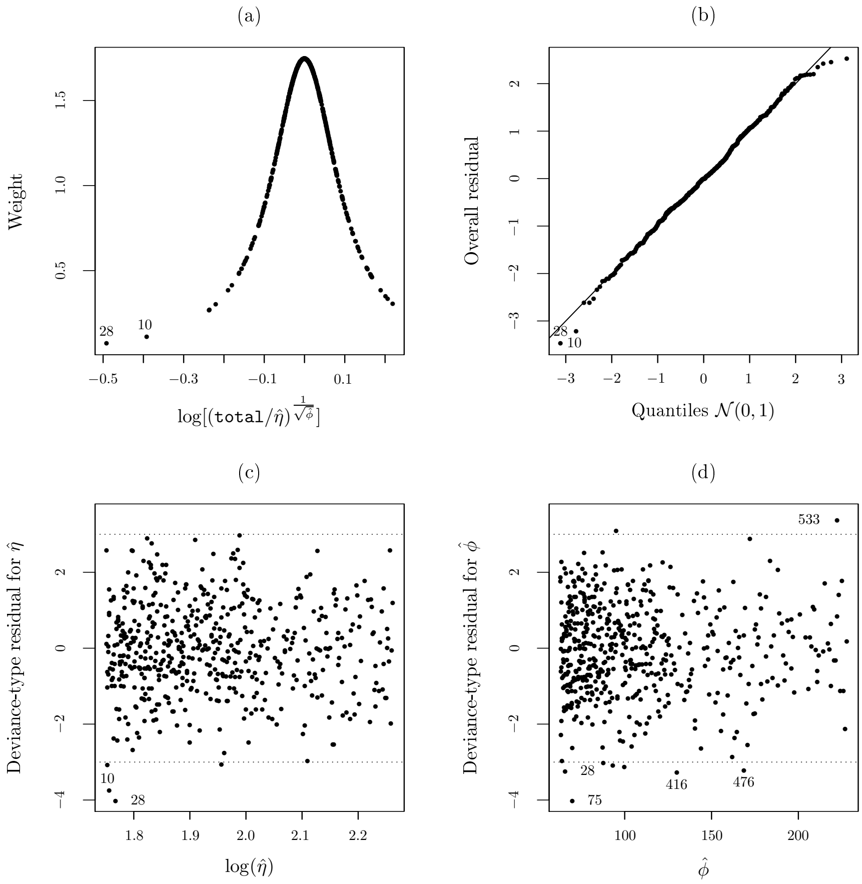

Figure 6.9: *Graphs of standardized individual-specific weights (a), overall goodness-of-fit statistic (b), deviance-type residuals for $\hat{\eta}$ (c) and $\hat{\phi}$ (d) of the Birnbaum-Saunders-t model fitted to Claims data.*

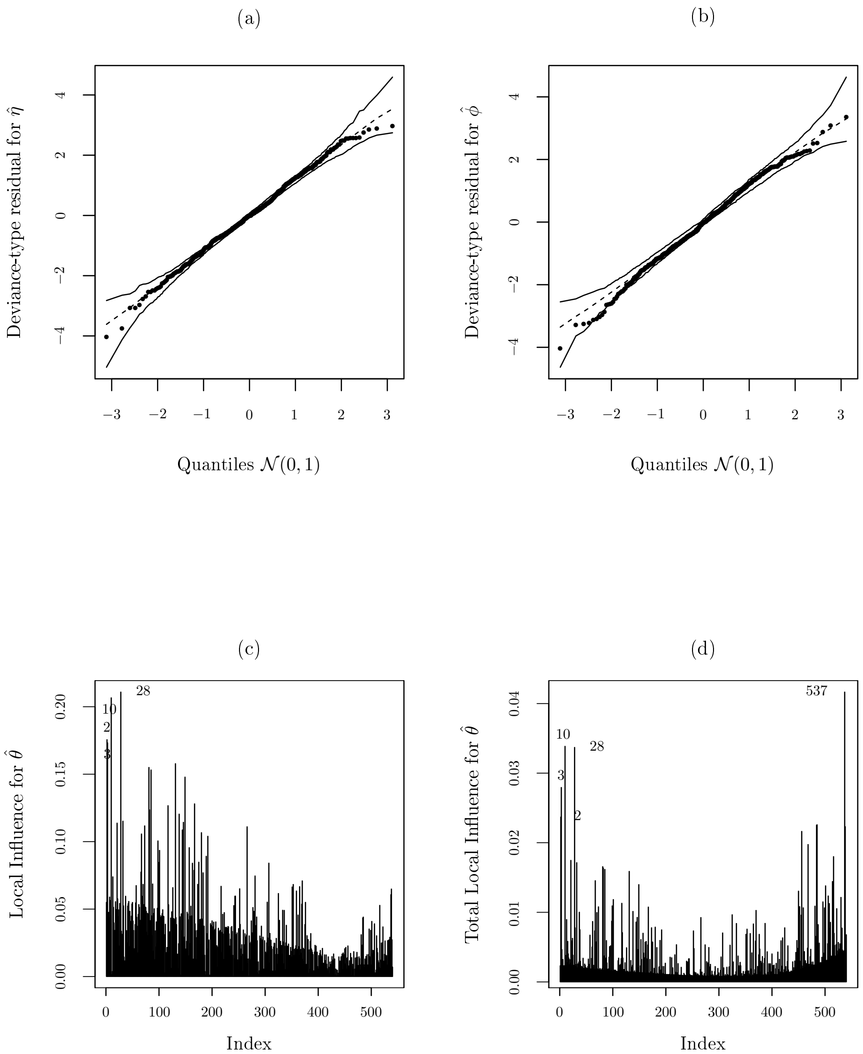

Figure 6.10: *Graphs of the deviance-type residuals with simulated envelope for $\hat{\eta}$ (a) and $\hat{\phi}$ (b), local influence for $\hat{\boldsymbol{\theta}}$ under the case-weight perturbation scheme (c), and total local influence for $\hat{\boldsymbol{\theta}}$ under the case-weight perturbation scheme (d) of the Birnbaum-Saunders-t model fitted to Claims data.*

(2011), the value of the skewness parameter ϕ is estimated from the data. The median of the T_k distribution is described using the following function:

$$\eta_k = \exp(\beta_1 x_k^{\beta_2}), \qquad\qquad k = 1, \ldots, 46.$$

The extra parameter $\zeta = 1.54$ of the Birnbaum-Saunders distribution was selected by minimizing the overall goodness-of-fit statistic Υ, as illustrated by the Figure 6.11(a). In **ssym**, the proposed model is fitted using the following instructions:

```
> data("Biaxial", package="ssym")
> fit <- ssym.nl(log(Life) ~ b1*Work^b2, start=c(b1=16, b2=-0.25),
+                 data=Biaxial, family='Sinh-normal', xi=1.54)
summary(fit)

     Family:  Sinh-normal ( 1.54 )
 Sample size:  46
 Quantile of the Weights
   0%  25%  50%  75% 100%
 0.23 0.29 0.41 0.53 1.28
 *********** Median/Location submodel ***********
 **** Parametric component
     Estimate Std.Err z-value Pr(>|z|)
 b1 15.82840  0.80425   19.68   <2e-16 ***
 b2 -0.26162  0.01475  -17.74   <2e-16 ***
 **** Deviance:  29.48
 ********* Skewness/Dispersion submodel *********
  **** Parametric component
              Estimate Std.Err z-value Pr(>|z|)
 (Intercept)  -1.0511  0.1533  -6.858    7e-12 ***
 **** Deviance:  48.86
 ***********************************************
 Overall goodness-of-fit statistic:  0.150182
                 -2*log-likelihood:  41.624
                               AIC:  47.624
                               BIC:  53.11
```

Figure 6.11(a) is reproduced by using the command `extra.parameter(fit,1.3,1.8)`. It can be seen that the approximate standard errors associated with $\hat{\beta}_1$ and $\hat{\beta}_2$ are lower than those obtained by Lemonte and Patriota (2011). In addition, the AIC and BIC values are 47.62 (`AIC(fit)`) and 53.11 (`BIC(fit)`), respectively, which are lower than those obtained by Rieck and Nedelman (1991) and Lemonte and Patriota (2011). Figure 6.11(b) presents the graph of the fitted median submodel. The instruction `envelope(fit,50)` produces the graphs of the deviance-type residuals with simulated envelope (6.11(c)-(d)), which reveal that the Birnbaum-Saunders model fits the data suitably.

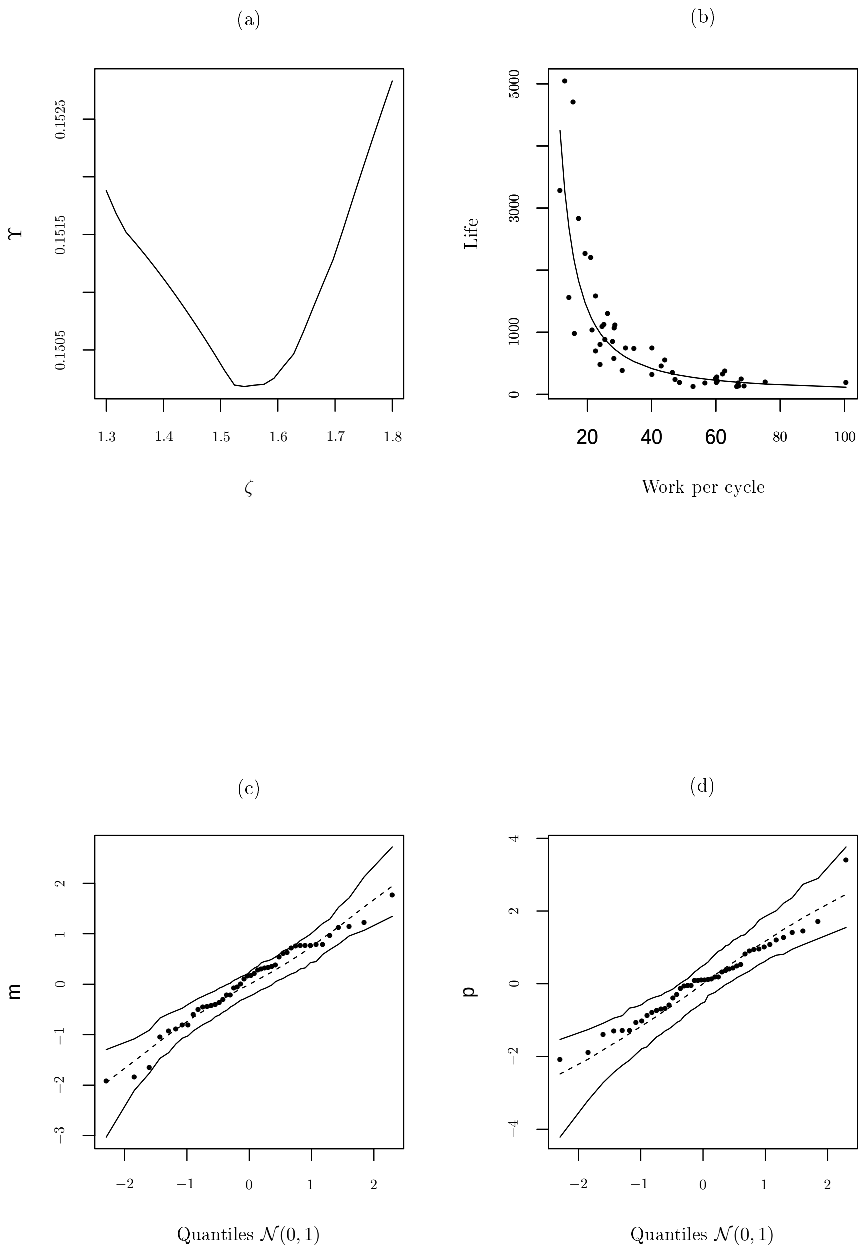

Figure 6.11: *Graphs of the fitted median submodel (a), fitted skewness submodel (b), deviance-type residuals with simulated envelope for $\hat{\eta}$ (c) and $\hat{\phi}$ (d) of the Birnbaum-Saunders model fitted to Biaxial Fatigue data.*

6.7 Primary Biliary Cirrhosis

These data are available in the object `pbc` of the `R` package **survival**. The data are from the Mayo Clinic trial in primary biliary cirrhosis (PBC) of the liver conducted between 1974 and 1984, and they consist of 412 observations of the dependent variable `time`, number of days between registration and the earlier of death, transplantation, or study analysis in July, 1986; `status`, 0 = alive at last contact, 1 = liver transplant, 2 = death; and other independent variables, which include `edema`: presence of edema; `stage`: histologic stage of disease; and `bili`: serum bilirunbin (mg/dl). It is proposed to analyze the dataset using a censored log-symmetric model where the median (η) and the skewness of the response distribution are given by

$$\begin{cases} \log(\eta_k) = \beta_1 + \beta_2\texttt{edema}_k + \beta_3\texttt{stage}_k + \mathrm{f}_\eta(\texttt{bili}_k), \\ \log(\phi_k) = \gamma_1, \qquad\qquad k = 1, \ldots, 412, \end{cases}$$

where $\mathrm{f}_\eta(\cdot)$ is a nonparametric function approximated by using a natural cubic spline. To describe the dataset, a Birnbaum-Saunders-t model is proposed, where the extra parameter vector was selected to be $\zeta = (0.65, 3)^\top$. The summary of the proposed model is obtained using the following instructions:

```
> data("pbc", package="survival")
> pbc2 <- data.frame(pbc[!is.na(pbc$edema) & !is.na(pbc$stage) & !is.na(pbc$bili),])
>
> fit <- ssym.l2(Surv(log(time),ifelse(status>=1,0,1) ) ~ factor(edema) +
+                stage + ncs(bili), data = pbc2, family="Sinh-t",
+                xi=c(0.65,3), local.influence=TRUE)
>summary(fit)

     Family:  Sinh-t ( 0.65 , 3 )
Sample size:  412
Censored %:  55.83
 ************************** Median/Location submodel **************************
 ******** Parametric component
                Estimate Std.Err z-value  Pr(>|z|)
(Intercept)       9.05527  0.2582 35.0674 < 2.2e-16 ***
factor(edema)0.5 -0.53409  0.1546 -3.4544 0.0005515 ***
factor(edema)1   -1.43010  0.2461 -5.8109 6.213e-09 ***
stage            -0.33485  0.0743 -4.5057 6.616e-06 ***
 ******** Nonparametric component
         Smooth.param Basis.dimen  d.f. Statis. p-value
ncs(bili)       2.527       9.000 4.671   113.4  <2e-16 ***
 **** Deviance:  436.03
 ************************* Skewness/Dispersion submodel **********************
 ******** Parametric component
           Estimate Std.Err z-value  Pr(>|z|)
(Intercept)   1.4799  0.1293 11.4477 < 2.2e-16 ***
 **** Deviance:  373.51
 ******************************************************************************
 Overall goodness-of-fit statistic:  0.036313
                -2*log-likelihood:  660.58
                              AIC:  679.922
                              BIC:  718.809
```

BIBLIOGRAPHY

Akaike H (1973). Information theory and an extension of the maximum likelihood principle. In BN Petrov, F Csaki (eds.), *2nd International Symposium on Information Theory*, pp. 267–281. Cited on page 17

Andrews DF, Mallows CL (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society Series B (Methodological)*, **36**, 99–102. Cited on page 11, 29, 51

Azzalini A, Cappello TD, Kotz S (2003). Log-skew-normal and log-skew-t distributions as model for family income data. *Journal of Income Distribution*, **11**, 12–20. Cited on page 3

Bagdonavičius V, Nikulin M (2001). *Accelerated Life Models, Modeling and Statistical Analysis.* Chapman & Hall, Boca Raton. Cited on page 47, 53

Balakrishnan N, Leiva V, Sanhueza A, Vilca F (2009). Estimation in the birnbaum-saunders distributions based on scale-mixture of normals and the em-algorithm. *Statistics and Operations Research Transactions (SORT)*, **33**, 171–192. Cited on page 7, 11, 30

Barndoff-Nielsen O (1977). Exponentially decreasing distributions for the logarithm of particle size. *Proceedings of the Royal Society of London Series A, Mathematical & Physical Sciences*, **353**, 401–419. Cited on page 6, 11, 29, 51

Barros M, Paula GA, Leiva V (2008). A new class of survival regression models with heavy-tailed errors: robustness and diagnostics. *Lifetime Data Analysis*, **14**, 316–332. Cited on page 3, 4, 7, 47, 50

Birnbaum ZW, Saunders SC (1969). A new family of life distributions. *Journal of Applied Probability*, **6**, 319–327. Cited on page 3, 4, 7

Bonett DG (2006). Confidence interval for a coefficient of quartile variation. *Computational Statistics and Data Analysis*, **50**, 2953–2957. Cited on page 9

Boor CD (1978). *A Practical Guide to Splines.* Applied Mathematical Sciences. Springer-Verlag, New York. Cited on page 26

Borgan O (1984). Maximum likelihood estimation in parametric counting process models, with applications tocensored failure time data. *Scandinavian Journal of Statistics*, **11**, 1–16. Cited on page 53

Box GEP, Tiao GC (1973). *Bayesian Inference in Statistical Analysis.* Addison-Wesley. Cited on page 5

Brostom G (2014). **eha**: *Event History Analysis.* R package version 2.4-2. URL http://CRAN.R-project.org/package=eha. Cited on page 59

Cancho VG, Lachos VH, Ortega EMM (2010). A nonlinear regression model with skew-normal errors. *Statistical Papers*, **51**, 547–558. Cited on page 23

Carrasco JMF, Ortega EMM, Cordeiro GM (2008). A generalized modified weibull distribution for lifetime modeling. *Computational Statistics & Data Analysis*, **53**, 450–462. Cited on page 3

Chib S, Greenberg E (1995). Understanding the metropolis-hasting algorithm. *The American Statistician*, **49**, 327–335. Cited on page 15

Conover WJ (1971). *Practical Nonparametric Statistics*. Wiley, New York. Cited on page 37, 55

Cook RD (1986). Assessment local influence (with discussion). *Journal of the Royal Statistical Society Series B (Methodological)*, **48**, 133–169. Cited on page 41, 58

Cordeiro GM, Andrade MG (2011). Transformed symmetric models. *Statistical Modelling*, **11**, 371–388. Cited on page 27

Cordeiro GM, Ferrari SLP, Uribe-Opazo MA, Vasconcellos KLP (2000). Corrected maximum-likelihood estimation in a class of symmetric nonlinear regression models. *Statistics and Probability Letters*, **46**, 317–328. Cited on page 10, 12

Cox DR, Hinkley DV (1974). *Theoretical Statistics*. John Wiley & Sons, New York. Cited on page 33

Cysneiros FJA, Cordeiro GM, Cysneiros AHMA (2010). Corrected maximum likelihood estimators in heteroscedastic symmetric nonlinear models. *Journal of Statistical Computation and Simulation*, **80**, 451–461. Cited on page 32

Cysneiros FJA, Vanegas LH (2008). Residuals and their statistical properties in symmetrical nonlinear models. *Statistics and Probability Letters*, **78**, 3269–3273. Cited on page 41

Davison AC, Gigli A (1989). Deviance residuals and normal score plots. *Biometrika*, **76**, 211–221. Cited on page 37

Díaz-García JA, Leiva V (2005). A new family of life distributions based on elliptically contoured distributions. *Journal of Statistical Planning and Inference*, **128**, 445–457. Cited on page 3, 4

de Jong P, Heller GZ (2008). *Generalized Linear Models for Insurance Data*. Cambridge University Press: Cambridge, England. Cited on page 63, 77

Dempster AP, Laird NM, Rubin DB (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society Series B (Methodological)*, **39**, 1–38. Cited on page 11, 29

Dudzinski ML, Mykytowycz R (1961). The eye lens as an indicator of age in the wild rabbit in australia. *CSIRO Wildlife Research*, **6**, 156–159. Cited on page 63

Dunn PK, Smith GK (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, **5**, 236–244. Cited on page 41

Eilers PHC, Marx BD (1996). Flexible smoothing with b-splines and penalties. *Statistical Science*, **11**, 89–121. Cited on page 23, 25, 35, 48, 52, 53

Fahrmeir L, Kneid T, Lang S, Marx B (2013). *Regression Models, Methods and Applications*. Springer-Verlag, Berlin. Cited on page 27

Fang KT, Kotz S, Ng KW (1990). *Symmetric Multivariate and Related Distributions*. Chapman and Hall, London. Cited on page 4

Fonseca TCO, Migon HS, Ferreira MAR (2012). Bayesian analysis based on the jeffreys prior for the hyperbolic distribution. *Brazilian Journal of Probability and Statistics*, **26**, 327–343. Cited on page 5

Garay AM, Prates M, Lachos V (2013). **nlsmsn**: *Fitting Nonlinear Models with Scale Mixture of Skew-Normal Distributions*. R package version 0.0-4. URL http://CRAN.R-project.org/package=nlsmsn. Cited on page 59

Gelfand AE, Smith AFM (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**, 398–409. Cited on page 14

Gelman A, Carlin JB, Stern HS, Rubin DB (2004). *Bayesian Data Analysis*. Chapman & Hall, Boca Raton. Cited on page 19

Green PJ, Silverman BW (1994). *Nonparametric Regression and Generalized Linear Models*. Chapman and Hall, Boca Raton. Cited on page 23, 25, 48

Groeneveld RA, Meeden G (1984). Measuring skewness and kurtosis. *The Statistician*, **33**, 391–399. Cited on page 9

Harrell FE (2015). **rms**: *Regression Modeling Strategies*. R package version 4.3-1. URL http://CRAN.R-project.org/package=rms. Cited on page 59

Hastie TJ, Tibshirani RJ (1990). *Generalized Additive Models*. Chapman and Hall, London. Cited on page 24, 28, 36, 54

Hinkley DV (1975). On power transformations to symmetry. *Biometrika*, **62**, 101–111. Cited on page 9

Hörmann W, Leydold J (2013). Generating generalized inverse gaussian random variates. *Statistics and Computing*, **24**, 547–557. Cited on page 14

Huet S, Bouvier A, Gruet MA, Jolivet E (1996). *Statistical Tools for Nonlinear Regression*. Springer, New York. Cited on page 76

Ibacache-Pulgar G, Paula GA, , Cysneiros FJA (2013). Semiparametric additive models under symmetric distributions. *TEST*, **22**, 103–121. Cited on page 24, 65

Jørgensen B (1982). *Statistical Properties of the Generalized Inverse Gaussian Distribution*. Springer, New York. Cited on page 3

Kalbfleisch JD, Prentice RL (2002). *The Statistical Analysis of Failure Time Data*. John Wiley & Sons, New York. Cited on page 47

Kano Y, Berkane M, Bentler P (1993). Statistical inference based on pseudo-maximum likelihood estimators in elliptical populations. *Journal of the American Statistical Association*, **88**, 135–143. Cited on page 12

Kaplan EL, Meier P (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, **53**, 457–481. Cited on page 57

Labra FV, Garay AM, Lachos VH, Ortega EMM (2012). Estimation and diagnostics for heteroscedastic nonlinear regression models based on scale mixtures of skew-normal distributions. *Journal of Statistical Planning and Inference*, **142**, 2149–2165. Cited on page 23, 70, 74

Lachos VH, Bandyopadhyay D, Garay AM (2011). Heteroscedastic nonlinear regression models based on scale mixtures of skew-normal distributions. *Statistics and Probability Letters*, **81**, 1208–1217. Cited on page 23, 59, 70, 74

Lancaster P, Salkauskas K (1986). *Curve and Surface Fitting: an introduction*. Academic Press, London. Cited on page 23, 25, 26, 48

LeGal F, Gasqui P, Renard JP (1984). Differential osmotic behavior of mammalian ovocytes before and after maturation: A quantitative analysis using goat oocytes as a model. *Cryobiology*, **31**, 154–170. Cited on page 63, 76

Leiva V, Riquelme M, Balakrishnan N, Sanhueza A (2008). Lifetime analysis based on the generalized birnbaum-saunders distribution. *Computational Statistics and Data Analysis*, **52**, 2079–2097. Cited on page 4, 7, 59

Lemonte AJ, Cordeiro GM (2009). Birnbaum-saunders nonlinear regression models. *Computational Statistics and Data Analysis*, **53**, 4441–4452. Cited on page 27

Lemonte AJ, Patriota AG (2011). Influence diagnostics in birnbaum-saunders nonlinear regression models. *Journal of Applied Statistics*, **38**, 871–884. Cited on page 80, 83

Li AP, Chen ZX, Xie FC (2012). Diagnostics analysis for heterogeneous log-birnbaum-saunders regression models. *Statistics and Probability Letters*, **82**, 1690–1698. Cited on page 50

Limpert E, Stahel WA, Abbt M (2001). Log-normal distributions across the sciences: key and clues. *BioScience*, **51**, 341–352. Cited on page 3

Lin JG, Xie FC, Wei BC (2009). Statistical diagnostics for skew-t-normal nonlinear models. *Communications in Statistics, Simulation and Computation*, **38**, 2096–2110. Cited on page 23, 70, 74

Lucas A (1997). Robustness of the student-t based m-estimator. *Communications in Statistics, Theory and Methods*, **26**, 1165–1182. Cited on page 12

Marchenko YV, Genton MG (2010). Multivariate log-skew-elliptical distributions with applications to precipitation data. *Environmetrics*, **21**, 318–340. Cited on page 3

Marshall AW, Olkin I (2007). *Life Distributions*. Springer, New York. Cited on page 4

Meng XL, Rubin DB (1993). Maximum likelihood estimation via the ecm algorithm: A general framework. *Biometrika*, **80**, 267–278. Cited on page 51

Moors JJA (1988). A quantile alternative for kurtosis. *Journal of the Royal Statistical Society D*, **37**, 25–32. Cited on page 9

Nadarajah S, Kotz S (2006). The exponentiated type distributions. *Acta Applicandae Mathematicae*, **92**, 97–111. Cited on page 14

Ortega JM (1970). *Iterative Solution of Nonlinear Equations in Several Variables*. Academic Press, New York. Cited on page 50

Paula GA (2013). On diagnostics in double generalized linear models. *Computational Statistics and Data Analysis*, **68**, 44–51. Cited on page 63, 66, 70

Paula GA, Leiva V, Barros M, Liu S (2012). Robust statistical modeling using birnbaum-saunders-t distribution applied to insurance. *Applied Stochastic Models in Business and Industry*, **28**, 16–34. Cited on page 3, 4, 7, 27, 47, 77, 80

Pierce DA, Shafer DW (1986). Residuals in generalized linear models. *Journal of the American Statistical Association*, **81**, 977–986. Cited on page 38

Poon WY, Poon YS (1999). Conformal normal curvature and assessment of local influence. *Journal of the Royal Statistical Society Series B (Methodological)*, **61**, 51–61. Cited on page 41, 58

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/. Cited on page 59

Puig P (2008). A note on the harmonic law: A two-parameter family of distributions for ratios. *Statistics and Probability Letters*, **78**, 320–326. Cited on page 3, 4, 7

Rieck JR (1999). A moment-generating function with application to the birnbaum-saunders distribution. *Communications in Statistics. Theory and Methods*, **28**, 2213–2222. Cited on page 6

Rieck JR, Nedelman JR (1991). A log-linear model for the birnbaum-saunders distribution. *Technometrics*, **33**, 51–60. Cited on page 7, 27, 63, 80, 83

Rigby RA, Stasinopoulos DM (2005). Generalized additive models for location, scale and shape. *Journal Applied Statistics*, **54**, 507–554. Cited on page 24, 36, 54, 59

Rigby RA, Stasinopoulos DM (2006). Using the box-cox t distribution in gamlss to model skewness and kurtosis. *Statistical Modelling*, **6**, 209–229. Cited on page 3, 17, 24

Rigby RA, Stasinopoulos DM (2007). Generalized additive models for location, scale and shape (gamlss) in R. *Journal of Statistical Software*, **23**, 1–46. Cited on page 59

Rigby RA, Stasinopoulos DM (2014). **gamlss***: Generalized Additive Models for Location Scale and Shape*. R package version 4.2-8. URL http://CRAN.R-project.org/package=gamlss. Cited on page 59

Rogers WH, Tukey JM (1972). Understanding some long-tailed symmetrical distributions. *Statistica Neerlandica*, **26**, 211–226. Cited on page 6

Schwarz GE (1978). Estimating the dimension of a model. *The Annals of Statistics*, **6**, 461–464. Cited on page 17

Stacy EW (1962). A generalization of the gamma distribution. *The Annals of Mathematical Statistics*, **33**, 1187–1192. Cited on page 3

Therneau T (2014). **survival***: A Package for Survival Analysis in S*. R package version 2.38-1. URL http://CRAN.R-project.org/package=survival. Cited on page 59

Vanegas LH, Paula GA (2014). **ssym***: Fitting Semiparametric Symmetric Regression Models*. R package version 1.5.2. URL http://CRAN.R-project.org/package=ssym. Cited on page 59

Vanegas LH, Paula GA (2014b). Log-symmetric distributions: Statistical properties and parameter estimation. *Brazilian Journal of Probability and Statistics*, **To appear**. Cited on page 41, 63

Vanegas LH, Paula GA (2015a). A semiparametric approach for joint modeling of median and skewness. *TEST*, **24**, 110–135. Cited on page 23

Villegas C, Paula GA, Cysneiros FJA, Galea M (2013). Influence diagnostics in generalized symmetric linear models. *Computational Statistics and Data Analysis*, **59**, 161–170. Cited on page 10

Waller LA, Turnbull BW (1992). Probability plotting with censored data. *The American Statistician*, **46**, 5–12. Cited on page 41, 57

West M (1987). On scale mixtures of normal distributions. *Biometrika*, **74**, 646–648. Cited on page 11, 29, 51

Wood SN (2006). *Generalized Additive Models: An Introduction with R*. Chapman and Hall, Boca Raton. Cited on page 24, 26, 27, 28, 36, 54

Wu C, Yu Y (2014). Partially linear modeling of conditional quantiles using penalized splines. *Computational Statistics and Data Analysis*, **77**, 170–187. Cited on page 24, 32, 36, 53, 54, 65

Yu Y, Ruppert D (2002). Penalized spline estimation for partially linear single-index models. *Journal of the American Statistical Association*, **97**, 1042–1054. Cited on page 32, 53

Zeileis A, Croissant Y (2010). Extended model formulas in R: Multiple parts and multiple responses. *Journal of Statistical Software*, **34**, 1–13. Cited on page 60

Zhou M (1991). Some properties of the kaplan-meier estimator for independent nonidentically distributed random variables. *The Annals of Statistics*, **19**, 2266–2274. Cited on page 57

Zwillinger D, Kokoska S (2000). *Standard Probability and Statistical Tables and Formula*. Chapman and Hall, Boca Raton. Cited on page 9