UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE CIÊNCIAS ECONÔMICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

HELTON SAULO BEZERRA DOS SANTOS

ESSAYS ON BIRNBAUM-SAUNDERS MODELS

Porto Alegre
2013

HELTON SAULO BEZERRA DOS SANTOS

ESSAYS ON BIRNBAUM-SAUNDERS MODELS

Tese submetida ao Programa de Pós-Graduação em Economia da Faculdade de Ciências Econômicas da UFRGS, como requesito parcial para obtenção do título de Doutor em Economia, com ênfase em Economia Aplicada.

Orientador: Prof. Dr. Flávio Augusto Ziegelmann
Coorientador: Prof. Dr. Víctor Leiva

Porto Alegre
2013

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)

Santos, Helton Saulo Bezerra dos
Essays on Birnbaum-Saunders Models / Helton Saulo Bezerra dos Santos. -- 2013.
87 f.

Orientador: Flávio Augusto Ziegelmann.
Coorientador: Víctor Leiva.

Tese (Doutorado) -- Universidade Federal do Rio Grande do Sul, Faculdade de Ciências Econômicas, Programa de Pós-Graduação em Economia, Porto Alegre, BR-RS, 2013.

1. Distribuição Birnbaum-Saunders. 2. Algoritmo EM. 3. Índices de capacidade do processo. 4. Estimador por função-núcleo. 5. Modelos autorregressivos de duração condicional. I. Ziegelmann, Flávio Augusto , orient. II. Leiva, Víctor, coorient. III. Título.

HELTON SAULO BEZERRA DOS SANTOS

ESSAYS ON BIRNBAUM-SAUNDERS MODELS

Tese submetida ao Programa de Pós-Graduação em Economia da Faculdade de Ciências Econômicas da UFRGS, como requesito parcial para obtenção do título de Doutor em Economia, com ênfase em Economia Aplicada.

Aprovada em: Porto Alegre, 13 de Setembro de 2013.

BANCA EXAMINADORA:

Prof. Dr. Flávio Augusto Ziegelmann (orientador) - UFRGS

Prof. Dr. Francisco Cribari Neto - UFPE

Prof. Dr. Miguel Angel Uribe Opazo - UNIOEST

Prof. Dr. Hudson da Silva Torrent - UFRGS

In Memory of my Grandfather Geraldo Barbosa

ACKNOWLEDGEMENTS

This thesis could never have been completed without ample guidance, assistance and encouragement. In this regard, my desire is to thank kindly the following persons:

My supervisors at UFRGS, Prof. Flavio Zielgelmann and Prof. Victor Leiva. I took great delight in having comments and support from them. They taught me how to guide my thoughts into appropriate statements. Thank you for sharing some of your knowledge and plenty of your time and wisdom.

My supervisor at McMaster University, Prof. N. Balakrishnan, who possesses an extraordinary knowledge in all fields of statistics. Thank you for sharing some of your knowledge and time.

The UFRGS community, Iara, Raquel, Delourdes and unknown servants, for countless assistances.

My old friends, Jeremias, Manoel, Marcelo, Austin, Agrinaldo, Lutemberg, Josimar, David and Heritier, for unceasing friendship.

The commitee members, my profound thanks.

My UFRGS friends, especially Pilar, Larissa, Vivian, Gabrielito, Julio Cesar, Paulo, Rodrigo, Eduardo, Prof. Flavio Zielgelmann, Prof. Hudson Torrent and Prof. Fabricio Tourrucoo, for providing me endless support both emotionally and professionally.

My POA brothers, Maciel and Diego, for sharing funny “chinelagens”. I also thank Ivan, Carol and Filipe.

My Chilean friends, Yesli Caro, Prof. Victor Leiva, Carolina Marchant, Margareth Cofre and Gonzalo Vicencio, for receiving me very well in Valpo/Vina.

My friends in Hamilton, Eric, Dr. Joseph, Cherry, Miercoles, Derek, Gry, Lene, Peter, Jack,

and so on, for sharing great moments.

My Canadian family, Natalie, Richard and Alejandro. Thank you for bearing with me.

A special word of thanks to my parents, Francisca and Hildomar, for all the effort you put in to my education, both as a person and a student. Thank you for your love and support.

My brothers, Kenny, Hildo and Hyata, for sharing good moments and giving me total support.

My sweet Alessandra Meneses (Sandrinha). I am very lucky and grateful.

My favorite composer, Johann Sebastian Bach, for the greatest pieces of music ever written.

The CAPES and CNPq for providing financial support.

Give Me Women, Wine, And Snuff
(John Keats)

Give me women, wine, and snuff
Until I cry out “hold, enough!”
You may do so sans objection
Till the day of resurrection;
For, bless my beard, they aye shall be
My beloved Trinity.

“26”
(Lawrence Ferlinghetti)

That “sensual phosphorescence
my youth delighted in”
now lies almost behind me
like a land of dreams
wherein an angel
of hot sleep
dances like a diva
in strange veils
thru which desire
looks and cries
And still she dances
dances still
and still she comes
at me
with breathing breasts
and secret lips
and (ah)
bright eyes

RESUMO

Nessa tese apresentamos três diferentes aplicações dos modelos Birnbaum-Saunders. No capítulo 2 introduzimos um novo método por função-núcleo não-paramétrico para a estimação de densidades assimétricas, baseado nas distribuições Birnbaum-Saunders generalizadas assimétricas. Funções-núcleo baseadas nessas distribuições têm a vantagem de fornecer flexibilidade nos níveis de assimetria e curtose. Em adição, os estimadores da densidade por função-núcleo Birnbaum-Saunders gene-ralizadas assimétricas são livres de viés na fronteira e alcançam a taxa ótima de convergência para o erro quadrático integrado médio dos estimadores por função-núcleo-assimétricas-não-negativos da densidade. Realizamos uma análise de dados consistindo de duas partes. Primeiro, conduzimos uma simulação de Monte Carlo para avaliar o desempenho do método proposto. Segundo, usamos esse método para estimar a densidade de três dados reais da concentração de poluentes atmosféricos. Os resultados numéricos favorecem os estimadores não-paramétricos propostos. No capítulo 3 propomos uma nova família de modelos autorregressivos de duração condicional baseados nas distribuições misturas de escala Birnbaum-Saunders (SBS). A distribuição Birnbaum-Saunders (BS) é um modelo que tem recebido considerável atenção recentemente devido às suas boas propriedades. Uma extensão dessa distribuição é a classe de distribuições SBS, a qual (i) herda várias das boas propriedades da distribuição BS, (ii) permite a estimação de máxima verossimilhança em uma forma eficiente usando o algoritmo EM, e (iii) possibilita a obtenção de um procedimento de estimação robusta, entre outras propriedades. O modelo autorregressivo de duração condicional é a família primária de modelos para analisar dados de duração de transações de alta frequência. A metodologia estudada aqui inclui estimação dos parâmetros pelo algoritmo EM, inferência para esses parâmetros, modelo preditivo e uma análise residual. Realizamos simulações de Monte Carlo para avaliar o desempenho da metodologia proposta. Ainda, avalia-mos a utilidade prática dessa metodologia usando dados reais de transações financeiras da bolsa de valores de Nova Iorque. O capítulo 4 trata de índices de capacidade do processo (PCIs), os quais são ferramentas utilizadas pelas empresas para determinar a qualidade de um produto e avaliar o desempenho de seus processos de produção.

Estes índices foram desenvolvidos para processos cuja característica de qualidade tem uma distribuição normal. Na prática, muitas destas ca-racterísticas não seguem esta distribuição. Nesse caso, os PCIs devem ser modificados considerando a não-normalidade. O uso de PCIs não-modificados podem levar a resultados inadequados. De maneira a estabelecer políticas de qualidade para resolver essa inadequação, transformação dos dados tem sido proposta, bem como o uso de quantis de distribuições não-normais. Um distribuição não-normal assimétrica o qual tem tornado muito popular em tempos recentes é a distribuição Birnbaum-Saunders (BS). Propomos, desenvolvemos, implementamos e aplicamos uma metodologia baseada em PCIs para a distribuição BS. Além disso, realizamos um estudo de simulação para avaliar o desempenho da metodologia proposta. Essa metodologia foi implementada usando o software estatístico chamado R. Aplicamos essa metodologia para um conjunto de dados reais de maneira a ilustrar a sua flexibilidade e potencialidade.

Palavras-chave: Dados de poluição atmosférica. Estimador por função-núcleo. Distribuições Birnbaum-Saunders generalizadas assimétricas. Modelos autorregressivos de duração condicional. Algoritmo EM. Método ML. Simulações de Monte Carlo. Distribuições misturas de escala Birnbaum-Saunders. Índices de capacidade do processo.

ABSTRACT

In this thesis, we present three different applications of Birnbaum-Saunders models. In Chapter 2, we introduce a new nonparametric kernel method for estimating asymmetric densities based on generalized skew-Birnbaum-Saunders distributions. Kernels based on these distributions have the advantage of providing flexibility in the asymmetry and kurtosis levels. In addition, the generalized skew-Birnbaum-Saunders kernel density estimators are boundary bias free and achieve the optimal rate of convergence for the mean integrated squared error of the nonnegative asymmetric kernel density estimators. We carry out a data analysis consisting of two parts. First, we conduct a Monte Carlo simulation study for evaluating the performance of the proposed method. Second, we use this method for estimating the density of three real air pollutant concentration data sets, whose numerical results favor the proposed nonparametric estimators. In Chapter 3, we propose a new family of autoregressive conditional duration models based on scale-mixture Birnbaum-Saunders (SBS) distributions. The Birnbaum-Saunders (BS) distribution is a model that has received considerable attention recently due to its good properties. An extension of this distribution is the class of SBS distributions, which allows (i) several of its good properties to be inherited; (ii) maximum likelihood estimation to be efficiently formulated via the EM algorithm; (iii) a robust estimation procedure to be obtained; among other properties. The autoregressive conditional duration model is the primary family of models to analyze high-frequency financial transaction data. This methodology includes parameter estimation by the EM algorithm, inference for these parameters, the predictive model and a residual analysis. We carry out a Monte Carlo simulation study to evaluate the performance of the proposed methodology. In addition, we assess the practical usefulness of this methodology by using real data of financial transactions from the New York stock exchange. Chapter 4 deals with process capability indices (PCIs), which are tools widely used by companies to determine the quality of a product and the performance of their production processes. These indices were developed for processes whose quality characteristic has a normal distribution. In practice, many of these characteristics do not follow this distribution. In such a case, the PCIs must be modified considering

the non-normality. The use of unmodified PCIs can lead to inadequacy results. In order to establish quality policies to solve this inadequacy, data transformation has been proposed, as well as the use of quantiles from non-normal distributions. An asymmetric non-normal distribution which has become very popular in recent times is the Birnbaum-Saunders (BS) distribution. We propose, develop, implement and apply a methodology based on PCIs for the BS distribution. Furthermore, we carry out a simulation study to evaluate the performance of the proposed methodology. This methodology has been implemented in a noncommercial and open source statistical software called R. We apply this methodology to a real data set to illustrate its flexibility and potentiality.

Keywords: Air pollutant data. Kernel estimator. Generalized skew-Birnbaum-Saunders distributions. Autoregressive conditional duration models. EM algorithm. ML method. Monte Carlo simulations. Scale-mixture Birnbaum-Saunders distributions. Process capability indices.

SUMÁRIO

1 INTRODUCTION

Nesta tese apresentamos três trabalhos abordando áreas distintas, a saber, estatística não-paramétrica, econometria financeira e controle de qualidade. Dessa maneira, cada capítulo pode ser lido de maneira independente, pois cada um é autocontido. O ponto comum reside no uso de modelos Birnbaum-Saunders (BS), em particular da distribuição BS e duas de suas generalizações, isto é, as distribuições BS generalizadas assimétricas (skew-GBS) e misturas de escala BS (SBS). Abaixo, fornecemos uma derivação sumária da distribuição BS.

Birnbaum e Saunders (1969) introduziram uma nova família de distribuições para modelar o tempo de vida de materiais e equipamentos sujeitos a cargas dinâmicas. A motivação dos autores advém de problemas de vibração em aviões comerciais e problemas de falhas de materiais. Basicamente, a distribuição é derivada a partir de um modelo cuja falhas acontecem em função do desenvolvimento e o crescimento de uma rachadura dominante. Considere, por exemplo, um material que é sujeito a um padrão cíclico de tensão e força. Defina um ciclo como m oscilações onde uma extensão aleatória da rachadura X_i é resultado da aplicação da i-ésima oscilação. Assim, a extensão da rachadura devido ao j-ésimo ciclo é dada por

$$Y_j = \sum_{i=1}^{m} X_i, \tag{1.1}$$

onde Y_j é uma variável aleatória com média μ e variância σ^2, para todo $j = 1, 2, 3, \ldots$ Depois de z ciclos, a extensão total da rachadura é dada por

$$W_z = \sum_{j=1}^{z} Y_j, \tag{1.2}$$

onde a função de distribuição é $H_z(\omega) = P(W_z \leq \omega)$, para $z = 1, 2, 3, \ldots$ Note que a falha acontece quando o comprimento da rachadura dominante ultrapassa um certo limiar ω.

Denotando C o número de ciclos até a falha, então sua distribuição é dada por

$$P(C \leq z) = P\left(\sum_{j=1}^{z} Y_j > \omega\right) = 1 - H_z(\omega). \tag{1.3}$$

Usando o Teorema Central do Limite e assumindo que os Y_j's são variáveis aleatórias independentes e identicamente distribuídas, temos

$$\begin{aligned} P(C \leq z) &= 1 - P\left(\sum_{j=1}^{z} \frac{Y_j - \mu}{\sigma\sqrt{z}} \leq \frac{\omega - z\mu}{\sigma\sqrt{z}}\right) \\ &= 1 - P\left(\sum_{j=1}^{z} \frac{Y_j - \mu}{\sigma\sqrt{z}} \leq \frac{\omega}{\sigma\sqrt{z}} - \frac{\mu\sqrt{z}}{\sigma}\right) \\ &\cong \Phi\left(\frac{\mu\sqrt{z}}{\sigma} - \frac{\omega}{\sigma\sqrt{z}}\right), \end{aligned} \tag{1.4}$$

onde $\Phi(\cdot)$ denota a função de distribuição acumulada normal padrão. Aqui z é substituída por uma variável aleatória real não-negativa t, tal que T seja a extensão contínua da variável aleatória discreta C. Logo, T pode ser considerada como o tempo até a falha, e segue uma sua distribuição BS bi-paramétrica, denotada por $T \sim BS(\alpha, \beta)$, cuja função de distribuição é dada por

$$F_T(t) = P(T \leq t) = \Phi\left(\frac{1}{\alpha}\left[\left(\frac{t}{\beta}\right)^{1/2} - \left(\frac{\beta}{t}\right)^{1/2}\right]\right), \quad t > 0, \tag{1.5}$$

onde $\alpha = \sigma/\sqrt{\omega\mu} > 0$ e $\beta = \omega/\mu > 0$. α e β são parâmetros de forma e escala, respectivamente. Para qualquer constante real $k > 0$, temos que $kT \sim BS(\alpha, k\beta)$. Quando α tende a zero, a distribuição BS tende para a distribuição normal de média β e variância τ, onde $\tau \to 0$ quando $\alpha \to 0$. Por outro lado, como β é um parâmetro de escala, segue que $T/\beta \sim BS(\alpha, 1)$. Adicionalmente, β é a mediana da distribuição, i.e., $F_T(\beta) = \Phi(0) = 1/2$. A distribuição BS possui a propriedade recíproca, i.e. $T^{-1} \sim BS(\alpha, \beta^{-1})$. Para mais detalhes ver Birnbaum e Saunders (1969).

A atratividade da distribuição BS para a análise de dados deve-se, entre outras coisas, a suas propriedades teóricas e sua relação com a distribuição normal. Dentre suas aplicações práticas, destacam-se as áreas de economia, engenharia, finanças, medicina, meio ambiante e negócios; ver, por exemplo, Jin e Kawczak (2003), Leiva et al. (2009, 2010, 2012), Ahmed et al. (2010), Bhatti (2010), Vilca et al. (2010, 2011), Ferreira et al. (2012), Paula et al. (2012), Marchant et al. (2013) e Leiva et al. (2013).

Recentemente, duas importantes generalizações da distribuição BS foram obtidas por

Vilca e Leiva (2006) e Balakrishnan et al. (2009) com os modelos skew-GBS e SBS, respectivamente. Por um lado, as distribuições skew-GBS são baseadas nos argumentos apresentados em Díaz-García e Leiva (2005) e nas distribuições elípticas assimétricas. As distribuições skew-GBS fornecem um grau maior de flexibilidade em relação à curtose e assimetria, sendo a última característica devido a insersão de um parâmetro de assimetria. Por outro lado, as distribuições SBS são baseadas na relação das distribuições BS e normal. Essa última generalização fornece propriedades interessantes como o estimação dos parâmetros de máxima verossimilhança em uma maneira eficiente, usando o algoritmo esperança-maximização (EM), e permite o procedimento de estimação robusta dos parâmetros.

Essa tese, como mencionado, explora o uso da distribuição BS e suas generalizações skew-GBS e SBS nas seguintes áreas: estatística não-paramétrica, econometria financeira e controle de qualidade. Assim, para cada área, há um respectivo capítulo.

O capítulo 2 tem por objetivo propor estimadores não-paramétricos por função-núcleo para densidades assimétricas baseados nas distribuições skew-GBS. Esses estimadores fornecem uma maior flexibilidade em termos de curtose e assimetria, são livres de viés na fronteira e alcançam a taxa ótima de convergência para o erro quadrático integrado médio dos estimadores por função-núcleo-assimétricas-não-negativos da densidade. Realizamos simulações de Monte Carlo e fazemos aplicações a dados reais de concentração de poluentes atmosféricos. Em particular, diferentes aspectos ambientais têm sido relacionados ao desenvolvimento e crescimento econômico, em especial, os efeitos de contaminantes atmosféricos na saúde humana e suas repercussões sobre a economia. Desse modo, um bom conhecimento, por exemplo, da distribuição estatística de dados ambientais, permitem descrever a qualidade atmosférica e por conseguinte seu impacto na economia.

No capítulo 3 propomos uma nova família de modelos autorregressivos de duração condicional (ACD) baseados nas distribuições SBS. Essa classe de distribuições (i) herda várias propriedades da distribuição BS, (ii) permite a estimação de máxima verossimilhança de forma eficiente usando o algoritmo EM, e (iii) possibilita a obtenção de um procedimento de estimação robusta. Os modelos ACD têm sido usados para analisar dados de duração de transações de alta frequência e são de grande relevância em alguns modelos da teoria de microestrutura, os quais são baseados em elementos da assimetria de informação. Em particular, uma alta frequência de transações implica transações feitas por negociadores informados. Realizamos simulações de Monte Carlo para avaliar o desempenho da metodologia proposta. Por fim, avaliamos a utilidade prática dessa metodologia usando dados reais de transações financeiras da bolsa de valores de Nova Iorque.

No capítulo 4 propomos índices de capacidade do processo (PCIs) baseados na distribuição BS. Os PCIs são ferramentas utilizadas pelas empresas para determinar a qualidade de

um produto e avaliar o desempenho de seus processos de produção. Em última instância, os PCIs são de extrema relevância para a produtividade de um país, e consequentemente para a economia. Estes índices foram desenvolvidos para processos cuja característica de qualidade tem uma distribuição normal. Na prática, muitas destas características não seguem esta distribuição. Nesse contexto, a distribuição BS se torna uma importante alternativa. Realizamos um estudo de simulação para avaliar o desempenho da metodologia proposta. Em adição, aplicamos essa metodologia para um conjunto de dados reais de maneira a ilustrar a sua flexibilidade e potencialidade.

Por fim, no capítulo 5 apresentamos algumas considerações finais sobre os trabalhos apresentados nessa tese.

2 A NONPARAMETRIC METHOD FOR ESTIMATING ASYMMETRIC DENSITIES

2.1 Introduction

Usually, air contaminant concentrations, such as particulate matter measuring less than 10 micrometers (PM_{10}), sulfur dioxide (SO_2) and tropospheric ozone (O_3), are considered as continuous non-negative random variables that can be modeled by a probability distribution. The probability density functions of these random variables (or simply densities) are often asymmetric and present positive skewness and high kurtosis. Therefore, because for instance the normal or Gaussian distribution is symmetrical, it is not a good model for describing the aforementioned environmental random variables.

Density estimation plays an important role in statistics, because it provides an idea about the shape of the probability distribution of the data. From a parametric density estimation perspective, one must rely on well-specified models, which depend on the data available. However, the lack of prior information usually makes nonparametric density estimation a better choice, which can be based on kernel methods; see seminal papers by Rosenblatt (1956) and Parzen (1962). For applications of the kernel method to environmental data; see Lorimer (1986), Haan (1999), Pagnini (2009), and Chang et al. (2012).

Classical kernel methods are based on distributions that are symmetric about zero, as in the case of the Gaussian kernel. However, a drawback arises when we try to estimate density functions with bounded supports via classical kernel methods, because this type of methods assign weight outside the support, when smoothing is performed near the boundary; see Fernandes and Monteiro (2005). Chen (1999, 2000) proposed two asymmetric kernel methods, which considerably increase the precision of the nonnegative density estimation near the boundary. He introduced the beta and gamma kernels with supports on $[0, 1]$ and $[0, \infty)$, respectively. Following the same line, Scaillet (2004) introduced kernel estimators based on inverse Gaussian distributions. All of these kernel estimators are boundary bias free and

achieve the optimal rate of convergence for the mean integrated squared error (MISE). The main advantages of the asymmetric kernel methods over the classical methods are that the formers have varying shape and permit flexibility in the smoothing throughout the support. In addition, the asymmetric kernel methods never assign weight outside the density support; see Fernandes and Monteiro (2005). Jin and Kawczak (2003) discussed asymmetric kernel density estimators based on the Birnbaum-Saunders (BS) and lognormal (LN) distributions; see Birnbaum and Saunders (1969). The BS and LN kernels possess the same properties as other asymmetric kernels, i.e., they are boundary bias free, nonnegative, of varying shape and achieve the optimal rate of convergence for the MISE. The authors found evidence that the BS and LN kernel density estimators outperform all other estimators based on asymmetric kernels. Abadir and Lawford (2004) provided a justification for the use of asymmetric kernels arguing that density estimators in moderately-sized samples tend to acquire salient properties of their kernels. Several other methods have been suggested to address the boundary problem in kernel density estimation; see Marchant et al. (2013) for a complete review about these methods. In view of these antecedents, effectively the use of asymmetric kernels for estimating densities produces good results. However, all the distributions employed for the mentioned kernels have little flexibility in their tails.

Díaz-García and Leiva (2005) generalized the BS distribution, obtaining a wider class of nonnegative densities that possesses either lighter or heavier tails than the BS density, allowing them to provide more flexibility. The generalized Birnbaum-Saunders (GBS) family has as particular cases the BS-classical, BS-power-exponential (BS-PE) and BS-Student-t (BS-t) distributions. Recently, Marchant et al. (2013) proposed GBS kernel density estimators, which generalize the BS kernel estimator, obtaining better results than those obtained by other nonparametric kernel methods proposed for estimating asymmetric densities.

Vilca and Leiva (2006) introduced skewed GBS distributions, providing a greater degree of flexibility due to the incorporation of a skewness or asymmetry parameter. Thus, such as mentioned by Vilca et al. (2011), data located at the tails (left or right) are accommodated in a better way by using generalized skew-BS (skew-GBS) distributions. Then, our conjecture is that, although density nonparametric estimators based on GBS kernels have good properties, skew-GBS kernel density estimators should provide better results. Another reason for considering BS distributions as kernels is that such models have been largely applied to environmental data; see Leiva et al. (2009, 2010, 2012), Vilca et al. (2010, 2011), and Ferreira et al. (2012).

The main goals of our work are (i) to propose new density estimators based on skew-GBS kernels, which should hold with the properties of nonnegative kernels, but in addition these should have a better behavior; and (ii) to apply the proposed kernel density estimators to real

environmental data, specifically, to O_3, PM_{10} and SO_2 air contaminant concentrations.

The remainder of this chapter unfolds as follows. In Section 2.2, we define the skew-GBS distributions. In Section 2.3, we propose new nonparametric kernel estimators associated with these distributions. In Section 2.4, we derive some statistical properties of the proposed estimators. In Section 2.5, we present some identification, selection and validation methods, and discuss a computational implementation of the proposed results. In Section 2.7, we perform a Monte Carlo (MC) simulation study for evaluating the proposed kernel density estimators. In Section 2.6, we carry out an empirical application of our method to the aforementioned environmental data. Finally, in Section 2.8, we provide some concluding remarks.

2.2 The Birnbaum-Saunders distribution and its generalizations

Birnbaum and Saunders (1969) introduced a two-parameter distribution for a positive random variable (RV) T with the following representation

$$T = \beta \left[\alpha\, Z/2 + (\{\alpha\, Z/2\}^2 + 1)^{1/2}\right]^2, \tag{2.1}$$

where $Z \sim \mathrm{N}(0,1)$, $\alpha > 0$ and $\beta > 0$ are shape and scale parameters, but β is also the median of this distribution. This is denoted by $\mathrm{BS}(\alpha, \beta)$. The BS distribution holds proportionality and reciprocal properties given by $b\,T \sim \mathrm{BS}(\alpha, b\,\beta)$, with $b > 0$, and $1/T \sim \mathrm{BS}(\alpha, 1/\beta)$. Díaz-García and Leiva (2005) postulated and characterized the GBS distribution assuming that Z given in (2.1) follows a symmetric distribution in $\mathbb{R}$, which is denoted by $T \sim \mathrm{GBS}(\alpha, \beta;\, g)$, where g is a density generator associated with a particular symmetric distribution. Vilca and Leiva (2006) obtained a greater generalization of the GBS distribution assuming that Z given in (2.1) follows a skew-symmetric distribution in $\mathbb{R}$, which is denoted by $T \sim$ skew-$\mathrm{GBS}(\alpha, \beta, \lambda;\, g)$, where now g is a density generator associated with a particular skew-symmetric distribution, and λ is a shape parameter, known as asymmetry or skewness parameter.

Standard skew-symmetric distributions are denoted by $Z \sim \mathrm{SS}(0, 1, \lambda;\, g)$, where λ is a skewness parameter and g is a density generator. In this case, the density of Z takes the form

$$f_Z(z; \lambda, g) = 2 f_S(z)\, F_S(\lambda z) = 2c\, g(z)\, F_S(\lambda z), \quad z \in \mathbb{R},\ \lambda \in \mathbb{R}, \tag{2.2}$$

with $f_S = c\, g$ being the density of a symmetric distribution in $\mathbb{R}$ and F_S its corresponding distribution function. Note that g is a real function that generates the density of Z and c its normalization constant, such that $\int_{-\infty}^{+\infty} g(z^2)\mathrm{d}z = 1/c$. The expression given in (2.2) permits

a large number of skew-symmetric distributions to be obtained. Nadarajah and Kotz (2003) considered f_S and G_S having the Cauchy, Laplace, logistic, normal, Student-t and uniform distributions. Other classes are the skew-Cauchy, skew-elliptic, skew-normal, skew-slash, and skew-t distributions; see Azzalini and Capitanio (1999), Arnold and Beaver (2000), Azzalini and Capitanio (2003), Wang and Genton (2006) and Gómez et al. (2007). All of these distributions provide wider asymmetry and kurtosis than the distributions proposed in Nadarajah and Kotz (2003).

Vilca and Leiva (2006) considered a RV Z such that $Z = [\sqrt{T/\beta} - \sqrt{\beta/T}]/\alpha \sim \text{SS}(0, 1, \lambda;\, g)$ and then

$$T = \beta \left[\alpha\, Z/2 + (\{\alpha\, Z/2\}^2 + 1)^{1/2}\right]^2 \sim \text{skew-GBS}(\alpha, \beta, \lambda;\, g). \tag{2.3}$$

From (2.2), the density of T in (2.3) is given by

$$\begin{aligned} f(t;\alpha,\beta,\lambda,g) &= 2c\, g\big(\tfrac{1}{\alpha^2}\big[\tfrac{t}{\beta} + \tfrac{\beta}{t} - 2\big]\big) F_S\big(\tfrac{\lambda}{\alpha}\big[\sqrt{t/\beta} - \sqrt{\beta/t}\big]\big) \frac{t^{-3/2}[t+\beta]}{2\alpha\sqrt{\beta}}, \\ & t > 0, \alpha, \beta > 0, \lambda \in \mathbb{R}. \end{aligned} \tag{2.4}$$

The mean and variance of T are, respectively,

$$\mathbb{E}[T] = \tfrac{\beta}{2}\left[2 + \alpha^2\gamma_2 + \alpha\,\omega_1\right] \text{ and } \text{Var}[T] = \tfrac{\beta^2}{4}\left[4\alpha^2\gamma_2 - \alpha^2\omega_1^2 + 2\alpha^3\omega_3 - 2\alpha^3\gamma_2\,\omega_1 - \alpha^4\gamma_2^2 + 2\alpha^4\gamma_4\right], \tag{2.5}$$

where $\gamma_r = \mathbb{E}[Z^r]$ and $\omega_r = \mathbb{E}[Z^r\sqrt{\alpha^2 Z^2 + 4}]$, with $Z \sim \text{SS}(0, 1, \lambda;\, g)$; see Vilca and Leiva (2006).

Table 2.1 presents density generators and their normalization constants.

Tabela 2.1: constant (c) and density generator (g) for the indicated distribution.

Distribution	c	$g = g(u), u > 0$
Normal	$\frac{1}{\sqrt{2\pi}}$	$\exp\left(-\frac{1}{2}u\right)$
PE	$\frac{\eta}{2^{\frac{1}{2\eta}}\Gamma\left(\frac{1}{2\eta}\right)}$	$\exp\left(-\frac{1}{2}\,u^{\eta}\right), \eta > 0$
t	$\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}$	$\left[1 + \frac{u}{\nu}\right]^{-\frac{\nu+1}{2}}, \nu > 0$

Based on Table 2.1, expressions for the skew-BS, skew-BS-PE and skew-BS-t densities

are

$$
\begin{aligned}
f_{\text{skew-BS}}(t) &= \tfrac{2}{\sqrt{2\pi}}\exp\Big(-\tfrac{1}{2\alpha^2}\big[\tfrac{t}{\beta}+\tfrac{\beta}{t}-2\big]\Big)\Phi\Big(\tfrac{\lambda}{\alpha}\Big[\sqrt{t/\beta}-\sqrt{\beta/t}\Big]\Big)\tfrac{t^{-3/2}[t+\beta]}{2\alpha\sqrt{\beta}},\\
f_{\text{skew-BS-PE}}(t) &= \frac{2\eta}{\Gamma\big(\frac{1}{2\eta}\big)2^{\frac{1}{2\eta}}}\exp\Big(-\tfrac{1}{2\alpha^{2\eta}}\big[\tfrac{t}{\beta}+\tfrac{\beta}{t}-2\big]^{\eta}\Big)\\
&\quad\times\Big[\tfrac{1}{2}+\tfrac{1}{2\Gamma\big(\frac{1}{2\eta}\big)}\Gamma\Big(\tfrac{1}{2\eta},\tfrac{1}{2}\big[\tfrac{\lambda}{\alpha}\big]^{2\eta}\big[\tfrac{t}{\beta}+\tfrac{\beta}{t}-2\big]^{\eta}\Big)\Big]\tfrac{t^{-3/2}[t+\beta]}{2\alpha\sqrt{\beta}},\\
f_{\text{skew-BS-}t}(t) &= \tfrac{2\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\Big[1+\tfrac{1}{\nu\alpha^2}\big\{\tfrac{t}{\beta}+\tfrac{\beta}{t}-2\big\}\Big]^{-\frac{\nu+1}{2}}\\
&\quad\times\Big(1-\tfrac{1}{2}\,\mathrm{I}_{\nu/[\nu+\{\lambda/\alpha\}^2\{t/\beta+\beta/t-2\}]}\big(\tfrac{\nu}{2},\tfrac{1}{2}\big)\Big)\tfrac{t^{-3/2}[t+\beta]}{2\alpha\sqrt{\beta}},\quad t>0,\alpha,\beta,\nu>0,\lambda\in\mathbb{R},
\end{aligned}
$$

where Φ is the N(0, 1) distribution function, $\Gamma(\alpha,y)=\int_0^y u^{\alpha-1}\exp(-u)\mathrm{d}u$ is the lower incomplete gamma function, and $\mathrm{I}_y(a,b)=\mathrm{B}_y(a,b)/\mathrm{B}_1(a,b)$ is the incomplete beta ratio, with $\mathrm{B}_y(a,b)=\int_0^y u^{a-1}[1-u]^{b-1}\mathrm{d}u$. Note that if $\eta=1$ or if $\nu\to\infty$, then we obtain the skew-BS distribution.

2.3 Skew-GBS kernel density estimators

A classical kernel estimate for an unknown density f with support on $[0,\infty)$, at a point x, based on observations $x_1,\ldots,x_n$, takes the form $\widehat{f}_n(x)=[1/(nh)]\sum_{i=1}^n K([x-x_i]/h)$, where h is a smoothing parameter (also known as bandwidth) and K is a symmetric kernel satisfying $\int K(x)\mathrm{d}x=1$. An asymmetric kernel density estimate can be expressed as

$$
\widehat{f}_n(x)=\frac{1}{n}\sum_{i=1}^{n}K_{L(h,\,x)}\left(x_i\right),\quad x\geq 0, \tag{2.6}
$$

where $K_{L(h,x)}$ is an asymmetric kernel associated with the distribution L, again h is the bandwidth, and x the point where the density is estimated. In general, the class of asymmetric kernel density estimators is boundary bias free and its bias is of order $O(h)$; see Chen (1999, 2000). Jin and Kawczak (2003) proposed density estimates based on the BS kernel given by

$$
\begin{aligned}
K_{\text{BS}(\sqrt{h},\,x)}(t) &= \tfrac{1}{\sqrt{2\pi}}\exp\Big(-\tfrac{1}{2h}\big[\tfrac{t}{x}+\tfrac{x}{t}-2\big]\Big)\tfrac{t^{-3/2}[t+x]}{\sqrt{4\,h\,x}}, &(2.7)\\
K_{\text{LN}(4\log(1+h),\,\log(x))}(t) &= \tfrac{1}{\sqrt{8\pi\log(1+h)}\;t}\exp\Big(-\tfrac{[\log(t)-\log(x)]^2}{8\log(1+h)}\Big), &(2.8)
\end{aligned}
$$

respectively, where h satisfies both $h\to 0$ and $nh\to\infty$, when $n\to\infty$. By using the estimate presented in (2.6) and the BS kernel provided in (2.7), Marchant et al. (2013) proposed a density estimate based on the GBS kernel defined by $K_{\text{GBS}(\sqrt{h},\,x;\,g)}(t)=c\,g([1/h][t/x+x/t-$

$2])\, t^{-3/2}[t+x]/\sqrt{4hx}$, where c and g are as given in (2.4). We extend the class of asymmetric kernel density estimators by proposing a new method based on skew-GBS distributions. As explained earlier, these distributions provide highly flexible densities with either heavier or lighter tails and more or less pronounced asymmetry than the BS distribution. From the density given in (2.4), the kernel estimate proposed in (2.6), and setting $\alpha = \sqrt{h}$ and $\beta = x$, a skew-GBS kernel estimate for the density f can be written as

$$\widehat{f}_{\text{skew-GBS}}(x) = \frac{1}{n}\sum_{i=1}^{n} K_{\text{skew-GBS}(\sqrt{h},\, x,\, \lambda;\, g)}(x_i), \tag{2.9}$$

where $K_{\text{skew-GBS}(\sqrt{h},\, x,\, \lambda;\, g)}(t) = 2c\, g([1/h]\,[t/x+x/t-2])F_S([\lambda/\sqrt{h}][\sqrt{t/x}-\sqrt{x/t}])\, t^{-3/2}[t+x]/\sqrt{4hx}$, c, g and F_S are as given in (2.4), h is the bandwidth, x the point where the density is estimated, and λ the skewness parameter. Note that, as $n \to \infty$, $h \to 0$ and $nh \to \infty$.

Specifically, skew-BS, skew-BS-PE and skew-BS-t kernels are given by

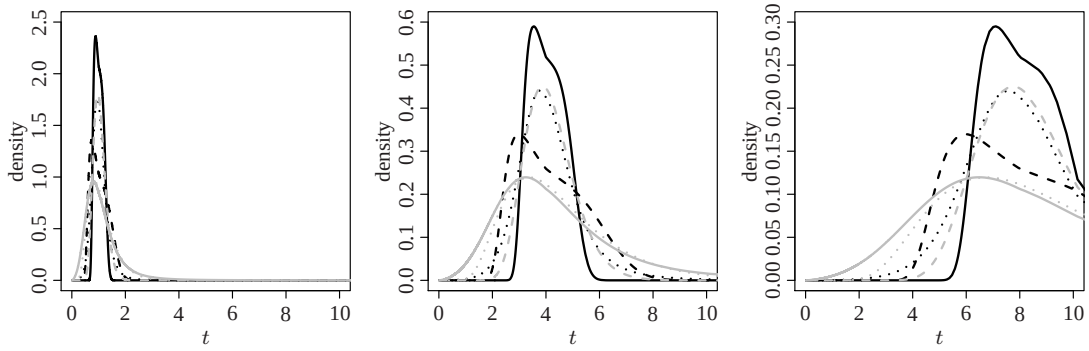
$$\begin{aligned}
K_{\text{skew-BS}(\sqrt{h},\, x,\, \lambda;\, g)}(t) &= \tfrac{2}{\sqrt{2\pi}} \exp\big(-\tfrac{1}{2h}\big[\tfrac{t}{x}+\tfrac{x}{t}-2\big]\big)\Phi\left(\tfrac{\lambda}{\sqrt{h}}\big[\sqrt{t/x}-\sqrt{x/t}\big]\right)\tfrac{t^{-3/2}[t+x]}{\sqrt{4hx}},\\
K_{\text{skew-BS-PE}(\sqrt{h},\, x,\, \lambda;\, g)}(t) &= \tfrac{2\eta}{\Gamma\left(\frac{1}{2\eta}\right)2^{\frac{1}{2\eta}}} \exp\big(-\tfrac{1}{2h^{\eta}}\big[\tfrac{t}{x}+\tfrac{x}{t}-2\big]^{\eta}\big)\\
&\quad \times\big[\tfrac{1}{2}+\tfrac{1}{2\Gamma\left(\frac{1}{2\eta}\right)}\Gamma\big(\tfrac{1}{2\eta},\tfrac{\lambda^{2\eta}}{2h^{\eta}}\big[\tfrac{t}{x}+\tfrac{x}{t}-2\big]^{\eta}\big)\big]\tfrac{t^{-3/2}[t+x]}{\sqrt{4hx}},\\
K_{\text{skew-BS-}t(\sqrt{h},\, x,\, \lambda;\, g)}(t) &= \tfrac{2\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\big[1+\tfrac{1}{\nu h}\big\{\tfrac{t}{x}+\tfrac{x}{t}-2\big\}\big]^{-\frac{\nu+1}{2}}\\
&\quad \times\big[1-\tfrac{1}{2}\mathrm{I}_{\nu/[\nu+\{\lambda^2/h\}\{t/x+x/t-2\}]}\big(\tfrac{\nu}{2},\tfrac{1}{2}\big)\big]\tfrac{t^{-3/2}[t+x]}{\sqrt{4hx}}.
\end{aligned}$$

Figure 2.1 shows two kernels comparative shapes for $h \in \{0.05, 0.2\}$. From this figure, note that, as x increases, the kernels shapes unfold in a similar way, and these kernels are sensitive to the choice of the bandwidth.

2.4 Properties of skew-GBS kernel density estimators

Let $X_1, \ldots, X_n$ be a random sample from a RV X distributed with unknown density f and support on $[0, \infty)$. For this true density f, suppose the following conditions: (C1) it is twice continuously differentiable; (C2) its second derivative is continuous and bounded; (C3) $\int_0^{\infty}[xf'(x)]^2\mathrm{d}x$
$< \infty$; (C4) $\int_0^{\infty}[x^2 f''(x)]^2\mathrm{d}x < \infty$; and (C5) $\int_0^{\infty} x^{-1}\, f(x)\mathrm{d}x < \infty$. Note that conditions (C1) and (C2) are common restrictions in many studies of nonparametric estimators. In particular, condition (C1) is necessary for the Taylor expansion. It is worthwhile to highlight that, assuming a higher order of regularity for f, does not improve the rate of convergence

Figura 2.1: plots of the skew-BS ($h = 0.05$, gray dashed; $h = 0.2$, gray dotted), skew-BS-PE ($h = 0.05$, black solid; $h = 0.2$, black dashed) and skew-BS-t ($h = 0.05$, black dotted; $h = 0.2$, gray solid) kernels for $\eta = 2$, $\nu = 5$, $\lambda = 0.1$ and $x = 1.0$ (a), $x = 4.0$ (b) and $x = 8.0$ (c).



of the skew-GBS kernel estimators, because the bias is of order h. Condition (C2) is quite restrictive, but we think that such a condition can be satisfied by various types of data, for example, environmental data. In addition, conditions (C3), (C4) and (C5) are used for computing the MISE of the proposed kernel estimators, and we think that such conditions can also be satisfied by various types of data. Note that conditions (C1)-(C5) can be verified for BS distributions, which have been successfully applied to environmental data by using theoretical arguments; see, e.g., Vilca et al. (2010). We present some statistical properties of the skew-GBS kernel density estimator in the following propositions, whose proofs are provided in the Appendix.

Proposition 2.4.1 *Let $\widehat{f}_{\text{skew-GBS}}$ be the skew-GBS kernel density estimator. Then, its bias is*

$$\text{Bias}[\widehat{f}_{\text{skew-GBS}}(x)] = h\left[\tfrac{1}{2}xf'(x)\gamma_2 + \tfrac{1}{2}x^2f''(x)\gamma_2\right] + o(h),$$

where γ_2 is given in (2.5)*.*

Proposition 2.4.2 *Let $\widehat{f}_{\text{skew-GBS}}$ be the skew-GBS kernel density estimator. Then, its variance is*

$$\text{Var}[\widehat{f}_{\text{skew-GBS}}(x)] = 2c^2C_\kappa^{-1}n^{-1}h^{-1/2}x^{-1}f(x) + o(n^{-1}h^{-1/2}),$$

where C_κ is a normalization constant such that $\int_{-\infty}^{\infty} 2g^2(z^2)F_S^2(\lambda z)\mathrm{d}z = 1/C_\kappa$.

Corollary 2.4.1 *Let $\widehat{f}_{\text{skew-GBS}}$ be the skew-GBS kernel density estimator. Then, its mean squa-*

red error (MSE) and MISE are, respectively,

$$\begin{aligned}\mathrm{MSE}[\widehat{f}_{\text{skew-GBS}}(x)] &= \tfrac{h^2}{4}\left[xf'(x)\gamma_2 + x^2f''(x)\gamma_2\right]^2 + 2c^2C_\kappa^{-1}n^{-1}h^{-1/2}x^{-1}f(x) \\ &\quad +o(n^{-1}h^{-1/2}+h^2), \\ \mathrm{MISE}[\widehat{f}_{\text{skew-GBS}}(x)] &= \int_0^\infty \mathrm{MSE}[\widehat{f}_{\text{skew-GBS}}(x)]\,\mathrm{d}x = \tfrac{h^2}{4}\gamma_2^2\int_0^\infty \left[xf'(x)+x^2f''(x)\right]^2\mathrm{d}x \\ &\quad +2c^2C_\kappa^{-1}n^{-1}h^{-1/2}\int_0^\infty x^{-1}f(x)\mathrm{d}x + o(n^{-1}h^{-1/2}+h^2).\end{aligned}$$

The value of the bandwidth that minimizes the MISE given in Corollary 2.4.1 is

$$h^*_{\text{skew-GBS}} = \frac{\left[2c^2C_\kappa^{-1}\int_0^\infty x^{-1}f(x)\,\mathrm{d}x\right]^{2/5}}{\left[\gamma_2^2\int_0^\infty\{xf'(x)+x^2f''(x)\}^2\,\mathrm{d}x\right]^{2/5}}n^{-2/5}. \tag{2.10}$$

Note that this optimal bandwidth is of order $O(n^{-2/5})$ for all the kernels. Inserting (2.10) into expression for the MISE given in Corollary 2.4.1, we obtain the optimal MISE as

$$\mathrm{MISE}^*[\widehat{f}_{\text{skew-GBS}}(x)] = \frac{5}{4}\left[\gamma_2^2\int_0^\infty\left\{xf'(x)+x^2f''(x)\right\}^2\mathrm{d}x\right]^{1/5}\left[2c^2C_\kappa^{-1}\int_0^\infty x^{-1}f(x)\mathrm{d}x\right]^{4/5}n^{-4/5}. \tag{2.11}$$

2.5 Identification, selection, validation and computational implementation

Some methods can be used (i) to identify the shape of the hazard rate (HR) function and, consequently, of the parametric distribution of the environmental data; (ii) to select the parameters η, ν and λ of the skew-GBS distributions; and (iii) to choose the bandwidth of the nonparametric method, validating this choice. Then, we discuss a implementation in the R software of the proposed methods; see `www.R-project.org`.

(i) A nice property of the HR is that it allows us to characterize the behavior of distributions. For example, the HR may have several shapes such as increasing, constant, decreasing, bathtub, inverse bathtub approaching to a non-null constant or to zero. An incorrect specification of the HR could have serious consequences in the analysis; see, e.g., Ferreira et al. (2012) and references therein. Specifically, let $h(t) = f(t)/[1-F(t)]$ be the HR of a RV T, where f and F are the density and distribution functions of T, respectively. The identification of the shape of the HR can be done by means of the scaled total time on test (TTT) method, which is given by $W(u) = H^{-1}(u)/H^{-1}(1)$, for $0 \leq u \leq 1$, where $H^{-1}(u) = \int_0^{F^{-1}(u)}[1-F(y)]dy$, and F^{-1} is the inverse function of F. Thus, $W(u)$ can be empirically approximated by constructing the empirical scaled TTT plot by depicting the consecutive points $[k/n, W_n(k/n)]$, with

$W_n(k/n) = [\sum_{i=1}^{k} t_{(i)} + \{n-k\}t_k]/\sum_{i=1}^{n} t_{(i)}$, for $k = 0, \ldots, n$, and $t_{(i)}$ being the observed ith order statistic; see Aarset (1987).

(ii) Consider $Z^2 = [T/\beta + \beta/T - 2]/\alpha^2 \sim \chi^2(1)$, with observations $z_i = [\{t_i/\widehat{\beta}^{(0)}\}^{1/2} - \{\widehat{\beta}^{(0)}/t_i\}^{1/2}]/\widehat{\alpha}^{(0)}$, for $i = 1, \ldots, n$, and starting values for α and β obtained from the modified moment estimates given by $\widehat{\alpha}^{(0)} = [2\{s/r\}^{1/2} - 1]^{1/2}$ and $\widehat{\beta}^{(0)} = [s\,r]^{1/2}$, where $s = [1/n]\sum_{i=1}^{n} t_i$ and $r = 1/[\{1/n\}\sum_{i=1}^{n}\{1/t_i\}]$. The maximum likelihood (ML) estimate of λ can be obtained by the method proposed by Azzalini (1985); see also Vilca et al. (2011). In order to obtain ν, we fix integer values for it within the interval $[1, 100]$, choosing the value of ν that maximizes the likelihood function, searching its optimal value by means of the following algorithm: (step 1) for ν from $\nu = 1$ to $\nu = 100$ by 1, estimate the parameters of the GBS or skew-GBS distributions using the profile ML method at the corresponding value of ν, and compute the value of the likelihood evaluating it at the ML estimates of the GBS or skew-GBS parameters previously obtained, and at the corresponding value of ν used in the profile ML method; and (step 2) choose the value of ν that maximizes the likelihood function and then consider the ML estimates of the GBS or skew-GBS parameters as result. A similar algorithm is used for the optimal searching of η.

(iii) The least squares cross validation (LSCV) method can be used to select the bandwidth h, which chooses it by minimizing the cross validation criterion defined as LSCV$(h) = \int_0^\infty \widehat{f}_h^2(x)$
$\mathrm{d}x - [2/n]\sum_{i=1}^{n}\widehat{f}_{h,-i}(x_i)$, where $\widehat{f}_h$ is a nonnegative asymmetric kernel density estimate of bandwidth h and $\widehat{f}_{h,-i}$ the corresponding estimate without the ith observation; see Rudemo (1982) and Bowman (1984). Note that the plug-in and bootstrapping methods, and the adaptive varying kernel size selection are not possible to adapt, because they use a pilot bandwidth and are symmetric kernel-driven; see Loader (1999) and Jin and Kawczak (2003). The methods proposed in the article, as well as the selection and validation tools discussed above, are implemented in `R` code and available to the interested readers upon request. In this code, we estimate λ using the command `sn.mle` of an `R` package named `sn`, whereas the parameters α, β and ν are determined by the command `mlegbs` of an `R` package named `gbs`; see Barros et al. (2009). In addition, we use the adjusted boxplot, which is a modified version of the usual boxplot for asymmetric data, that was constructed by the command `adjbox` of an `R` package named `robustbase`; see Hubert and Vandervieren (2008). The LSCV and TTT methods are also implemented in the computer code.

2.6 Monte Carlo simulation study

We conducted Monte Carlo (MC) experiments and compute the minimum average integrated squared error (ISE) for evaluating the performance of the BS, BS-PE, BS-t, LN, skew-BS, skew-BS-PE, skew-BS-t kernel estimators. The bandwidth h is chosen by minimizing $\text{ISE}(h) = \int_B [\widehat{f}_h(x) - f(x)]^2 \mathrm{d}x$, where $B \in [0, \infty)$ and $\widehat{f}_h$ is a nonnegative asymmetric kernel density estimate of f. We use $M = 500$ MC replications and sample sizes $n = 100, 200, 500, 1000$ from BS(1.5,1), Burr(1,3,1) and GG(3,1,0.9) (generalized gamma) distributions; see Table 2.2. From this table, note that $U \sim \text{U}(0,1)$, $Z \sim \text{N}(0, 9/16)$ and $Y \sim \text{Gamma}(3,1)$. The GG distribution has, as special cases, the exponential distribution ($\kappa = \theta = 1$), the gamma distribution ($\theta = 1$), the LN distribution ($\kappa \to \infty$), and the Weibull distribution ($\kappa = 1$). Hereafter, we use $\eta = 2$ (BS-PE, skew-BS-PE), $\nu = 5$ (BS-t, skew-BS-t), and $\lambda = 0.1$ (skew-BS-PE, skew-BS-t).

Tabela 2.2: characteristics of the indicated distributions used for generation of random numbers.

Distribution	Density	Transformation
BS($\alpha = 1.5, \beta = 1$)	$\frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{1}{2\alpha^2}\big[\frac{t}{\beta} + \frac{\beta}{t} - 2\big]\Big) \frac{[t+\beta]}{2\alpha\sqrt{\beta t^3}}, \quad t, \alpha, \beta > 0$	$T = 1 + 2Z^2 + 2Z[1+Z^2]^{1/2}$
Burr($\mu = 1, k = 3, r = 1$)	$\frac{kt^{k-1}}{[1+rt^k]^{\frac{1}{r}+1}}, \quad t > 0, k, r > 0$	$T = \Big[\frac{U}{1-U}\Big]^{1/3}$
GG($\kappa = 3, \sigma = 1, \theta = 0.9$)	$\frac{\theta}{\sigma\Gamma(\kappa)}\big[\frac{t}{\sigma}\big]^{\kappa\theta-1} \exp\Big(-\big[\frac{t}{\sigma}\big]^{\theta}\Big), \quad t, \kappa, \sigma, \theta > 0$	$T = Y^{1/0.9}$

Tables 2.3, 2.4 and 2.5 report the minimum average ISEs and bandwidths for the BS(1.5,1), Burr(1,3,1) and GG(3,1,0.9) distributions, respectively. We can observe that the BS and LN kernel estimators display similar performance with a slight advantage of the former over the latter. Another important point is that the skew-BS kernel estimator performs better than its closest competitor, i.e., the BS kernel estimator. An important result from Tables 2.3, 2.4 and 2.5 is that the skew-BS-PE kernel estimator outperforms all the other estimators (except when $n = 500, 1000$ for the GG(3,1,0.9) distribution, i.e., the skew-BS kernel estimator has a superior performance) for the sample sizes considered. Observing the achieved bandwidths in Tables 2.3, 2.4 and 2.5, we see that, in general, and as expected, the bandwidths decrease as the sample size increases. Table 2.6 provides the minimum average ISEs and bandwidths considering different values for the skewness parameter λ, when the data are generated from the Burr(1,3,1) distribution. Note that again the skew-BS-PE kernel estimator outperforms all the other estimators when $\lambda = 0.1$. This suggests that when we introduce a positive skewness on the GBS kernel, the observations located at the tails (left or right) are accommodated in a better way, providing better estimates.

We now give particular attention to the simulated data from the Burr distribution. We generate a 10000 random sample ($t \in [0.04, 17.50]$) from a Burr$(1,3,1)$ distribution, estimate

Tabela 2.3: average ISE (with h in parentheses) of the density estimator for the BS(1.5,1) distribution by using the indicated kernel ($\lambda = 0.1$).

n	BS	LN	BS-PE(5)	BS-$t(2)$	skew-BS	skew-BS-PE(5)	skew-BS-$t(2)$
100	0.021239	0.021430	0.213729	0.217255	0.019382	0.012943	0.018082
	(0.177)	(0.045)	(0.206)	(0.086)	(0.206)	(0.351)	(0.156)
200	0.009271	0.009387	0.080318	0.082340	0.008451	0.006350	0.010531
	(0.138)	(0.037)	(0.167)	(0.068)	(0.156)	(0.282)	(0.110)
500	0.004016	0.004076	0.023754	0.024753	0.003813	0.003266	0.006189
	(0.098)	(0.031)	(0.131)	(0.053)	(0.108)	(0.196)	(0.072)
1000	0.001922	0.001961	0.008262	0.008816	0.001838	0.001678	0.003998
	(0.075)	(0.030)	(0.109)	(0.043)	(0.081)	(0.142)	(0.050)

Tabela 2.4: average ISE (with h in parentheses) of the density estimator for the Burr(1,3,1) distribution by using the indicated kernel ($\lambda = 0.1$).

n	BS	LN	BS-PE(5)	BS-$t(2)$	skew-BS	skew-BS-PE(5)	skew-BS-$t(2)$
100	0.014887	0.015219	0.012650	0.014310	0.013836	0.010706	0.055511
	(0.054)	(0.036)	(0.107)	(0.041)	(0.057)	(0.113)	(0.074)
200	0.008703	0.009003	0.007754	0.008870	0.008209	0.007019	0.052801
	(0.043)	(0.036)	(0.083)	(0.031)	(0.044)	(0.088)	(0.065)
500	0.004346	0.004781	0.003811	0.004532	0.004048	0.003566	0.050176
	(0.033)	(0.032)	(0.057)	(0.025)	(0.031)	(0.061)	(0.058)
1000	0.002744	0.003345	0.002184	0.002655	0.002346	0.002064	0.049065
	(0.030)	(0.030)	(0.043)	(0.026)	(0.029)	(0.047)	(0.056)

Tabela 2.5: average ISE (with average h in parentheses) of the density estimator for the GG(3,1,0.9) distribution by using the indicated kernel ($\lambda = 0.1$).

n	BS	LN	BS-PE(5)	BS-$t(2)$	skew-BS	skew-BS-PE(5)	skew-BS-$t(2)$
100	0.021711	0.021800	0.214910	1.186278	0.019973	0.015753	0.019785
	(0.070)	(0.033)	(0.194)	(0.485)	(0.085)	(0.141)	(0.061)
200	0.013219	0.013291	0.144340	1.031841	0.012058	0.011055	0.014331
	(0.044)	(0.037)	(0.167)	(0.499)	(0.054)	(0.089)	(0.037)
500	0.008232	0.008363	0.109601	0.947764	0.007435	0.007737	0.010525
	(0.032)	(0.031)	(0.205)	(0.500)	(0.035)	(0.055)	(0.024)
1000	0.006866	0.007075	0.098854	0.913920	0.006139	0.006724	0.009256
	(0.030)	(0.030)	(0.239)	(0.500)	(0.030)	(0.049)	(0.036)

the densities and evaluate the performance of each kernel estimator, computing the point-wise bias, variance and MSE. The bandwidths are chosen by minimizing the corresponding ISE. Table 2.7 provides some descriptive measures for Burr$(1, 3, 1)$ data, which include central tendency statistics, the standard deviation (SD), and coefficients of variation (CV), of skewness (CS) and of kurtosis (CK), among others. It is noteworthy that the positive asymmetry and high kurtosis are evidenced by Table 2.3 and Figure 2.2. The TTT plot and the usual and adjusted boxplots displayed in Figure 2.2(a)-(c) show a HR with a unimodal shape and heavy tails.

Figure 2.3 depicts the Burr(1,3,1) true density with BS, BS-PE, skew-BS and skew-BS-PE kernel density estimates. To make the figure less crowded and considering that the BS-t, LN, and skew-BS-t kernel estimators have poorer performances, we omit the results of these kernel estimators throughout this section. The results from Figure 2.3 show an excellent agreement between the true density and its kernel estimates. Figure 2.4 presents the point-

Tabela 2.6: average ISE (with h in parentheses) of the density estimator for the Burr(1,3,1) distribution by using the indicated λ and kernel.

n	λ	skew-BS	skew-BS-PE(2)	skew-BS-$t(5)$
100	0.01	0.014742	0.010814	0.012269
		(0.054)	(0.110)	(0.043)
	0.05	0.014279	0.010746	0.022638
		(0.055)	(0.112)	(0.055)
	0.1	0.013836	0.010706	0.055511
		(0.057)	(0.113)	(0.074)
	0.5	0.015353	0.012195	0.019935
		(0.057)	(0.126)	(0.221)
	1.0	0.022032	0.017358	0.028850
		(0.045)	(0.143)	(0.053)
200	0.01	0.008601	0.007116	0.008148
		(0.042)	(0.085)	(0.034)
	0.05	0.008377	0.007052	0.019062
		(0.043)	(0.086)	(0.046)
	0.1	0.008209	0.007019	0.052801
		(0.044)	(0.088)	(0.065)
	0.5	0.010614	0.008573	0.015187
		(0.040)	(0.102)	(0.168)
	1.0	0.015969	0.013823	0.018211
		(0.030)	(0.120)	(0.042)
500	0.01	0.004166	0.003639	0.004401
		(0.030)	(0.058)	(0.026)
	0.05	0.004078	0.003595	0.015878
		(0.030)	(0.060)	(0.036)
	0.1	0.004048	0.003566	0.050176
		(0.031)	(0.061)	(0.058)
	0.5	0.006226	0.005252	0.010976
		(0.035)	(0.075)	(0.117)
	1.0	0.009765	0.010637	0.009701
		(0.032)	(0.095)	(0.041)
1000	0.01	0.002386	0.002114	0.002722
		(0.025)	(0.044)	(0.024)
	0.05	0.002341	0.002092	0.014562
		(0.025)	(0.046)	(0.033)
	0.1	0.002346	0.002064	0.049065
		(0.029)	(0.047)	(0.056)
	0.5	0.004047	0.003841	0.009219
		(0.061)	(0.061)	(0.094)
	1.0	0.006520	0.009302	0.006060
		(0.043)	(0.082)	(0.045)

Tabela 2.7: descriptive measures for Burr$(1, 3, 1)$ data.

n	Min.	Max.	Median	Mean	SD	CV	CS	CK
10000	0.04	17.50	1.00	1.21	0.95	78.76%	4.84	43.62

wise bias, variance and MSE of the four kernel estimators for the Burr(1,3,1) density, where the bandwidths were selected by minimizing the ISE. We observe that most of the point-wise bias, variance and MSE go to zero, when x is greater than 10. We also note that the considered kernel estimates (except the skew-BS case) present similar results for the bias, and MSE. A relevant bias-variance tradeoff can be observed with the BS kernel estimator, i.e., the variance can be reduced at the expense of the bias, and vice versa; see Figure 2.4(a)-(c).

Figura 2.2: TTT plot (a), histogram (b), and usual and adjusted boxplots (c) of the Burr$(1, 3, 1)$ data.

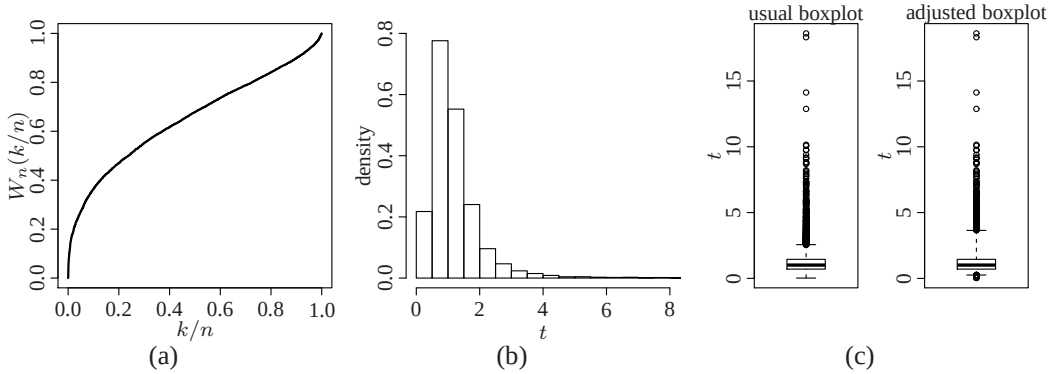


Figura 2.3: estimation of Burr(1,3,1) density (black solid) via BS (black dashed, $h^* = 0.011$), BS-PE (black dotted, $h^* = 0.021$), skew-BS (gray dashed, $h^* = 0.011$), skew-BS-PE (gray dotted, $h^* = 0.023$) kernels for $\eta = 2$ and $\lambda = 0.1$ (a), and zoom on the corresponding left tail (b).

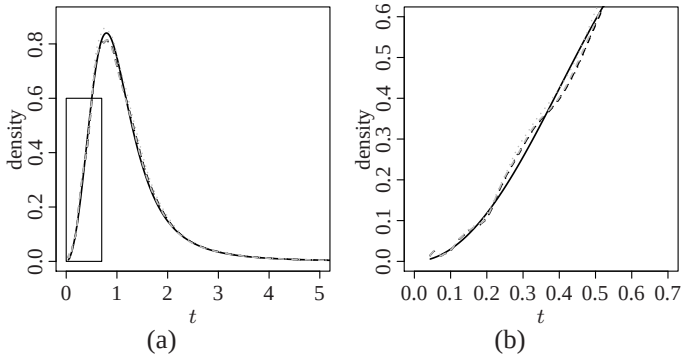
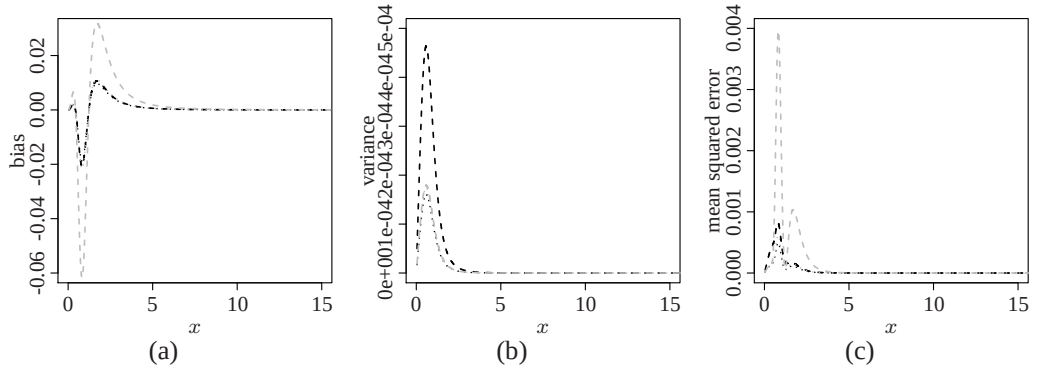


Figura 2.4: point-wise bias (a), variance (b) and MSE (c) of the BS (black dashed, $h^* = 0.011$), BS-PE (black dotted, $h^* = 0.021$), skew-BS (gray dashed, $h^* = 0.011$), skew-BS-PE (gray dotted, $h^* = 0.023$) kernel density estimators for the Burr(1,3,1) distribution.



2.7 Application to real environmental data

We estimate the unknown true density of the three following real environmental data sets.

- O_3: daily tropospheric ozone concentrations (in ppb = ppm $\times$ 1000) observed in New York during May-September, 1973, provided by the New York State Department of Conservation; see Nadarajah (2008), Leiva et al. (2010), Vilca et al. (2011), Ferreira et al. (2012).
- PM_{10}: hourly particulate matter concentrations (in micrometers/normalized cubic meters, μg/m^{3}N) observed in Santiago, Chile, during April, 2003, provided by the Chilean environment commission and metropolitan health authority.
- SO_2: hourly sulfur dioxide concentrations (in ppb = ppm $\times$ 1000) observed in Santiago during March, 2002, provided by the Chilean environment commission; see Vilca et al. (2010).

Table 2.8 provides some descriptive measures for O_3, PM_{10} and SO_2 data. Also, the histograms and (usual and adjusted) boxplots are presented in Figures 2.5, 2.6 and 2.7. The TTT plots suggest that these data have an increasing HR, such as indicated in Vilca et al. (2011); see Figures 2.5(a), 2.6(a) and 2.7(a). The descriptive summaries effectively show a positive skewness and a high kurtosis (which is an indication of heavy tails) for these data sets. This is corroborated by histograms displayed in Figures 2.5(b), 2.6(b) and 2.7(b). Note that the usual boxplots present some atypical observations lying on the right-tail of the distributions of O_3, PM_{10} and SO_2 data. However, it is well known that, in cases where the data follow a skewed distribution, a significant number of observations can be classified as atypical when they are not. Figures 2.5(c), 2.6(c) and 2.7(c) show boxplots that confirm such a situation, i.e., most of the observations considered as potential outliers by the usual boxplot are not outliers when we consider the adjusted boxplot.

Tabela 2.8: descriptive statistics for the indicated data.

Data set	n	Min.	Max.	Median	Mean	SD	CV	CS	CK
O_3	116	1.00	168.00	31.50	42.13	32.99	78.30%	1.21	4.11
PM_{10}	717	1.00	230.00	66.00	71.72	39.30	54.80%	0.64	3.06
SO_2	744	1.00	25.00	2.00	2.93	2.02	68.87%	4.32	37.57

Next, we use kernel estimation for determining O_3, PM_{10} and SO_2 data density. Figures 2.5(b), 2.6(b) and 2.7(b) show the histograms with kernel estimates and bandwidths selected by the LSCV method for O_3, PM_{10} and SO_2 data. Note from these figures that all the density estimates based on the BS, BS-PE, BS-t, LN, skew-BS, skew-BS-PE and skew-BS-t kernels seem to be quite reasonable to the environmental data, despite the difficulty of bandwidth selection; see Loader (1999). We are selecting the bandwidth through the LSCV method, which is somewhat unstable and can underestimate the density. However, as highlighted by Loader (1999), this instability is not a problem, but a symptom stemmed from the difficulty of bandwidth selection. This problem is emphasized in the case of O_3 and PM_{10}

data, because, in the first bin of the corresponding histograms, few observations are registered, but it is not the case of SO_2 data.

Figura 2.5: TTT plot (a), density estimation via BS (black solid, $h^* = 0.206$), BS-PE (black dotted, $h^* = 0.368$), BS-t (black dotdash, $h^* = 0.159$), LN (black dashed, $h^* = 0.051$), skew-BS (gray dotted, $h^* = 0.219$), skew-BS-PE (gray solid, $h^* = 0.370$), and skew-BS-t (gray dashed, $h^* = 0.159$) kernels for $\eta = 2$, $\nu = 5$ and $\lambda = 0.1$ (b), and usual and adjusted boxplots (c) of O_3 data.

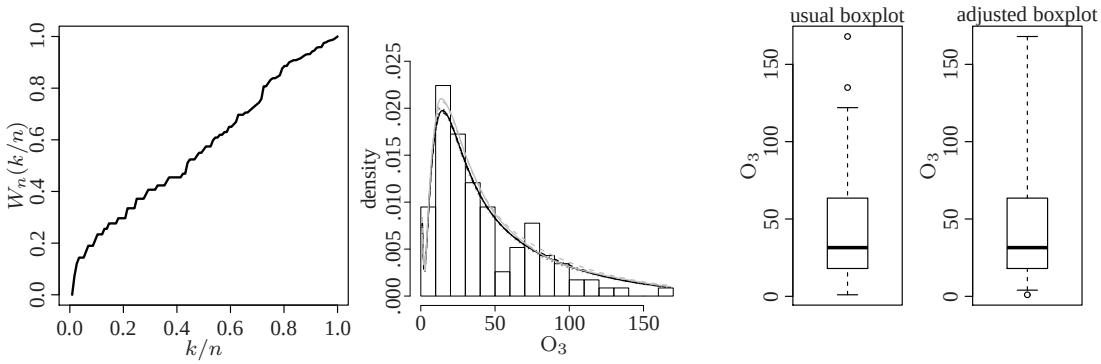
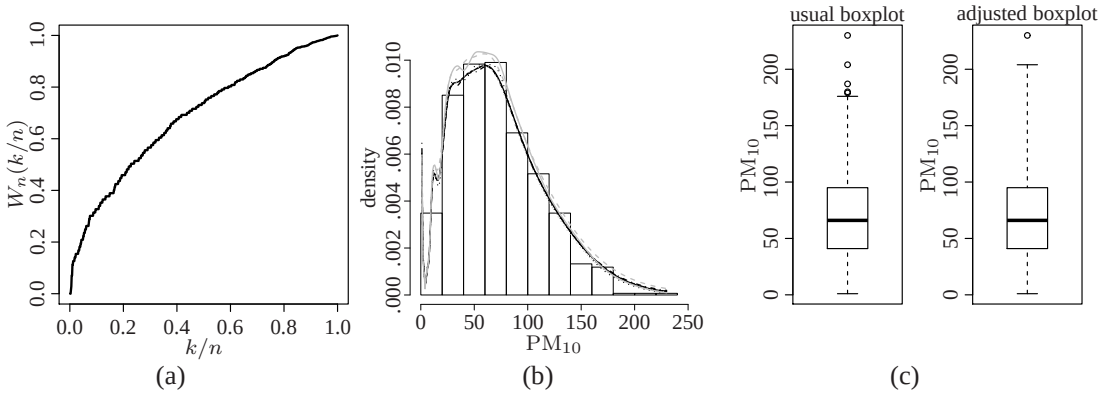


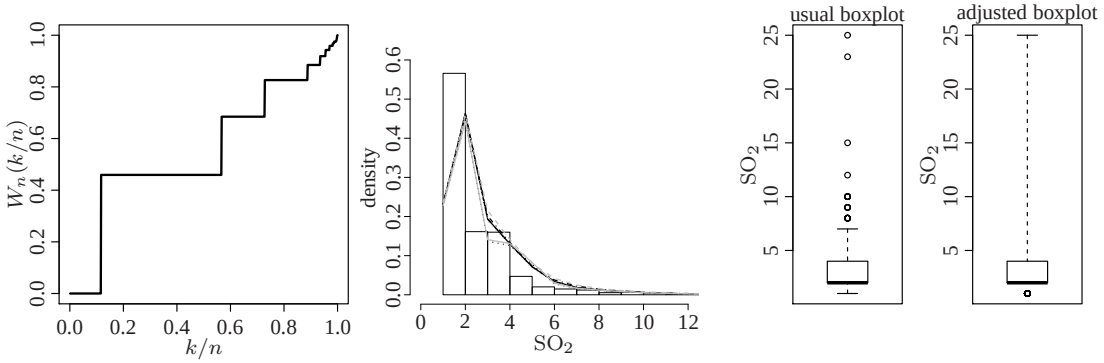
Figura 2.6: TTT plot (a), density estimation (b) via BS (black solid, $h^* = 0.034$), BS-PE (black dotted, $h^* = 0.046$), BS-t (black dotdash, $h^* = 0.027$), LN (black dashed, $h^* = 0.009$), skew-BS (gray dotted, $h^* = 0.035$), skew-BS-PE (gray solid, $h^* = 0.049$), and skew-BS-t (gray dashed, $h^* = 0.032$) kernels for $\eta = 2$, $\nu = 5$ and $\lambda = 0.1$, and usual and adjusted boxplots (c) of PM_{10} data.



2.8 Concluding remarks

In this chapter, we have introduced new asymmetric kernel density estimators based on generalized skew-Birnbaum-Saunders distributions. These distributions provide densities with either heavier or lighter tails and more or less pronounced asymmetry than the Birnbaum-Saunders distribution. The kernel estimators that we have proposed are boundary

Figura 2.7: TTT plot (a), density estimation via BS (black solid, $h^* = 0.040$), BS-PE (black dotted, $h^* = 0.056$), BS-t (black dotdash, $h^* = 0.039$), LN (black dashed, $h^* = 0.010$), skew-BS (gray dotted, $h^* = 0.041$), skew-BS-PE (gray solid, $h^* = 0.056$), and skew-BS-t (gray dashed, $h^* = 0.039$) kernels for $\eta = 2$, $\nu = 5$ and $\lambda = 0.1$ (b), and usual and adjusted boxplots (c) of SO_2 data.



(a) (b) (c)

bias free, with shapes that vary according to the data point location, and achieving the optimal rate of convergence for the mean integrated squared error of nonnegative asymmetric kernel density estimators. In addition, we have presented some identification, selection and validation methods. A computational implementation in the R statistical software of the results derived in this work has been discussed. Furthermore, we have compared the proposed density estimators based on generalized skew-Birnbaum-Saunders kernels with some nonparametric density estimators existing in the literature. Monte Carlo simulation results have indicated that the skew-Birnbaum-Saunders-power-exponential kernel density estimator we have derived outperforms most other estimators, in all sample sizes considered. We have used our methodology for estimating the density of real air pollutant concentration data. The numerical results have shown the flexibility and good performance of the proposed nonparametric estimators.

3 A NEW FAMILY OF AUTOREGRESSIVE CONDITIONAL DURATION MODELS

3.1 Introduction

High-frequency data have gained an increasing attention with the advances in computer technology and data recording and storage; see Engle (2000). The availability of high-frequency financial transaction (trade duration –TD–) data has given an impulse to the research in business, economics and finance. The autoregressive conditional duration (ACD) model proposed by Engle and Russell (1998) has been the primary class of models for analyzing TD data, which are irregularly time-spaced and convey meaningful information. The importance of this type of data, and of its modeling, is stressed by the relatively recent market microstructure literature; see Diamond and Verrechia (1987), Easley and O'Hara (1992), Easley et al. (1997), Meitz and Teräsvirta (2006) and Pacurar (2008) .

Although TD data (i) have usually a unimodal hazard rate (HR) and (ii) follow an asymmetric distribution with heavy tails (see GRAMMIG and MAURER, 2000; BHATTI, 2010), generalizations of the original ACD model are based on assumptions that do not necessarily comply with (i) and (ii). Thus, generalizations of the ACD model should be based on assumptions that take into account (A1) the shape of the HR of TD data; (A2) the symmetry or asymmetry of the distribution of these data; and (A3) the conditional dynamics established in terms of the mean or median of TD data, depending of their symmetry or asymmetry; see Bauwens and Giot (2000), Luca and Zuccolotto (2006), Fernandes and Grammig (2006) and Allen et al. (2008).

Birnbaum and Saunders (1969) introduced a two-parameter distribution for modeling failure times of a material exposed to fatigue. They assumed that the failure due to fatigue follows from the development and growth of a dominant crack produced by cyclic stress. The Birnbaum-Saunders (BS) distribution has been widely studied because of its interesting properties and its relation with the normal distribution; see Kotz et al. (2010). In addition,

although it has its genesis from engineering, its applications have been considered in several other fields, including business, economics and finance; see, e.g., Jin and Kawczak (2003), Ahmed et al. (2010), Bhatti (2010), Paula et al. (2012), Marchant et al. (2013) and Leiva et al. (2013).

The BS distribution is asymmetrical, has positive skewness and a unimodal HR, and has been successfully applied to model lifetime data. Thus, it can be a good model for describing TD data. Bhatti (2010) proposed a generalization of the ACD model based on the BS distribution (in short BS-ACD model), which provides (B1) a realistic distributional assumption (in terms of the shape of its probability density function, in short PDF, and of its HR); (B2) an easy parameter estimation (because it is simple and converges fast); and (B3) a natural parameterization in terms of a conditional median duration, which is expected to improve the model fit, instead of using the conditional mean duration; see (A3). This is because the median is often considered as a better measure of central tendency than the mean, when the data follow asymmetric and heavy-tailed distributions, such as is the case of TD data.

Recently, based on the relationship between the BS and normal distributions, Balakrishnan et al. (2009) introduced the scale-mixture BS (SBS) distributions; see also Díaz-García and Leiva (2005), and Marchant et al. (2013) for a recent TD analysis using kernel estimation based on SBS models, with independent data. The class of SBS distributions (C1) inherits the good properties of the BS distribution discussed in (B1)-(B3); (C2) permits the maximum likelihood (ML) estimates of the model parameters to be computed in an efficient way, using the expectation-maximization (EM) algorithm; and (C3) allows a robust estimation procedure of parameters to be obtained, which is not possible with the BS distribution; among other properties.

The main aim of this work is to propose a new methodology based on ACD models generated from SBS distributions, in short SBS-ACD. This methodology includes efficient estimation of the SBS-ACD model parameters via the EM algorithm, inference about these parameters, the predictive model and a residual analysis for model checking. SBS-ACD models should hold with the properties of the BS-ACD model, but further properties should be also obtained. We apply the new methodology to TD data, which have unique features absent in data with low frequencies. For example, as mentioned, TD data (D1) are collected in irregular time intervals; (D2) possess a large number of observations; (D3) their trading activities exhibit some diurnal pattern, that is, activity is higher near the beginning and closing than in the middle of the trading day; and (D4) present a unimodal HR; see Engle and Russell (1998) and Bhatti (2010).

The rest of the chapter unfolds as follows. In Section 3.2, we introduce SBS distributions. In Sections 3.3 and 3.4, we propose a methodology based on the SBS-ACD models, which

includes formulation, estimation of their parameters by means of the EM algorithm, the corresponding inference, which is useful for obtaining the predictive model, and a residual for this model, which can be used for checking model adequacy in practice. In Section 3.5, we perform a Monte Carlo (MC) study for evaluating the proposed methodology. In Section 3.6, we present an application of this methodology to six real data sets of NYSE securities. Finally, in Section 3.7, we discuss some conclusions and futures studies.

3.2 Scale-mixture Birnbaum-Saunders distributions

In this section, we present some useful results about SBS distributions.

A random variable (RV) X follows a two-parameter BS distribution if it can be represented by

$$X = \sigma\big[\kappa Z/2 + (\{\kappa Z/2\}^2 + 1)^{1/2}\big]^2, \tag{3.1}$$

where $Z \sim \mathrm{N}(0,1)$, and $\kappa > 0$, $\sigma > 0$ are shape and scale parameters, respectively.

In this case, the notation $X \sim \mathrm{BS}(\kappa, \sigma)$ is used and the corresponding PDF is given by

$$f_{\mathrm{BS}}(x;\kappa,\sigma) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2\kappa^2}\left[\frac{x}{\sigma} + \frac{\sigma}{x} - 2\right]\right) \frac{x^{-3/2}[x+\sigma]}{2\kappa\sigma^{1/2}}, \quad x > 0.$$

Note that, as the shape parameter κ goes to zero, the BS distribution tends to be symmetrical, degenerating at σ, when $\kappa = 0$ (see KUNDU et al., 2008), whereas the scale parameter σ is also the median of the distribution. The BS model holds proportionality and reciprocal properties given by $b\,X \sim \mathrm{BS}(\kappa, b\,\sigma)$, with $b > 0$, and $1/X \sim \mathrm{BS}(\kappa, 1/\sigma)$, respectively.

A RV Y follows a scale mixture of normal (SMN) distribution (symmetric) if it can be represented by $Y = \mu + \{g(U)\}^{1/2}V$, where $V \sim \mathrm{N}(0, \vartheta^2)$, U is a positive RV independent of V, with cumulative distribution function (CDF) $H(\cdot)$, and $g(\cdot)$ is a positive function associated with $H(\cdot)$. This is denoted by $Y \sim \mathrm{SMN}(\mu, \vartheta^2, H)$ and the corresponding PDF is given by

$$\phi_{\mathrm{SMN}}(y;\mu,\vartheta,H) = \int_0^\infty \phi(y;\mu, g(u)\vartheta^2)\,\mathrm{d}H(u), \tag{3.2}$$

where $\phi(\cdot;\mu, g(u)\vartheta^2)$ denotes the PDF of the normal distribution with mean μ and variance $g(\cdot)\vartheta^2$. Note that the distribution of Y given U is $Y|U = u \sim \mathrm{N}(\mu, g(u)\vartheta^2)$.

Following Díaz-García and Leiva (2005), Balakrishnan et al. (2009) obtained a class of BS distributions replacing Z in (3.1) by $Y = \{g(U)\}^{1/2}Z \sim \mathrm{SMN}(0,1,H) \equiv \mathrm{SMN}(H)$, with $Z \sim \mathrm{N}(0,1)$, such that

$$X = \sigma\big[\kappa\{g(U)\}^{1/2}Z/2 + \{(\kappa\{g(U)\}^{1/2}Z/2)^2 + 1\}^{1/2}\big]^2. \tag{3.3}$$

Thus, the RV X given in (3.3) has now a SBS distribution, which is denoted by $X \sim \text{SBS}(\kappa, \sigma, H)$, with PDF

$$f_{\text{SBS}}(x;\kappa,\sigma,H) = \phi_{\text{SMN}}\big(a(x;\kappa,\sigma);0,1,H\big)A(x), \quad x>0,\kappa>0,\sigma>0, \tag{3.4}$$

where $\phi_{\text{SMN}}(\cdot)$ is as given in (3.2), with $\mu = 0$ and $\vartheta^2 = 1$,

$$a(x;\kappa,\sigma) = a(x) = \frac{1}{\kappa}\left[\{x/\sigma\}^{1/2} - \{\sigma/x\}^{1/2}\right] \text{ and } A(x) = \frac{\text{d}}{\text{d}x}a(x) = \frac{x^{-3/2}[x+\sigma]}{2\kappa\sigma^{1/2}}. \tag{3.5}$$

Then, the CDF of X is $F_{\text{SBS}}(x;\kappa,\sigma,H) = \Phi_{\text{SMN}}(a(x))$, where $\Phi_{\text{SMN}}(\cdot)$ is the CDF associated with the PDF given in (3.2). From the SBS CDF, we can obtain the quantile function (QF) of X as $x_{\text{SBS}}(q;\kappa,\sigma,H) = F^{-1}_{\text{SBS}}(q;\kappa,\sigma,H) = \sigma[\kappa y(q)/2 + \{\big(\kappa y(q)/2\big)^2+1\}^{1/2}]^2$, for $0<q<1$, where $F^{-1}_{\text{SBS}}(\cdot)$ is the inverse SBS CDF and $y(q)$ is the qth quantile of the SMN distribution. Note that $x_{\text{SBS}}(0.5;\kappa,\sigma,H) = \sigma$, because, due to the symmetry of the SMN distributions, $y(0.5) = 0$, so that the parameter σ is the median of the distribution. The mean, variance and coefficients of skewness (CS) and kurtosis (CK) of X are

$$\begin{aligned}
\text{E}[X] &= \frac{\sigma}{2}\left[2+w_1\kappa^2\right], \text{V}[X] = \frac{\sigma^2\kappa^2}{4}\left[w_1 + \{2w_2 - w_1^2\}\kappa^2\right], \\
\text{CS}[X] &= \frac{4\kappa\left[\{3w_2 - 3w_1^2\} + \frac{1}{2}\{2w_3 - 3w_1w_2 + w_1^3\}\kappa^2\right]}{[4w_1 + \{2w_2 - w_1^2\}\kappa^2]^{3/2}} \quad \text{and} \\
\text{CK}[X] &= \frac{16w_2 + [32w_3 - 48w_1w_2 + 24w_1^3]\kappa^2 + [8w_4 - 16w_1w_3 + 12w_1^2w_2 - 3w_1^4]\kappa^4}{[4w_1 + \{2w_2 - w_1^2\}\kappa^2]^2},
\end{aligned}$$

respectively, where $w_r = \text{E}[\{g(U)\}^r]$, with $g(U)$ being as given in (3.3).

Note that, if $X \sim \text{SBS}(\kappa,\sigma,H)$, its distribution conditional to U is $X|U = u \sim \text{BS}(\{g(u)\}^{1/2}$
$\kappa,\sigma)$, that is, $X|U$ follows a classical BS distribution. Thus, from this result, we can obtain the conditional distribution of $U|X$, allowing the expected value $\text{E}[1/g(U)|X=x]$ to be determined, which is useful for implementing the EM algorithm employed in Section 3.4 of this work.

Three members of the SBS family are the BS (classical), power exponential BS (BS-PE) and Student-t BS (BS-t) distributions; see Table 3.1 for explicit expressions of their PDFs. To obtain the BS distribution, it is assumed in (3.3) that $g(U) = 1$, that is, U is a RV degenerate at this value. For the BS-t distribution, it is assumed in (3.3) that $g(U) = 1/U$ and $U \sim \text{Gamma}(\nu/2,\nu/2)$, whereas the BS-PE distribution is obtained when $U \sim \text{St}(\varrho)$, for $0.5 < \varrho \leq 1$, that is, when U follows a positive stable (St) distribution with index 2ϱ, whose PDF cannot be analytically expressed, but it is denoted by $p_{2\varrho}(\cdot)$, and, in this case, $g(U)$ does

not have a specific form; see details in West (1987) and Ferreira et al. (2011). In addition, note that BS-PE and BS-t distributions reduce to the BS case when $\varrho = 1$ and $\nu \to \infty$, respectively (see FERREIRA et al., 2011; AZEVEDO et al., 2012), whereas the BS-PE and BS-t distributions reduce to the BS-Laplace and BS-Cauchy cases when $\varrho = 1/2$ and $\nu = 1$, respectively; see Barros et al. (2009). BS-PE and BS-t distributions are often considered as alternative models to the BS distribution due to the BS-t distribution has a kurtosis level greater (heavier tails) than the BS distribution, whereas the BS-PE distribution has either less or greater (lighter or heavier tails) kurtosis levels than the BS distribution, which is useful for accommodating different types of data. Furthermore, the BS-t distribution provides a robust estimation procedure of its parameters; see Paula et al. (2012).

Tabela 3.1: associated PDFs of some members of the SBS family.

Family member	PDF associated with $H(u)$	PDF of $X \sim \text{SBS}(\kappa, \sigma, H)$
BS	1	$\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2\kappa^2}\left[\frac{x}{\sigma} + \frac{\sigma}{x} - 2\right]\right) \frac{x^{-3/2}[x+\sigma]}{2\kappa\sqrt{\sigma}}$, $x > 0$
BS-PE	$p_{2\varrho}(u)$, $u > 0,\ 1/2 < \varrho \leq 1$	$\frac{\varrho}{\sqrt{2^{\varrho}}\Gamma\left(\frac{1}{2\varrho}\right)} \exp\left(-\frac{1}{2\kappa^{2\varrho}}\left[\frac{x}{\sigma} + \frac{\sigma}{x} - 2\right]^{\varrho}\right) \frac{x^{-3/2}[x+\sigma]}{2\kappa\sqrt{\sigma}}$, $x > 0$
BS-t	$\frac{[\frac{\nu}{2}]^{\frac{\nu}{2}} u^{\frac{\nu}{2}-1}}{\Gamma(\frac{\nu}{2})} \exp\left(-\frac{\nu u}{2}\right)$, $u > 0, \nu > 0$	$\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi}\sqrt{\nu}\Gamma(\frac{\nu}{2})} \left[1 + \frac{1}{\nu\kappa^2}\left\{\frac{x}{\sigma} + \frac{\sigma}{x} - 2\right\}\right]^{-\frac{\nu+1}{2}} \frac{x^{-3/2}[x+\sigma]}{2\kappa\sqrt{\sigma}}$, $x > 0$

3.3 SBS-ACD models

Let $X_i = T_i - T_{i-1}$ denote the TD, that is, the time elapsed between two consecutive financial transactions, $(i-1)$th and ith transactions say, at times T_i and T_{i-1}, respectively. Considering serial dependence usually found in financial duration data, Engle and Russell (1998) assumed that this dependence can be described by a function $\psi_i = \mathrm{E}[X_i|\Omega_{i-1}]$, denoting the conditional mean of the ith TD based on the past information set Ω_{i-1}, which includes all information available at time T_{i-1} (past).

The usual ACD class of models assumes that

$$X_i = \psi_i \, \varepsilon_i, \tag{3.6}$$

where $\psi_i = \alpha + \sum_{j=1}^{r} \beta_j \psi_{i-j} + \sum_{j=1}^{s} \gamma_j X_{i-j}$, for $i = 1, \ldots, n$, which is called the ACD(r, s) model, where r and s refer to the orders of the lags, and $\{\varepsilon_i; i = 1, \ldots, n\}$ is a sequence of independent and identically distributed positive RVs with PDF $f(\cdot)$.

Note that a wide range of ACD(r, s) models may be defined by switching the distribution of ε_i and the specification of ψ_i given in (3.6); see Fernandes and Grammig (2006) and

Pathmanathan et al. (2009). For example, an ACD(r, s) model based on the generalized gamma (GG) distribution with PDF given by

$$f_{\mathrm{GG}}(x;\upsilon,\eta,\omega)=\frac{\eta}{\omega\Gamma(\upsilon)}\left[\frac{x}{\omega}\right]^{\upsilon\eta-1}\exp\Big(-\left[\frac{x}{\omega}\right]^{\eta}\Big),\quad x,\upsilon,\eta,\omega>0, \tag{3.7}$$

can be formulated, with shape (υ, η) and scale (ω) parameters and $\Gamma(x)=\int_0^\infty x^{\upsilon-1}\exp(-x)\mathrm{d}x$ being the gamma function.

Special cases of the GG distribution are the exponential ($\upsilon=\eta=1$), gamma ($\eta=1$), lognormal ($\upsilon\to\infty$) and Weibull ($\upsilon=1$) distributions. Thus, a GG-ACD(r, s) model can be obtained by the conditional mean of the ith TD, $\mathrm{E}[X_i|\Omega_{i-1}]=\psi_i=\omega\Gamma(\upsilon+1/\eta)/\Gamma(\upsilon)$ say, so that one can reparameterize the GG distribution by $\omega=\psi_i\Gamma(\upsilon)/\Gamma(\upsilon+1/\eta)=\psi_i\varphi(\upsilon,\eta)$, obtaining from (3.7) the PDF

$$f_{\mathrm{GG}}(x_i;\upsilon,\eta,\psi_i)=\frac{\eta}{\varphi(\upsilon,\eta)\psi_i\Gamma(\upsilon)}\left[\frac{x_i}{\varphi(\upsilon,\eta)\psi_i}\right]^{\upsilon\eta-1}\exp\left(-\left[\frac{x_i}{\varphi(\upsilon,\eta)\psi_i}\right]^{\eta}\right),\quad i=1,\ldots,n, \tag{3.8}$$

for $x_i,\upsilon,\eta,\psi_i>0$. Note that this way of representing the PDF given in (3.8) is equivalent to the model given (3.6), with a dynamic structure for $\mathrm{E}[X_i|\Omega_{i-1}]=\psi_i$ expressed in a general fashion, which must be specified.

An alternative approach to the existing ACD models was proposed by Bhatti (2010). This approach takes into account the scale parameter of the BS distribution to specify the BS-ACD(r, s) model in terms of a time-varying conditional median duration, $\sigma_i=F^{-1}_{\mathrm{BS}}(0.5|\Omega_{i-1})$ say, where $F^{-1}_{\mathrm{BS}}(\cdot)$ denotes the inverse CDF or QF of the BS distribution. This specification has several advantages over the existing ACD models, as mentioned in (B1)-(B3). From Table 3.1, the PDF associated with the BS-ACD(r, s) model can be obtained in an analogous way to that provided for the GG-ACD(r, s) model given in (3.8) as

$$f_{\mathrm{BS}}(x_i;\kappa,\sigma_i)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2\kappa^2}\left[\frac{x_i}{\sigma_i}+\frac{\sigma_i}{x_i}-2\right]\right)\frac{x_i^{-3/2}[x_i+\sigma_i]}{2\kappa\sigma_i^{1/2}},\ x_i,\kappa,\sigma_i>0,\ i=1,\ldots,n, \tag{3.9}$$

We extend the class of BS-ACD(r, s) models by using SBS distributions. As mentioned in (C1)-(C3), these distributions have good properties. Such as in the BS-ACD(r, s) case, the SBS-ACD(r, s) model is not an ACD(r, s) model in the usual sense. This model is specified in terms of a time-varying conditional median, $\sigma_i=F^{-1}_{\mathrm{SBS}}(0.5|\Omega_{i-1})$ say, which also corresponds to the scale parameter of the SBS(κ,σ_i,H) distribution. Specifically, from the PDF given in (3.4), the SBS-ACD(r, s) model can be obtained in an analogous way to that

provided for the BS-ACD(r, s) model given in (3.9) with an associated PDF expressed as

$$f_{\text{SBS}}(x_i;\kappa,\sigma_i,H)=\phi_{\text{SMN}}(a(x_i);0,1,H)A(x_i),\quad x_i,\kappa,\sigma_i>0,\ \ i=1,\ldots,n, \tag{3.10}$$

where $a(\cdot)$ and $A(\cdot)$ are as given in (3.5). Note that (3.10) is equivalent to the formulation $X_i=\sigma_i\,\epsilon_i$, such that $Y_i=\log(\sigma_i)+\varepsilon_i$, where $Y_i=\log(X_i)$, $\varepsilon_i=\log(\epsilon_i)$ and

$$\sigma_i=\exp\left(\alpha+\sum_{j=1}^{r}\beta_j\log(\sigma_{i-j})+\sum_{j=1}^{s}\gamma_j\frac{x_{i-j}}{\sigma_{i-j}}\right), \tag{3.11}$$

with $X_i \overset{\text{ind}}{\sim}$ SBS(κ,σ_i,H) and $\varepsilon_i \overset{\text{ind}}{\sim}$ SBS$(\kappa,1,H)$, for $i=1,\ldots,n$; see Bauwens and Giot (2000). From Table 3.1, the PDFs associated with the BS-PE-ACD(r, s) and BS-t-ACD(r, s) models are given by

$$f_{\text{BS-PE}}(x_i;\kappa,\sigma_i,\varrho)=\frac{\varrho}{\sqrt{2^{\varrho}}\,\Gamma\left(1/[2\varrho]\right)}\exp\left(-\frac{1}{2\kappa^{2\varrho}}\left[\frac{x_i}{\sigma_i}+\frac{\sigma_i}{x_i}-2\right]^{\varrho}\right)\frac{x_i^{-3/2}[x_i+\sigma_i]}{2\kappa\sqrt{\sigma_i}}\quad\text{and}$$

$$f_{\text{BS-}t}(x_i;\kappa,\sigma_i,\nu)=\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi}\sqrt{\nu}\Gamma(\frac{\nu}{2})}\left[1+\frac{1}{\nu\kappa^2}\left\{\frac{x_i}{\sigma_i}+\frac{\sigma_i}{x_i}-2\right\}\right]^{-\frac{\nu+1}{2}}\frac{x_i^{-3/2}[x_i+\sigma_i]}{2\kappa\sqrt{\sigma_i}},$$

respectively, for $x_i,\kappa,\nu,\sigma_i>0, 1/2<\varrho\leq 1$, and $i=1,\ldots,n$.

3.4 Estimation, inference and checking

In this section, we use the ML method to estimate the SBS-ACD(r, s) parameters. EM algorithm is used for facilitating the implementation of this method. Inference and model checking are also considered here.

Let $\boldsymbol{X}=(X_1,\ldots,X_n)^\top$ be a sample from $X_i\sim\text{SBS}(\kappa,\sigma_i,H)$, for $i=1,\ldots,n$, with PDF as given in (3.10), and $\boldsymbol{x}=(x_i,\ldots,x_n)^\top$ be its observed TDs. As mentioned, we assume a time-varying conditional median duration given as in (3.11). Such as in Engle and Russell (1998), the simplest member of the ACD family, and often useful for empirical applications, is the SBS-ACD($r=1,s=1$) model, that we abbreviate as SBS-ACD model. Thus, in the sequel, any ACD($r=1,s=1$) model is simply denoted as ACD model. It is worthwhile to highlight that autoregressive structures of higher order do not increase the distributional fit of the residuals; see Bhatti (2010). We estimate the parameters of the SBS-ACD model by maximizing the log-likelihood function for $\boldsymbol{\xi}=(\boldsymbol{\delta}^\top,\boldsymbol{\zeta}^\top)^\top$ obtained from

(3.10) and expressed as

$$\ell_{\mathrm{SBS-ACD}}(\boldsymbol{\xi};\boldsymbol{x}) = \sum_{i=1}^{n}\left\{\log(\phi_{\mathrm{SMN}}(a(x_i;\kappa,\sigma_i);0,1,H)) - \frac{3}{2}\log(x_i) + \log\left(\frac{x_i+\sigma_i}{2\,\kappa\,\sqrt{\sigma_i}}\right)\right\}, \tag{3.12}$$

with $\log(\sigma_i) = \alpha + \beta\log(\sigma_{i-1}) + \gamma[x_{i-1}/\sigma_{i-1}]$, where $\boldsymbol{\delta} = (\alpha,\beta,\gamma,\kappa)^\top$ and $\boldsymbol{\zeta}$ denotes the additional parameters corresponding to the PDF given in (3.10) associated with $H(\cdot)$. Note that the vector of parameters $\boldsymbol{\zeta}$ that indexes the PDF associated with $H(\cdot)$ can be assumed to be known or obtained from the data. We select the log-linear form σ_i given as in (3.11) with $r = 1$ and $s = 1$, because it allows an unconstrained parameter estimation to be considered. Otherwise, one should add constrains in the parameter estimation, so that one arrives at a local extreme with positive parameter values; see Bhatti (2010).

The EM algorithm is a widely applicable approach for iterative computation of ML estimates, useful when unobserved data or latent variables are present. Because this algorithm encompasses the expectation step (E-step) and the maximization step (M-step), Dempster et al. (1977) coined this name in their seminal paper. According to Ferreira et al. (2011), the EM algorithm is popular due to (i) its computational simplicity in the M-step, because it involves only complete data ML estimation; and (ii) its stable and straightforward implementation, because the iterations converge monotonically and there are no need for second derivatives. Specifically, let $\boldsymbol{x} = (x_1,\ldots,x_n)^\top$ and $\boldsymbol{u} = (u_1,\ldots,u_n)^\top$ denote observed and unobserved data, respectively, and $\boldsymbol{X}$ and $\boldsymbol{U}$ their corresponding random vectors. Then $\boldsymbol{x}_c = (\boldsymbol{x}^\top,\boldsymbol{u}^\top)^\top$ stands for the original data $\boldsymbol{x}$ augmented with $\boldsymbol{u}$. In general, the two steps of the EM algorithm are:

Algorithm 1 EM algorithm

1: **E-step.** Compute $Q(\boldsymbol{\delta}|\widehat{\boldsymbol{\delta}}^{(r)})$, for $r = 1,2,\ldots$; and
2: **M-step.** Find $\boldsymbol{\delta}^{(r+1)}$ such that $Q(\boldsymbol{\delta}^{(r+1)}|\widehat{\boldsymbol{\delta}}^{(r)}) = \max_{\boldsymbol{\delta}\in\boldsymbol{\Delta}} Q(\boldsymbol{\delta}|\widehat{\boldsymbol{\delta}}^{(r)})$, for $r = 1,2,\ldots$, where $Q(\boldsymbol{\delta}|\widehat{\boldsymbol{\delta}}^{(r)})$ is the expectation of the complete log-likelihood function conditioned to the observed data $\boldsymbol{x}$ and evaluated at the rth estimation of $\boldsymbol{\delta}$.

From Theorem 4 in Balakrishnan et al. (2009), note that

$$\begin{aligned} X_i|(U_i = u_i) &\overset{\text{ind}}{\sim} \mathrm{BS}(\sqrt{g(u_i)}\,\kappa,\sigma_i), \\ U_i &\overset{\text{ind}}{\sim} H(u_i), \quad i = 1,\ldots,n. \end{aligned} \tag{3.13}$$

Thus, the complete data log-likelihood function for SBS-ACD models, associated with $\boldsymbol{x}_c =$

$(\boldsymbol{x}^\top, \boldsymbol{u}^\top)^\top$, taken from (3.12) and (3.13), is given by

$$\ell(\boldsymbol{\xi}; \boldsymbol{x}_c) = c_1 - \frac{1}{2\kappa^2}\sum_{i=1}^{n}\frac{1}{g(u_i)}\left[\frac{x_i}{\sigma_i} + \frac{\sigma_i}{x_i} - 2\right] + \sum_{i=1}^{n}\log\left(x_i + \sigma_i\right) - n\log(\kappa) - \frac{n}{2}\log(\sigma_i), \tag{3.14}$$

where c_1 is a constant that depends on known values and on $\boldsymbol{\zeta}$, but does not on $\boldsymbol{\delta}$.

In order to implement Algorithm 1, we need to state the conditional expectation of the complete log-likelihood function given in (3.14), $Q(\boldsymbol{\delta}|\widehat{\boldsymbol{\delta}})$ say, as

$$Q(\boldsymbol{\delta}|\widehat{\boldsymbol{\delta}}) = c_1 - \frac{1}{2\kappa^2}\sum_{i=1}^{n}\widehat{u}_i\left[\frac{x_i}{\sigma_i} + \frac{\sigma_i}{x_i} - 2\right] + \sum_{i=1}^{n}\log\left(x_i + \sigma_i\right) - n\log(\kappa) - \frac{n}{2}\log(\sigma_i), \tag{3.15}$$

where σ_i is as given in (3.11) with $r = 1$ and $s = 1$, and

$$\widehat{u}_i = \mathrm{E}\left[\frac{1}{g(U_i)}\Big| X_i = x_i, \boldsymbol{\delta} = \widehat{\boldsymbol{\delta}}\right], \quad i = 1, \ldots, n. \tag{3.16}$$

Specific expressions for $Q(\boldsymbol{\delta}|\widehat{\boldsymbol{\delta}})$ given in (3.5) depend on each SBS-ACD model considered. Thus, the log-likelihood function for the BS-PE-ACD model, $\ell_{\text{BS-PE-ACD}}(\boldsymbol{\delta}; \boldsymbol{x}_c)$ say, is obtained in an analogous way as in (3.14), with c_1 being a constant that depends on known values and on the parameter ϱ of $U \sim \mathrm{St}(2\varrho)$, but does not on $\boldsymbol{\delta}$. Then, the conditional expectation of the complete log-likelihood function $\ell_{\text{BS-PE-ACD}}(\boldsymbol{\delta}; \boldsymbol{X}_c)$ given $\boldsymbol{x}$ is obtained as in (3.15), with $\widehat{u}_i$ defined in (3.16) specified as $\widehat{u}_i = \varrho_{\text{k}}\, a(x_i, \widehat{\kappa}, \widehat{\sigma}_i)^{2[\varrho_{\text{k}}-1]}$, for $i = 1, \ldots, n$, where $\boldsymbol{X}_c = (\boldsymbol{X}^\top, \boldsymbol{U}^\top)^\top$ and ϱ_{k}, as mentioned, can be a known value or obtained from the data in an optimal way; see details in Saulo et al. (2013). For the BS-t-ACD model, we have a similar expression to that given in (3.15) for $Q(\boldsymbol{\delta}|\widehat{\boldsymbol{\delta}})$, but now $\widehat{u}_i$ given in (3.16) is

$$\widehat{u}_i = \mathrm{E}\left[\frac{1}{g(U_i)}\Big| X_i = x_i, \boldsymbol{\delta} = \widehat{\boldsymbol{\delta}}\right] = \mathrm{E}[U_i | X_i = x_i, \boldsymbol{\delta} = \widehat{\boldsymbol{\delta}}] = \frac{\nu_{\text{k}} + 1}{\nu_{\text{k}} + a(x_i, \widehat{\kappa}, \widehat{\sigma}_i)}, \quad i = 1, \ldots, n, \tag{3.17}$$

where, in this case, c_1 is a constant that depends on the parameter ν of $U \sim \mathrm{Gamma}(\nu/2, \nu/2)$, but does depend not on $\boldsymbol{\delta}$, and, such as for ϱ_{k} of the BS-PE-ACD model, ν_{k} can be a known value or obtained from the data in an optimal way. Therefore, in both of these cases (BS-PE and BS-t), based on Algorithm 1, the steps to obtain the ML estimates of the SBS-ACD model parameters are summarized in the following algorithm:

Algorithm 2 EM algorithm for SBS-ACD models

1: **E-step.** Compute $\widehat{u}_i^{(r)}$ given $\boldsymbol{\delta} = \widehat{\boldsymbol{\delta}}^{(r)}$, for $i = 1, \ldots, n$, $r = 1, 2, \ldots$; and
2: **M-step.** Update $\widehat{\boldsymbol{\delta}}^{(r)}$ by maximizing (3.15) over $\boldsymbol{\delta}$.

Note that the EM approach presented in Algorithm 2 permits the ML estimates of the parameters to be computed in an efficient way, such as mentioned in (C2). Note also that $\widehat{u}_i$ given in (3.17) works in Algorithm 2 as a weight function, such as mentioned in (C3), assigning small values to large values of the data, which allows us to conjecture the robustness proposed of the estimation procedure; see Paula et al. (2012).

As is well known, under some regularity conditions, ML estimators are asymptotically normal distributed. Thus, based on this, for the case of the SBS-ACD model, we have

$$\sqrt{n}\,[\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}] \xrightarrow{\mathcal{D}} \mathrm{N}_4(0, \boldsymbol{\Sigma}_{\boldsymbol{\delta}}), \quad \text{as } n \to \infty, \tag{3.18}$$

where $\Sigma_{\boldsymbol{\delta}}$ is the variance-covariance matrix of $\widehat{\boldsymbol{\delta}}$ and $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution. An approximate $100 \times [1-a]\%$ confidence region for $\boldsymbol{\delta}$ obtained from (3.18) is

$$\mathcal{R} = \{\boldsymbol{\delta} \in \mathbb{R}^4 : [\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}]^\top \widehat{\Sigma}_{\boldsymbol{\delta}}^{-1} [\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}] \leq \chi^2_{4;1-a}\}, \quad 0 < a < 1, \tag{3.19}$$

where $\chi^2_{4;1-a}$ denotes the $[1-a] \times 100$th percentile of the chi-squared distribution with 4 degrees of freedom and $\widehat{\Sigma}_{\boldsymbol{\delta}}$ is an estimate of $\Sigma_{\boldsymbol{\delta}}$.

In order to check the SBS-ACD model for a data set in practice, as usual in lifetime and duration analysis, we propose a generalized Cox-Snell residual (COX and SNELL, 1968) given by

$$r_i^{\mathrm{cs}} = \log(\widehat{S}(x_i|\Omega_{i-1})), \quad i = 1, \ldots, n, \tag{3.20}$$

where $\widehat{S}(x_i|\Omega_{i-1})$ is the fitted conditional survival function of the ith TD based on the past information set Ω_{i-1}.

Survival functions for BS, BS-PE, BS-t, and GG distributions useful for obtaining (3.20) are given by

$$\begin{aligned}
S_{\mathrm{BS}}(x_i;\kappa,\sigma_i) &= 1 - \Phi\left(a(x_i)\right), \quad x_i,\kappa,\sigma_i > 0,\\
S_{\mathrm{BS-PE}}(x_i;\kappa,\sigma_i,\varrho) &= 1 - [1/2]\left[1 + \Gamma\left(\tfrac{1}{2}\{a(x_i)\}^{2\varrho}, \tfrac{1}{2\varrho}\right)/\Gamma\left(1/2\varrho\right)\right],\\
&\quad x_i,\kappa,\sigma_i > 0,\ 1/2 < \varrho \leq 1,\\
S_{\mathrm{BS}-t}(x_i;\kappa,\sigma_i,\nu) &= 1 - [1/2]\left[1 + \mathrm{I}_{[\{a(x_i)\}^2]/[\{a(x_i)\}^2+\nu]}\left(1/2, \nu/2\right)\right],\\
&\quad x_i,\kappa,\sigma_i,\nu > 0,\\
S_{\mathrm{GG}}(x_i;\upsilon,\eta,\psi_i) &= 1 - \Gamma(x_i^\eta[\psi_i\varphi(\upsilon,\eta)]^{-\eta},\upsilon)/\Gamma(\upsilon), \quad x_i,\upsilon,\eta,\psi_i > 0,
\end{aligned}$$

where $\Phi(\cdot)$ is the N(0, 1) CDF and $\Gamma(\cdot)$, $a(x)$ and $\varphi(\upsilon,\eta)$ are as given in (3.5), (3.7) and (3.8), respectively. In addition, $\Gamma(x,\upsilon)$ is the lower incomplete gamma function defined as $\Gamma(y,\iota) = \int_0^y u^{\iota-1}\exp(-u)\mathrm{d}u$, and $\mathrm{I}_y(a,b) = \int_0^y u^{a-1}[1-u]^{b-1}\,\mathrm{d}u / \int_0^1 u^{a-1}[1-u]^{b-1}\,\mathrm{d}u$ is

the incomplete beta function ratio. Note that, when the ACD model is correctly specified, the Cox-Snell residual has a unit exponential (EXP(1)) distribution; see Bhatti (2010).

3.5 Simulation study

In this section, we carry out two simulation studies, one for evaluating the behavior of the ML estimators of the SBS-ACD models, and other one for detecting the performance of the residual proposed.

3.5.1 Estimation

We use the MC method for carrying out this first study. We focus on the BS-PE-ACD model, because, as mentioned, it has either less or greater (lighter or heavier tails) kurtosis levels than the BS distribution. However, in Section 3.6 (application with real TD data), we consider BS, BS-PE and BS-t and GG-ACD models. The scenario of this study is the following: the simulated sample sizes from the BS-PE-ACD model are considered to be as $n = \{500, 1000, 3000, 5000\}$ and the vector of true parameters as $\boldsymbol{\delta} = (\alpha, \beta, \gamma, \kappa)^\top = (0.1, 0.9, 0.1, 1.1)^\top$, whereas the number of MC replications is $B = 1000$. The BS-PE-ACD samples are generated by the transformation given in (3.3), considering $Y_i = \{g(U_i)\}^{1/2} Z_i \sim \text{PE}(\varrho = 1)$, with $U_i \sim \text{St}(\varrho = 1)$ and $Z_i \sim \text{N}(0, 1)$, for $i = 1, \ldots, n$; see Leiva et al. (2008) for details about how generating random numbers from SBS distributions. We estimate the SBS-ACD model parameters by using Algorithm 2, with starting values for (α, β, γ) to be considered as $(0.01, 0.70, 0.01)$, for κ as $\kappa_0 = \sqrt{2\,[\overline{x}/\text{Med}[x] - 1]}$, where $\overline{x}$ and $\text{Med}[x]$ are the sample mean and median based on observations (data) $\boldsymbol{x} = (x_1, \ldots, x_n)^\top$, and for σ as $\sigma_0 = \text{Med}[x]$. For each parameter and each sample size, we report the empirical mean, CS, CK, relative bias (RB) in absolute value and root of the mean squared error ($\sqrt{\text{MSE}}$) of the ML estimators in Table 3.2. The RB is defined as $\text{RB}[\widehat{\tau}] = |(\text{E}[\widehat{\tau}] - \tau)/\tau|$, where $\widehat{\tau}$ is the ML estimator of a parameter τ, and the sample CS and CK are, respectively, calculated by

$$\text{CS}[x] = \frac{\sqrt{n[n-1]}}{[n-2]} \frac{n^{-1}\sum_{i=1}^{n}[x_i - \bar{x}]^3}{[n^{-1}\sum_{i=1}^{n}\{x_i - \bar{x}\}^2]^{3/2}} \quad \text{and} \quad \text{CK}[x] = \frac{n^{-1}\sum_{i=1}^{n}[x_i - \bar{x}]^4}{[n^{-1}\sum_{i=1}^{n}\{x_i - \bar{x}\}^2]^2}, \tag{3.21}$$

where, as mentioned, $\boldsymbol{x} = (x_1, \ldots, x_n)^\top$ denotes the observations from a sample.

The definition of $\text{CK}[x]$ given in (3.21) is the raw measure, not excess kurtosis, which is obtained subtracting 3 from $\text{CK}[x]$ in (3.21). In Table 3.2, we observe that, as the sample size increases, the RB and $\sqrt{\text{MSE}}$ of the all the estimators decrease, tending them to be unbiased, as expected. From this table, observe also that the empirical distributions of all the parameter estimators are somewhat skewed and with high kurtosis, but tending to the normal

case, as the sample size increases. This is corroborated by Figures 3.1, 3.2, 3.3 and 3.4, which show the empirical distributions of the parameter estimators from the simulation study. These graphical plots suggest symmetric empirical distributions for these ML estimators as n increases, as expected. In addition, these figures provides confidence intervals (CIs) for the corresponding parameters, obtained from (3.18), considering several values for the sample size (n) and confidence level (ϖ).

Figura 3.1: empirical distribution of $\widehat{\alpha}$ and CI($\alpha; [1-\varpi] \times 100\%$) for the indicated n and ϖ.

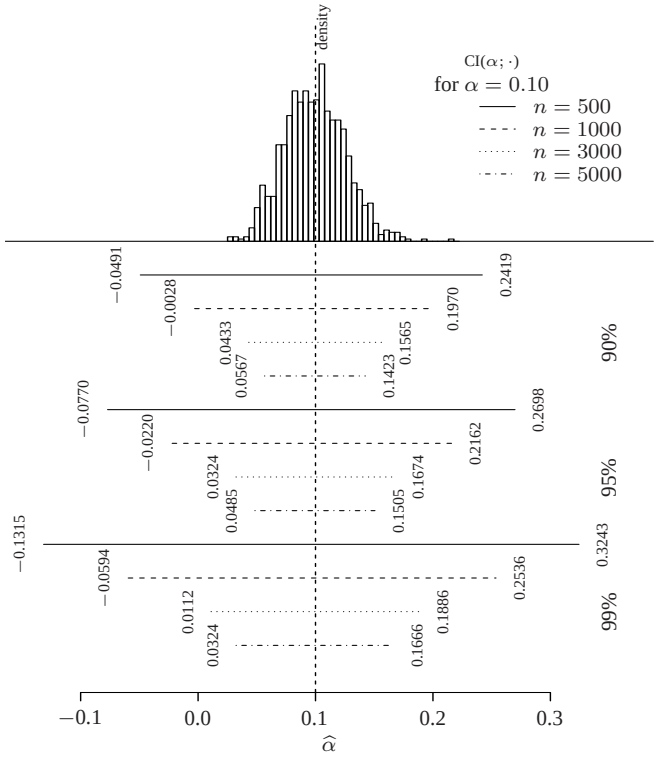
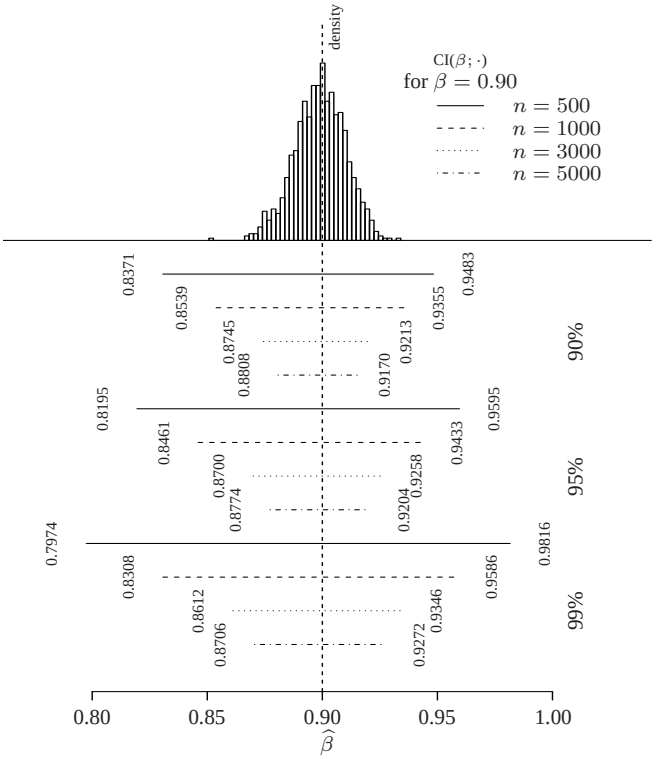


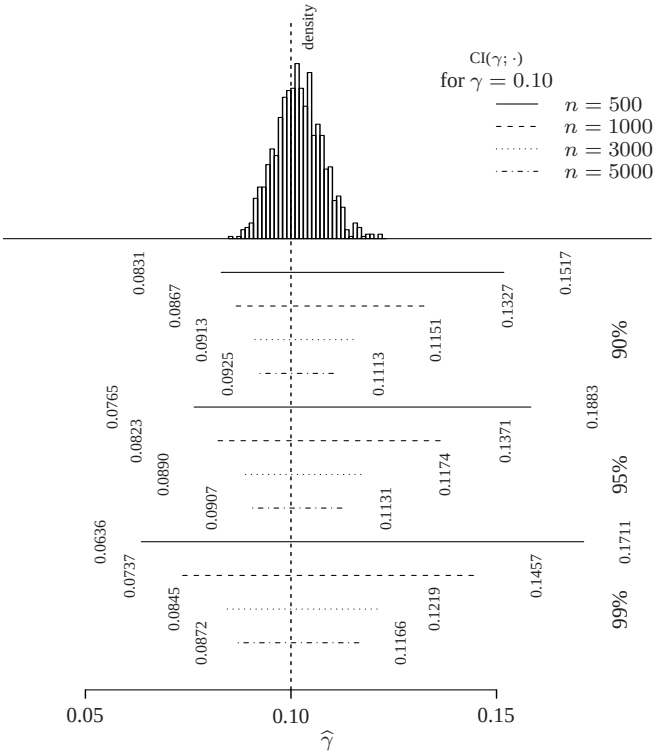
Figura 3.2: empirical distribution of $\widehat{\beta}$ and CI($\beta; [1-\varpi] \times 100\%$) for the indicated n and ϖ.



3.5.2 Residuals

Now, we conduct a second MC simulation study to examine the performance of the Cox-Snell residual r^{cs} defined in (3.20); see Bauwens and Giot (2000). Similarly to Subsection 3.5.1, we simulate $B = 1000$ MC samples of size $n = 200$ from the BS-PE-ACD model and, once again, we use Algorithm 2 to estimate the model parameters. Figure 3.5(a) shows a plot of the time against the residual r^{cs}. This figure does not show unusual features. Figure 3.5(b) displays the empirical autocorrelation function (ACF) of the residual r^{cs}. Note that the empirical ACFs are averages, over 1000 samples, of the ACFs associated with each sample of size 200. From Figure 3.5(b), the BS-PE-ACD model seems to be adequately specified, because the residual r^{cs} behaves as a sequence of independent and identically distributed RVs and there is no indication of serial correlation. Figure 3.5(c) presents a quantile against quantile (QQ) plot with simulated envelope, which allows us to compare the empirical distribution of the residuals for a simulated sample and the EXP(1) distribution; for more details on the envelope plot, see Atkinson (1985). From Figure 3.5(c), note that the Cox-Snell residual presents a linear behavior and has an excellent agreement with the EXP(1) distribution, which

Figura 3.3: empirical distribution of $\widehat{\gamma}$ and CI($\gamma; [1-\varpi] \times 100\%$) for the indicated n and ϖ.



confirms the adequacy of the BS-PE-ACD model.

3.6 Analysis of high frequency financial transaction data

In this section, we apply existing and proposed ACD models to high frequency financial transaction data. We consider here six data sets, studied in Bhatti (2010), corresponding to the time elapsed between two consecutive transactions, which contain forty trading days since 01-January-2002 until 28-February-2002. These data correspond to General Motors (GM), International Business Machines (IBM), Johnson & Johnson (JJ), McDonald (MD), Proctor & Gamble (PG), and Schlumberger Limited (SL) corporations. Note that, as mentioned in (D3), this type of data present an active trading in opening and closing hours and a dormant trading around noon. This trading is named the time-of-day effect. As explained by Engle and Russell (1998), it is necessary to transform the data in order to extract this effect from the raw durations. We apply the approach suggested in Tsay (2002) to obtain the time-of-day adjusted durations $\bar{x}_i$ as

$$\bar{x}_i = \frac{x_i}{\widehat{\phi}}, \quad i = 1, \ldots, n, \tag{3.22}$$

where x_i is the raw duration and $\widehat{\phi}$ is the time-of-day effect. This effect is estimated using a

Figura 3.4: empirical distribution of $\widehat{\kappa}$ and CI(κ; $[1-\varpi]\times 100\%$) for the indicated n and ϖ.

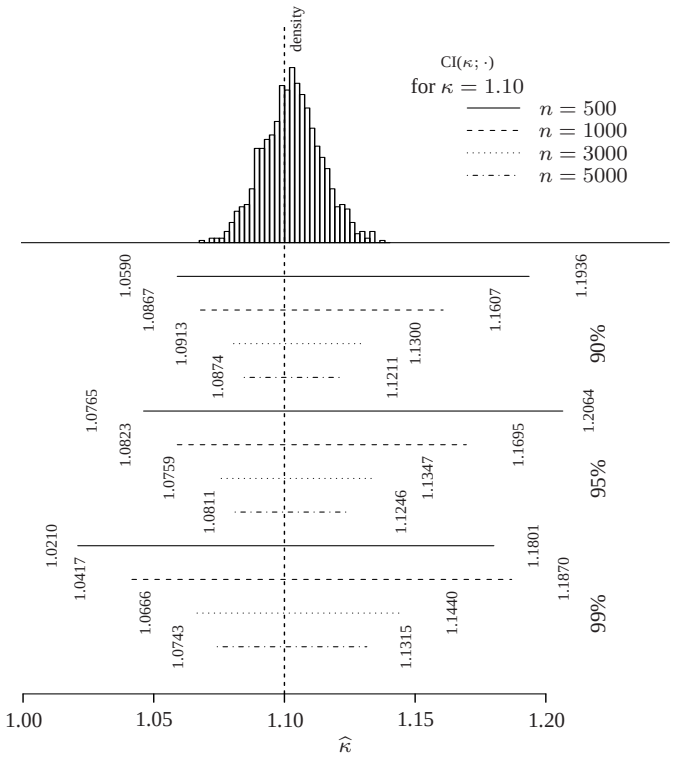
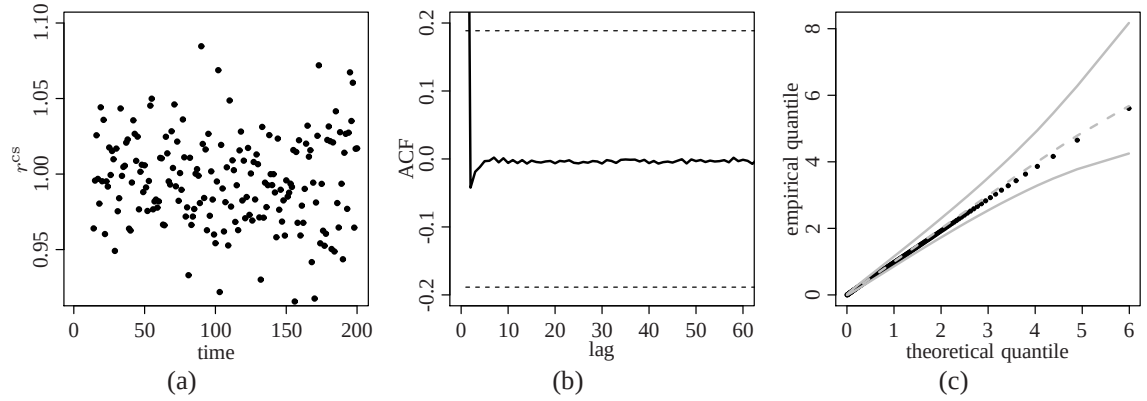


Figura 3.5: Cox-Snell residual against time (a), ACF (b) and QQ plot with envelope (c).



set of quadratic functions and indicator variables for each half hour interval of the trading day from 9:30am to 4:00pm; for more details on this and alternative procedures, see Giot (2000), Tsay (2002), Drost and Werker (2004) and Bhatti (2010). In the final detrended TD data sets, we use only observations in the period 10:00am to 4:00pm.

Tabela 3.2: summary statistics from simulated BS-PE-ACD data for the indicated estimators and n.

	n			
Statistic	500	1000	3000	5000
		$\widehat{\alpha}$		
True value	0.1000	0.1000	0.1000	0.1000
Mean	0.0964	0.0971	0.0999	0.0995
CS	1.1885	0.8571	0.4547	0.2516
CK	6.5193	4.3830	3.6492	3.2585
RB	0.0361	0.0290	0.0014	0.0049
$\sqrt{\text{MSE}}$	0.0885	0.0608	0.0344	0.0260
		$\widehat{\beta}$		
True value	0.9000	0.9000	0.9000	0.9000
Mean	0.8895	0.8947	0.8979	0.8989
CS	-1.0616	-0.6945	-0.3834	-0.2153
CK	5.5867	3.6278	3.5492	3.1182
RB	0.0116	0.0058	0.0024	0.0012
$\sqrt{\text{MSE}}$	0.0372	0.0254	0.0144	0.0110
		$\widehat{\gamma}$		
True value	0.1000	0.1000	0.1000	0.1000
Mean	0.1174	0.1097	0.1032	0.1019
CS	0.2109	0.0525	0.0398	0.1968
CK	3.0894	3.1472	3.1859	3.0387
RB	0.1737	0.0966	0.0325	0.0190
$\sqrt{\text{MSE}}$	0.0272	0.0170	0.0079	0.0060
		$\widehat{\kappa}$		
True value	1.1000	1.1000	1.1000	1.1000
Mean	1.1263	1.1143	1.1053	1.1029
CS	0.2882	0.1934	0.1094	0.0775
CK	3.7016	3.5011	3.0083	2.9696
RB	0.0239	0.0130	0.0048	0.0027
$\sqrt{\text{MSE}}$	0.0486	0.0316	0.0159	0.0115

3.6.1 Exploratory data analysis

It is well known that high frequency financial data have serial dependence. Figure 3.6 shows graphical plots of the ACF and partial ACF for the GM, IBM and MD data sets, from where it is noted the presence of serial correlation. A similar behavior is detected for the JJ, PG and SL data sets (omitted here).

Table 3.3 provides some descriptive measures for TD data adjusted by (3.22), which include central tendency statistics, as well as the standard deviation (SD), coefficient of variation (CV), CS and CK, among others. These measures indicate the positively skewed nature and the high kurtosis level of the data distribution.

Figura 3.6: autocorrelation and partial autocorrelation functions of the indicated data set.

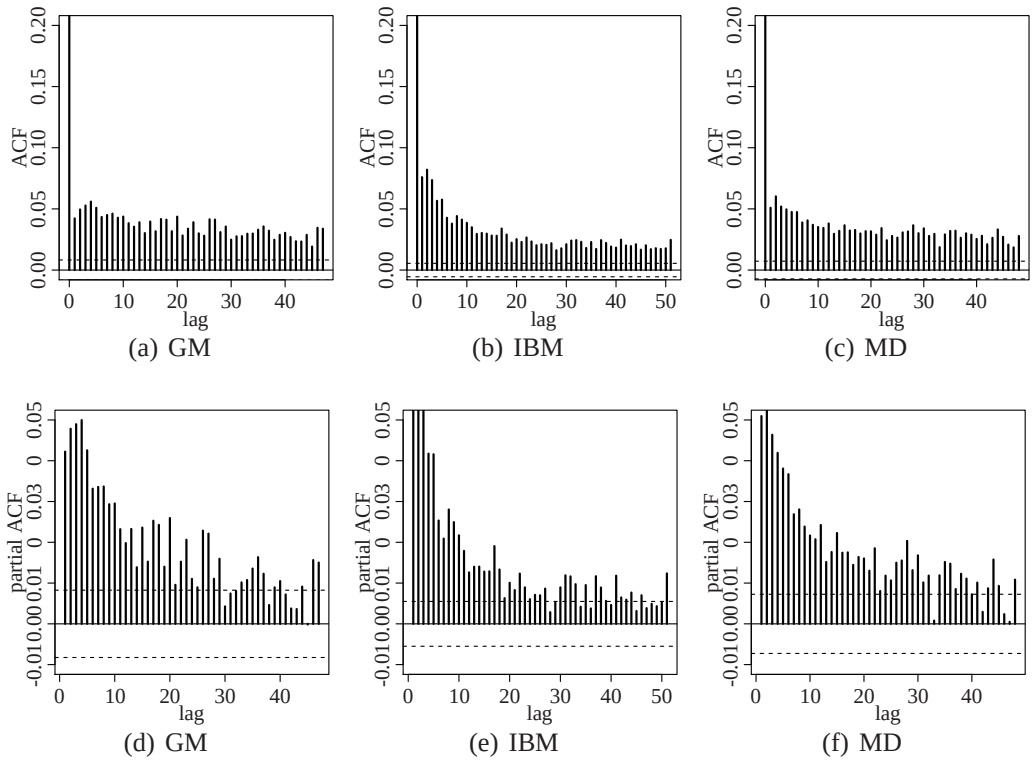


Tabela 3.3: summary statistics for the indicated data set.

Data set	n	Min.	Median	Mean	Max.	SD	CV	CS	CK
GM	56408	0.092	1.008	1.740	26.467	2.020	116.13%	2.782	11.951
IBM	127309	0.169	1.038	1.384	32.523	1.252	90.43%	3.023	19.802
JJ	82938	0.131	0.976	1.557	33.973	1.680	107.91%	3.135	18.463
MD	72979	0.121	1.006	1.752	47.163	2.081	118.75%	3.090	19.164
PG	78933	0.121	0.985	1.582	26.327	1.718	108.58%	2.865	13.311
SL	90694	0.143	0.996	1.708	31.143	2.011	117.72%	2.980	14.571

The HR of a positive RV X is $h(x) = f(x)/[1 - F(x)]$, where $f(\cdot)$ and $F(\cdot)$ are the PDF and CDF of X, respectively. One simple manner to characterize the HR is by the scaled total time on test (TTT) function. We can detect the type of HR that the data have and then choose a suitable distribution. The TTT function is given by $W(u) = H^{-1}(u)/H^{-1}(1)$, for $0 \leq u \leq 1$, where $H^{-1}(u) = \int_0^{F^{-1}(u)}[1 - F(y)]\mathrm{d}y$, with $F^{-1}(\cdot)$ being the inverse CDF of X. By plotting the points $[k/n, W_n(k/n)]$, with $W_n(k/n) = [\sum_{i=1}^{k} x_{(i)} + \{n-k\}x_k]/\sum_{i=1}^{n} x_{(i)}$, for $k = 1, \ldots, n$, and $x_{(i)}$ being the ith observed order statistic, it is possible to approximate $W(\cdot)$; see Aarset (1987) and Azevedo et al. (2012).

From Figure 3.7, we detect that the TTT plots suggest a unimodal HR for the GM, IBM and MD data. The same results are obtained for the other data sets (omitted here). To confirm

this detection, we did a simple simulation study from the BS distribution. First, we generate BS data with a similar framework to that found in real TD data. Then, with the generated data, we plot the theoretical BS HR and the TTT curve. The BS HR plot showed a totally unimodal shape, whereas the empirical TTT plot was very similar to that found in the GM, IBM, JJ, MD, PG, and SL data sets. This is an indication that TD data have usually a unimodal HR, as indicated by several authors; see, e.g., Grammig and Maurer (2000) and Bhatti (2010). From Figure 3.7, we also observe that the histograms, as well as the nonparametric density estimates based on asymmetric kernels proposed by Saulo et al. (2013) and Marchant et al. (2013), evidence a positive skewness and heavy tails for the data PDF. This ratifies the results shown in Table 3.3. Hubert and Vandervieren (2008) pointed out that, in cases where the data follow a skewed distribution, a significant number of observations can be classified as atypical when they are not. The boxplots depicted in Figure 3.7 confirm such a situation, that is, several of the cases considered as potentially outliers by the usual boxplot are not outliers when we consider the adjusted boxplot, although atypical data still remain. In summary, the conducted exploratory data analysis has shown the different conjectures mentioned in (A1)-(A3), (B1)-(B3) and (D1)-(D4). All these conjectures allow us to propose SBS-ACD models for analyzing the GM, IBM, JJ, MD, PG, and SL data sets.

3.6.2 Estimation

We now estimate the parameters of the new BS-PE-ACD and BS-t-ACD models via the EM algorithm described in Section 3.4; see Algorithm 2. Estimation of (α, β, γ) is initiated at the same values as in Subsection 3.5.1, whereas starting values for σ and ψ are considered to be the sample median and mean of TDs respectively, over the period [9:30, 10:00), so that spillovers of information from one trading day to the next trading day are avoided. In addition, we consider the existing BS-ACD and GG-ACD models, which parameters are estimated through a sequence of stages (BHATTI, 2010) based on the Nelder-Mead (NM) and Broyden-Fletcher-Goldfarb-Shanno (BFGS) approaches; for more details, see Nelder and Mead (1965) and Mittelhammer et al. (2000, p. 199). Notice that, in the last reference, the BFGS quasi Newton approach is considered as the best-performing algorithm.

As mentioned earlier, we consider lags of order $r = 1$ and $s = 1$, because a higher order for ACD models does not improve the model fit. Thus, the ACD(1,1) specification is sufficient for capturing the usual dynamics of TD data; see Bhatti (2010).

The standard errors (SEs) of the ML estimators of the ACD model parameters can be obtained using the White covariance (WC) matrix (see ENGLE and RUSSELL, 1998) given by

$$\mathrm{WC}(\boldsymbol{\varsigma}) = \left(\nabla^2 \ell(\boldsymbol{\varsigma})\right)^{-1} \left\{ \left(\nabla \ell(\boldsymbol{\varsigma})\right) \left(\nabla \ell(\boldsymbol{\varsigma})\right)^{\top} \right\} \left(\nabla^2 \ell(\boldsymbol{\varsigma})\right)^{-1}, \tag{3.23}$$

where $\nabla\ell(\boldsymbol{\varsigma})$ and $\nabla^2\ell(\boldsymbol{\varsigma})$ stands respectively for the gradient vector and the Hessian matrix after we replace parameters by estimates of the corresponding log-likelihood function. Specifically, the SEs are obtained by the square roots of the diagonal elements of the WC. Hypothesis H_0: $\tau = 0$ against H_1: $\tau \neq 0$ can be tested by using the Wald statistic given by $[\widehat{\tau} - \tau_0]/\mathrm{SE}(\widehat{\tau}) \overset{\mathrm{H}_0}{\sim} \mathrm{N}(0,1)$, where $\widehat{\tau}$ and τ_0 are the corresponding estimator and its proposed value in H_0, respectively.

Estimated SEs can be obtained from (3.23) evaluating it at $\widehat{\boldsymbol{\varsigma}}$. Table 3.4 reports the model parameter estimates and the estimated SEs of the corresponding ML estimators. All the estimates are statistically significant at a level of 1%. Note that the ACD parameter estimates are very similar across the models, independently of the assumed distribution.

Figura 3.7: TTT plots, histograms with asymmetric kernel, and usual and adjusted boxplots for the indicated data sets.

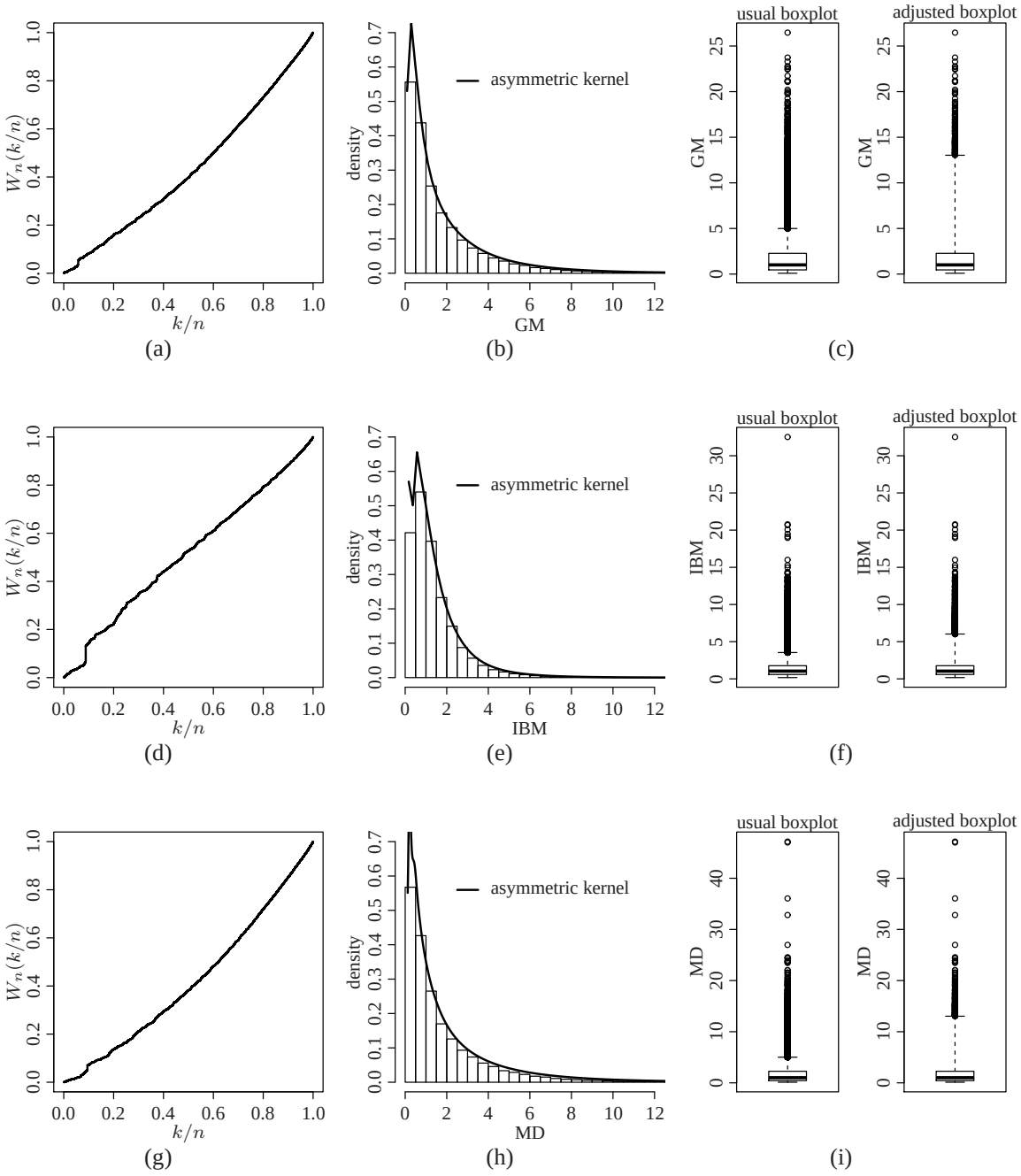


Tabela 3.4: ML estimates (with estimated SEs of the corresponding ML estimators in parenthesis) for the indicated parameter, model and data set, where ℓ is the maximum value of the log-likelihood function.

	Data set					
Parameter	GM	IBM	JJ	MD	PG	SL
	BS-ACD model					
α	-0.018	-0.045	-0.017	-0.017	-0.018	-0.019
	(0.0023)	(0.0032)	(0.0010)	(0.0043)	(0.0060)	(0.0092)
β	0.986	0.939	0.974	0.984	0.985	0.980
	(0.0302)	(0.0253)	(0.0272)	(0.0214)	(0.0253)	(0.0093)
γ	0.010	0.032	0.011	0.010	0.011	0.012
	(0.0001)	(0.0020)	(0.0006)	(0.0003)	(0.0007)	(0.0016)
κ	1.213	0.873	1.042	1.204	1.063	1.147
	(0.0022)	(0.0002)	(0.0012)	(0.0011)	(0.0016)	(0.0007)
ℓ	83518.5	153644.7	112581.4	107508.0	108461.2	130066.5
	BS-PE-ACD model					
α	-0.026	-0.047	-0.024	-0.024	-0.024	-0.021
	(0.0029)	(0.0011)	(0.0059)	(0.0034)	(0.0019)	(0.0013)
β	0.968	0.923	0.950	0.965	0.968	0.969
	(0.0094)	(0.0097)	(0.0207)	(0.0117)	(0.0080)	(0.0088)
γ	0.008	0.021	0.008	0.008	0.009	0.007
	(0.0016)	(0.0055)	(0.0013)	(0.0014)	(0.0018)	(0.0016)
κ	1.046	0.753	0.891	1.032	0.912	0.981
	(0.2188)	(0.1601)	(0.1848)	(0.2118)	(0.1923)	(0.1978)
ϱ	0.592	0.590	0.592	0.592	0.591	0.593
ℓ	70373.2	123469.9	92717.3	90602.2	89545.1	130066.5
	BS-t-ACD model					
α	-0.018	-0.045	-0.017	-0.017	-0.018	-0.019
	(0.0000)	(0.0004)	(0.0001)	(0.0000)	(0.0001)	(0.0001)
β	0.986	0.938	0.974	0.984	0.985	0.980
	(0.0002)	(0.0005)	(0.0006)	(0.0002)	(0.0003)	(0.0002)
γ	0.010	0.032	0.011	0.010	0.011	0.012
	(0.0000)	(0.0003)	(0.0001)	(0.0000)	(0.0000)	(0.0000)
κ	1.209	0.870	1.036	1.201	1.059	1.145
	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0001)
ν	292	246	167	341	242	341
ℓ	83523.2	153635.7	112581.9	107523	108461.5	130087
	GG-ACD model					
α	-0.015	-0.031	-0.010	-0.013	-0.018	-0.017
	(0.0017)	(0.0011)	(0.0002)	(0.0063)	(0.0015)	(0.0027)
β	0.988	0.939	0.976	0.983	0.980	0.981
	(0.0037)	(0.0181)	(0.0198)	(0.0279)	(0.0005)	(0.0404)
γ	0.022	0.050	0.020	0.022	0.027	0.026
	(0.0008)	(0.0037)	(0.0004)	(0.0036)	(0.0008)	(0.0024)
υ	37.671	37.78	38.449	37.791	38.255	38.038
	(0.0408)	(0.4146)	(0.8190)	(0.6603)	(0.4642)	(0.9821)
η	0.152	0.2010	0.170	0.151	0.168	0.156
	(0.0000)	(0.0001)	(0.0006)	(0.0001)	(0.0005)	(0.0000)
ℓ	84201.8	153818.5	113326.7	109037.1	109025.7	132047.8

3.6.3 Model checking

We now analyze the goodness-of-fit of the SBS-ACD model to the data through the residual defined in (3.20); see Section 3.4. We recall that, when the model is correctly specified, the Cox-Snell residual has an EXP(1) distribution. Figure 3.8 depicts the QQ plots of the Cox-Snell residual for the indicated models and data sets. These QQ plots allow us to check graphically whether such residuals follow the EXP(1) distribution or not. These graphical plots show a superiority, in terms of fitting to the data, of the BS-PE-ACD model over the GG-ACD model, followed by the BS-t-ACD and BS-ACD models, although the GG-ACD model fits some data well; see Figure 3.8(b). Considering all six real data sets, the BS-PE-ACD model fits the data adequately to furnish effective based-ML inference.

Figura 3.8: QQ plots of Cox-Snell residuals for the indicated fitted model and data set.

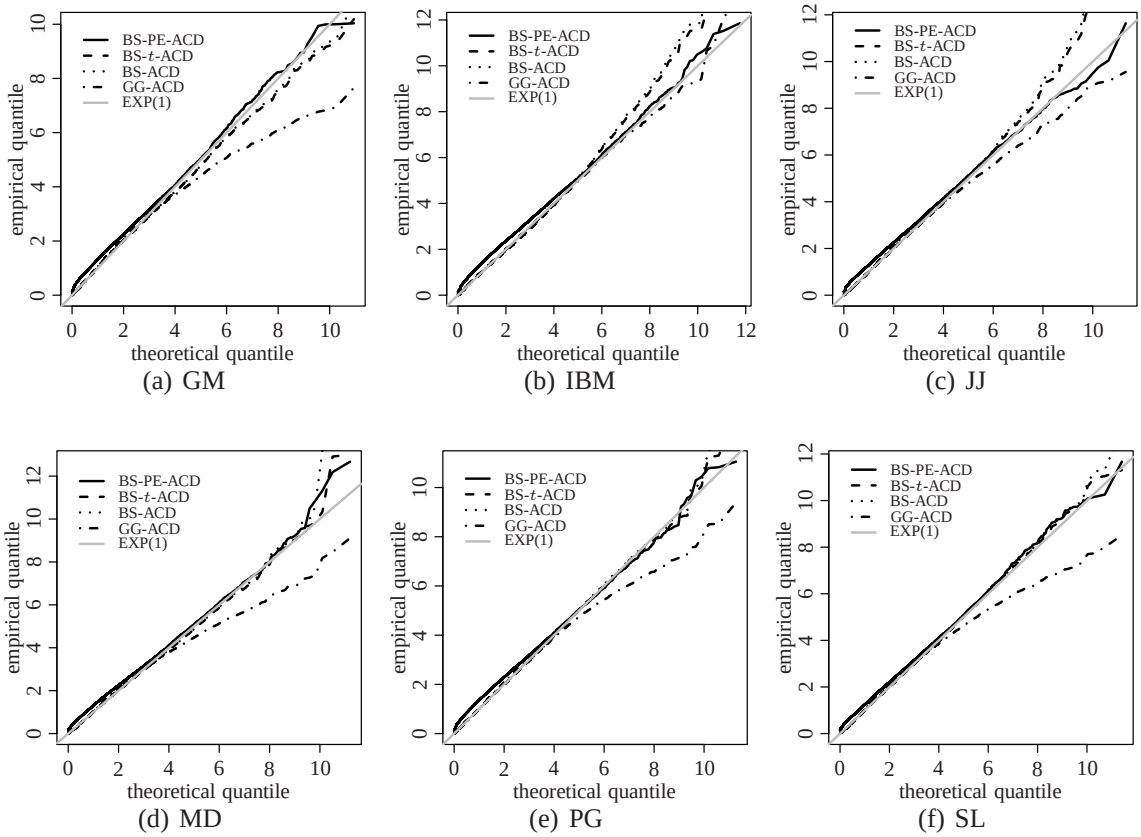


Table 3.5 reports the Akaike and Bayesian information criteria, given by AIC $= -2\ell + 2p$ and BIC $= -2\ell + p\log(n)$, respectively, where ℓ is the maximum value of the corresponding log-likelihood function, p denotes the number of model parameters and n the number of observations. Also, the maximum and minimum values of the sample ACF from order 1 to 60, and the mean magnitude of autocorrelation $\bar{\rho}$ for the first 15 lags, namely

$\bar{\rho} = [1/15]\sum_{i=1}^{15} |\widehat{\rho}_k|$, where $\widehat{\rho}_k = \widehat{\text{cor}}(r_i^{\text{cs}}, r_{i+k}^{\text{cs}})$, are reported. The statistic $\bar{\rho}_k$ is relevant to separate the influence of the sample size on the measure of the degree of autocorrelation in the residuals; see Bhatti (2010). From Table 3.5, we observe that, in terms of AIC values, the BS-PE-ACD model is the best one, whereas the BS-t-ACD and BS-ACD models provide close values, and the GG-ACD model is the worst one. Considering the BIC values, we once again notice that the BS-PE-ACD model outperforms all the considered models.

Turning now to check for misspecification, we look at the sample ACF of lth order with l varying from 1 to 60. Table 3.5 reports that there is no sample autocorrelation greater than 0.05 (in magnitude) throughout the models and residuals.

Tabela 3.5: AIC and BIC values, sample autocorrelations, and the mean magnitude of autocorrelation for the indicated data set and model.

	Data set					
Indicator	GM	IBM	JJ	MD	PG	SL
	BS-PE-ACD model					
AIC	140754.4	246947.8	185442.7	181212.4	179098.3	218204.2
BIC	140790.2	246986.8	185480.0	181249.2	179135.4	218241.9
max ACF	0.019	0.031	0.024	0.048	0.031	0.026
min ACF	−0.009	−0.000	−0.011	−0.001	−0.009	−0.006
$\bar{\gamma}$	0.009	0.015	0.011	0.017	0.013	0.009
	BS-t-ACD model					
AIC	167054.4	307279.4	225171.8	215054.0	216931.0	260182.0
BIC	167090.2	307318.4	225209.1	215090.8	216968.1	260219.7
max ACF	0.021	0.026	0.019	0.031	0.039	0.035
min ACF	−0.008	−0.007	−0.010	−0.007	−0.014	−0.003
$\bar{\gamma}$	0.008	0.006	0.007	0.009	0.011	0.011
	BS-ACD model					
AIC	167045.1	307297.4	225170.8	215024.0	216930.4	260141.0
BIC	167080.9	307336.0	225208.1	215060.8	216967.5	260178.7
max ACF	0.021	0.025	0.019	0.031	0.0396	0.034
min ACF	−0.008	−0.007	−0.010	−0.007	−0.014	−0.003
$\bar{\gamma}$	0.008	0.006	0.007	0.009	0.011	0.011
	GG-ACD model					
AIC	168411.7	307645.0	226661.4	218082.2	218059.4	264103.6
BIC	168447.5	307684.0	226698.7	218119.0	218096.5	264141.3
max ACF	0.017	0.022	0.015	0.028	0.032	0.029
min ACF	−0.010	−0.009	−0.014	−0.008	−0.014	−0.006
$\bar{\gamma}$	0.006	0.005	0.006	0.006	0.007	0.007

3.6.4 Predictive model

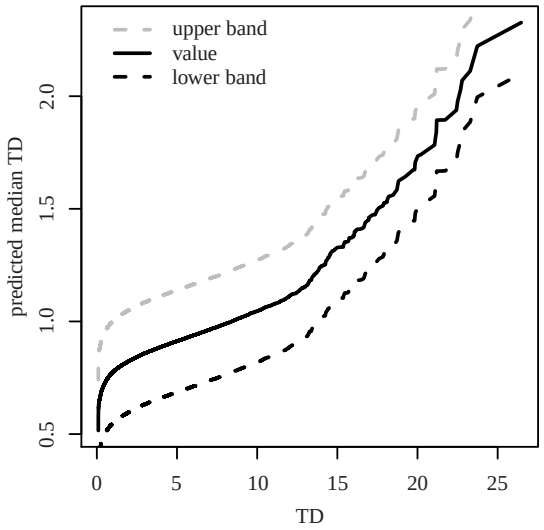
The analysis performed in Subsection 3.6.3 suggests that the BS-PE-ACD model is the most appropriate for fitting the TD data. Assuming, for example, GM data, the ML es-

timates of the parameters are (with estimated SEs in parenthesis): $\widehat{\alpha} = -0.026(0.0029)$, $\widehat{\beta} = 0.968(0.0094)$, $\widehat{\gamma} = 0.008(0.0016)$, and $\widehat{\kappa} = 1.046(0.2188)$. All estimates are statistically significant at the 1% level. Then, the predictive model is

$$\widehat{x}_{\text{pred}}^{\text{BS-PE}} = \exp\left(-0.018 + 0.9860\log(\widehat{\sigma}_{i-1}) + 0.0100[x_{i-1}/\widehat{\sigma}_{i-1}]\right), \tag{3.24}$$

where the initial value $\widehat{\sigma}_0$ is the median over the period [9:30, 10:00). We may interpret the expression given in (3.24) as the median TD in seconds. Figure 3.9 depicts approximate 95% confidence bands for the median TD based on the BS-PE-ACD model according to the result given in (3.19).

Figura 3.9: approximate 95% confidence bands for the predicted median TD based on GM data.



3.7 Concluding remarks

In this chapter, we have introduced a new methodology based on scale-mixture Birnbaum-Saunders autoregressive conditional duration models. These models allow us to obtain an efficient computation of the maximum likelihood estimators of the model parameters, by using the expectation and maximization algorithm. The new methodology includes a formulation of the model, estimation of the model parameters, inference for these parameters, the predictive model and a residual analysis for checking model adequacy in practice. We have conducted a Monte Carlo simulation study to evaluate the performance of the proposed methodology, which indicated its adequacy. We have also compared the proposed autoregressive conditional duration models based on the Birnbaum-Saunders-exponential-power and

Birnbaum-Saunders-Student-t distributions with some of the existing autoregressive conditional duration models, through an analysis by using real data of financial transactions from the New York stock exchange, which has shown the superiority of the Birnbaum-Saunders-exponential-power case.

As part of future research, out-of-sample forecasting ability of the model deserves special attention. Also, it is important to consider different estimation procedures, such as the estimating functions and the method of modified moments, as well as semiparametric structures and influence diagnostics.

4 PROCESS CAPABILITY INDICES FOR THE BIRNBAUM-SAUNDERS DISTRIBUTION

4.1 Introduction

Quality control is extremely relevant to companies that want to differentiate themselves from competitors in highly competitive markets. When quality parameters or specifications in product manufacture are planned, it is important to ensure that the production process is capable of maintaining tolerances in the design stage of these products. In this sense, the concept of process capability or process ability provides a quantitative tool to establish how suitable a production process can be. The capability of a manufacturing process is the ability it has to generate a result that meets a set of specifications established by the company so that a product can be considered of quality.

In statistical terms, the process capability is defined as the ratio between the allowable variation (based on specifications) and the natural variation of the production process (based on the data) due to non-assignable causes. There are several ways to measure process capability, including graphical methods, design of experiments, and process capability indices (PCIs). These indices were developed for processes whose quality characteristic to be studied is normally distributed; see Montgomery (2005, pp. 334). However, in many cases, production processes follow non-normal distributions. In these cases, the PCIs for the normal distribution should not be used, because the obtained results using them, about the performance of the process in question, could be inaccuracy, misleading and unreliable; see Kane (1986) and Somerville and Montgomery (1996). Gunter (1989) emphasized the difference between perfect (when the quality characteristic is normally distributed) and occasionally erratic processes. Gunter (1989), McCoy (1991), and Johnson (1992) investigated the properties of PCIs and their estimators, when the distribution of the data is normal.

The literature on PCIs under non-normality concerning the construction of new PCIs and/or development of new approaches, can be categorized into five groups (see KOTZ

and JOHNSON, 2002): (G1) data transformation methods (see JOHNSON, 1949; BOX and COX, 1964; RIVERA et al., 1995; SOMERVILLE and MONTGOMERY, 1996; HOSSEINIFARD et al., 2009); (G2) quality control procedures for non-normal distributions (see SHORE, 1998; LOVELACE and SWAIN, 2009); (G3) distribution fitting for empirical data (see CLEMENTS, 1989; KOTZ and JOHNSON, 2002); (G4) distribution-free procedures (see CHAN et al, 1988); and (G5) construction of new PCIs (see CLEMENTS, 1989; GILCHRIST, 1993; JOHNSON et al., 1994; WRIGHT, 1995; PEARN and CHEN, 1995, 1997, 1998; LIU and CHEN, 2006; VÄNNMAN and ALBING, 2007; HSU et al., 2008). For a review of PCIs, see Tang and Tang (1999), Spiring et al. (2003) and Yum and Kim (2011).

The method introduced by Clements (1989) employs both a distribution fitting approach and a new PCI based on percentiles. In this method, the mean may or not coincide with the center of the specification limits, when Pearson distributions are used. Gilchrist (1993) introduced a quantile transformation similar to the Clements method, but based on a standardized distribution, instead of Pearson curves. Johnson et al. (1994) applied the Clements method to obtain estimators of two other PCIs. Pearn and Chen (1995) introduced a new method for estimating PCIs, which can be viewed as a modification of the Clements method. The authors found that estimators based on the new method can differentiate on-target processes from off-target processes better than those obtained applying the Clements method; see Pearn and Chen (1995). The idea of all these authors was to reproduce the property of the normal distribution to yield a nonconformity proportion about 0.27%.

The Birnbaum-Saunders (BS) model is a two-parameter, unimodal distribution with positive skewness and non-negative support, which was originated from a material fatigue problem; see Birnbaum and Saunders (1969). The BS model describes the total time elapsed until a type of cumulative damage induced by stress exceeds the resistance threshold of a material thereby producing its rupture; see Johnson et al. (1995, pp. 651-663). This model has received considerable attention over the last decade, due mainly to its properties and recent applications in various fields; see Leiva et al. (2012). One of these properties is that a random variable (RV) following a BS model can be seen as a transformation of a normal RV. Despite the widespread use of the BS distribution, including some studies applied to quality (see BALAKRISHNAN et al., 2007; LIO et al., 2010; LEIVA et al., 2011), there is no study of process capability when the quality characteristic follows a BS distribution. Therefore, the aim of this chapter is to develop and implement computationally a methodology for process capability analysis based on the BS distribution and on the Clements-type PCIs proposed by Pearn and Chen (1995).

The rest of this chapter unfolds as follows. In Section 4.2, we provide a background about some results that are useful for developing our methodology. In Section 4.3, we derive this

methodology for PCIs based on the BS distribution. In Section 4.4, we perform a study of Monte Carlo (MC) simulation for evaluating the proposed methodology. In Section 4.5, we carry out an empirical application of the methodology by using real data and a computational code implemented in a noncommercial and open source statistical software called `R`. The `R` software can be freely downloaded from `www.r-project.org`. Finally, in Section 4.6, we sketch conclusions and recommendations for futures works.

4.2 Background

In this section, we provide some results which allow PCIs for the BS distribution to be developed.

4.2.1 Process capability indices

In general, the PCIs, denoted by $\mathrm{C_p}$, are defined as the ratio between a specification range and the process variation, that is, as

$$\mathrm{C_p} = \frac{\text{Specification range}}{\text{Statistical variability}}. \tag{4.1}$$

It is common to define the numerator range in (4.1) as the difference between the upper specification limit (USL) and the lower specification limit (LSL), which are predetermined by the company.

The PCIs are used when it is desired to study the process ability over time, taking into account the influence of several operation conditions (for example, shift, batch and raw materials). The PCIs are defined depending on several situations, which are: (i) whether the mean coincides with the center of the specification limits (USL and LSL); (ii) by considering only the USL; (iii) by using just the LSL; and (iv) whether the mean does not coincides with the center of the specification limits. In these situations, the PCIs are defined, respectively, as

$$\mathrm{C_p} = \frac{\mathrm{USL}-\mathrm{LSL}}{R},\ \mathrm{C_{pl}} = \frac{2[\theta-\mathrm{LSL}]}{R},\ \mathrm{C_{pu}} = \frac{2[\mathrm{USL}-\theta]}{R},\ \text{and}\ \ \mathrm{C_{pk}} = \min\{\mathrm{C_{pl}}, \mathrm{C_{pu}}\},$$

where R is the process variability and θ is a parameter that represents the center of it, which is associated with the distribution that governs the quality characteristic or RV under study. We denote this RV as X, its mean by $\mathrm{E}[X] = \mu$, its standard deviation (SD) by $\sqrt{\mathrm{Var}[X]} = \sigma$, and its quantile function (QF) by $x(p)$, for $0 < p < 1$.

4.2.2 PCI under normality

As aforementioned, PCIs were originally developed for normal process. Therefore, it is assumed here that $X \sim \mathrm{N}(\mu, \sigma^2)$. Given the LSL and USL, we have the corresponding PCIs are

$$\mathrm{C_p} = \frac{\mathrm{USL} - \mathrm{LSL}}{6\sigma},\ \mathrm{C_{pl}} = \frac{\mu - \mathrm{LSL}}{3\sigma},\ \mathrm{C_{pu}} = \frac{\mathrm{USL} - \mu}{3\sigma},\ \text{and}\ \ \mathrm{C_{pk}} = \min\{\mathrm{C_{pl}}, \mathrm{C_{pu}}\}. \quad (4.2)$$

Assume that μ and σ are estimated from the sample mean and SD as $\widehat{\mu} = \bar{x} = [1/n]\sum_{i=1}^{n} x_i$ and $\widehat{\sigma} = S = [1/(n-1)]\sum_{i=1}^{n}[x_i - \bar{x}]^2$, respectively, for a random sample of size n, say $X_1, \ldots, X_n$, with observations, $x_1, \ldots, x_n$. Therefore, natural estimates for the PCIs given in (4.2) are obtained by substituting μ and σ by $\widehat{\mu}$ and $\widehat{\sigma}$, respectively. The $\mathrm{C_p}$ and $\mathrm{C_{pk}}$ are the most commonly utilized PCIs in industry; see Kotz and Johnson (2002) and Anis (2008).

4.2.3 PCI under non-normality

If the characteristic of the production process X follows a non-normal distribution, then the PCIs should be modified. Although now the process is non-normal, we still use the notation μ, σ and $x(p)$ for the mean, SD and QF, respectively. In this case, a widely adopted procedure to construct PCIs is to substitute 6σ in expression (4.2) for a range R covering a similar percentage to that of 6σ in a normal distribution, that is, $R = \mathrm{U_p} - \mathrm{L_p}$ covering a 99.73%, such that $\mathrm{U_p}$ is the 99.865th quantile and $\mathrm{L_p}$ is the 0.135th quantile, from the corresponding non-normal distribution. The idea behind these substitutions is to mimic the normal distribution property, so that the output percentage falling outside the $\mu \pm 3\sigma$ limits is 0.27%. This ensures that, if the process is well centered (that is, the mean coincides with the midpoint of the specifications), the probability that the process is outside the specification range (LSL, USL) is negligible; see Clements (1989), Pearn and Chen (1997) and Hsu et al. (2008).

In the method introduced by Clements (1989), for computing $\mathrm{C_p}$ and $\mathrm{C_{pk}}$, the center of the process is based on the median $x(0.5)$, instead of using the mean, because the median is a robust measure of the central tendency of the process, particularly for skewed heavy-tailed distributions. Thus, $\mu - \mathrm{LSL}$ and $\mathrm{USL} - \mu$ given in (4.2) are replaced by $x(0.5) - \mathrm{L_p}$ and $\mathrm{U_p} - x(0.5)$, respectively; see Johnson et al. (1994). A modification of the Clements method to obtain $\mathrm{C_p}$ and $\mathrm{C_{pk}}$ was approached by Pearn and Chen (1995). They replaced the two 3σ's by $[\mathrm{U_p} - \mathrm{L_p}]/2$, so that $\mathrm{C_p}$ and $\mathrm{C_{pk}}$ can be written as

$$\mathrm{C'_p} = \frac{\mathrm{USL} - \mathrm{LSL}}{x(p_2) - x(p_1)},\ \mathrm{C'_{pl}} = \frac{2[x(0.5) - \mathrm{LSL}]}{x(p_2) - x(p_1)},\ \mathrm{C'_{pu}} = \frac{2[\mathrm{USL} - x(0.5)]}{x(p_2) - x(p_1)},\ \text{and}\ \ \mathrm{C'_{pk}} = \min\{\mathrm{C'_{pl}}, \mathrm{C'_{pu}}\}, \quad (4.3)$$

where p_1 and p_2 are fixed percentages, which can be chosen using an optimal statistical

criterion, $x(p)$ is the corresponding pth quantile, and $x(p_2) - x(p_1)$ is responsible for the non-normality and/or asymmetry of the distribution.

4.2.4 The Birnbaum-Saunders distribution

A RV X with BS distribution has two parameters, one of shape ($\alpha > 0$) and another of scale ($\beta > 0$), where β is also the median of the distribution. This is denoted by $X \sim \text{BS}(\alpha, \beta)$. BS and standard normal RVs, denoted respectively by X and Z, are related by

$$X = \beta\big[\alpha Z/2 + \sqrt{\{\alpha Z/2\}^2 + 1}\big]^2 \text{ and } Z = \big[\sqrt{X/\beta} - \sqrt{\beta/X}\big]/\alpha.$$

Then, it is possible to obtain $W = [1/\alpha^2][X/\beta + \beta/X - 2] \sim \chi^2(1)$, which is useful for goodness of fit and detecting outliers using the Mahalanobis distance. If $X \sim \text{BS}(\alpha, \beta)$, the following properties hold. The probability density function (PDF) of X is given by

$$f_{\text{BS}}(x; \alpha, \beta) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2\alpha^2}\left[\frac{x}{\beta} + \frac{\beta}{x} - 2\right]\right) \frac{x^{-\frac{3}{2}}[x+\beta]}{2\alpha\sqrt{\beta}},\ x > 0, \alpha > 0, \beta > 0. \quad (4.4)$$

The cumulative distribution function (CDF) of X is $F_X(x) = \text{P}(X \le x) = \Phi([1/\alpha][\sqrt{x/\beta} - \sqrt{\beta/x}])$, for $x > 0$, where $\Phi(\cdot)$ is the CDF of $Z \sim \text{N}(0, 1)$. Therefore, the pth quantile or QF of X is given by

$$x(p) = F_X^{-1}(p) = \beta\big[\alpha z(p)/2 + \sqrt{\{\alpha z(p)/2\}^2 + 1}\big]^2,\ 0 < p < 1, \quad (4.5)$$

where $F_X^{-1}(\cdot)$ is the inverse function of $F_X(\cdot)$ and $z(p)$ is the pth quantile of $Z \sim \text{N}(0, 1)$. Hence, from (4.5), $x(0.5) = \beta$, and so it is the median of the BS model, as aforementioned.

Let $X_1, \ldots, X_n$ be a random sample of size n from $X \sim \text{BS}(\alpha, \beta)$, with observations, $x_1, \ldots, x_n$. Then, the log-likelihood function for α and β is given by

$$\ell(\alpha, \beta) = k + \frac{n}{\alpha^2} - n\log(\alpha) - \frac{n}{2}\log(\beta) + \sum_{i=1}^{n}\left\{\log(x_i + \beta) - \frac{1}{2\alpha^2}\left[\frac{x_i}{\beta} + \frac{\beta}{x_i}\right]\right\}, \quad (4.6)$$

where k is a constant that does not depend on neither α or β. Taking derivatives of (4.6) with respect to the parameters α and β and equating them to zero, we obtain the maximum likelihood (ML) estimates of α and β, $\widehat{\alpha}$ and $\widehat{\beta}$ say, as $\widehat{\alpha} = [s/\widehat{\beta} + \widehat{\beta}/r - 2]^{1/2}$, where s and r are arithmetic and harmonic means of $x_1, \ldots, x_n$, given by $s = [1/n]\sum_{i=1}^{n} x_i$ and $r = 1/[\{1/n\}\sum_{i=1}^{n} 1/x_i]$, respectively. However, $\widehat{\beta}$ must be obtained by using an iterative numerical method, which can employ the sample median as starting value or the mean-mean estimation for β given by $\overline{\overline{\beta}} = [sr]^{1/2}$.

The asymptotic distribution of the ML estimators $\widehat{\alpha}$ and $\widehat{\beta}$ is normal and therefore

$$\sqrt{n}\left[\begin{pmatrix}\widehat{\alpha}\\ \widehat{\beta}\end{pmatrix}-\begin{pmatrix}\alpha\\ \beta\end{pmatrix}\right] \quad \stackrel{\mathcal{D}}{\to} \quad \mathrm{N}\left(0, I^{-1}(\alpha,\beta)\right), \quad \text{as } n\to\infty,$$

where $I(\alpha,\beta)$ is the Fisher information matrix and "$\stackrel{\mathcal{D}}{\to}$" means "convergence in distribution to".

The Fisher information matrix for α and β can be given by

$$I(\alpha,\beta)=\begin{bmatrix}\frac{\alpha^2}{2n} & 0\\ 0 & \frac{\beta^2}{n[1/4+1/\alpha^2+P(\alpha)]}\end{bmatrix},$$

where

$$P(\alpha)=2\int_0^\infty\left[\frac{1}{1+h(\alpha x)}-\frac{1}{2}\right]^2 \mathrm{d}\Phi(x), \tag{4.7}$$

with $h(y)=1+y^2/2+y[1+y^2/4]^{1/2}$.

4.3 Birnbaum-Saunders process capability

In this section, we propose and develop PCIs in cases where the quality characteristic follows a BS distribution. We focus on some results for the PCI $\mathrm{C_p}$, which we denote as $\mathrm{C_p^{BS}}$. However, similar results can be obtained for the indices $\mathrm{C_{pl}^{BS}}$, $\mathrm{C_{pu}^{BS}}$ and $\mathrm{C_{pk}^{BS}}$.

4.3.1 PCI for the BS distribution

We propose a PCI for the BS distribution comparing the specification limits with some range that covers a high percentage, which must be specified, $1-[p_1+p_2]$ say, for the distribution of X. Consider the QF of the BS distribution given in expression (4.5) and the PCI presented in (4.3). Then, we propose

$$\mathrm{C_p^{BS}}=\frac{\mathrm{USL}-\mathrm{LSL}}{x(p_2)-x(p_1)}=\frac{\mathrm{USL}-\mathrm{LSL}}{\beta\alpha\left[z_2\left\{\frac{\alpha z_2}{2}+\sqrt{\left(\frac{\alpha z_2}{2}\right)^2+1}\right\}-z_1\left\{\frac{\alpha z_1}{2}+\sqrt{\left(\frac{\alpha z_1}{2}\right)^2+1}\right\}\right]}, \tag{4.8}$$

where $z_1=z(p_1)$ and $z_2=z(p_2)$. Now, in order to compare a product with a LSL, we use

$$\mathrm{C_{pl}^{BS}}=\frac{2[\beta-\mathrm{LSL}]}{x(p_2)-x(p_1)}=\frac{2[\beta-\mathrm{LSL}]}{\beta\alpha\left[z_2\left\{\frac{\alpha z_2}{2}+\sqrt{\left(\frac{\alpha z_2}{2}\right)^2+1}\right\}-z_1\left\{\frac{\alpha z_1}{2}+\sqrt{\left(\frac{\alpha z_1}{2}\right)^2+1}\right\}\right]}, \tag{4.9}$$

recall that β is the median of the BS distribution, we then have

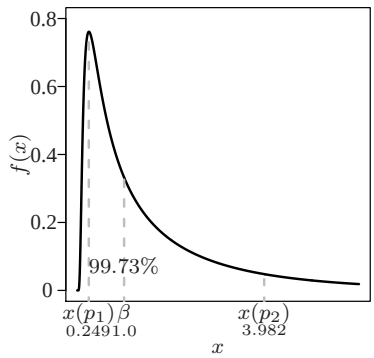
$$\mathrm{C}_{\mathrm{pu}}^{\mathrm{BS}} = \frac{2[\mathrm{USL}-\beta]}{x(p_2)-x(p_1)} = \frac{2[\mathrm{USL}-\beta]}{\beta\alpha\left[z_2\left\{\frac{\alpha z_2}{2}+\sqrt{\left(\frac{\alpha z_2}{2}\right)^2+1}\right\} - z_1\left\{\frac{\alpha z_1}{2}+\sqrt{\left(\frac{\alpha z_1}{2}\right)^2+1}\right\}\right]}. \tag{4.10}$$

When the median moves away from the midpoint of the specification limits, we consider $\mathrm{C}_{\mathrm{pk}}^{\mathrm{BS}} = \min\{\mathrm{C}_{\mathrm{pl}}^{\mathrm{BS}}, \mathrm{C}_{\mathrm{pu}}^{\mathrm{BS}}\}$, where $\mathrm{C}_{\mathrm{pl}}^{\mathrm{BS}}$ and $\mathrm{C}_{\mathrm{pu}}^{\mathrm{BS}}$ are given in (4.9) and (4.10), respectively. Rewriting expression in (4.8), we have

$$\mathrm{C}_{\mathrm{p}}^{\mathrm{BS}} = \frac{\mathrm{USL}-\mathrm{LSL}}{x(p_2)-x(p_1)} = \frac{2[\mathrm{USL}-\mathrm{LSL}]/\beta}{\left[\alpha^2{z_2}^2+\alpha z_2\sqrt{\{\alpha z_2\}^2+4}\right] - \left[\alpha^2{z_1}^2+\alpha z_1\sqrt{\{\alpha z_1\}^2+4}\right]}. \tag{4.11}$$

Figure 4.1 depicts a process for which the quality characteristic follows a BS distribution, with $[1-p_1-p_2]\times 100\% = 99.73\%$ and $p_1 = p_2 = 0.00135$.

Figura 4.1: upper and lower limits of natural tolerance for the BS distribution.



4.3.2 Estimation and inference for the BS PCI

This section presents the ML estimator of the BS PCI as well as its asymptotic distribution. Also, interval estimation and hypothesis test are presented.

4.3.2.1 ML estimator of the BS PCI

Due to the invariance property of the ML estimators and using (4.11), an estimator of the PCI is given by

$$\widehat{\mathrm{C}}_{\mathrm{p}}^{\mathrm{BS}} = \frac{2[\mathrm{USL}-\mathrm{LSL}]/\widehat{\beta}}{\left[\widehat{\alpha}^2{z_2}^2+\widehat{\alpha}z_2\sqrt{\{\widehat{\alpha}z_2\}^2+4}\right] - \left[\widehat{\alpha}^2 z_1^2+\widehat{\alpha}z_1\sqrt{\{\widehat{\alpha}z_1\}^2+4}\right]}, \tag{4.12}$$

where $\widehat{\alpha}$ and $\widehat{\beta}$ are the ML estimators of α and β, and z_1 and z_2 are the $p_1 \times 100$th and $p_2 \times 100$th quantiles of the N(0, 1) distribution, respectively.

4.3.2.2 Asymptotic distribution of the ML estimator of the BS PCI

Under some regularity conditions (see COX and HINKLEY, 1974), $\widehat{\mathrm{C}}_{\mathrm{p}}^{\mathrm{BS}}$ has a distribution that is asymptotically normal. Then,

$$\sqrt{n}\big[\widehat{\mathrm{C}}_{\mathrm{p}}^{\mathrm{BS}} - \mathrm{C}_{\mathrm{p}}^{\mathrm{BS}}\big] \xrightarrow{\mathcal{D}} \mathrm{N}\big(0, \mathrm{Var}[\widehat{\mathrm{C}}_{\mathrm{p}}^{\mathrm{BS}}]\big), \quad \text{as } n \to \infty, \tag{4.13}$$

where $\widehat{\mathrm{C}}_{\mathrm{p}}^{\mathrm{BS}}$ is given in (4.12). Note that, by using the delta method, we have

$$\mathrm{Var}[\widehat{\mathrm{C}}_{\mathrm{p}}^{\mathrm{BS}}] = \begin{pmatrix} \frac{\partial}{\partial\alpha}\mathrm{PCI}(\alpha,\beta) & \frac{\partial}{\partial\beta}\mathrm{PCI}(\alpha,\beta) \end{pmatrix} \begin{pmatrix} \frac{\alpha^2}{2n} & 0 \\ 0 & \frac{\beta^2}{n[1/4+1/\alpha^2+I(\alpha)]} \end{pmatrix} \begin{pmatrix} \frac{\partial}{\partial\alpha}\mathrm{PCI}(\alpha,\beta) \\ \frac{\partial}{\partial\beta}\mathrm{PCI}(\alpha,\beta) \end{pmatrix},$$

that is,

$$\mathrm{Var}[\widehat{\mathrm{C}}_{\mathrm{p}}^{\mathrm{BS}}] = \Big[\frac{\partial}{\partial\alpha}\mathrm{PCI}(\alpha,\beta)\Big]^2\frac{\alpha^2}{2n} + \Big[\frac{\partial}{\partial\beta}\mathrm{PCI}(\alpha,\beta)\Big]^2\frac{\beta^2}{n[1/4+1/\alpha^2+I(\alpha)]}, \tag{4.14}$$

where $I(\alpha)$ is given in (4.7) and

$$\frac{\partial}{\partial\alpha}\mathrm{PCI}(\alpha,\beta) = -\frac{2\beta^{-1}[\mathrm{USL}-\mathrm{LSL}]\left[2\alpha z_2^2 + z_2\sqrt{\{\alpha z_2\}^2+4} + \frac{z_2^3\alpha^2}{\sqrt{\{\alpha z_2\}^2+4}} - 2\alpha z_1^2 - z_1\sqrt{\{\alpha z_1\}^2+4} - \frac{z_1^3\alpha^2}{\sqrt{\{\alpha z_1\}^2+4}}\right]}{\left[\alpha^2 z_2^2 + z_2\alpha\sqrt{\{\alpha z_2\}^2+4} - \alpha^2 z_1^2 - z_1\alpha\sqrt{\{\alpha z_1\}^2+4}\right]}, \tag{4.15}$$

$$\frac{\partial}{\partial\beta}\mathrm{PCI}(\alpha,\beta) = -\frac{2\beta^{-2}[\mathrm{USL}-\mathrm{LSL}]}{\left[\alpha^2 z_2^2 + z_2\alpha\sqrt{\{\alpha z_2\}^2+4} - \alpha^2 z_1^2 - z_1\alpha\sqrt{\{\alpha z_1\}^2+4}\right]}. \tag{4.16}$$

4.3.2.3 Parametric confidence interval for the BS PCI

Based on Equation (4.13), an approximate $100 \times [1-\omega]\%$ confidence interval (CI) for the BS PCI is

$$\mathrm{CI}_{100\times[1-\omega]\%}(\mathrm{C}_{\mathrm{p}}^{\mathrm{BS}}) = \Big[\widehat{\mathrm{C}}_{\mathrm{p}}^{\mathrm{BS}} \pm z(1-\omega/2)]\,\widehat{\mathrm{SE}}[\widehat{\mathrm{C}}_{\mathrm{p}}^{\mathrm{BS}}]\Big],$$

where $\mathrm{SE}[\widehat{\mathrm{C}}_{\mathrm{p}}^{\mathrm{BS}}] = \sqrt{\mathrm{Var}[\widehat{\mathrm{C}}_{\mathrm{p}}^{\mathrm{BS}}]}$.

4.3.2.4 Nonparametric confidence interval for the BS PCI

Bootstrap techniques have the advantage of being free from assumptions of the distribution of the estimator of the PCI; see Efron and Tibshirani (1986) and Franklin and Gary (1991). Specifically, let $X_1, \ldots, X_n$ be the original random sample from a process

with distribution $F(\cdot)$. A bootstrap sample $X_1^*, \ldots, X_n^*$ is obtained from a drawn with replacement from the original sample. Here, we set the number of bootstrap replications equal to 1000. In order to calculate a CI for $\mathrm{C_p^{BS*}}$, we consider the two following methods:

(i) Normal-approximation bootstrap method: first, we compute the estimator of $\mathrm{C_p^{BS}}$ from (4.12) based on the jth bootstrap sample, namely, $\widehat{\mathrm{C}}_{\mathrm{p}}^{\mathrm{BS*}}(\mathrm{j})$, with $j = 1, \ldots, B$. We assume that the distribution of $\widehat{\mathrm{C}}_{\mathrm{p}}^{\mathrm{BS}} - \mathrm{C_p^{BS}}$ is $\mathrm{N}(0, [\mathrm{SE}_{\widehat{\mathrm{C}}_{\mathrm{p}}^{\mathrm{BS}}}]^2)$. Since $\mathrm{SE}_{\widehat{\mathrm{C}}_{\mathrm{p}}^{\mathrm{BS}}}$ is unknown, we can estimate it as

$$\widehat{\mathrm{SE}}_{\widehat{\mathrm{C}}_{\mathrm{p}}^{\mathrm{BS*}}} = \sqrt{\frac{1}{B}\sum_{j=1}^{B}\left[\widehat{\mathrm{C}}_{\mathrm{p}}^{\mathrm{BS*}}(j) - \bar{\widehat{\mathrm{C}}}_{\mathrm{p}}^{\mathrm{BS*}}\right]^2},$$

where $\bar{\widehat{\mathrm{C}}}_{\mathrm{p}}^{\mathrm{BS*}} = [1/B]\sum_{j=1}^{B}\widehat{\mathrm{C}}_{\mathrm{p}}^{\mathrm{BS*}}(j)$. Thus, a CI of the $100\times[1-\omega]\%$ for $\mathrm{C_p^{BS*}}$ can be obtained using a normal approximation

$$\mathrm{CI}_{100\times[1-\omega]\%}(\mathrm{C_p^{BS}}) = \left[\widehat{\mathrm{C}}_{\mathrm{p}}^{\mathrm{BS}} \pm z(1-\omega/2)\,\widehat{\mathrm{SE}}_{\widehat{\mathrm{C}}_{\mathrm{p}}^{\mathrm{BS*}}}\right].$$

(ii) Bootstrap percentile method: let $\mathrm{C_p^{BS*}}(1) \leq \cdots \leq \mathrm{C_p^{BS*}}(B)$ be the bootstrap estimates in ascending order. Then, $\mathrm{C_p^{BS*}}(B[\omega/2])$ and $\mathrm{C_p^{BS*}}(B[1-\omega/2])$ are the $[\omega/2]\times 100$th and $[1-\omega/2]\times 100$th percentile of the distribution of $\widehat{\mathrm{C}}_{\mathrm{p}}^{\mathrm{BS*}}(j)$. A CI of the $100\times[1-\omega]\%$ for $\mathrm{C_p^{BS*}}$ is

$$\mathrm{CI}_{100\times[1-\omega]\%}(\mathrm{C_p^{BS}}) = \left[\widehat{\mathrm{C}}_{\mathrm{p}}^{\mathrm{BS*}}(B[\omega/2]), \widehat{\mathrm{C}}_{\mathrm{p}}^{\mathrm{BS*}}(B[1-\omega/2])\right].$$

4.3.2.5 Hypothesis testing for the BS PCI

Hypothesis $\mathrm{H_0}$: $\mathrm{C_p^{BS}} \leq C_0$ (the process is not capable) against $\mathrm{H_1}$: $\mathrm{C_p^{BS}} > C_0$ (the process is capable) can be tested by using

$$W = \frac{[\widehat{\mathrm{C}}_{\mathrm{p}}^{\mathrm{BS}} - C_0]}{\widehat{\mathrm{SE}}[\widehat{\mathrm{C}}_{\mathrm{p}}^{\mathrm{BS}}]} \overset{\mathrm{H_0}}{\sim} \mathrm{N}(0,1),$$

where C_0 is a predetermined capability requirement.

4.3.3 Selecting the optimal percentage specified for the BS PCI

Optimum selection of p_1 and $p_2 = [1-\varrho] + p_1$, given ϱ, can be done by minimizing the variance of the estimator of the BS PCI, that is, by minimizing

$$\mathrm{Var}[\widehat{\mathrm{C}}_{\mathrm{p}}^{\mathrm{BS}}] = V(p_1, p_2), \tag{4.17}$$

where $\mathrm{Var}[\widehat{\mathrm{C}}_{\mathrm{p}}^{\mathrm{BS}}]$ is given in (4.14). We now must find the optimal values of p_1 and p_2. First, we need the derivatives

$$\frac{\partial}{\partial p_1}V(p_1,p_2) = \frac{\alpha^2\psi(p_1,p_2)}{n}\frac{\partial}{\partial p_1}\psi(p_1,p_2)+\frac{2\beta^2\omega(p_1,p_2)}{n[1/4+1/\alpha^2+I(\alpha)]}\frac{\partial}{\partial p_1}\omega(p_1,p_2), \quad (4.18)$$

and

$$\frac{\partial}{\partial p_2}V(p_1,p_2) = \frac{\alpha^2\psi(p_1,p_2)}{n}\frac{\partial}{\partial p_2}\psi(p_1,p_2)+\frac{2\beta^2\omega(p_1,p_2)}{n[1/4+1/\alpha^2+I(\alpha)]}\frac{\partial}{\partial p_2}\omega(p_1,p_2), \quad (4.19)$$

where $V(p_1,p_2)$ is given in (4.17), $\psi(p_1,p_2)=\partial\mathrm{PCI}(\alpha,\beta)/\partial\alpha$ and $\omega(p_1,p_2)=\partial\mathrm{PCI}(\alpha,\beta)/\partial\beta$, are as given in (4.15) and (4.16), respectively, and

$$\begin{aligned}
\frac{\partial}{\partial p_1}\psi(p_1,p_2) &= -\frac{2[\mathrm{USL}-\mathrm{LSL}]\left[-4\alpha z_1 z_1' + \frac{\alpha^4 {z_1}^4 z_1'}{\{\alpha^2{z_1}^2+4\}^{3/2}} - \frac{4\alpha^2{z_1}^2 z_1'}{\sqrt{\alpha^2{z_1}^2+4}} - \sqrt{\alpha^2{z_1}^2+4}z_1'\right]}{\beta\left[-\alpha^2{z_1}^2-\alpha z_1\sqrt{\alpha^2{z_1}^2+4}+\alpha^2{z_2}^2+\alpha z_2\sqrt{\alpha^2{z_2}^2+4}\right]} \\
&\quad + \left[2\{\mathrm{USL}-\mathrm{LSL}\}\left\{-2\alpha{z_1}^2 - \frac{\alpha^2{z_1}^3}{\sqrt{\alpha^2{z_1}^2+4}} - z_1\sqrt{\alpha^2{z_1}^2+4}+2\alpha{z_2}^2+\frac{\alpha^2{z_2}^3}{\sqrt{\alpha^2{z_2}^2+4}}+z_2\sqrt{\alpha^2{z_2}^2+4}\right\}\right. \\
&\quad \left.\left\{-2\alpha^2 z_1 z_1' - \frac{\alpha^3{z_1}^2 z_1'}{\sqrt{\alpha^2{z_1}^2+4}} - \alpha\sqrt{\alpha^2{z_1}^2+4}z_1'\right\}\right] \Big/ \left[\beta\left\{-\alpha^2{z_1}^2-\alpha z_1\sqrt{\alpha^2{z_1}^2+4}+\alpha^2{z_2}^2\right.\right. \\
&\quad \left.\left.+\alpha z_2\sqrt{\alpha^2{z_2}^2+4}\right\}^2\right], \\
\frac{\partial}{\partial p_1}\omega(p_1,p_2) &= \frac{2[\mathrm{USL}-\mathrm{LSL}]\left[-2\alpha^2 z_1 z_1' - \frac{\alpha^3{z_1}^2 z_1'}{\sqrt{\alpha^2{z_1}^2+4}} - \alpha\sqrt{\alpha^2{z_1}^2+4}z_1'\right]}{\beta^2\left[-\alpha^2{z_1}^2-\alpha z_1\sqrt{\alpha^2{z_1}^2+4}+\alpha^2{z_2}^2+\alpha z_2\sqrt{\alpha^2{z_2}^2+4}\right]^2},
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial p_2}\psi(p_1,p_2) &= -\frac{2[\mathrm{USL}-\mathrm{LSL}]\left[4\alpha z_2 z_2' - \frac{\alpha^4 {z_2}^4 z_2'}{\{\alpha^2{z_2}^2+4\}^{3/2}} + \frac{4\alpha^2{z_2}^2 z_2'}{\sqrt{\alpha^2{z_2}^2+4}} + \sqrt{\alpha^2{z_2}^2+4}z_2'\right]}{\beta\left[-\alpha^2{z_1}^2-\alpha z_1\sqrt{\alpha^2{z_1}^2+4}+\alpha^2{z_2}^2+\alpha z_2\sqrt{\alpha^2{z_2}^2+4}\right]} \\
&\quad + \left[2\{\mathrm{USL}-\mathrm{LSL}\}\left\{-2\alpha{z_1}^2 - \frac{\alpha^2{z_1}^3}{\sqrt{\alpha^2{z_1}^2+4}} - z_1\sqrt{\alpha^2{z_1}^2+4}+2\alpha{z_2}^2+\frac{\alpha^2{z_2}^3}{\sqrt{\alpha^2{z_2}^2+4}}+z_2\sqrt{\alpha^2{z_2}^2+4}\right\}\right. \\
&\quad \left.\left\{2\alpha^2 z_2 z_2' + \frac{\alpha^3{z_2}^2 z_2'}{\sqrt{\alpha^2{z_2}^2+4}} + \alpha\sqrt{\alpha^2{z_2}^2+4}z_2'\right\}\right] \Big/ \left[\beta\left\{-\alpha^2{z_1}^2-\alpha z_1\sqrt{\alpha^2{z_1}^2+4}+\alpha^2{z_2}^2\right.\right. \\
&\quad \left.\left.+\alpha z_2\sqrt{\alpha^2{z_2}^2+4}\right\}^2\right], \\
\frac{\partial}{\partial p_2}\omega(p_1,p_2) &= \frac{2[\mathrm{USL}-\mathrm{LSL}]\left[2\alpha^2 z_2 z_2' + \frac{\alpha^3{z_2}^2 z_2'}{\sqrt{\alpha^2{z_2}^2+4}} + \alpha\sqrt{\alpha^2{z_2}^2+4}z_2'\right]}{\beta^2\left[-\alpha^2{z_1}^2-\alpha z_1\sqrt{\alpha^2{z_1}^2+4}+\alpha^2{z_2}^2+\alpha z_2\sqrt{\alpha^2{z_2}^2+4}\right]^2},
\end{aligned}$$

where $z_1' = \partial z(p_1)/\partial p_1$ and $z_2' = \partial z(p_2)/\partial p_2$. Now, in order to select the optimal values of p_1 and $p_2 = [1-\varrho]+p_1$, given ϱ, we must solve the system of two equations in p_1 and p_2 formed by (4.18) and (4.19). Note that closed-form expressions for optimal values of p_1 and p_2 cannot be obtained, so that iterative numerical methods must be used for determining them. We use the built-in `optim` function of the `R` software to solve this minimization problem.

4.4 Simulation

In this section, we carry out MC simulations to analyze the performance of C'_{pk} index defined in (4.3) (with $C'_{pk} = C'_p = 1$, that is, the process is centered at the midpoint of the specifications) in the case of non-normal distributions. We consider the BS distribution with density (4.4), and the log-normal (LN), three-parameter gamma (3-gamma) and three-parameter Weibull (3-Weibull) distributions with densities given by

$$\begin{aligned}
f_{\mathrm{LN}}(x;\mu,\upsilon) &= \frac{1}{x\upsilon\sqrt{2\pi}}\exp\left(-\frac{[\ln(x)-\mu]^2}{2\upsilon^2}\right), \quad -\infty<\mu<\infty, x, \upsilon>0,\\
f_{\mathrm{3GA}}(x;r,\lambda,\gamma) &= \frac{[(x-\gamma)^{r-1}]}{\lambda^r\Gamma(r)}\exp\left(-\frac{[x-\gamma]}{\lambda}\right), \quad x>0 \text{ or } \gamma, \gamma\in\mathbb{R}, r, \lambda>0,\\
f_{\mathrm{3WE}}(x;\delta,\eta,\gamma) &= \left[\frac{\eta(x-\gamma)^{\eta-1}}{\delta^\eta}\right]\exp\left(-\frac{[x-\gamma]^\eta}{\delta^\eta}\right), \quad x>0 \text{ or } \gamma, \gamma\in\mathbb{R}, \eta, \delta>0,
\end{aligned}$$

respectively, where $\Gamma(\cdot)$ is the gamma function. The latter three distributions are chosen due to a variety of asymmetrical and positively skewed shapes. Considering the fact that `R` does not provide in-built functions for the BS, 3-gamma and 3-Weibull random number generators, the `rbs()` function in the `bs` package, and the `rweibull3()` and `rgamma3()` functions in the `FAdist` package are utilized, respectively. The simulation scenario assumes sample sizes $n \in \{10, 25, 50, 100, 200\}$. For the BS distribution, we consider high, moderate and low symmetry, that is, $(\alpha, \beta) \in \{(0.2, 4.88), (0.5, 1.75), (1.0, 0.67)\}$, respectively. Note that the considered value of α constrains the value of β to obtain $C_{pk}^{BS} = C_p^{BS} = 1$; see BS PCI given in (4.11). Figure 4.2 shows a graphical plot for the PDF of these three BS distributions, from which we note that, although all of them have $C_{pk}^{BS} = C_p^{BS} = 1$, the curves of the corresponding PDFs are very different, and each of them presenting a distinct shape. We generate 10,000 samples for each of different sizes from BS, log-normal, 3-gamma and 3-Weibull distributions. For each distribution and each sample size, we report the empirical mean, bias, SD, and the root of the mean squared error ($\sqrt{\mathrm{MSE}}$) of $\widehat{C}'_{pk}$ in Table 4.1.

From Table 4.1, we observe that, as the sample size increases, the empirical bias and $\sqrt{\mathrm{MSE}}$ of all the estimators decrease as expected. BS distributions with parameters $(\alpha, \beta) \in \{(0.2, 4.88), (0.5, 1.75)\}$, which are the low and moderate asymmetric cases, respectively, present the lowest empirical $\sqrt{\mathrm{MSE}}$s ($n \in \{50, 100, 200\}$) in relation to the other distributions considered. In particular, the Weibull case has a good performance for small sample sizes ($n \in \{10, 25\}$).

Figura 4.2: PDF of the indicated BS distributions for $C_{pk}^{BS} = C_{p}^{BS} = 1$.

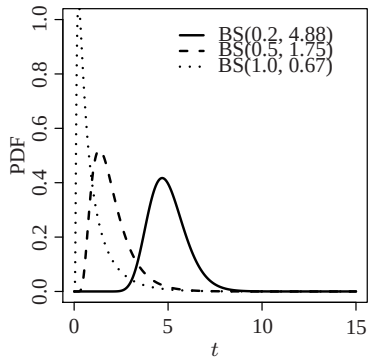


Tabela 4.1: simulation results for the indicated distributions

		log-normal	3-Weibull	3-gamma	BS		
					$\alpha = 0.20$	$\alpha = 0.50$	$\alpha = 1.00$
	n				$\beta = 4.88$	$\beta = 1.75$	$\beta = 0.67$
Mean	10	2.1720	1.8423	2.1319	2.2180	2.0565	4.7586
	25	1.6428	1.4501	1.6138	1.5262	1.4458	2.6209
	50	1.4179	1.2950	1.4004	1.2265	1.2102	1.9512
	100	1.2625	1.1830	1.2516	1.0481	1.0616	1.5584
	200	1.1537	1.1044	1.1466	0.9405	0.9593	1.3264
Bias	10	1.1720	0.8423	1.1319	1.2180	1.0565	3.7586
	25	0.6428	0.4501	0.6138	0.5262	0.4458	1.6209
	50	0.4179	0.2950	0.4004	0.2265	0.2102	0.9512
	100	0.2625	0.1830	0.2516	0.0481	0.0616	0.5584
	200	0.1537	0.1044	0.1466	0.0594	0.0406	0.3264
SD	10	0.7386	0.6049	0.7872	0.9627	0.8570	3.4667
	25	0.3914	0.3099	0.4034	0.4540	0.3937	1.3526
	50	0.2750	0.2190	0.2839	0.2707	0.2623	0.7316
	100	0.2069	0.1669	0.2128	0.1705	0.1900	0.4501
	200	0.1608	0.1256	0.1635	0.1176	0.1444	0.3102
$\sqrt{\text{MSE}}$	10	1.3853	1.0371	1.3787	1.5525	1.3604	5.1133
	25	0.7527	0.5465	0.7345	0.6950	0.5947	2.1111
	50	0.5003	0.3675	0.4909	0.3529	0.3361	1.2000
	100	0.3343	0.2477	0.3295	0.1772	0.1998	0.7172
	200	0.2225	0.1634	0.2196	0.1318	0.1500	0.4503

4.5 Application to real data

To illustrate the methodology developed in this work, we apply it to a real data set presented in Hsu et al. (2008).

4.5.1 Problem: manufacture of integrated circuits

Below, we detail the problem upon study, present the data and analyze them.

4.5.1.1 Description of the problem

The manufacture of integrated circuits comprises the initial process of wafer and the final process of packaging. In an integrated circuit packaging factory, the manufacturing process generally includes the following principal steps: die sawing, die mounting, wire bonding, molding, trimming and forming, marking, plating and testing. The wire bonding is the most common way to provide an electrical connection from the integrated circuit apparatus to the lead-frame. It is done by using an ultra-thin gold or aluminum wire to form the electrical interconnection between the chip and package leads. The high-speed wire bonding equipment consists of a control system to feed the lead frame towards the work area. The image recognition system guarantees that the die is oriented to match the bonding diagram for a particular device. The wires are bonded one at a time with two bonds for each connection: one in the die (first bond) and the other in the lead frame (second bond). The first bond requires a ball formation that is put within the bond pad opening on the die, under load and ultrasonic energy for a few milliseconds, forming a ball bond to the bond pad metal. In the wire bonding process, one of the most important factors that is directly related to the quality level, is the ball size. Because the process can be interrupted and shut down when the width between the two bond balls is too small, the bond ball size must be considered. The goal of this application is to determine the production process capability of balls for electrical connections. To achieve this objective, it is established that the proposed LSL and USL for the ball size are 0.5 mil and 8 mil (1 mil = 1/1000 in. = 0.0254 mm), respectively.

4.5.1.2 Data

The quality characteristic X under study to determine the production process capability is the ball size (in mm). A portion of historical data has been collected and shown in Table 4.2.

4.5.1.3 Data analysis

In order to implement the PCIs for the BS distribution, we check that the data of Table 4.2 fits this distribution. First, we provide a descriptive statistics of the data using the function `descriptiveSummary()` of the `bs` 2.0 package. Descriptive statistics displayed in Table 4.3 allows a BS distribution to be reasonably assumed for modeling ball size data, due to their asymmetric nature and their level of kurtosis. Figure 4.3 shows graphical plots of the autocorrelation function (ACF) and partial ACF for the ball size data set, from where is not

Tabela 4.2: 100 observations collected from historical data of the ball size (in mm) for electrical connections.

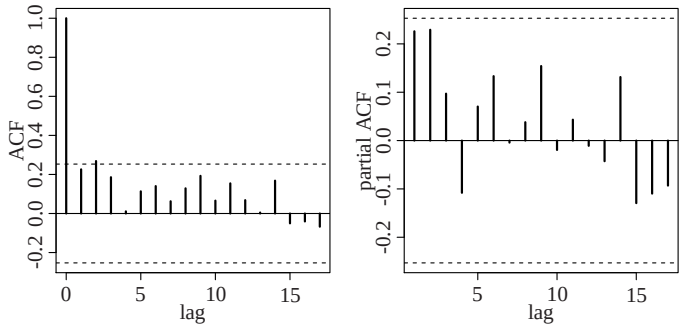
2.891	4.035	4.495	2.890	2.312	3.158	5.228	3.334	5.896	5.639
3.842	1.590	1.954	1.842	0.680	2.752	1.301	2.260	0.889	2.381
0.619	2.788	1.050	3.750	3.508	6.123	6.549	5.954	2.207	4.417
4.805	1.516	2.227	2.797	1.636	1.066	0.940	4.101	4.542	1.295
1.770	3.492	5.706	3.722	6.644	2.472	1.383	4.494	1.694	2.892
2.111	3.591	2.093	3.222	2.891	2.582	0.665	3.234	1.102	1.083
1.508	1.811	2.803	6.659	0.923	6.229	3.177	2.333	1.311	4.419
2.495	0.921	4.061	9.725	1.600	4.281	3.360	1.131	1.618	4.489
3.696	1.982	2.413	5.480	1.992	2.573	1.845	4.620	6.221	1.694
4.882	1.380	3.982	2.260	2.366	2.899	3.782	2.336	1.175	3.055

noted the presence of serial correlation.

Tabela 4.3: descriptive statistics of the ball size for electrical connections (in mm).

Median	Mean	SD	CS	CK	Min	Max	n
2.77	3.036	1.715	0.984	4.048	0.619	9.725	100

Figura 4.3: autocorrelation and partial autocorrelation functions of the ball size data set.



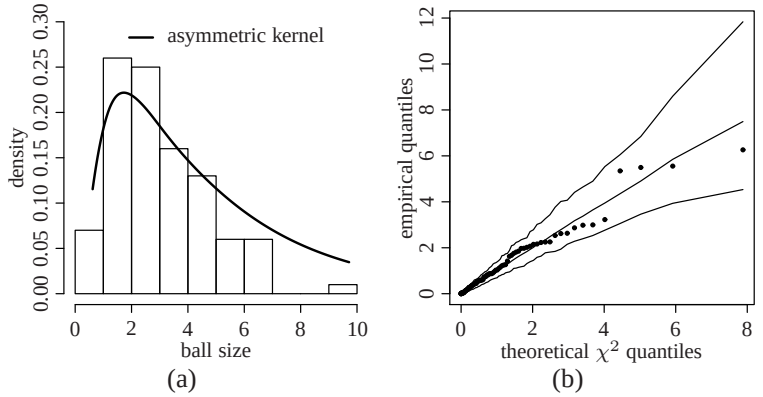
(a) (b)

Second, in order to detect adequacy of the model to the data, we apply the Kolmogorov-Smirnov (KS) goodness-of-fit test by using the function `ksbs()` of the `bs` 2.0 package. The result of the application of the KS test by using the package is presented next:

```
One-sample Kolmogorov-Smirnov test
data: x
D = 0.0591,
p-value = 0.8758
alternative hypothesis: two-sided
```

The KS test indicates that there is not sufficient statistical evidence as for supporting that the data do not follow a BS distribution (p-value $= 0.8758$). Thus, the BS distribution is a very good option to model these data. In Figure 4.4, we provide a histogram with the nonparametric density estimate based on asymmetric kernels proposed by Saulo et al. (2013) and Marchant et al. (2013), and a probability plot with envelope obtained with the function `envelopeBS()` of the `bs` 2.0 package. These graphical goodness-of-fit methods support the result obtained by the KS test.

Figura 4.4: histogram and envelope of the data set under study



Third, the parameters of the BS distribution are estimated via the function `mlebs()` from the `bs` 2.0 package. Using this function, we have the following estimates of α and β:

```
$alpha
[1] 0.6144122
$beta
[1] 2.55106
```

With the estimated parameters and the information provided in the description of the problem given in Section 4.5.1.1, we compute the PCIs for the BS distribution using the expression given in (4.3). In addition, we compute the CI bootstrap as described in Section 4.3.2.4 using the function `boot.ci()` from the `boot` package. Note that $\widehat{C}_p^{BS} = 2.847$ and its CIs are given by

```
Intervals :
Level      Normal              Percentile
95%   ( 2.528,  3.157 )   ( 2.551,  3.176 )
Calculations and Intervals on Original Scale
```

Note, however, that the production process is not centered relative to the specifications. For this reason, we must estimate $\widehat{C}_{pk}^{BS} = \min\{\widehat{C}_{pl}^{BS}, \widehat{C}_{pu}^{BS}\}$, that is, $\widehat{C}_{pk}^{BS} = 1.734$. The corresponding CIs are given by

```
Intervals :
Level      Normal              Percentile
95%   ( 1.684,  1.784 )   ( 1.681,  1.784 )
Calculations and Intervals on Original Scale
```

Note also that a gauge of how off-center the process is operating is given by the magnitude of $\widehat{C}_{pk}^{BS}$ relative to $\widehat{C}_{p}^{BS}$. According to Table 4.4, which presents some recommended minimum values of the PCI (see MONTGOMERY, 2005, p. 337), we can conclude that the integrated circuit manufacturing process is capable, since the lower confidence limit is greater than 1.5, that is, the recommended minimum value for existing processes with two-sided specifications and involving critical parameters.

Tabela 4.4: recommended minimum values of the PCI.

Situation	Two-sided specifications	One-sided specifications
Existing process	1.33	1.25
New process	1.50	1.45
Existing process including safety, strength, or critical parameters	1.50	1.45
New process including safety, strength, or critical parameters	1.67	1.60

4.6 Conclusions and future works

Process capability analysis has become an important tool integral to applications of statistical process control for continuous improvement of quality and productivity. The use of process capability indices under normality as adequate as well as the ignorance of the effect of asymmetric distributions can lead to misinterpretation of process capability. It is known that the use of symmetric distributions is common in practical situations, but the asymmetric distribution must also be considered in the manufacturing industry. That is why several authors have proposed indices for non-normal processes. Several approaches have been analyzed to address the problem of process capability for non-normally distributed data. In this chapter, we proposed a methodology to analyze productive process capability indices when the quality characteristic follows a Birnbaum-Saunders distribution. These indices are based on the interquartile process rather than the process variability, as is the case for symmetric data.

We also proposed a technique for obtaining the optimum quantile values used to minimize the estimation error. In addition, we have performed a simulation study using presented process capability indices and the results showed that these indices are better for processes with higher asymmetric. Moreover, we have implemented a code in R language to calculate the four proposed indices. In order to illustrate the development of these indices we performed an application using real data. This application showed the convenience of using the proposed methodology.

5 CONCLUDING REMARKS

No decorrer dessa tese exploramos algumas aplicações da distribuição Birnbaum-Saunders e duas de suas generalizações, isto é, as distribuições Birnbaum-Saunders generalizadas assimétricas e misturas de escala Birnbaum-Saunders. De forma sucinta, nossos principais resultados são elencados abaixo.

1. Estudos de Monte Carlo indicaram que dentre os estimadores não-paramétricos por função-núcleo estudados, aquele baseado na distribuição exponential potência Birnbaum-Saunders assimétrica apresenta os melhores resultados.

2. Uma análise usando dados reais de transações financeiras da bolsa de valores de Nova Iorque mostrou que o modelo autoregressivo de duração condicional baseado na distribuição exponential potência Birnbaum-Saunders é superior aos demais modelos.

3. Estudos de simulação mostraram que índices de capacidade do processo baseado em uma distribuição mais assimétrica, nesse caso a distribuição Birnbaum-Saunders, apresenta um melhor desempenho.

Considerando o que foi desenvolvido nessa tese, surgem perspectivas de desenvolvimento de algumas linhas de pesquisa, as quais são expostas abaixo.

1. O uso da função-núcleo, basedo nas distribuições Birnbaum-Saunders generalizadas assimétricas, para estimação da função de risco. Tal perspectiva já está sendo elaborada.

2. É importante considerar modelos autorregressivos de duração condicional não-lineares, em que a duração mediana depende não-linearmente de variáveis ââde informações passadas.

3. O uso da metodologia proposta a diferentes características do processo, tal como processos com apenas limite inferior (superior). Também, o uso de distribuições mais

robustas, como a Birnbaum-Saunders t-Student, para os índices de capacidade do processo. Tais perspectivas também já estão sendo elaboradas.

REFERENCES

AARSET, M. V. How to identify a bathtub hazard rate. **IEEE Trans. Rel.**, v.36, p.106–108, 1987.

ABADIR, K. M.; LAWFORD, S. Optimal asymmetric kernels. **Econ. Letters**, v.83, p.61–68, 2004.

AHMED, S. E. et al. A truncated version of the Birnbaum-Saunders distribution with an application in financial risk. **Pak. J. Statist.**, v.26, p.293–311, 2010.

ALLEN, D. et al. Finite sample properties of the QMLE for the log-ACD model: application to Australian stocks. **J. Econometrics**, v.147, p.163–183, 2008.

ANIS, M. Z. Basic process capability indices: an expository review. **Intern. Statist. Rev.**, v.76, p.347–367, 2008.

ARNOLD, B. C.; BEAVER, R. J. The skew-Cauchy distribution. **Statist. Prob. Letters**, v.49, p.285–290, 2000.

ATKINSON, A. C. **Plots, Transformations and Regression**. Oxford: Oxford University Press, 1985.

AZEVEDO, C. et al. Shape and change point analyses of the Birnbaum-Saunders-t hazard rate and associated estimation. **Comp. Statist. Data Anal.**, v.56, p.3887–3897, 2012.

AZZALINI, A. A class of distributions which includes the normal ones. **Scand. J. Statist.**, v.12, p.171–178, 1985.

AZZALINI, A.; CAPITANIO, A. Statistical applications of the multivariate skew normal distribution. **J. R. Statist. Soc. B**, v.61, p.579–602, 1999.

AZZALINI, A.; CAPITANIO, A. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew-t distribution. **J. R. Statist. Soc. B**, v.65, p.367–389, 2003.

BALAKRISHNAN, N. et al. Mixture inverse Gaussian distribution and its transformations, moments and applications. **Statistics**, v.43, p.91–104, 2009.

BALAKRISHNAN, N. et al. Estimation in the Birnbaum-Saunders distribution based on scale-mixture of normals and the EM-algorithm. **Statist. Oper. Res. Trans.**, v.33, p.171–192, 2009.

BALAKRISHNAN, N.; LEIVA, V.; LÓPEZ, J. Acceptance sampling plans from truncated life tests from generalized Birnbaum-Saunders distribution. **Comm. Statist. Simul. Comp.**, v.36, p.643–656, 2007.

BARROS, M.; PAULA, G. A.; LEIVA, V. An R implementation for generalized Birnbaum-Saunders distributions. **Comp. Statist. Data Anal.**, v.53, p.1511–1528, 2009.

BAUWENS, L.; GIOT, P. The logarithmic ACD model: an application to the bid-ask quote process of three NYSE stocks. **Ann. Econ. Statist.**, v.60, p.117–149, 2000.

BHATTI, C. R. The Birnbaum-Saunders autoregressive conditional duration model. **Math. Comp. Simul.**, v.80, p.2062–2078, 2010.

BIRNBAUM, Z. W.; SAUNDERS, S. C. A new family of life distributions. **J. Appl. Prob.**, v.6, p.319–327, 1969.

BOWMAN, A. W. An alternative method of cross-validation for the smoothing of density estimates. **Biometrika**, v.71, p.353–360, 1984.

BOX, G. E.; COX, D. R. An analysis of transformations. **J. R. Statist. Soc. B**, v.26, p.211–246, 1964.

CHAN, L. K.; CHENG, S. W.; SPIRING, F. A. The robustness of process capability index Cp to departures from normality. ***In*** **Statistical Theory and Data Analysis**, v.II, p.223–229, 1988. ed. K. Matusita, North-Holland, Amsterdam.

CHANG, T.-J.; K-H. CHANG, H.-M. K.; CHANG, Y.-S. Comparison of a new kernel method and a sampling volume method for estimating indoor particulate matter concentration with Lagrangian modeling. **Build. Environ.**, v.54, p.20–28, 2012.

CHEN, S. X. Beta kernel estimators for density functions. **Comp. Statist. Data Anal.**, v.31, p.131–145, 1999.

CHEN, S. X. Probability density function estimation using gamma kernels. **Ann. Inst. Statist. Math.**, v.52, p.471–480, 2000.

CLEMENTS, J. A. Process capability calculations for non-normal distributions. **Qual. Prog.**, v.22, p.95–100, 1989.

COX, D. R.; HINKLEY, D. V. **Theoretical Statistics**. London: Chapman & Hall, 1974.

COX, D.; SNELL, E. A general definition of residuals. **J. R. Statist. Soc. B**, v.30, p.248–275, 1968.

DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via EM algorithm. **J. R. Statist. Soc. B**, v.39, p.1–38, 1977.

DIAMOND, D. W.; VERRECHIA, R. E. Constraints on short-selling and asset price adjustments to private information. **J. Finan. Econ.**, v.18, p.277–311, 1987.

DÍAZ-GARCÍA, J. A.; LEIVA, V. A new family of life distributions based on the contoured elliptically distributions. **J. Statist. Plann. Infer.**, v.128, p.445–457, 2005.

DROST, F. C.; WERKER, B. J. M. Semiparametric duration models. **J. Bus. Econ. Stat.**, v.22, p.40–50, 2004.

EASLEY, D.; KIEFER, N. M.; O'HARA, M. The information content of the trading process. **J. Empir. Fin.**, v.4, p.159–186, 1997.

EASLEY, D.; O'HARA, M. Time and the process of security price adjustment. **J. Fin.**, v.47, p.577–605, 1992.

EFRON, B.; TIBSHIRANI, R. J. Bootstrap method for standard errors, confidence intervals, and other measures of statistical accuracy. **Statist. Sc.**, v.1, p.54–77, 1986.

ENGLE, R. F. The econometrics of ultra high frequency data. **Econometrica**, v.68, p.1–22, 2000.

ENGLE, R.; RUSSELL, J. Autoregressive conditional duration: a new method for irregularly spaced transaction data. **Econometrica**, v.66, p.1127–1162, 1998.

FERNANDES, M.; GRAMMIG, J. A family of autoregressive conditional duration models. **J. Econometrics**, v.130, p.1–23, 2006.

FERNANDES, M.; MONTEIRO, P. K. Central limit theorem for asymmetric kernel functionals. **Ann. Inst. Statist. Math.**, v.57, p.425–442, 2005.

FERREIRA, C. S.; BOLFARINEA, H.; LACHOS, V. H. Skew scale mixtures of normal distributions: properties and estimation. **Statist. Method.**, v.8, p.154–171, 2011.

FERREIRA, M.; GOMES, M. I.; LEIVA, V. On an extreme value version of the Birnbaum-Saunders distribution. **RevStat Statist. J.**, v.10, p.181–210, 2012.

FRANKLIN, L. R.; GARY, W. Bootstrap confidence interval estimates of Cpk: an introduction. **Comm. Statist. Simul. Comp.**, v.20, p.231–242, 1991.

GILCHRIST, W. Modeling capability. **J. Oper. Res. Soc.**, v.44, p.909–923, 1993.

GIOT, P. Time transformations, intraday data and volatility models. **J. Comp. Finance**, v.4, p.31–62, 2000.

GÓMEZ, H. W.; VENEGAS, O.; BOLFARINE, H. Skew-symmetric distributions generated by the normal distribution function. **Environmetrics**, v.18, p.395–407, 2007.

GRAMMIG, J.; MAURER, K. Non-monotonic hazard functions and the autoregressive conditional duration model. **Economet. J.**, v.3, p.16–38, 2000.

GUNTER, B. The use and abuse of Cpk: part i. **Qual. Prog.**, v.22, p.79–80, 1989.

HAAN, P. de. On the use of density kernels for concentration estimations within particle and puff dispersion models. **Atmosp. Environ.**, v.33, p.2007–2021, 1999.

HOSSEINIFARD, S. Z. et al. A transformation technique to estimate the process capability index for non-normal processes. **Int. J. Adv. Manuf. Techn.**, v.40, p.512–517, 2009.

HSU, Y.; PEARN, W.; WU, P. Capability adjustment for gamma processes with mean shift consideration in implementing six sigma program. **Europ. J. Oper. Res.**, v.119, p.517–529, 2008.

HUBERT, M.; VANDERVIEREN, E. An adjusted boxplot for skewed distributions. **Comp. Statist. Data Anal.**, v.52, p.5186–5201, 2008.

JIN, X.; KAWCZAK, J. Birnbaum-Saunders and lognormal kernel estimators for modelling durations in high frequency financial data. **Ann. Econ. Finance**, v.4, p.103–124, 2003.

JOHNSON, M. Statistics simplified. **Qual. Prog.**, v.25, p.10–11, 1992.

JOHNSON, N. L. System of frequency curve generated by methods of translation. **Biometrika**, v.36, p.149–176, 1949.

JOHNSON, N. L.; KOTZ, S.; BALAKRISHNAN, N. **Continuous Univariate Distributions**. New York: Wiley, 1995. v.2.

JOHNSON, N. L.; KOTZ, S.; PEARN, W. L. Flexible process capability indices. **Pak. J. Statist.**, v.10, p.23–31, 1994.

KANE, V. E. Process capability indices. **J. Qual. Techn.**, v.18, p.41–52, 1986.

KOTZ, S.; JOHNSON, N. L. Process capability indices - a review, 1992-2000. **J. Qual. Techn.**, v.34, p.2–53, 2002.

KOTZ, S.; LEIVA, V.; SANHUEZA, A. Two new mixture models related to the inverse Gaussian distribution. **Meth. Comp. Appl. Prob.**, v.12, p.199–212, 2010.

KUNDU, D.; KANNAN, N.; BALAKRISHNAN, N. On the hazard function of Birnbaum-Saunders distribution and associated inference. **Comp. Statist. Data Anal.**, v.52, p.2692–2702, 2008.

LEIVA, V. et al. Random number generators for the generalized Birnbaum-Saunders distribution. **J. Statist. Comp. Simul.**, v.78, p.1105–1118, 2008.

LEIVA, V. et al. A skewed sinh-normal distribution and its properties and application to air pollution. **Comm. Statist. Theor. Meth.**, v.39, p.426–443, 2010.

LEIVA, V. et al. Modeling wind energy flux by a Birnbaum-Saunders distribution with unknown shift parameter. **J. Appl. Statist.**, v.38, p.2819–2838, 2011.

LEIVA, V. et al. New control charts based on the Birnbaum-Saunders distribution and their implementation. **Col. J. Statist.**, v.34, p.147–176, 2011.

LEIVA, V. et al. Fatigue statistical distributions useful for modeling diameter and mortality of trees. **Col. J. Statist.**, v.35, p.349–367, 2012.

LEIVA, V. et al. Birnbaum-Saunders statistical modelling: a new approach. **Statist. Mod.**, p.(in press), 2013.

LEIVA, V.; PAULA, M. B. G. A.; SANHUEZA, A. Generalized Birnbaum-Saunders distributions applied to air pollutant concentration. **Environmetrics**, v.19, p.235–249, 2008.

LEIVA, V.; SANHUEZA, A.; ANGULO, J. M. A length-biased version of the Birnbaum-Saunders distribution with application in water quality. **Stoch. Environ. Res. Risk. Assess.**, v.23, p.299–307, 2009.

LIO, Y. L.; TSAI, T.-R.; WU, S.-J. Acceptance sampling plans from truncated life tests based on the Birnbaum-Saunders distribution for percentiles. **Comm. Statist. Simul. Comp.**, v.39, p.1–18, 2010.

LIU, P. H.; CHEN, F. L. Process capability analysis of non-normal process data using the Burr XII distribution. **Int. J. Adv. Manuf. Techn.**, v.27, p.975–984, 2006.

LOADER, C. R. Bandwidth selection: classical or plug-in? **Ann. Statist.**, v.27, p.415–438, 1999.

LORIMER, G. S. The kernel method for air quality modelling–I: mathematical foundation. **Atmosp. Environ.**, v.20, p.1447–1452, 1986.

LOVELACE, C. R.; SWAIN, J. J. Process capability analysis methodologies for zero-bound, non-normal process data. **Qual. Eng.**, v.21, p.190–202, 2009.

LUCA, G. D.; ZUCCOLOTTO, P. Regime-switching Pareto distributions for ACD models. **Comp. Statist. Data Anal.**, v.51, p.2179–2191, 2006.

MARCHANT, C. et al. Air contaminant statistical distributions with application to PM10 in Santiago, Chile. **Rev. Env. Cont. Tox.**, v.223, p.1–31, 2013.

MARCHANT, C. et al. Generalized Birnbaum-Saunders kernels to estimate densities applied to financial data. **Comp. Statist. Data Anal.**, v.63, p.1–15, 2013.

MCCOY, P. F. Using performance indexes to monitor production processes. **Qual. Prog.**, v.2, p.49–55, 1991.

MEITZ, M.; TERÄSVIRTA, T. Evaluating models of autoregressive conditional duration. **J. Bus. Econ. Stat.**, v.24, p.104–124, 2006.

MITTELHAMMER, R. C.; JUDGE, G. G.; MILLER, D. J. **Econometric Foundations**. 2000.

MONTGOMERY, D. **Statistical Quality Control**. New York: Wiley, 2005.

NADARAJAH, S. A truncated inverted beta distribution with application to air pollution data. **Stoch. Environ. Res. Risk. Assess.**, v.22, p.285–289, 2008.

NADARAJAH, S.; KOTZ, S. Skewed distributions generated by the normal kernel. **Statist. Prob. Letters**, v.65, p.269–277, 2003.

NELDER, J. A.; MEAD, R. A simplex method for function minimization. **Comp. J.**, v.7, p.308–313, 1965.

PACURAR, M. Autoregressive conditional durations models in finance: a survey of the theoretical and empirical literature. **J. Econ. Surv.**, v.22, p.711–751, 2008.

PAGNINI, G. The kernel method to compute the intensity of segregation for reactive pollutants: mathematical formulation. **Atmosp. Environ.**, v.43, p.3691–3698, 2009.

PARZEN, E. On the estimation of a probability density function and the mode. **Ann. Math. Statist.**, v.33, p.1065–1076, 1962.

PATHMANATHAN, D.; NG, K. H.; PEIRIS, M. S. S. On estimation of autoregressive conditional duration (ACD) models based on different error distributions. **S. Lankan J. Appl. Statist.**, v.10, p.251–269, 2009.

PAULA, G. A. et al. Robust statistical modeling using the Birnbaum-Saunders-t distribution applied to insurance. **Appl. Stoch. Mod. Bus. Ind.**, v.28, p.16–34, 2012.

PEARN, W. L.; CHEN, K. S. Estimating process capability indices for non-normal Pearson populations. **Qual. Rel. Eng. Int.**, v.11, p.386–388, 1995.

PEARN, W. L.; CHEN, K. S. Capability indices for non-normal distributions with an application in electrolytic capacitor manufacturing. **Microel. Rel.**, v.37, p.1853–1858, 1997.

PEARN, W. L.; CHEN, K. S. New generalizations of the process capability index Cpk. **J. Appl. Statist.**, v.25, p.801–810, 1998.

RIVERA, L. A. R.; HUBELE, N. F.; LAWRENCE, F. P. Cpk index estimation using data transformation. **Comp. Ind. Engin.**, v.29, p.55–58, 1995.

ROSENBLATT, M. Remarks on some nonparametric estimates of a density function. **Ann. Math. Statist.**, v.27, p.832–837, 1956.

RUDEMO, M. Empirical choice of histograms and kernel density estimators. **Scand. J. Statist.**, v.9, p.65–78, 1982.

SAULO, H. et al. A nonparametric method for estimating asymmetric densities based on skewed Birnbaum-Saunders distributions applied to environmental data. **Stoch. Envir. Res. Risk Ass.**, v.27, p.1479–1491, 2013.

SCAILLET, O. Density estimation using inverse and reciprocal inverse Gaussian kernels. **J. Nonpar. Statist.**, v.16, p.217–226, 2004.

SHORE, H. A new approach to analysing non-normal quality data with application to process capability analysis. **Intern. J. Prod. Res.**, v.36, p.1917–1933, 1998.

SOMERVILLE, S.; MONTGOMERY, D. Process capability indices and non-normal distributions. **Qual. Eng.**, v.19, p.305–316, 1996.

SPIRING, F. et al. A bibliography of process capability papers. **Qual. Rel. Eng. Int.**, v.19, p.445–460, 2003.

TANG, L. C.; TANG, S. E. Computing process capability indices for non-normal data: a review and comparative study. **Qual. Rel. Eng. Int.**, v.15, p.339–353, 1999.

TSAY, R. S. **Analysis of Financial Time Series**. 2002.

VäNNMAN, K.; ALBING, M. Process capability indices for one-sided specification intervals and skewed distributions. **Qual. Rel. Eng. Int.**, v.23, p.755–765, 2007.

VILCA, F. et al. An extended Birnbaum-Saunders model and its application in the study of environmental quality in Santiago, Chile. **Stoch. Environ. Res. Risk. Assess.**, v.24, p.771–782, 2010.

VILCA, F. et al. Estimation of extreme percentiles in Birnbaum-Saunders distributions. **Comp. Statist. Data Anal.**, v.55, p.1665–1678, 2011.

VILCA, F.; LEIVA, V. A new fatigue life model based on the family of skew-elliptical distributions. **Comm. Statist. Theor. Meth.**, v.35, p.229–244, 2006.

WANG, J.; GENTON, M. The multivariate skew-slash distribution. **J. Statist. Plann. Infer.**, v.136, p.209–220, 2006.

WEST, M. On scale mixtures of normal distributions. **Biometrika**, v.74, p.646–648, 1987.

WRIGHT, P. A. A process capability index sensitive to skewness. **J. Statist. Comp. Simul.**, v.52, p.195–203, 1995.

YUM, B. J.; KIM, K. W. A bibliography of the literature on process capability indices: 2000-2009. **Qual. Rel. Eng. Int.**, v.27, p.251–268, 2011.

APPENDIX

Proof of Proposition 2.4.1

Consider f is the density to be estimated and $\widehat{f}_{\text{skew-GBS}}$ is its skew-GBS kernel estimator. Then,

$$\mathbb{E}[\widehat{f}_{\text{skew-GBS}}(x)] = \int_0^\infty K_{\text{skew-GBS}(\sqrt{h},\, x,\, \lambda;\, g)}(u) f(u) \mathrm{d}u = \mathbb{E}[f(\xi_x)], \tag{5.1}$$

where $\xi_x \sim \text{skew-GBS}(\sqrt{h}, x, \lambda; g)$. Taking the expressions for the mean and variance of a skew-GBS distributed RV given in Vilca and Leiva (2006) and expressed in (2.5), we obtain

$$\begin{aligned} \mu_{\xi_x} = \mathbb{E}[\xi_x] &= x + \tfrac{xh}{2}\mathbb{E}[Z^2] + \tfrac{x\sqrt{h}}{2}\mathbb{E}\left[Z\{hZ^2+4\}^{1/2}\right], \\ \sigma_{\xi_x} = \mathrm{Var}[\xi_x] &= x^2 h\mathbb{E}[Z^2] - \tfrac{x^2h}{4}\left(\mathbb{E}\left[Z\{hZ^2+4\}^{1/2}\right]\right)^2 + \tfrac{x^2h^{3/2}}{2}\mathbb{E}\left[Z^3\{hZ^2+4\}^{1/2}\right] \\ &\quad - \tfrac{x^2h^{3/2}}{2}\mathbb{E}[Z^2]\mathbb{E}\left[Z\{hZ^2+4\}^{1/2}\right] - \tfrac{x^2h^2}{4}(\mathbb{E}[Z^2])^2 + \tfrac{x^2h^2}{2}\mathbb{E}\left[Z^4\right], \end{aligned} \tag{5.2}$$

where $Z \sim \text{SS}(0, 1, \lambda;\, g^2)$. Applying a Taylor-Lagrange expansion in (5.2) for the density f, we have $\mathbb{E}[f(\xi_x)] = f(\mu_{\xi_x}) + \frac{1}{2}f''(\mu_{\xi_x})\sigma_{\xi_x} + A$, where $\theta \in (0,1)$ and $A = \mathbb{E}[\{\xi_x - \mu_{\xi_x}\}^2 \{f''(\mu_{\xi_x} + \theta[\xi_x - \mu_{\xi_x}]) - f''(\mu_{\xi_x})\}]$.

Lemma 1

Under conditions (C1) and (C2), we have $A = \mathbb{E}[\{\xi_x - \mu_{\xi_x}\}^2\{f''(\mu_{\xi_x} + \theta[\xi_x - \mu_{\xi_x}]) - f''(\mu_{\xi_x})\}] = o(h)$. Then, since the density f is a twice differentiable function with continuous derivatives, we obtain

$$\mathbb{E}[f(\xi_x)] = f(x) + h\left[\tfrac{1}{2}xf'(x)\mathbb{E}[Z^2] + \tfrac{1}{2}x^2 f''(x)\mathbb{E}[Z^2]\right] + o(h). \tag{5.3}$$

Hence, (5.1) and (5.3) leads to $\mathrm{Bias}[\widehat{f}_{\text{skew-GBS}}(x)] = \mathbb{E}[\widehat{f}_{\text{skew-GBS}}(x)] - f(x) = h[\frac{1}{2}xf'(x)\gamma_2 + \frac{1}{2}x^2 f''(x)\gamma_2] + o(h)$, where γ_2 is given in (2.5). ■

Proof of Lemma 1

Let $A(h) = \mathbb{E}[\{\xi_x - \mu_{\xi_x}\}^2\{f''(\mu_{\xi_x} + \theta[\xi_x - \mu_{\xi_x}]) - f''(\mu_{\xi_x})\}]$, where $\theta \in (0,1)$. First, using (5.2), we have that $\mathrm{Var}[\xi_x] = \mathbb{E}[\{\xi_x - \mu_{\xi_x}\}^2] \to 0$, as $h \to 0$. This implies that

$\{\xi_x - \mu_{\xi_x}\}$ converges in probability to zero, i.e., $\{\xi_x - \mu_{\xi_x}\} \xrightarrow{\mathbb{P}} 0$, as $h \to 0$, or more specifically, $\forall\, \epsilon > 0$, $\lim_{h\to 0} \mathbb{P}(|\xi_x - \mu_{\xi_x}| > \epsilon) = 0$.

Let $\epsilon > 0$, $|\theta\,\{\xi_x - \mu_{\xi_x}\}| \leq |\theta||\xi_x - \mu_{\xi_x}|$, and $|\mu_{\xi_x} + \theta\,\{\xi_x - \mu_{\xi_x}\} - \mu_{\xi_x}| \leq |\theta||\xi_x - \mu_{\xi_x}| \leq |\xi_x - \mu_{\xi_x}|$. Then, $0 < \mathbb{P}(|\mu_{\xi_x} + \theta\,\{\xi_x - \mu_{\xi_x}\} - \mu_{\xi_x}| > \epsilon) \leq \mathbb{P}(|\xi_x - \mu_{\xi_x}| > \epsilon) \to 0$, as $h \to 0$, and, therefore, $\mu_{\xi_x} + \theta\,\{\xi_x - \mu_{\xi_x}\} \xrightarrow{\mathbb{P}} \mu_{\xi_x}$, as $h \to 0$. Using (C1) and (C2) and as f'' is continuous, $f''(\mu_{\xi_x} + \theta\,[\xi_x - \mu_{\xi_x}]) \xrightarrow{\mathbb{P}} f''(\mu_{\xi_x})$, as $h \to 0$. Let again $\epsilon > 0$ and $\mathbb{P}(|f''(\mu_{\xi_x} + \theta\,[\xi_x - \mu_{\xi_x}]) - f''(\mu_{\xi_x})| > \epsilon) \to 0$, as $h \to 0$, and $B_\epsilon = \{|f''(\mu_{\xi_x} + \theta\,[\xi_x - \mu_{\xi_x}]) - f''(\mu_{\xi_x})| > \epsilon\}$. Then, we have

$$\begin{aligned} A(h) &\leq \mathbb{E}[\{\xi_x - \mu_{\xi_x}\}^2\,|f''(\mu_{\xi_x} + \theta\,[\xi_x - \mu_{\xi_x}]) - f''(\mu_{\xi_x})|\,1_{B_\epsilon}] \\ &\quad + \mathbb{E}[\{\xi_x - \mu_{\xi_x}\}^2\,|f''(\mu_{\xi_x} + \theta\,[\xi_x - \mu_{\xi_x}]) - f''(\mu_{\xi_x})|\,1_{B_\epsilon^c}]. \end{aligned}$$

Since f'' is bounded, i.e., $|f''(x)| \leq M$, we deduce that

$$A(h) \leq 2M\,\mathbb{E}[\{\xi_x - \mu_{\xi_x}\}^2\,1_{B_\epsilon}] + \epsilon\,\mathbb{E}[\{\xi_x - \mu_{\xi_x}\}^2\,1_{B_\epsilon^c}] \leq 2M\,\mathbb{E}[\{\xi_x - \mu_{\xi_x}\}^2\,1_{B_\epsilon}] + c\,\epsilon\,h.$$

Using the Cauchy-Schwarz inequality, we have that $h \leq h_0$ and $\mathbb{E}[\{\xi_x - \mu_{\xi_x}\}^2 1_{B_\epsilon}] \leq (\mathbb{E}[\{\xi_x - \mu_{\xi_x}\}^4])^{1/2} \leq c\,h\,\mathbb{P}(B_\epsilon)$. Then, $\forall\, h \leq h_0$, we get that $A(h)/h \leq \widetilde{c}\,\epsilon$, with $\widetilde{c} > 0$. Therefore, $\lim_{h\to 0} A(h)/h = 0$. ■

Proof of Proposition 2.4.2

Consider f is the density to be estimated and $\widehat{f}_{\text{skew-GBS}}$ is its skew-GBS kernel estimator. Then, $\text{Var}[\widehat{f}_{\text{skew-GBS}}(x)] = (1/n)\mathbb{E}[K^2_{\text{skew-GBS}(\sqrt{h},\,x,\,\lambda;\,g)}(X_i)] + O(n^{-1})$. Note that

$$\mathbb{E}[K^2_{\text{skew-GBS}(\sqrt{h},\,x,\,\lambda;\,g)}(X_i)] = \tfrac{c^2}{C_\kappa\sqrt{hx}}\,\mathbb{E}[\psi_x^{-1/2} f(\psi_x)] + \tfrac{c^2}{C_\kappa\sqrt{h/x}}\,\mathbb{E}[\psi_x^{-3/2} f(\psi_x)]. \tag{5.4}$$

We point out that the integral needed for computing C_k cannot be solved analytically and then numerical integration must be applied. Using Taylor expansion in the first expectation of (5.4) and assuming that the function $\kappa(\psi_x) = \psi_x^{-1/2} f(\psi_x)$ is continuous and bounded, we obtain $\mathbb{E}[\psi_x^{-1/2} f(\psi_x)] = x^{-1/2} f(x) + O(h)$, where $\psi_x \sim \text{skew-GBS}(\sqrt{h}, x, \lambda; g^2)$. Therefore, we have

$$\begin{aligned} \mu_{\psi_x} = \mathbb{E}[\psi_x] &= \frac{x}{2}\left[2 + h\mathbb{E}[Z^2] + \sqrt{h}\mathbb{E}[Z\{hZ^2+4\}^{1/2}]\right], \\ \sigma_{\psi_x} = \text{Var}[\psi_x] &= x^2 h\mathbb{E}[Z^2] - \frac{x^2 h}{4}\left(\mathbb{E}[Z\{hZ^2+4\}^{1/2}]\right)^2 + \frac{x^2 h^{3/2}}{2}\mathbb{E}[Z^3\{hZ^2+4\}^{1/2}] \\ &\quad - \frac{x^2 h^{3/2}}{2}\mathbb{E}[Z^2]\mathbb{E}[Z\{hZ^2+4\}^{1/2}] - \frac{x^2 h^2}{4}(\mathbb{E}[Z^2])^2 + \frac{x^2 h^2}{2}\mathbb{E}[Z^4], \end{aligned}$$

where $Z \sim \text{SS}(0, 1, \lambda;\, g^2)$. Once again, using Taylor expansion now in the second expectation of (5.4) and assuming that the function $\rho(\psi_x) = \psi_x^{-3/2}\, f(\psi_x)$ is continuous and bounded, we obtain $\mathbb{E}[\psi_x^{-3/2} f(\psi_x)] = x^{-3/2} f(x) + O(h)$. Therefore, we prove that $\text{Var}[\widehat{f}_{\text{skew-GBS}}(x)] = 2c^2 C_\kappa^{-1} n^{-1} h^{-1/2} x^{-1} f(x) + o(n^{-1}h^{-1/2})$. ■