Discussion of "Birnbaum-Saunders distribution: A review of models, analysis and applications"

Víctor Leiva[1,*], **Camilo Lillo**[2], **M. Ivette Gomes**[3], **Marta Ferreira**[4]

[1]School of Industrial Engineering, Pontificia Universidad Católica de Valparaíso, Chile

[2]Department of Computer Science, Faculty of Engineering, Universidad de Playa Ancha, Chile

[3]Centre of Statistics and Applications, Universidade de Lisboa, Portugal

[4]Centre of Mathematics, Universidade do Minho, Portugal

1 Introduction

The univariate Birnbaum-Saunders (BS) distribution was first postulated to model failure times in material science; see Birnbaum and Saunders (1969). In this modeling, a cumulative damage exceeds a threshold to produce the failure; see Leiva and Saunders (2015). The univariate BS distribution is unimodal, positively skewed (although closely symmetrical in some cases), supported over a positive range of values and possessor of diverse and attractive properties; see Johnson et al. (1995). The univariate BS distribution has been extensively studied and applied with its recent developments until 2016 published in the book of Leiva (2016).

Multivariate BS distributions were derived as a natural extension to the univariate case, based on mathematical methods, with no fatigue theoretical arguments, different to the univariate BS distribution. Aykroyd et al. (2018) published recently a review on multivariate BS distributions with some applications. In addition, cumulative damage models and their relation to times of occurrence were recently modeled in a multivariate setting for multicomponent systems by Fierro et al. (2018).

Balakrishnan and Kundu (2018) conducted a complete and interesting review of the BS distribution, which considered physical justifications, mathematical and statistical issues, shape analysis and links to other models, as well as formulations and generalizations for the univariate case. Furthermore, extensions to multivariate and matrix-variate versions of the BS distribution were also included. This review provides a complete and updated list of references on the topic. However, in the book of Leiva (2016) and in the review of Balakrishnan and Kundu (2018), no attention was paid to an extreme values BS (EVBS) distribution, which has several attractive properties and a different conception to its standard version. Indeed, the standard BS distribution cannot be obtained as a particular case of the EVBS distribution, as it happens with other generalizations and extensions of the BS distribution.

There are several reasons to justify the use of the BS distribution in the modeling of data with heavy tails. In particular, the EVBS distribution developed by Ferreira et al. (2012) provides some of these justifications. The EVBS distribution is based on the generalized extreme values (GEV) distribution. As mentioned, the standard BS distribution was derived to solve a problem of material fatigue, but the BS distribution has been also employed to solve problems in areas as diverse as environment, finance and medicine, where heavy-tailed phenomena are often detected. This kind of phenomena is frequently described by extreme values models. Then, the EVBS distribution can be a good alternative to BS distributions because of its good properties to model heavy-tailed phenomena.

*Corresponding author: Víctor Leiva. Email: victorleivasanchez@gmail.com, URL: www.victorleiva.cl

We discuss some aspects of the BS distribution reviewed by Balakrishnan and Kundu (2018) and delve into the EVBS distribution, considering its features, modeling properties and new scopes and applications.

2 EVBS distributions

Consider an independent identically distributed sequence of random variates (RVs) $\{X_n, n \geq 1\}$, with marginal cumulative distribution function (CDF) F, and the statistic $M_n = \max\{X_1, \ldots, X_n\}$. Then, for the constants $a_n > 0$ and $b_n \in \mathbb{R}$, we assume that $\mathrm{P}((M_n - b_n)/a_n \leq z) \to G(z)$, as $n \to \infty$, with G being a non-degenerate CDF. Then, G is a GEV CDF, which depends on a parameter $\xi \in \mathbb{R}$, known as extreme value index (EVI). The notation $Z \sim \mathrm{GEV}(\xi)$ is used in this case, with CDF given by

$$G(z;\xi) = \begin{cases} \exp(-(1+\xi z)^{-1/\xi}), & \text{if } \xi \neq 0 \text{ and } 1+\xi z > 0; \\ \exp(-\exp(-z)), & \text{if } \xi = 0. \end{cases}$$

An RV with a BS distribution can be considered as a transformation of another RV with standard normal distribution. An RV T with BS distribution and parameters of shape α and of scale ϱ (which is also the median) is denoted by $T \sim \mathrm{BS}(\alpha, \varrho)$. Note that T can be represented as $T = \varrho(\alpha Z/2 + ((\alpha Z/2)^2 + 1)^{1/2})^2$, where $Z \sim \mathrm{N}(0,1)$. Then, it is possible to relax the normality assumption of the representation above described assuming that $Z \sim \mathrm{GEV}(\xi)$ and obtaining the EVBS CDF expressed as

$$F(x;\alpha,\varrho,\xi) = \begin{cases} \exp(-(1+\xi a(x))^{-1/\xi}), & \text{if } \xi \neq 0 \text{ and } 1+\xi a(x) > 0; \\ \exp(-\exp(-a(x))), & \text{if } \xi = 0. \end{cases}$$

where $a(x) = (1/\alpha)((x/\varrho)^{1/2} - (\varrho/x)^{1/2})$. In this case, we use the notation $T \sim \mathrm{EVBS}(\alpha, \varrho, \xi)$

Figure 1 shows some graphs of the EVBS probability density function (PDF) given by $f(x;\alpha,\varrho,\xi) = \mathrm{d}F(x;\alpha,\varrho,\xi)/\mathrm{d}x$, for different values of α and ξ, considering $\varrho = 1.0$, without loss of generality. Note that the EVBS distribution is widely flexible to model diverse types of data. Even the EVBS distribution is more flexible than the GEV distribution, because the GEV distribution only admits unimodality, but the EVBS distribution can have unimodal and bimodal behaviors (depending on the values of ξ). In addition, the EVBS distribution is more flexible than the BS distribution, because the EVBS distribution has negative, zero and positive skewness, while the BS distribution only allows for positive skewness. In addition, Gomes et al. (2012) compared the right tail of the EVBS, BS and normal distributions, confirming that the EVBS distribution has the heaviest tails among these three families of distributions, for almost all $\xi \geq 0$. Further comparisons on the weight of the right-tail within BS, generalized BS and EVBS distributions can be seen in Ferreira (2013). Balakrishnan and Kundu (2018) showed some generalizations of the BS distribution, but the flexibility of the EVBS is higher than such generalizations. In addition, as mentioned, the standard BS distribution is not a particular case of the EVBS distribution.

Other features to discuss about the EVBS distribution are found in the work of Gomes et al. (2012). For example, the first four moments of the EVBS distribution are restricted by the distribution tail. Particularly, the mean, variance and coefficients of skewness and kurtosis can only be computed under the restrictions $\xi < 1$, $\xi < 1/2$, $\xi < 1/3$ and $\xi < 1/4$, respectively. Lillo et al. (2018) derived L-moments of the EVBS distribution, that is, linear combinations of order statistics that allows us to characterize a distribution, and mentioned its L-moments can only be calculated if $\xi < 1/2$. This has an impact on the estimation of the parameters of the EVBS distribution, since the method for estimating parameters through L-moments would be restricted by ξ. Computational details on the maximum likelihood (ML) estimation of the GEV parameters can be found, for example, in Coles et al. (2001). Note that there are no analytical explicit

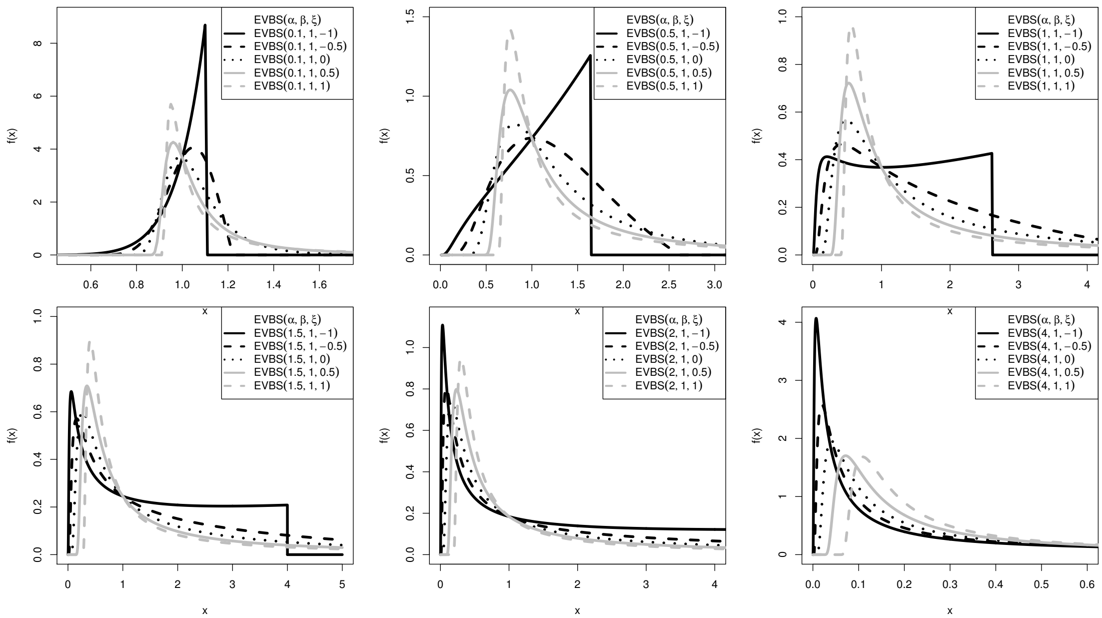

Figure 1: PDF plots of the EVBS distribution for $\varrho = 1.0$ and the indicated values of (α, ξ).

expressions for the ML estimators of GEV parameters and besides they can be numerically difficult to compute, just as it happens with the EVBS distribution. Specifically, Coles et al. (2001) mentioned that, for certain values of $\xi > 0.5$, often the GEV parameter estimates cannot be obtained, which also occurs with the EVBS distribution. Then, other methods of estimation, such as the probability weighted moments method may be used for estimating EVBS parameters. Furthermore, it is an opportunity to investigate and develop more applications about the EVBS distribution. According to the statistical properties discussed by Balakrishnan and Kundu (2018) in their review, the BS distribution and the presented generalizations and extensions do not have restriction problems in the estimation of parameters.

3 EVBS regression models

Leiva et al. (2016) introduced EVBS regression models based on the formulation

$$T_i = \varrho_i \delta_i = \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta})\delta_i, \quad i = 1, \ldots, n, \tag{3.1}$$

where T_i is the response variable, $\boldsymbol{x}_i^\top = (1, x_{i2}, \ldots, x_{ip})$ are the observations of $p-1$ explanatory variables, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_{p-1})^\top$ is a vector of unknown parameters to be estimated and δ_i is the error term of the model. Note that $\varrho_i = \exp(\boldsymbol{x}_i^\top)$ acts as a scale parameter, which can be modeled by regression in a similar way as in Marshall and Olkin (2007, pp. 533-540). Thus, by using proportionality property of the EVBS distribution presented in Ferreira et al. (2012), we have that $\delta_i \sim \text{EVBS}(\alpha, 1, \xi)$ and therefore $T_i \sim \text{EVBS}(\alpha, \exp(\boldsymbol{x}_i^\top), \xi)$. Based on the model defined in (3.1), we have an EVBS log-linear regression model given by $Y_i = \eta_i + \varepsilon_i$ for $i = 1, \ldots, n$, where $Y_i = \log(T_i)$ is the log-response variable, $\eta_i = \log(\varrho_i) = \boldsymbol{x}_i^\top \boldsymbol{\beta}$ is a linear predictor, with $\boldsymbol{x}_i$ and $\boldsymbol{\beta}$ defined in (3.1), and $\varepsilon_i = \log(\delta_i)$ is the model error; see more details of the linear model above defined and its diagnostics in Leiva et al. (2016). As in Santana et al. (2011), note that η_i is a location parameter, which is used in this model, as mentioned, in the form of a linear predictor, but it is not the mean as in other BS regressions. Then, in this form, it is not possible to present models for the median/mean of EVBS/log-EVBS distributions. However, we could utilize the parameterization of the BS distribution proposed in Leiva et al. (2014) and Santos-Neto et al. (2014, 2016) of the type $T \sim \mathrm{BS}(\alpha, \mu)$, where $\mathrm{E}[T] = \mu$, for introducing new EVBS regression models. This will allow the development of GLM-type models for the EVBS distribution and the corresponding derivation of influence methods, as well as the incorporation of temporal and spatial components for the modeling in a standard framework; see Garcia-Papani et al. (2018) for BS spatial regression models. An alternative form of modeling, such as quantile regression, can be used; see Furno and Vistocco (2018). In the linear model above presented, $\mathrm{Var}[\varepsilon]$ is constant and $\mathrm{Cov}[\varepsilon_l, \varepsilon_k] = 0$, for $l \neq k$ (uncorrelated errors). This implies that a regression model describes the conditional mean $\mathrm{E}[T|\boldsymbol{X} = \boldsymbol{x}] = \boldsymbol{x}_i^\top\boldsymbol{\beta}$. In this perspective, one can even postulate a general expression as $\varrho_i = h(\boldsymbol{x}_i^\top\boldsymbol{\beta})$, where h is an invertible function, such as in GLM. Now, considering the EVBS as the conditional distribution, we may establish the formulation $T_i|\boldsymbol{X}_i = \boldsymbol{x}_i \sim f(t; \alpha, h(\boldsymbol{x}_i^\top\boldsymbol{\beta}), \xi)$, for $i = 1, \ldots, n$, where f is now the EVBS PDF. Thus, we can model the median of the conditional distribution above presented instead of the mean (due to the asymmetry of extreme value data). Then, it is possible to use quantile regression models, which offer a mechanism to estimate and predict the median response, as well as other quantiles. This class of regression models is based on the quantile function given by $Q(\tau; \boldsymbol{\theta}) = \inf\{t \colon F(t; \boldsymbol{\theta}) \geq \tau\}$, where F is the EVBS CDF, $\boldsymbol{\theta}^\top = (\alpha, \boldsymbol{\beta}, \xi)$ and $0 < \tau < 1$; see Sánchez et al. (2018) for BS quantile regression models.

4 An application in finance

Extreme value theory is commonly based on maximum (or minimum) values taken from an RV; for example, maximum daily coastal water level. This theory is also associated with rare extreme events, commonly related to climate change, such as earthquakes, floods, hurricanes and tsunamis. In this illustration, we focus on economic loss after earthquakes in various countries. This type of macroeconomic variable needs to be predicted, because it is difficult to measure, since it is related to the loss of real estate and industrial properties, expenses for associated health problems, costs for emergency installations and other economic impacts. A data set of this type is difficult to obtain, since the data must be collected from different sources. We use data presented in Dunbar et al. (2003) corresponding to: economic loss (response, T, in USD million at the time of the earthquake) and the explanatory variables "gross domestic product of a country" (X_1), as an indicator of total exposure (see World Bank Group, 2014), "amount of deaths after earthquake" (X_2) and "Mercalli's intensity of the earthquake" (X_3).

Next, we provide an exploratory analysis of the economic loss data due to earthquakes with a sample of $n = 25$ countries. A descriptive summary of these data indicates that the range of loss was between 0.92 and 8500 USD millions, with median and mean losses of 143 and 1094.33 USD millions, respectively, showing a strong asymmetry to the right, supported by a coefficient skewness of 2.49, a high coefficient of kurtosis of 7.75, indicating a heavy-tailed phenomenon, in addition to a high variability with a coefficient of variation of 214.77. Figure 2 sketches log-linear relations among the response and explanatory variables, but no collinearity problems. Commonly, regression models related to economic variables are presented in logarithmic scale for response and explanatory variables. All the findings from the exploratory data analysis support the use of extreme values and, particularly, of EVBS models. Similarly as in Galea et al. (2004) for BS regression models, where response and explanatory variables were transformed to a logarithmic scale, and employing the log-EVBS distribution presented in Leiva et al. (2016), we formulate an EVBS log-linear regression model given by $Y_i = \log(T_i) = \beta_0 + \beta_1 x^\star_{1_i} + \beta_2 x^\star_{2_i} + \beta_3 x^\star_{3_i} + \varepsilon_i$, where $x^\star_{j_i} = \log(x_{j_i})$, for $j = 1, 2, 3$, $i = 1, \ldots, 25$, and $\varepsilon \sim \text{log-EVBS}(\alpha, 0, \xi)$, with α being the EVBS shape parameter and ξ being its EVI. The ML estimates of the model parameters, with estimated (approximate) standard errors in parenthesis, are: $\widehat{\beta}_0 = -5.8035(1.1707)$, $\widehat{\beta}_1 = 0.2645(0.2088)$, $\widehat{\beta}_2 = 0.5195(0.1125)$ and $\widehat{\beta}_3 =$

$2.9455(1.2472)$. Note that all coefficients are significant at 5%. Also, notice that $\widehat{\alpha} = 1.0027(0.0015)$ and $\widehat{\xi} = -0.3599(0.0699)$. Now, we verify the model assumptions by a residual analysis for economic loss data. We use the quantile residual because this showed the best performance for EVBS models among several types of residuals as proven in Leiva et al. (2016). Figure 3(a)-(b) displays an index plot of quantile residuals and its PP plot with 95% acceptance bands to verify the distributional assumption of the regression model. From this figure, no outlying observations nor heterogeneity problems are detected, whereas a KS p-value = 0.4071 supports to a 5% of significance the normality assumption of the quantile residuals obtained from the EVBS regression model proposed by us. Therefore, once detected the good performance of our model, we are able to propose a predictive model for the economic loss based on

$$\widehat{T}_{\text{pred}} = 0.003\ x_{1_{\text{pred}}}^{0.2645}\ x_{2_{\text{pred}}}^{0.5195}\ x_{3_{\text{pred}}}^{2.9455},$$

whereas confidence bands (omitted here) can be produced using also $\widehat{\alpha} = 1.0027$ and $\widehat{\xi} = -0.3599$.

Figure 2: scatter plots and their correlations for the indicated variables with economic loss data.

Figure 3: Index plot of quantile residual (a) and PP plot with 95% acceptance bands for quantile residuals (b).

5 Conclusions

In this discussion, we have compared the BS distribution reviewed by Balakrishnan and Kundu (2018) with the EVBS distribution. The EVBS distribution was developed in 2012 and is therefore a relatively new distribution. The EVBS model is a flexible, possibly heavy-tailed distribution, with nice properties when compared to the BS distribution. We have introduced a new application of the EVBS distribution in finance, an area where extreme value data following asymmetric heavy-tailed distributions often occur. With respect to future research on the EVBS distribution, some new ideas on estimation and modeling, which can be exploited in further research on BS and EVBS distributions, have been put forward throughout our discussion.

Acknowledgement We thank Editor-in-Chief Prof. Fabrizio Ruggeri for inviting us to discuss this interesting review prepared by Profs. Balakrishnan and Kundu. This work was partially supported by the Chilean government under the project grant FONDECYT 1160868 and by Portuguese funds through FCT –Fundação para a Ciência e a Tecnologia– within the Projects UID/MAT/00013/2013, UID/MAT/00006/2013 and by the research center CEMAT (Instituto Superior Técnico, Universidade de Lisboa) through the Project UID/Multi/04621/2013.

References

Aykroyd, R. G., Leiva, V., and Marchant, C. (2018). Multivariate birnbaum-saunders distributions: Modelling and applications. *Risks*, 6-1(article 21):1–25.

Balakrishnan, N. and Kundu, D. (2018). Birnbaum-saunders distribution: A review of models, analysis and applications. *Applied Stochastic Models in Business and Industry*, page in press.

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

Coles, S., Bawa, J., Trenner, L., and Dorazio, P. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer, New York, US.

Dunbar, P.K., Bilham, R.G., and Laituri, M.J. (2003). *Earthquake loss estimation for India based on macroeconomic indicators*. In Risk science and sustainability (pp. 163–180). Springer, Dordrecht.

Ferreira, M. (2013). A study of exponential-type tails applied to Birnbaum-Saunders models. *Chilean Journal of Statistics*, 4:87–97.

Ferreira, M., Gomes, M. I., and Leiva, V. (2012). On an extreme value version of the Birnbaum-Saunders distribution. *REVSTAT Statistical Journal*, 10:181–210.

Fierro, R., Leiva, V., and Maidala, J. P. (2018). Cumulative damage and times of occurrence for a multicomponent system: A discrete time approach. *Journal of Multivariate Analysis*, page in press.

Furno, M. and Vistocco, D. (2018). *Quantile Regression: Estimation and Simulation*. Wiley, New York, US.

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

Garcia-Papani, F., Leiva, V., Ruggeri, F., and Uribe-Opazo, M. A. (2018). Kriging with external drift in a Birnbaum-Saunders geostatistical model. *Stochastic Environmental Research and Risk Assessment*, 32:1517–1530.

Gomes, M., Ferreira, M., and Leiva, V. (2012). The extreme value Birnbaum-Saunders model and its moments and application in biometry. *Biometrical Letters*, 49:81–94.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1995). *Continuous Univariate Distributions*, volume 2. Wiley, New York, US.

Leiva, V. (2016). *The Birnbaum-Saunders Distribution*. Academic Press, New York, US.

Leiva, V., Ferreira, M., Gomes, M. I., and Lillo, C. (2016). Extreme value Birnbaum-Saunders regression models applied to environmental data. *Stochastic Environmental Research and Risk Assessment*, 30:1045–1058.

Leiva, V., Santos-Neto, M., Cysneiros, F. J. A., and Barros, M. (2014). Birnbaum-Saunders statistical modelling: A new approach. *Statistical Modelling*, 14:21–48.

Leiva, V. and Saunders, S. C. (2015). Cumulative damage models. *Wiley StatsRef: Statistics Reference Online*, pages 1–10.

Lillo, C., Leiva, V., Nicolis, O., and Aykroyd, R. G. (2018). L-moments of the Birnbaum-Saunders distribution and its extreme value version: Estimation, goodness of fit and application to earthquake data. *Journal of Applied Statistics*, 45:187–209.

Marshall, A. and Olkin, I. (2007). *Life Distributions*. Springer, New York, US.

Sánchez, L., Leiva, V., and Galea, M. (2018). On quantile regression for the Birnbaum-Saunders distribution. *Under review*.

Santana, L., Vilca, F., and Leiva, V. (2011). Influence analysis in skew-Birnbaum-Saunders regression models and applications. *Journal of Applied Statistics*, 38:1633–1649.

Santos-Neto, M., Cysneiros, F. J. A., Leiva, V., and Barros, M. (2014). On new parameterizations of the Birnbaum-Saunders distribution and its moments, estimation and application. *REVSTAT Statistical Journal*, 12:247–272.

Santos-Neto, M., Cysneiros, F. J. A., Leiva, V., and Barros, M. (2016). Reparameterized Birnbaum-Saunders regression models with varying precision. *Electronic Journal of Statistics*, 10:2825–2855.

World Bank Group. (2014). *World development indicators 2014*. World Bank Publications.