TWO-SIDED GENERALIZED HYPERBOLIC SECANT DISTRIBUTION WITH REAL DATA APPLICATIONS AND MODEL COMPARISONS

Mustafa Ç. Korkmaz[1], Christophe Chesneau[2] and Julien Marie[3]

ABSTRACT

In this paper, we use the structure of the standard two-sided power distribution to implement a new version of the hyperbolic secant distribution. This new distribution is more flexible than the hyperbolic secant distribution when it comes to interpreting data presenting an abrupt change in values. In the first part of the paper, we show some of its properties, such as the shape behavior of the probability density and hazard rate functions, and the analysis of moment-type measures. Then, the statistical side of the underlying model is explored. We provide the maximum likelihood estimates for the model parameters, as well as an efficient algorithm to calculate them. After this, to demonstrate the potential of the proposed modeling strategy, we present three real data applications. The beta-normal, power-normal, Kuramaswamy-normal and two-sided generalized normal distribution models are considered as competitors. The results are favorable to the proposed model.

Keywords: Applications, General Family of Distribution, Model Comparisons, Standard Two-sided Power Distribution

[1] Corresponding Author, Assoc. Prof. Dr., Department of Measurement and Evaluation, Artvin Çoruh University, City Campus, Artvin, 08000, Turkey, ORCID ID: https://orcid.org/0000-0003-3302-0705
[2] Assoc. Prof. Dr., Université de Caen Normandie, LMNO, Campus II, Science 3, 14032, Caen, France, ORCID ID: https://orcid.org/0000-0002-1522-9292
[3] Université de Caen Normandie, LMNO, Campus II, Science 3, 14032, Caen, France, ORCID ID: https://orcid.org/0000-0001-6171-6369

GERÇEK VERİ UYGULAMALARI VE MODEL KARŞILAŞTIRMALARI İLE İKİ YÖNLÜ GENELLEŞTİRİLMİŞ HİPERBOLİK SEKANT DAĞILIMI

ÖZ

Bu çalışmada, hiperbolik sekant dağılımının yeni bir versiyonunu elde etmek için standart iki yönlü kuvvet dağılım yapısı kullanılmıştır. Bu yeni dağılım, değerlerde ani bir değişiklik gösteren verilerin yorumlanması söz konusu olduğunda, hiperbolik sekant dağılımından daha esnektir. Makalenin ilk bölümünde, olasılık yoğunluğu ve tehlike oranı fonksiyonlarının biçim davranışı ve moment ölçümlerin analizi gibi bazı özelliklerine değinilmiştir. Sonrasında, baz alınan modelin istatistiksel yönü araştırılmıştır. Model parametreleri için en çok olabilirlik tahmin yönteminin yanı sıra bunları hesaplamak için kullanışlı bir algoritma sağlanmıştır. Tahminlerin ardından, önerilen modelleme stratejisinin potansiyelini göstermek için üç gerçek veri uygulaması sunulmuştur. Beta-normal, kuvvet-normal, Kuramaswamy-normal ve iki taraflı genelleştirilmiş normal dağılım modelleri rakip olarak kabul edilmiş olup, elde edilen sonuçların önerilen model için daha uygun olduğu görülmüştür.

Anahtar Kelimeler: Uygulamalar, Dağılımların Genel Ailesi, Model Karşılaştırmaları, Standart İki Yönlü Güç Dağılımı

1. INTRODUCTION

This paper deals with a new modeling strategy for the analysis of data that present an abrupt change in terms of values, breaking a certain monotonic tendency. In order to fully motivate the findings, a retrospective on the standard two-sided power (STSP) distribution is necessary. To begin, the STSP distribution was introduced by van Dorp and Kotz (2002a). It is mathematically defined by the cumulative distribution function (cdf) and probability density function (pdf) given as

$$F(x;\alpha,\beta) = \begin{cases} 0 & for\ x < 0 \\ \beta\left(\frac{x}{\beta}\right)^{\alpha} & for\ 0 \le x < \beta \\ 1 - (1-\beta)\left(\frac{1-x}{1-\beta}\right)^{\alpha} & for\ \beta \le x \le 1 \\ 1 & for\ x > 1 \end{cases} \tag{1}$$

and

$$f(x;\alpha,\beta) = \begin{cases} \alpha\left(\frac{x}{\beta}\right)^{\alpha-1} & for\ 0 \le x < \beta \\ \alpha\left(\frac{1-x}{1-\beta}\right)^{\alpha-1} & for\ \beta \le x \le 1 \\ 0 & otherwise \end{cases} \tag{2}$$

respectively. It is supposed that $\alpha > 0$ and $0 \le \beta \le 1$. The parameter α shapes the distribution, whereas β can be viewed as the distributional change-point, called the reflection point (or corner point). As a basic but important comment, the STSP distribution is very flexible. It can be unimodal, U-shaped, uniform or triangular. The distribution corresponds to the power distribution when $\beta = 1$. The corresponding model is designed to analyze data with values in (0,1), corresponding to proportional or percentage-type data, that also present a sudden change in values (abrupt decay, peaks, etc.). When it comes to the analysis of financial data, where peaked shapes are more common, the STSP model is particularly helpful (see van Dorp and Kotz (2002a) and Pérez et al. (2005)). An extension of the STSP distribution to four parameters has been developed by van Dorp and Kotz (2002b) for more perspectives on applications. Further, with the idea of combining the change-point strategy with any parental distribution, a new family of continuous distributions was proposed in Korkmaz and Genc (2017), providing an alternative to the beta generated Eugene et al. (2002) and Kumaraswamy generated Cordeiro and Castro (2011) families, among others. The cdf and pdf of this family can be expressed as

$$F(x;\alpha,\beta,\varepsilon) = \begin{cases} 0 & for\ x < G^{-1}(0;\varepsilon) \\ \beta^{1-\alpha}[G(x;\varepsilon)]^{\alpha} & for\ G^{-1}(0;\varepsilon) \le x < G^{-1}(\beta;\varepsilon) \\ 1 - (1-\beta)^{1-\alpha}g(x;\varepsilon)[1 - G(x;\varepsilon)]^{\alpha} & for\ G^{-1}(\beta;\varepsilon) \le x \le G^{-1}(1;\varepsilon) \\ 1 & for\ x > 1 \end{cases} \tag{3}$$

and

$$f(x; \alpha, \beta, \varepsilon) = \begin{cases} \alpha\beta^{1-\alpha}g(x; \varepsilon)[G(x; \varepsilon)]^{\alpha-1} & for \ G^{-1}(0; \varepsilon) \le x < G^{-1}(\beta; \varepsilon) \\ \alpha(1-\beta)^{1-\alpha}g(x; \varepsilon)[1-G(x; \varepsilon)]^{\alpha-1} & for \ \beta \le x \le 1 \\ 0 & otherwise \end{cases} \quad (4)$$

respectively, where $G(x; \varepsilon)$ and $g(x; \varepsilon)$ are the cdf and pdf of the baseline distribution, respectively, with ε a certain parameter vector, and $G^{-1}(\beta; \varepsilon)$ is the quantile function (qf) of the baseline model taken at the reflection parameter. In fact, this family consists of a mixture distribution structure based on the two components, one of which is a truncated exponentiated family and the other one is a truncated Kumaraswamy family. Authors have studied the basic distributional properties, estimation procedure, and real data application of this family with a detailed example of the normal distribution, called the TSGN distribution. Other members have been studied in the literature. We cite the two-sided generalized exponential distribution by Korkmaz and Genc (2015b), two-sided generalized Weibull distribution by Korkmaz and Genc (2015a) and two-sided generalized Gumbel distribution by Korkmaz (2015). It may be seen Korkmaz and Genc (2017) for other examples of two-sided distributions and models.

Secondly, always for motivational reasons, we need to present the fundamentals of the hyperbolic secant (HS) distribution. It was introduced by Perks (1932), and was recently reviewed by Fischer (2013). It is mathematically defined by the cdf and pdf given as

$$G(x) = \frac{2}{\pi} arctan\left(e^{\frac{\pi x}{2}}\right), \ x \in R, \quad (5)$$

and

$$g(x) = \frac{1}{2} sech\left(\frac{\pi}{2}x\right), \ x \in R, \quad (6)$$

respectively, where "sech" denotes the hyperbolic secant function defined by $sech(x) = \frac{1}{\cosh(x)} = \frac{2}{(e^x + e^{-x})}$. As for the normal distribution, the HS distribution is symmetric and bell-shaped but has the feature of being leptokurtic. As a matter of fact, there are a few generalizations of the HS distribution in the literature. The most popular are those created by Harkness and Harkness (1968) and Morris (1982). More recently, Vaughan (2002) introduced a new generalized HS distribution which showed nice properties but didn't take into account

different levels of skewness. This was reworked by Fischer and Vaughan (2002) to correct it. Later, Fischer and Vaughan (2004) introduced the beta HS (BHS) distribution, which was shown to be quite flexible.

In the light of the above information, the main aim of the paper is to propose an alternative unbounded distribution with two-sided structure using the HS distribution. The name of this distribution is the two-sided generalized hyperbolic secant (TSGHS) distribution. It is intended to generate a flexible model specified over the entire real line that is suitable for dealing with data that has a sudden change in terms of values, thereby breaking a monotonic trend. Furthermore, it provides a new extension of the hyperbolic secant distribution, which remains rare enough for a thorough investigation.

In the first part, we highlight the flexible aspect of the TSGHS distribution by showing some plots of the pdf and hazard rate function (hrf) of the newly defined distribution. We also study the qf and various moment measures. The second part of the paper is about the modeling capability of the TSGHS model.

We consider the maximum likelihood estimates of the parameters and prove their efficiency via a complete simulation study. We end this paper with a comparison between the TSGHS model and other existing models, like the beta-normal model (see Eugene et al. (2002)), power-normal model (see Gupta and Gupta (2008)), Kumaraswamy-normal model (see Cordeiro and Castro (2011)), TSGN model (see Korkmaz and Genc (2017)) and BHS model (see Fischer and Vaughan (2004)) with three real data applications.

The precise organization of the paper is as follows: Section 2 focuses on the main functions of the TSGHS distribution. The quantile and moment properties are examined in Section 3. The inference on the TSGHS model parameter is the subject of Section 4. Applications to real data are given in Section 5. Finally, Section 6 concludes the paper.

2. THE TSGHS DISTRIBUTION

In this section, we define the two-sided generalized HS distribution by its main functions, and study their shape behavior.

2.1. Main functions

From Equations (3) and (4) on the one hand, and Equations (5) and (6) with the use of a central parameter $\mu \in R$ and a scale parameter $\sigma > 0$ on the other hand, by applying the

method described in Korkmaz and Genc (2017), the cdf and pdf of the TSGHS distribution are obtained as

$$F(x; \alpha, \beta, \mu, \sigma) =$$

$$
\begin{cases}
\beta \left(\dfrac{\frac{2}{\pi} arctan\left(e^{\pi \frac{(x-\mu)}{2\sigma}} \right)}{\beta} \right)^{\alpha} & for -\infty \le x \le \mu + \sigma \frac{2}{\pi} ln\left[tan\left(\frac{\pi}{2}\beta \right) \right] \\[4mm]
1 - (1-\beta) \left(\dfrac{1 - \frac{2}{\pi} arctan\left(e^{\pi \frac{(x-\mu)}{2\sigma}} \right)}{1-\beta} \right)^{\alpha} & for \ \mu + \sigma \frac{2}{\pi} ln\left[tan\left(\frac{\pi}{2}\beta \right) \right] < x \le +\infty
\end{cases}
\tag{7}
$$

and

$$f(x; \alpha, \beta, \mu, \sigma) =$$

$$
\begin{cases}
\dfrac{\alpha}{2\sigma} sech\left[\frac{\pi}{2}\left(\frac{x-\mu}{\sigma} \right) \right] \left(\dfrac{\frac{2}{\pi} arctan\left(e^{\pi \frac{(x-\mu)}{2\sigma}} \right)}{\beta} \right)^{\alpha-1} & for -\infty \le x \le \mu + \sigma \frac{2}{\pi} ln\left[tan\left(\frac{\pi}{2}\beta \right) \right] \\[4mm]
\dfrac{\alpha}{2\sigma} sech\left[\frac{\pi}{2}\left(\frac{x-\mu}{\sigma} \right) \right] \left(\dfrac{1 - \frac{2}{\pi} arctan\left(e^{\pi \frac{(x-\mu)}{2\sigma}} \right)}{1-\beta} \right)^{\alpha-1} & for \ \mu + \sigma \frac{2}{\pi} ln\left[tan\left(\frac{\pi}{2}\beta \right) \right] < x \le +\infty
\end{cases}
\tag{8}
$$

respectively. To specify the parameters, we will sometimes denote the proposed distribution by $TSGHS(\alpha, \beta, \mu, \sigma)$. When $\alpha = 1$, the cdf and pdf given in Equations (6) and (7), respectively, represents the HS distribution. So the TSGHS distribution is a generalization of the HS distribution. The pdf changes of defintion in function of the position of x with the reflection point: $\mu + \sigma \frac{2}{\pi} ln\left[tan\left(\frac{\pi}{2}\beta \right) \right]$. When $x \le \mu + \sigma \frac{2}{\pi} ln\left[tan\left(\frac{\pi}{2}\beta \right) \right]$, a truncated exponentiated version of the HS distribution is activated, whereas when $x > \mu + \sigma \frac{2}{\pi} ln\left[tan\left(\frac{\pi}{2}\beta \right) \right]$, a truncated Kumaraswamy version of the HS distribution is considered. The shape behavior of this pdf is of importance to understand the modeling capability of the TSGHS distribution. Due to its high level of complexity, we prefer to perform a graphical study. Thus, Figures 1 and 2 display $f(x; \alpha, \beta, \mu, \sigma)$ for some combinasions of the parameters.
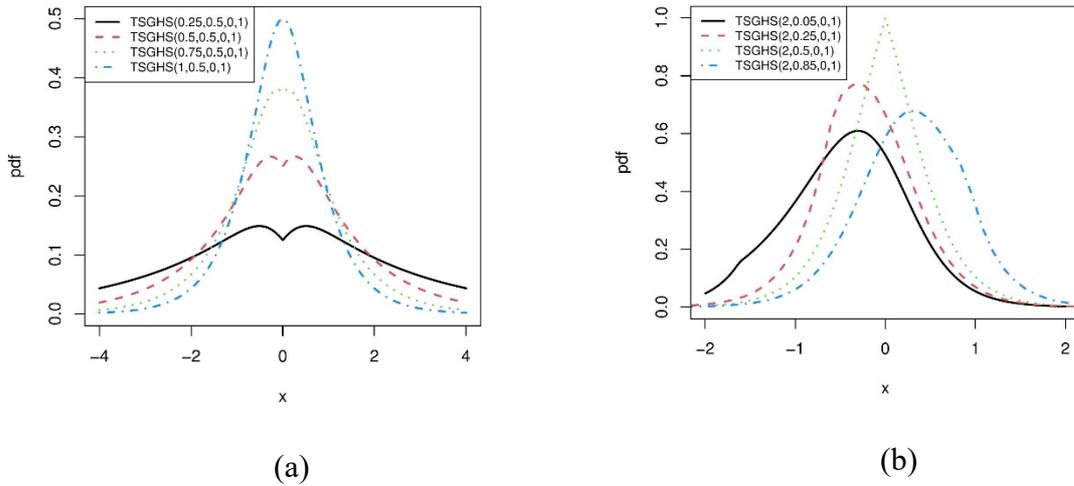
Figure 1. Curves of the pdf of the TSGHS distribution with (a) $\beta = 0.5$ and α changes, and
(b) $\alpha = 2$ and β changes
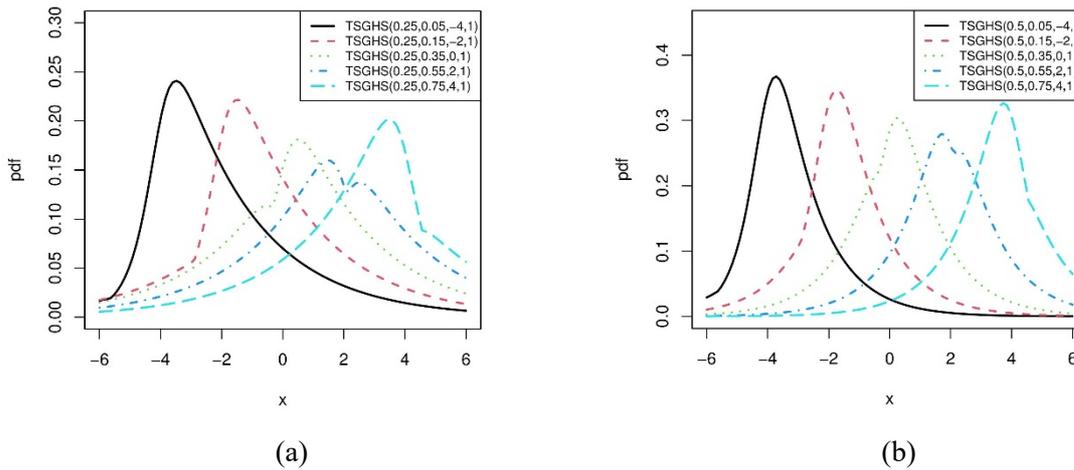
Figure 2. Curves of the pdf of the TSGHS distribution with (a) $\alpha = 0.25$ and β changes, and
(b) $\alpha = 0.5$ and β changes

The curves in Figure 1(a) are symmetric as $\beta = 0.5$ with any α. We can also note that when $\alpha < 1$, the distribution is bimodal. The bimodality is more apparent when α gets closer to 0. Figure 1(b) shows the possible right and left skewness of the distribution with $\alpha = 2$ and β varying. When $\beta = 0.5$, the pdf is symmetric and has a sharp peak, which can be useful for modeling some data (see section 6). Figures 2(a) and (b) show interesting plots with the corner point ($\alpha = 0.25$ and $\alpha\_0.5$, respectively).

2.2. Hazard rate function

The hrf is an important tool in survival analysis, and may be of interest for distribution defined on the whole real line. Its shape behavior is informative on the model capacity. In a condensed form, the hrf of the TSGHS distribution is indicated as

$$
h(x; \alpha, \beta, \mu, \sigma) = \begin{cases} \dfrac{\alpha}{\sigma} \dfrac{g\left(\frac{x-\mu}{\sigma}\right)\left[G\left(\frac{x-\mu}{\sigma}\right)\right]^{\alpha-1}}{\beta^{\alpha-1}-\left[G\left(\frac{x-\mu}{\sigma}\right)\right]^{\alpha}} & for -\infty \le x \le \mu + \sigma\frac{2}{\pi}ln\left[tan\left(\frac{\pi}{2}\beta\right)\right] \\ \dfrac{\alpha}{\sigma} h_{HS}\left(\dfrac{x-\mu}{\sigma}\right) & for \mu + \sigma\frac{2}{\pi}ln\left[tan\left(\frac{\pi}{2}\beta\right)\right] \le x \le +\infty \end{cases}, \tag{9}
$$

where $g(x)$ and $G(x)$ are the pdf and cdf of the HS distribution as given in Equations (5) and (6), respectively, and $h_{HS}(x)$ is the hrf of the HS distribution defined as follows:

$$
h_{HS}(x) = \frac{\frac{1}{2}sech\left(\frac{\pi}{2}x\right)}{1-\frac{2}{\pi}arctan\left(e^{\frac{\pi x}{2}}\right)}, \qquad x \in R.
$$

Similarly to the pdf of the TSGHS distribution, the hrf is two-sided with the changing point being $\mu + \sigma\frac{2}{\pi}ln\left[tan\left(\frac{\pi}{2}\beta\right)\right]$. However, the complexity level of this function is too high for an analytical study; we prefer to give plots of it to understand its shape behavior. Figures 3 and 4 are made in this regard.

(a)

(b)
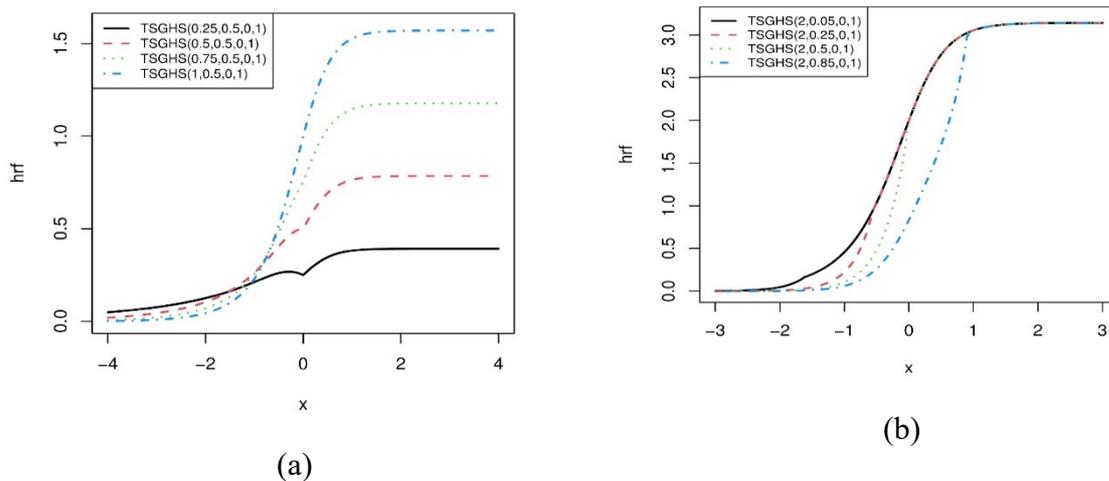
Figure 3. Curves of the hrf of the TSGHS distribution with (a) $\beta = 0.5$ and α changes, and (b) $\alpha = 2$ and β changes

(a)   (b)
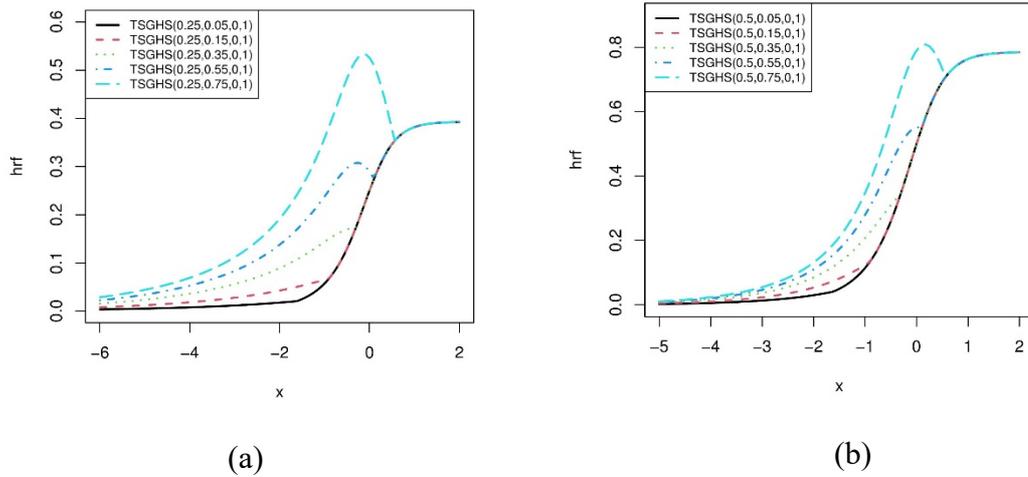
Figure 4. Curves of the hrf of the TSGHS distribution with (a) $\alpha = 0.25$ and β changes, and (b) $\alpha = 0.5$ and β changes

Figure 3 shows increasing hrf first and then it stays constant. As for Figure 4, for some parameter values, the hrf increases then decreases before the changing point. Then it increases again to stay constant after. For the other parameter values (lower β) the hrf is increasing or slightly increasing up to the changing point where it sharply increases to then stay constant.

3.   QUANTILE AND MOMENTS

Quantile and moment analyses of the TSGHS distribution are now examined.

3.1.   Quantile

The qf of a distribution is as useful as its cdf. It can be used to generate random numbers from the distribution, among other things. We may refer to the book by Gilchrist (2000) for a complete overview of the role of the qf in probability and statistics. In our setting, the qf of the TSGHS distribution is defined by

$$F^{-1}(q; \alpha, \beta, \mu, \sigma) = \begin{cases} \mu + \sigma G^{-1}\left(\left(q\beta^{\alpha-1}\right)^{\frac{1}{\alpha}}\right) & for\ 0 < q \leq \beta \\ \mu + \sigma G^{-1}\left(1 - \left((1-q)(1-\beta)^{\alpha-1}\right)^{\frac{1}{\alpha}}\right) & for\ \beta < q < 1 \end{cases}, \tag{10}$$

where $G^{-1}(x)$ is the qf of the HS distribution defined as follows:

$$G^{-1}(x) = \frac{2}{\pi} \log\left(\tan\left(\frac{\pi}{2}x\right)\right), \quad 0 < x < 1.$$

To generate values from a random variable X with the TSGHS distribution, we can use this function. The following result may be used: For any uniform random variable U defined on $(0,1)$, the random variable $X = F^{-1}(U; \alpha, \beta, \mu, \sigma)$ follows the TSGHS distribution. This technique will be used later for simulation purposes.

3.2. Moments

We now explore the moments of the TSGHS distribution. Let X be a random variable with the TSGHS distribution. Then, for any integer r, the rth non central moment of X is specified by

$$E(X^r) = \int_{-\infty}^{+\infty} x^r f(x; \alpha, \beta, \mu, \sigma) dx$$

$$= \frac{\alpha}{\sigma \beta^{\alpha-1}} \int_{-\infty}^{\mu+\sigma G^{-1}(\beta)} x^r g\left(\frac{x-\mu}{\sigma}\right) \left[G\left(\frac{x-\mu}{\sigma}\right)^{\alpha-1}\right] dx$$

$$= \frac{\alpha}{\sigma(1-\beta)^{\alpha-1}} \int_{\mu+\sigma G^{-1}(\beta)}^{+\infty} x^r g\left(\frac{x-\mu}{\sigma}\right) \left[1 - G\left(\frac{x-\mu}{\sigma}\right)\right]^{\alpha-1} dx,$$

where $g(x)$ and $G(x)$ are the pdf and cdf of the HS distribution as given in Equations (5) and (6), respectively.

From Korkmaz and Genc (2017) and with the change of variable $u = G((x - \mu)\sigma^{-1})$, we get

$$E(X^r) = \frac{\alpha}{\beta^{\alpha-1}} \int_0^\beta u^{\alpha-1} [\sigma G^{-1}(u) + \mu]^r du$$

$$+ \frac{\alpha}{(1-\beta)^{\alpha-1}} \sum_{j=0}^{+\infty} (-1)^j \binom{\alpha-1}{j} \int_\beta^1 u^j [\sigma G^{-1}(u) + \mu]^r du$$

Using the classical binomial expansion of $(\sigma G^{-1}(u) + \mu)^r$, and expressing $G^{-1}(u) = \frac{2}{\pi} log\left(tan\left(\frac{\pi}{2}u\right)\right)$, we obtain

$$E(X^r) = \frac{\alpha}{\beta^{\alpha-1}} \sum_{i=0}^{r} \binom{r}{i} \mu^{r-i} \sigma^i \left(\frac{2}{\pi}\right)^i \mathfrak{I}(\alpha-1, 0, \beta, i)$$

$$+ \frac{\alpha}{(1-\beta)^{\alpha-1}} \sum_{j=0}^{+\infty} \sum_{i=0}^{r} (-1)^j \binom{\alpha-1}{j} \binom{r}{i} \mu^{r-i} \sigma^i \left(\frac{2}{\pi}\right)^i \mathfrak{I}(j, \beta, 1, i),$$

where $\mathfrak{I}(\gamma, a, b, i)$ denotes the integral operator defined by

$$\mathfrak{I}(\gamma, a, b, i) = \int_{a}^{b} u^{\gamma} \left[log\left(tan\left(\frac{\pi}{2}u\right)\right)\right]^i du.$$

This integral has no closed form expression, to our knowledge, but can be computed quite easily with the help of any mathematical software. As an illustration, in Table 1, numerical work on the moments of X is performed. More precisely, we determine the mean, variance, skewness $\delta_1 = \frac{(\mu_3 - 3\mu_1\mu_2 + 2\mu_1^3)}{(\mu_2 - \mu_1^2)^{3/2}}$ and kurtosis $\delta_2 = \frac{(\mu_4 - 4\mu_1\mu_3 + 6\mu_2\mu_1^2 - 3\mu_1^4)}{(\mu_2 - \mu_1^2)^2}$, where $\mu_r = E(X^r)$, for selected values of the parameters of the TSGHS distribution. We fixed $\mu = 0$ and $\sigma = 1$.

Table 1. Mean, variance, skewness and kurtosis of the TSGHS distribution for some values of α and β with $\mu = 0$ and $\sigma = 1$

β	α	Mean	Variance	Skewness	Kurtosis
0.1	1.5	-0.2317	0.5854	-0.0181	4.1140
	2	-0.3661	0.4154	0.0168	3.5986
	5	-0.6906	0.1549	0.3168	2.8459
0.2	1.5	-0.1639	0.5284	0.0471	4.4035
	2	-0.2563	0.3438	0.1135	4.0420
	5	-0.4659	0.0934	0.4639	3.5664
0.5	1.5	-6.0871e-17	0.4795	-2.6998e-15	4.7905
	2	-4.9021e-17	0.2867	-6.8582e-16	4.6793
	5	-6.9645e-17	0.0563	2.2525e-16	4.6522
0.75	1.5	0.1340	0.5114	-0.0536	4.5225
	2	0.2088	0.3235	-0.1196	4.2334
	5	0.3752	0.0791	-0.4552	3.8825
0.95	1.5	0.2734	0.6385	0.1150	3.9887
	2	0.4358	0.4871	0.1288	3.4235
	5	0.8489	0.2349	-0.0866	2.5291

According to Table 1, the distribution is near symmetrical when $\beta = 0.5$. When $\beta < 0.5$ is fixed and $\alpha > 1$, α increases, the mean, variance and kurtosis decrease and the skewness increases. When $\beta > 0.5$ is fixed and $\alpha > 1$ increases, the mean increases as well as the variance, but the kurtosis decreases. For the HS distribution, when $\alpha = 1$, the mean, variance, skewness, and kurtosis values are 0, 1, 0 and 5 respectively. We can see that the distribution can be right-skewed, left-skewed, or symmetrical and usually has a heavy tail.

We illustrate the skewness and kurtosis of the TSGHS distribution in Figure 5 for $\beta = 0.1$, $\mu = 0$ and $\sigma = 1$.

(a)	(b)
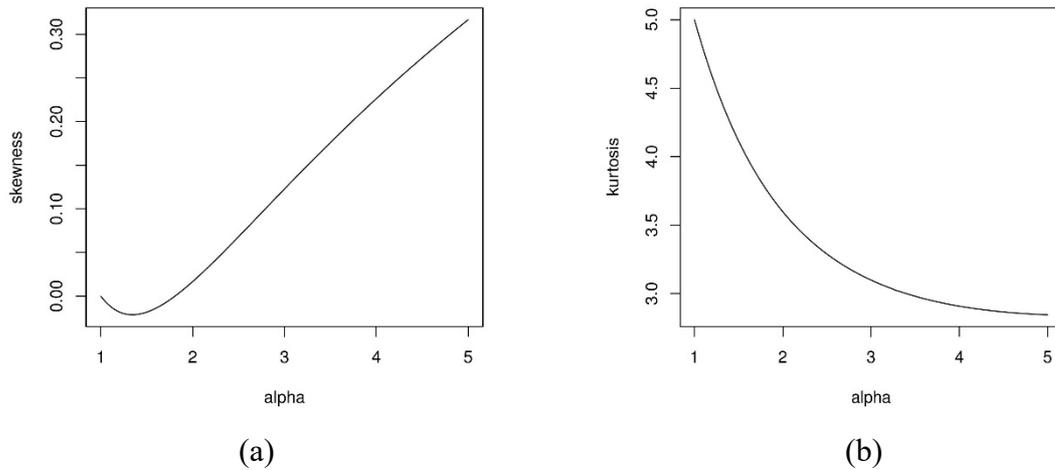
Figure 5. Curves of the (a) skewness and (b) kurtosis of the TSGHS distribution for selected parameters value

Figure 5(a) shows that, for $1 < \alpha < 2$, the distribution is slightly left-skewed, it becomes fairly symmetrical near $\alpha = 2$ and then becomes right-skewed. It is worth noting that, when $\beta = 0.9$, it is the opposite, the distribution is slightly right-skewed but becomes left-skewed when α exceeds 2. In Figure 5(b), the distribution is leptokurtic at first but is closer to the normal kurtosis, mesokurtic, when α approaches 5. When $\alpha > 5$, the kurtosis of the distribution gradually increases.

4.   ESTIMATION

The inference of the TSGHS model is the main line of this section.

4.1.   Maximum Likelihood Estimates

Let $X_1, X_2, \ldots, X_n$ be independent random variables following the $TSGHS(\alpha, \beta, \mu, \sigma)$ distribution, with unknown parameters. Let $x_1, x_2, \ldots, x_n$ be observations of these variables and $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ denote the corresponding order values. From Korkmaz and Genc (2017), the log-likelihood function of the TSGHS model is defined by

$$\ell(\alpha, \beta, \mu, \sigma) = \tilde{\ell}(\alpha, \beta, \mu, \sigma, r) = nlog(\alpha) - nlog(\sigma) + \sum_{i=1}^{n} log\left[g\left(\frac{x_i - \mu}{\sigma}\right)\right] + (\alpha -$$

$$1)log\left[\frac{\prod_{i=1}^{r} G\left(\frac{x_i - \mu}{\sigma}\right)\prod_{i=r+1}^{n}\left(1 - G\left(\frac{x_i - \mu}{\sigma}\right)\right)}{\beta^r (1-\beta)^{n-r}}\right], \tag{11}$$

where $x_{(r)} \leq \mu + \sigma G^{-1}(\beta) \leq x_{(r+1)}$ for $r = 1,2,\dots,n$, $x_{(0)} = -\infty$, $x_{(n+1)} = +\infty$, $g(x)$ and $G(x)$ are the pdf and cdf of the HS distribution as given in Equations (5) and (6), respectively. Then, the maximum likelihood estimates (MLEs) of α, β, μ and σ, say, $\hat{\alpha}, \hat{\beta}, \hat{\mu}$ and $\hat{\sigma}$ are given by

$$\left(\hat{\alpha},\hat{\beta},\hat{\mu},\hat{\sigma}\right) = \underset{(\alpha,\beta,\mu,\sigma)}{argmax}\ell(\alpha,\beta,\mu,\sigma)$$

Based on van Dorp and Kotz (2002a) and Korkmaz and Genc (2017), provided that μ and σ are fixed, the MLEs of α and β are

$$\hat{\alpha} = -\frac{n}{\log[M(\hat{r},\mu,\sigma)]},$$

and

$$\hat{\beta} = \frac{2}{\pi}arctan\left(e^{\pi\frac{x_i-\mu}{2\sigma}}\right),$$

respectively, where $\underset{r\in\{1,2,\dots,n\}}{argmax}M(r,\mu,\sigma)$ with

$$M(r,\mu,\sigma) = \prod_{i=1}^{r-1}\frac{G\left(\frac{x_i-\mu}{\sigma}\right)}{G\left(\frac{x_r-\mu}{\sigma}\right)}\prod_{i=r+1}^{n}\frac{1-G\left(\frac{x_i-\mu}{\sigma}\right)}{1-G\left(\frac{x_r-\mu}{\sigma}\right)}.$$

To find the estimates of the parameters μ and σ, we can use an algorithm described as follows:

Step 1: Set $k = 0$ and put initial values $\hat{\mu}^{(0)}$ and $\hat{\sigma}^{(0)}$ for μ and σ in the log-likelihood function.

Step 2: Compute the following estimates:

$$\hat{\beta}^{(k+1)} = \frac{2}{\pi}arctan\left(e^{\pi\left(\frac{x_{\hat{r}(k)}-\hat{\mu}^{(k)}}{2\hat{\sigma}^{(k)}}\right)}\right)$$

$$\hat{\alpha}^{(k+1)} = -\frac{n}{\log[M(\hat{r}^{(k)},\hat{\mu}^{(k)},\hat{\sigma}^{(k)})]}$$

where $\underset{r\in\{1,2,\dots,n\}}{argmax}M\left(r,\hat{\mu}^{(k)},\hat{\sigma}^{(k)}\right)$ with

$$M\left(r, \hat{\mu}^{(k)}, \hat{\sigma}^{(k)}\right) = \prod_{i=1}^{r-1} \frac{G\left(\frac{x_i - \hat{\mu}^{(k)}}{\hat{\sigma}^{(k)}}\right)}{G\left(\frac{x_r - \hat{\mu}^{(k)}}{\hat{\sigma}^{(k)}}\right)} \prod_{i=r+1}^{n} \frac{1 - G\left(\frac{x_i - \hat{\mu}^{(k)}}{\hat{\sigma}^{(k)}}\right)}{1 - G\left(\frac{x_r - \hat{\mu}^{(k)}}{\hat{\sigma}^{(k)}}\right)}$$

Step 3: Update μ and σ by using the following system of equations: $\frac{\partial \tilde{\ell}\left(\alpha^{(k)}, \beta^{(k)}, \mu, \sigma, \hat{r}^{(k)}\right)}{\partial \mu} = 0$ and

$\frac{\partial \tilde{\ell}\left(\alpha^{(k)}, \beta^{(k)}, \mu, \sigma, \hat{r}^{(k)}\right)}{\partial \sigma} = 0$ to find $\hat{\mu}^{(k+1)}$ and $\hat{\sigma}^{(k+1)}$, also giving $\hat{r}^{(k+1)}$.

Step 4: If $\left|\tilde{\ell}\left(\hat{\alpha}^{(k+1)}, \hat{\beta}^{(k+1)}, \hat{\mu}^{(k+1)}, \hat{\sigma}^{(k+1)}, \hat{r}^{(k+1)}\right) - \tilde{\ell}\left(\hat{\alpha}^{(k+1)}, \hat{\beta}^{(k+1)}, \hat{\mu}^{(k)}, \hat{\sigma}^{(k)}, \hat{r}^{(k)}\right)\right|$ is less

than a given tolerance, 10^{-2} for example, then stop, and consider $\hat{\alpha}^{(k)}, \hat{\beta}^{(k)}, \hat{\mu}^{(k)}$ and $\hat{\sigma}^{(k)}$ as

the MLEs of $\hat{\alpha}, \hat{\beta}, \hat{\mu}$ and $\hat{\sigma}$.

Else $k = k + 1$ then go back to Step 2.

4.2. Simulation Study

In this section, empirical results are given to see the performance of the established MLEs. The simulation study is summarized via graphics.

To begin, $N = 1000$ samples of size $n = 20, 30, \ldots, 1000$ from a random variable following the TSGHS distribution are generated. The true values of the parameters are taken as $\alpha = 5$, $\beta = 0.5$, $\mu = 0$ and $\sigma = 1$. The qf described in Equation (10) is used for the generation of random numbers. For evaluating the performance of the estimates, the empirical mean, bias, and mean square error (MSE) are obtained for comparisons. By setting $\epsilon = \alpha, \beta, \mu$ or σ, the related empirical mean, bias, and MSE are calculated by
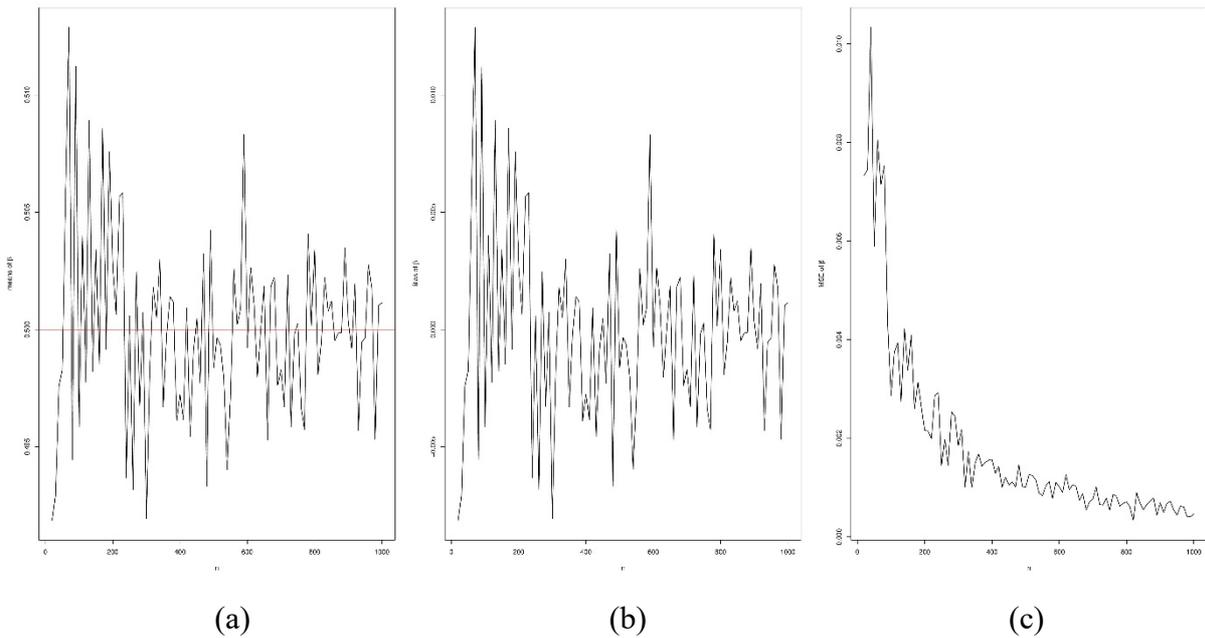
$$E_n(\epsilon) = \frac{1}{N}\sum_{i=1}^{N} \hat{\epsilon}_i, \quad Bias_n(\epsilon) = \frac{1}{N}\sum_{i=1}^{N}(\epsilon - \hat{\epsilon}_i), \quad MSE_n(\epsilon) = \frac{1}{N}\sum_{i=1}^{N}(\epsilon - \hat{\epsilon}_i)^2,$$

respectively, where $\hat{\epsilon}_i$ denotes the MLE of ϵ obtained at the i$^{\text{th}}$ sample.
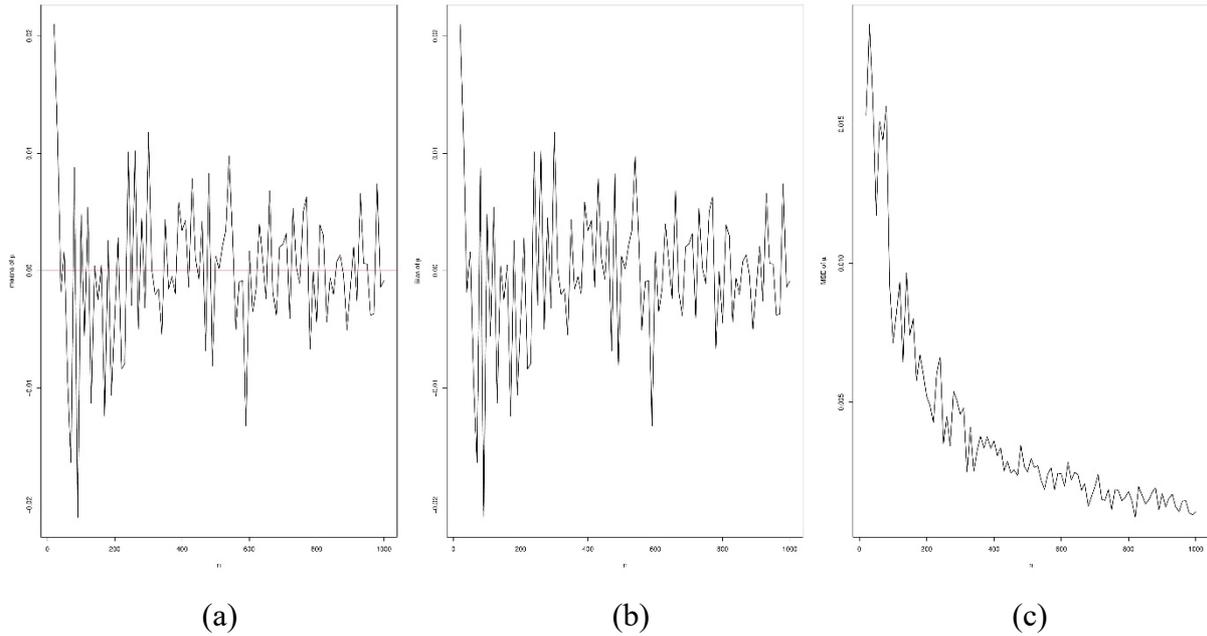
Figure 6, 7, 8 and 9 give a visualization of the results of the simulation studies related to α, β, μ or σ, respectively.

(a)   (b)   (c)
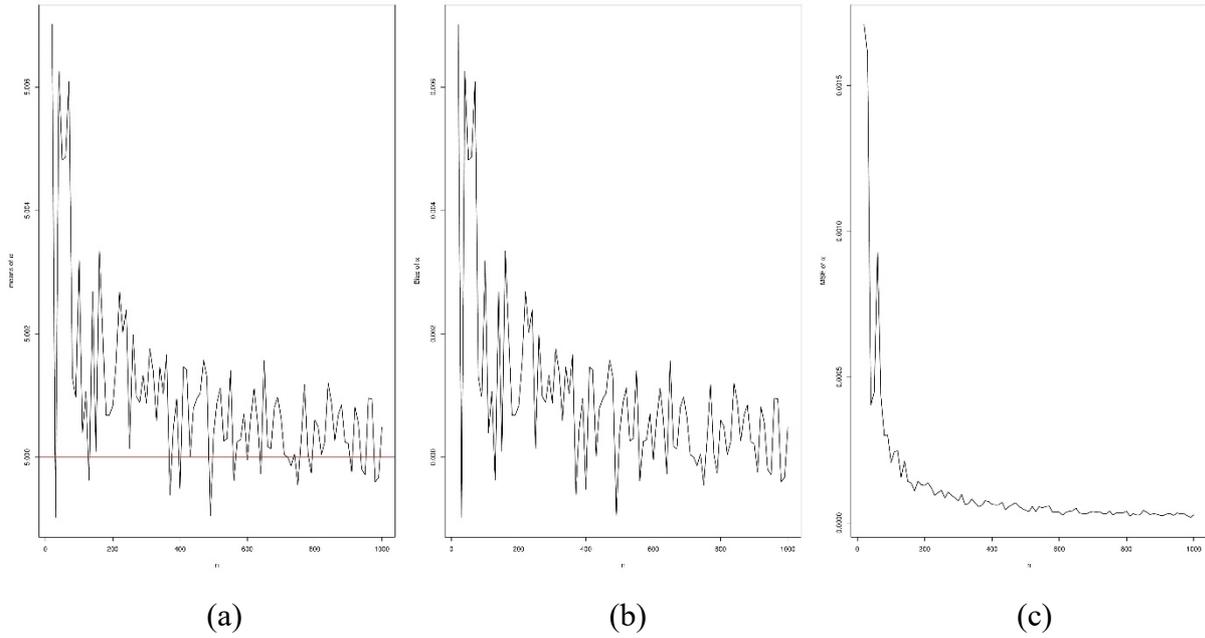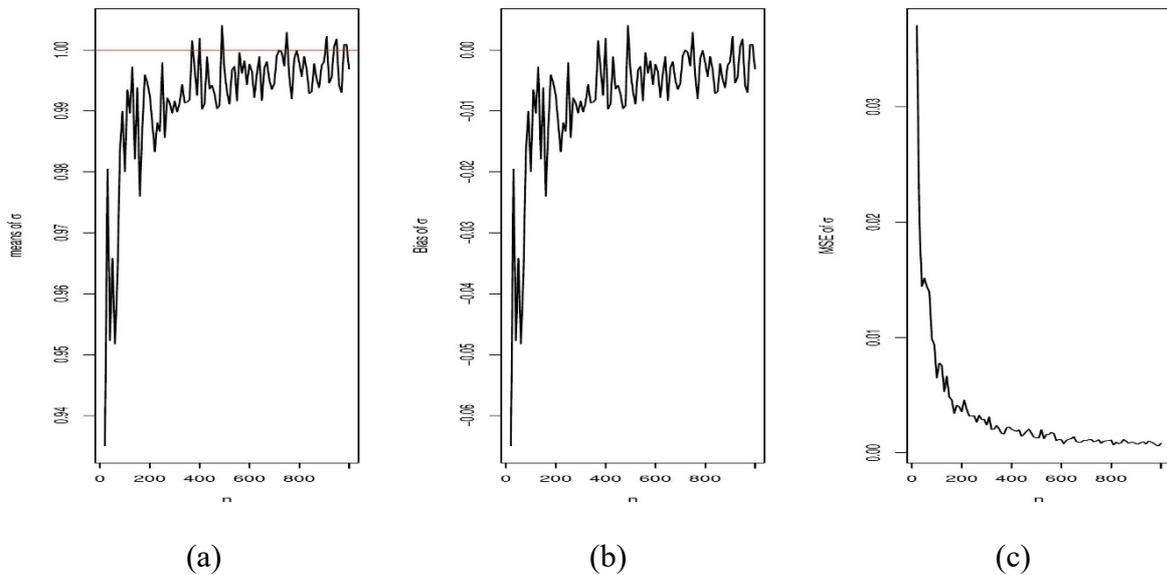
Figure 6. Graphical results of the parameter α for the simulation study: (a) empirical mean, (b) bias and (c) MSE

(a)   (b)   (c)

Figure 7. Graphical results of the parameter β for the simulation study: (a) empirical mean, (b) bias and (c) MSE

| (a) | (b) | (c) |

Figure 8. Graphical results of the parameter μ for the simulation study: (a) empirical mean, (b) bias and (c) MSE

| (a) | (b) | (c) |

Figure 9. Graphical results of the parameter σ for the simulation study: (a) empirical mean, (b) bias and (c) MSE

These figures show that the expected result with increasing sample size n is that empirical means are close to true values, MLEs are asymptotically unbiased, and all MSEs go to zero. At the same time, the empirical results are close, when the sample size increases.

5. REAL DATA APPLICATIONS

This section is devoted to the applicability of the TSGHS model for data analysis.

5.1. Method

We fit three data sets using the TSGHS model, and compare it to other existing models, namely the beta-normal (BN), power-normal (PN), Kuramaswamy-normal (KwN) and two-sided generalized normal (TSGN) models. To compare them, the unknown parameters of the models are estimated by the maximum likelihood method, and we use the Akaike information criterion and its correction (AIC and CAIC), the Bayesian information criterion (BIC) and the Hannan-Quinn information criterion (HQIC).

### 5.1.1. First Application: Time to Failure of Turbochargers

The data are made up of 40 measurements of the time to failure of turbochargers diesel engine. These data were used by Noor and Mundher (2020) and Nasir et al. (2019). The data are Data = (1.6, 3.5, 4.8, 5.4, 6.0, 6.5, 7.0, 7.3, 7.7, 8.0, 8.4, 2.0, 3.9, 5.0, 5.6, 6.1, 6.5, 7.1, 7.3, 7.8, 8.1, 8.4, 2.6, 4.5, 5.1, 5.8, 6.3, 6.7, 7.3, 7.7, 7.9, 8.3, 8.5, 3.0, 4.6, 5.3, 6.0, 8.7, 8.8, 9.0). We report the estimates found using R program and criterions of the TSGHS model and other competitors in Table 2.

Table 2. MLEs, AIC, CAIC, BIC and HQIC values related to the TSGHS model and other models

	α	β	μ	σ	AIC	CAIC	BIC	HQIC
TSGHS	38.1109	0.9687	-6.6773	8.0658	164.0750	165.2179	170.8305	166.5176
BN	38.6150	0.0333	10.1861	0.5534	165.0574	166.2003	171.8129	167.5000
PN	0.0060	-	9.5288	0.2113	165.8376	166.5043	170.9042	167.6695
KwN	0.0121	99.9305	21.7733	0.5376	169.6174	170.7603	176.3730	172.0600
TSGN	19.6023	0.9423	-6.0820	9.3878	165.1690	166.3119	171.9245	167.6116
BHS	0.3212	12.3274	9.8663	0.6817	167.8785	169.0213	174.6340	170.3211

Table 2 indicates that our model is the better fit regarding the criteria used, although very close to the other competitors.

Figure 10 shows the fit of the estimated pdf and cdf of the TSGHS model to the data.

(a)   (b)
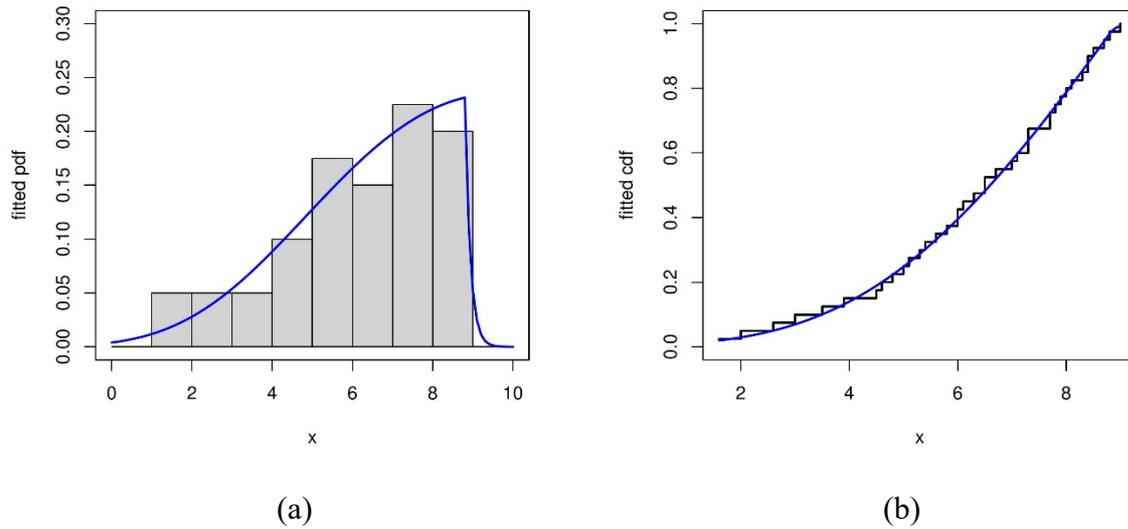
Figure 10. Graphical representations of the time to failure of turbochargers data set with (a) the estimated pdf and (b) the estimated cdf of the TSGHS model

In particular, from Figure 10(a), we see that the estimated pdf of the TSGHS model captures both the left skewness of the data and the sudden drop in values around $x = 9$.

For a visual comparison with the other models, Figure 11 presents the corresponding estimated pdfs.
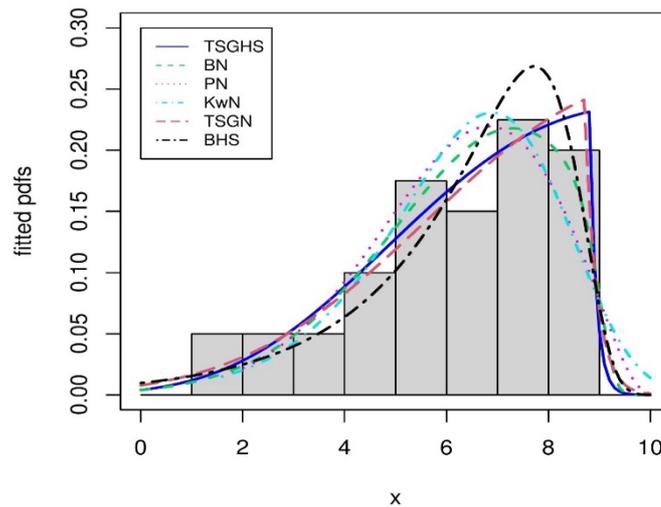
Figure 11. Graphical representation of the time to failure of turbochargers data set and the estimated pdfs of the TSGHS, TSGN, PN, BN, KwN and BHS models

Based on Figure 11, the PN, BN, KwN, and BHS models can't seem to adjust well to the sharp fall of the data. The TSGN model, which comes from a generalized normal distribution using the two-sided distribution as a generator, looks very similar to the fitting behavior of our model. This example of application shows that our distribution has some potential to fit data.

5.1.2. Second Application: Remission Times (in months) of 128 Bladder Cancer Patients

The considered data set is a random sample of 128 bladder cancer patients with their remission times (in months).

The data set comes from Lee and Wang (2003) and has been used by Reyes et al. (2017). The data are Data = (0.08, 2.09, 3.48, 4.87, 6.94, 8.66, 13.11, 23.63, 0.20, 2.23, 3.52, 4.98, 6.97, 9.02, 13.29, 0.40, 2.26, 3.57, 5.06, 7.09, 9.22, 13.80, 25.74, 0.50, 2.46, 3.64, 5.09, 7.26, 9.47, 14.24, 25.82, 0.51, 2.54, 3.70, 5.17, 7.28, 9.74, 14.76, 26.31, 0.81, 2.62, 3.82, 5.32, 7.32, 10.06, 14.77, 32.15, 2.64, 3.88, 5.32, 7.39, 10.34, 14.83, 34.26, 0.90, 2.69, 4.18, 5.34, 7.59, 10.66, 15.96, 36.66, 1.05, 2.69, 4.23, 5.41, 7.62, 10.75, 16.62, 43.01, 1.19, 2.75, 4.26, 5.41, 7.63, 17.12, 46.12, 1.26, 2.83, 4.33, 5.49, 7.66, 11.25, 17.14, 79.05, 1.35, 2.87, 5.62, 7.87, 11.64, 17.36, 1.40, 3.02, 4.34, 5.71, 7.93, 11.79, 18.10, 1.46, 4.40, 5.85, 8.26, 11.98, 19.13, 1.76, 3.25, 4.50, 6.25, 8.37, 12.02, 2.02, 3.31, 4.51, 6.54, 8.53, 12.03, 20.28, 2.02, 3.36, 6.76, 12.07, 21.73, 2.07, 3.36, 6.93, 8.65, 12.63, 22.69).

The results of the criteria used to compare models are reported in Table 3.

Table 3. MLEs, AIC, CAIC, BIC and HQIC values related to the TSGHS model and other models

|	α	β	μ	σ	AIC	CAIC	BIC	HQIC
TSGHS	264.3606	0.0689	399.5727	282.0842	841.1545	841.4797	852.5626	845.7896
BN	0.5154	17.7705	-17.1429	11.7873	907.0064	907.3316	918.4145	911.6415
PN	1.6016e+11	-	258.1246	38.9431	876.5680	876.7616	885.1241	880.0444
KwN	1.5108	0.5635	1.4179	9.5096	952.6799	953.0051	964.0880	957.3151
TSGN	133.1274	0.1110	310.5948	252.7084	843.5995	843.9247	855.0076	848.2347
BHS	13.7641	0.3056	-3.6112	2.3304	843.4384	843.7636	854.8465	848.0736

Table 3 exposes that our proposed model has the best fit, with a very small AIC for these data: AIC = 341.1545. Figure 12 isolates the fit of the TSGHS model via complementary graphical approaches.

(a)   (b)
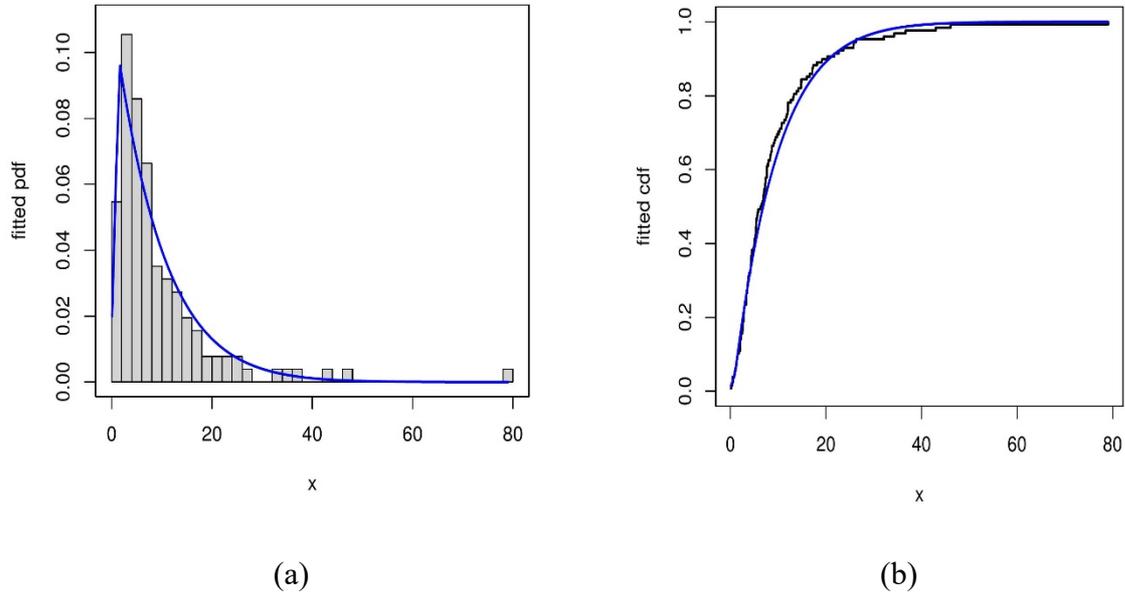
Figure 12. Graphical representations of the remission times of bladder cancer patients data set with (a) the estimated pdf and (b) the estimated cdf of the TSGHS model

Figure 12(a) shows right skewed data with a heavy tail and the fit of the TSGHS model which captures quite well the sharp peak of the data. In Figure 12(b), we can see that the estimated cdf of our model has also a good fit.

Visual comparisons with the other models are given in Figure 13.
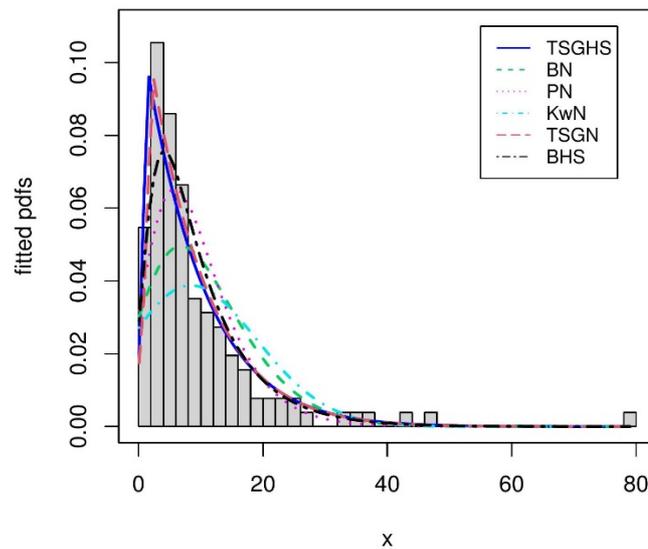
Figure 13. Graphical representation of the remission times of bladder cancer patients data set and the estimated pdfs of the TSGHS, TSGN, PN, BN, KwN and BHS models

When we look at the other competitors in Figure 13, we can see that the PN, BN, and KwN models struggle to adjust to the fairly-peaked data. The BHS model goes higher than the three models, but not as high as the TSGHS model. As 112ort he TSGN model, it looks very similar to our model. Indeed, the HS distribution, which is the baseline of the TSGHS distribution, is a bell-shaped distribution like the normal distribution, which is the base distribution of the TSGN distribution. Both distributions are very flexible when it comes to data fitting, especially high-peaked data.

### 5.1.3. Third Application: Repair Lifetimes of an Airborne Transceiver

The data set represents 46 repair times in hours for an airborne communication transceiver from Chhikara and Folks (1977). The data are Data = (0.2, 0.3, 0.5, 0.5, 0.5, 0.5, 0.6, 0.6, 0.7, 0.7, 0.7, 0.8, 0.8, 1.0, 1.0, 1.0, 1.0, 1.1, 1.3, 1.5, 1.5, 1.5, 1.5, 2.0, 2.0, 2.2, 2.5, 2.7, 3.0, 3.0, 3.3, 3.3, 4.0, 4.0, 4.5, 4.7, 5.0, 5.4, 5.4, 7.0, 7.5, 8.8, 9.0, 10.3, 22.0, 24.5).

Table 3 reports the results of the criteria used to compare models.

Table 4. MLEs, AIC, CAIC, BIC and HQIC values related to the TSGHS model and other models

TSGHS	α	β	μ	σ	AIC	CAIC	BIC	HQIC
BN	159.0335	0.0758	106.9668	78.7840	220.5556	221.5312	227.8702	223.2957
PN	2.9437	79.0839	-30.6061	18.3417	274.1300	275.1056	281.4446	276.8701
KwN	1.1564e+12	-	-123.9995	17.8545	244.8911	245.4625	250.3770	246.9461
TSGN	63.1287	0.2995	-8.4773	4.0249	250.8490	251.8246	258.1636	253.5891
BHS	89.6610	0.0603	67.0336	42.8625	224.2761	225.2518	231.5907	227.0162
TSGHS	76.0448	0.2356	-2.8074	0.7683	225.6571	226.6327	232.9716	228.3972

When we compare the criteria of the TSGHS model with those of the other competitors in Table 4, we can see that our model has the best fit. This fit is illustrated in Figure 14.

(a) (b)
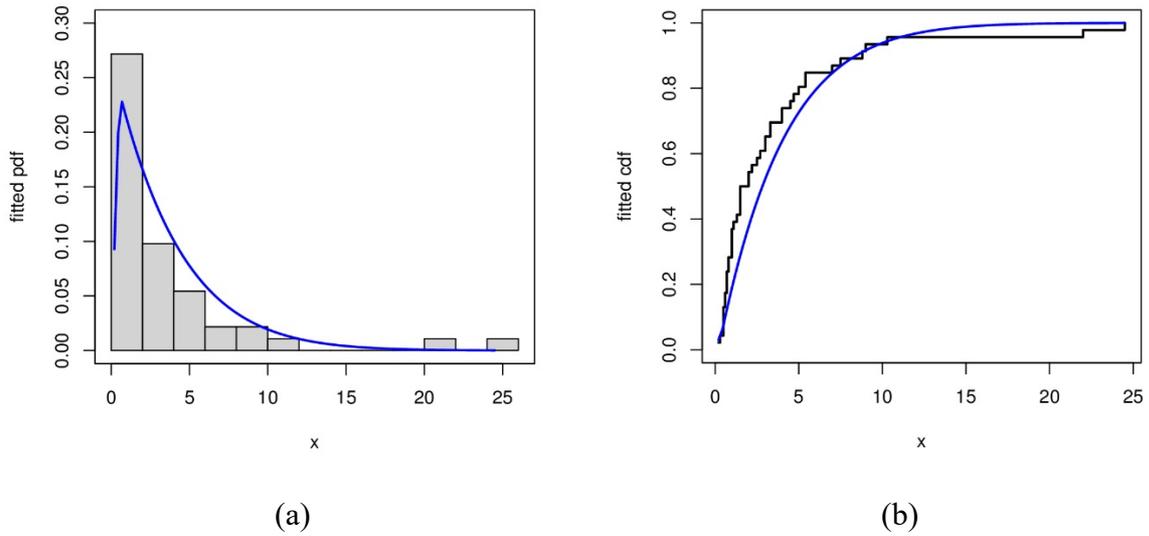
Figure 14. Graphical representations of the repair lifetimes of an airborne transceiver data set with (a) the estimated pdf and (b) the estimated cdf of the TSGHS model

The data set is similar to the previous one, right skewed with a heavy tail as we can see on Figure 14(a). The estimated pdf of the TSGHS model fits the data quite well as the estimated cdf plotted in Figure 14(b). Figure 15 performs a visual estimated model comparison based on the fitted pdf.
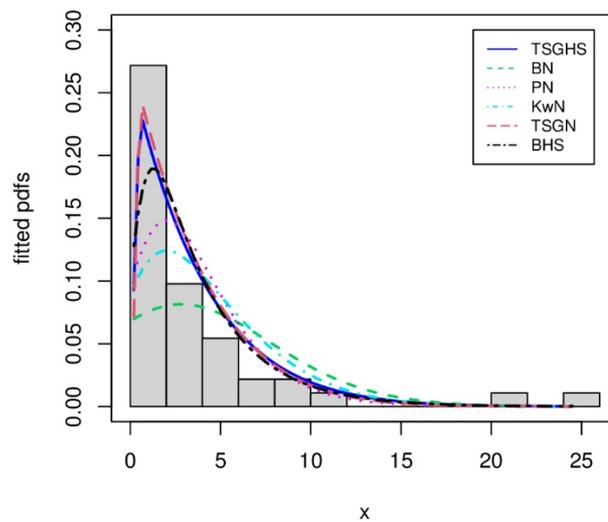
Figure 15. Graphical representation of the repair lifetimes of an airborne transceiver data set and the estimated pdfs of the TSGHS, TSGN, PN, BN, KwN and BHS models

From Figure 15, it is clear that the PN, BN, KwN, and BHS models can't reach the peak of the data, unlike the TSGHS and TSGN models. This example shows that the TSG-G family allows the creation of more flexible models for such a data analysis scenario.

6.   CONCLUSION

In this paper, we introduced a generalized hyperbolic secant distribution using the standard two-sided power distribution as a generator. The corresponding model is made to fit data that has an abrupt change in values. We studied some of its properties and estimated the parameters using the maximum likelihood estimation procedure. In this regard, an efficient algorithm is given, and a simulation study showing the behavior of the estimates is performed. The obtained distribution is very flexible as it can be unimodal, right or left skewed, symmetric or bimodal with equal or different heights. This flexibility is illustrated for the analysis of three different data sets of importance, with fair comparison to important models of the literature. As a result, we contribute to the evolution of the so-called two-sided family of distributions, demonstrating that its members have a lot of modeling potential.

ETHICAL DECLARATION

In the writing process of the study titled "Two-sided Generalized Hyperbolic Secant Distribution with Real Data Applications and Model Comparisons", there were followed the scientific, ethical and the citation rules; was not made any falsification on the collected data and this study was not sent to any other academic media for evaluation.

REFERENCES

Chhikara, R. S. and Folks, J. L. (1977), The inverse gaussian distribution as a lifetime model, *Technometrics*, 9(4), 461-468.

Cordeiro, G. and Castro, M. (2011), A new family of generalized distributions, *Journal of Statistical Computation and Simulation*, 81, 883-898.

Eugene, N., Lee, C. and Famoye, F. (2002), Beta-normal distribution and its application, *Communications in Statistics-theory and Methods*, 31, 497-512.

Fischer, M. (2013), Generalized hyperbolic secant distributions: with Applications to Finance, Springer.

Fischer, M. and Vaughan, D. (2002), Classes of skew generalized hyperbolic secant distributions, Friedrich-Alexander-University Erlangen-Nuremberg, Chair of Statistics and Econometrics, *Discussion Papers*.

Fischer, M. and Vaughan, D. (2004), The Beta-Hyperbolic secant (BHS) distribution, Friedrich-Alexander-UniversityErlangen-Nuremberg, Chair of Statistics and Econometrics, *Discussion Papers*, 39.

Gilchrist, W. (2000), Statistical modelling with quantile functions, CRC Press, Abingdon.

Gupta, R. C. and Gupta, R. D. (2008), Analyzing skewed data by power normal model, *Test* 17, 197- 210.

Harkness, W. L. and Harkness, M. L. (1968), Generalized hyperbolic secant distributions, *Journal of the American Statistical Association*, 63(321), 329-337.

Korkmaz, M. Ç. (2015), Two-sided generalized Gumbel distribution with application to air pollution data, *International Journal of Statistical Distributions and Applications*, 1, 19-26.

Korkmaz, M. Ç. and Genç, A. İ. (2017), A new generalized two-sided class of distributions with an Emphasis on Two-sided Generalized Normal Distribution, *Communications in Statistics - Simulation and* Computation, doi: 10.1080/03610918.2015.1005233.

Korkmaz, M. Ç. and Genç, A. İ. (2015b), Two-sided generalized exponential distribution, *Communications in Statistics - Theory and Methods*, 44(23), 5049-5070.

Korkmaz, M. Ç. and Genç, A. İ. (2015a). A lifetime distribution based on a transformation of a two-sided power variate, *Journal of Statistical Theory and Applications*, 14(3), 265–280.

Lee, E. T. and Wang, J. W. (2003), Statistical methods for survival data analysis, John Wiley & Sons, Inc.

Morris, C. N. (1982), Natural exponential families with quadratic variance functions, *The Annals of Statistics*, 10(1), 65-80.

Nasir, A., Jamal, F. and Chesneau, C. (2019), The odd generalized gamma-G family of distributions: Properties, regressions and applications. 2019. hal-01916828v4.

Noor, I. and Mundher, K. (2020), Generalizations of Burr type X distribution with applications, *ASM Science Journal*. 1-8.

Perez, J., Rambaud, S. and Garcia, C. (2005), The two-sided power distribution for the treatment of the uncertainty in PERT, *Statistical Methods and Applications*, 14, 209-222.

Perks, W. F. (1932), On some experiments in the graduation of mortality statistics. *Institute of Actuaries Journals*, Series B, 58, 12-57.

Reyes, J., Venegas, O. and G´omez G. H. (2017), Modified Slash Lindley distribution, *Journal of Probability and Statistics*.1-9.

Van Dorp, J. R. and Kotz, S. (2002a), The standard two-sided power distribution and its properties: With applications in financial engineering. *The American Statistician,* 56, 90-99.

Van Dorp, J. R. and Kotz, S. (2002b), A novel extension of the triangular distribution and its parameter estimation, *The American Statistician*, 51, 63-79.

Vaughan, D. C. (2002). The generalized secant hyperbolic distribution and its properties, *Communications in Statistics - Theory and Methods*, 31(2), 219-238.