Modern Stochastics: Theory and Applications 10 (3) (2023) 287–312
https://doi.org/10.15559/23-VMSTA227

On some composite Kies families: distributional properties and saturation in Hausdorff sense

Tsvetelin Zaevski[a,*], **Nikolay Kyurkchiev**[a,b]

[a]*Institute of Mathematics and Informatics, Bulgarian Academy of Sciences,Acad. Georgi Bonchev Str., Block 8, 1113 Sofia, Bulgaria*
[b]*Faculty of Mathematics and Informatics, University of Plovdiv Paisii Hilendarski, Bulgaria*

t_s_zaevski@math.bas.bg, t_s_zaevski@abv.bg (T. Zaevski), nkyurk@math.bas.bg (N. Kyurkchiev)

Received: 9 November 2022, Revised: 7 January 2023, Accepted: 12 March 2023,
Published online: 21 March 2023

Abstract The stochastic literature contains several extensions of the exponential distribution which increase its applicability and flexibility. In the present article, some properties of a new power modified exponential family with an original Kies correction are discussed. This family is defined as a Kies distribution which domain is transformed by another Kies distribution. Its probabilistic properties are investigated and some limitations for the saturation in the Hausdorff sense are derived. Moreover, a formula of a semiclosed form is obtained for this saturation. Also the tail behavior of these distributions is examined considering three different criteria inspired by the financial markets, namely, the VaR, AVaR, and expectile based VaR. Some numerical experiments are provided, too.

Keywords Exponential distribution, Weibull distribution, Kies distribution, estimator, tail behavior, Hausdorff saturation

MSC2020 MSC 33C15, 60E05, 60E10

1 Introduction

The Weibull distribution is one of the most important generalizations of the exponential distribution. Despite of the loss of the important memorylessness feature, the

*Corresponding author.

Weibull distribution has several advantages which determine its wide use in many real life areas including engineering sciences [29], meteorology and hydrology [28], communications and telecommunications [26], energetics [22], chemical and metallurgical industry [4], epidemiology [16], insurance and banking [6], etc. For more theoretical and practical details related to this distribution we refer to the original work [27] as well as the recently published books [18, 20, 17], and [15].

A significant modification of the Weibull distribution known as a Kies distribution was firstly proposed in [9] and recently considered in many studies. It reduces the positive real half-line support of the Weibull to the interval (0, 1) changing the variables as $t = \frac{x}{x+1}$. Later, many authors discussed several modifications. Refs. [11, 21] and [30] examine four parameter distributions whose domain is translated to an arbitrary positive interval. Refs. [12, 13] take a power of the Kies cumulative distribution function (CDF, hereafter) to define a new family.

A composition approach for the Kies distribution is presented in [2]. This way the new distribution is defined after the change of variables $t = H(x)$ where $H(x)$ is the CDF of an auxiliary distribution. Ref. [25] introduces the Fréchet distribution for this purpose and later in [24] the resulting Kies-Fréchet distribution is applied to model the COVID 19 mortality. Alternatively, [1] and [3] use the exponential and Lomax distributions, respectively. See also [5] for another composition based on the generalized uniform distribution.

In the present article we define a new Kies family by its CDFs which are constructed as a composition of two other Kies CDFs on the interval (0, 1), $G(H(t))$. Note that this definition is possible due to this domain. We name the distributions $G(\cdot)$ and $H(\cdot)$ the original one and the correction. Thus we have a four-parameter family – two parameters for each initial distribution. We investigate the probabilistic properties of the resulting distribution in the light of its compositional essence. We derive many relations between the corresponding terms – CDF, probability density function, quantile function, mean residual life function, different expectations and moments – of the resulting and the initial distributions. Also we investigate the tail behavior by the use of three risk measures arising in the modern capital markets, namely VaR (abbreviated from Value-at-Risk), AVaR (Average-Value-at-Risk, also known as CVAR and TVAR), and expectile based VaR.

Other important results we derive are related to the so-called Hausdorff saturation. It presents the distance between the CDF and a Γ-shaped curve connecting its endpoints. In fact, the saturation measures how the distribution mass is located in the domain – the distribution is more left-placed when the saturation is lower, and vice versa. Also, when studying specific classes of cumulative distribution functions it is important to know their intrinsic characteristics – the saturation in the Hausdorff sense is namely such one. This characteristic is important for researchers in choosing an appropriate model for approximating specific data from different branches of scientific knowledge such as Biostatistics, Population dynamics, Growth theory, Debugging and Test theory, Computer viruses propagation, Insurance mathematics. In addition, the use of composite Kiess families can also be useful in the study of reaction-kinetic models – a similar study of the dynamics of the classical Kiess model is discussed in [14]. We obtain in this paper an interval evaluation of the saturation and investigate its power w.r.t. the four distribution parameters. Moreover, we prove a formula of a semiclosed form for the Hausdorf saturation. Many numerical experiments are provided.

The next question we discuss is related to the inverse problem – the calibration w.r.t. some empirical data. It is accepted in the present literature that the maximum likelihood estimation is a good approach for the Kies style distributions due to the available closed form estimator. However, our numerical simulations do not support this opinion. For this we construct an algorithm based on the least square errors. It turns out that this method produces very believable outcomes.

To illustrate our results we explore an empirical data from the real financial markets, namely, for the S&P500 index. It is well known that there are high- and low-volatility periods. In fact, this is the ubiquitous phenomenon of volatility clustering. We extract the periods between two market shocks and examine the distribution of their lengths. We compare the results which the corrected Kies distribution returns with the outcome of its ancestors, namely, the exponential, the Weibull, and the original Kies distributions.

The paper is organized as follows. Section 2 defines the new class of distributions and discusses their probabilistic properties. The tail behavior is examined in Section 3 trough the measures VaR, AVaR, and expectile based VaR. The Hausdorff distance and the related saturation are considered in Section 4. We discuss the calibration problem in Section 5. Finally, we present a numerical example based on the S&P500 index in Section 6.

2 Definitions and distributional properties

We shall use for convenience the following notations in the whole paper. The cumulative distribution function (CDF, as we mentioned above) of a distribution will be denoted by an uppercase letter, the overlined letter will be used for the complementary cumulative distribution function (CCDF), the corresponding lowercase letter is preserved for the probability density function (PDF), and finaly the letter Q indexed by the CDF's letter will mean the quantile function (QF). For example, if $F(t)$ is the CDF, then $\overline{F}(t)$, $f(t)$, and $Q_F(t)$ are the corresponding CCDF, PDF, and QF, respectively. Also, we shall use the Greek letter ξ for random variables, and we shall mark it by the corresponding CDF letter in the subscript, i.e. ξ_F means a random variable the CDF of which is $F(t)$.

The standard Kies distribution is defined on the domain $(0, 1)$ by its CDF

$$H(t) := 1 - e^{-k\left(\frac{t}{1-t}\right)^b} \tag{1}$$

for some positive parameters b and k. Inverting CDF (1) we can derive the quantile function for $t \in (0, 1)$:

$$Q_H(t) = \frac{(-\ln(1-t))^{\frac{1}{b}}}{k^{\frac{1}{b}} + (-\ln(1-t))^{\frac{1}{b}}}. \tag{2}$$

Differentiating equation (1) we obtain the probability density function

$$h(t) = bke^{-k\left(\frac{t}{1-t}\right)^b} \frac{t^{b-1}}{(1-t)^{b+1}}. \tag{3}$$

The following proposition describes the shape of PDF (3).

Proposition 2.1. *The value of the PDF at the right end of the distribution domain is zero,* $h(1) = 0$. *Let the function* $\alpha(t)$ *for* $t \in (0, 1)$ *be defined as*

$$\alpha(t) := kb\left(\frac{t}{1-t}\right)^b - (2t + b - 1). \tag{4}$$

The following statements for PDF (3) *w.r.t. the position of the power* b *w.r.t.* 1 *hold.*

1. *If* $b > 1$, *then PDF* (3) *is zero in the left domain's endpoint,* $h(0) = 0$. *Function* (4) *has a unique root for* $t \in (0, 1)$, *we denote it by* t_2. *The PDF increases for* $t \in (0, t_2)$ *having a maximum for* $t = t_2$ *and decreases for* $t \in (t_2, 1)$.

2. *If* $b = 1$, *then the left limit of the PDF is* $h(0) = k$. *If* $k \geq 2$, *then the PDF is a function decreasing from* k *to* 0. *Otherwise, if* $k < 2$, *then we introduce the value* $t_2 = 1 - \frac{k}{2}$; *note that* $t_2 \in (0, 1)$. *The PDF starts from the value* k *for* $t = 0$, *increases to a maximum for* $t = t_2$, *and decreases to zero.*

3. *If* $b < 1$, *then* $h(0) = \infty$. *The derivative of function* (4) *is*

$$\alpha'(t) = kb^2\frac{t^{b-1}}{(1-t)^{b+1}} - 2. \tag{5}$$

Let $\bar{t}$ *be defined as* $\bar{t} := \frac{1-b}{2}$. *The PDF is a decreasing function when* $\alpha'(\bar{t}) \geq 0$. *Suppose that* $\alpha'(\bar{t}) < 0$. *In this case derivative* (5) *has two roots in the interval* $(0, 1)$; *we denote them by* $\bar{t}_1$ *and* $\bar{t}_2$. *If* $\alpha(\bar{t}_2) \geq 0$, *then the PDF decreases in the whole distribution domain. Otherwise, if* $\alpha(\bar{t}_2) < 0$, *then function* (4) *has two roots in the interval* $(0, 1)$, *too; we notate them by* t_1 *and* t_2. *The PDF starts from infinity, decreases in the interval* $(0, t_1)$ *having a local minimum for* $t = t_1$, *increases for* $t \in (t_1, t_2)$ *having a local maximum for* $t = t_2$, *and decreases to zero for* $t \in (t_2, 1)$.

Proof. We have that $h(1) = 0$ due to the exponential decay of PDF (3).The value $h(0)$ and the shape of PDF (3) is derived in Appendix A. □

We introduce and investigate a new class of distributions for which the correction is presented by another Kies distribution (1).

Definition 2.2. Let a, b, λ, and k be positive constants. Let two Kies distributed random variables be defined by their CDFs

$$\begin{aligned} H(t) &:= 1 - e^{-k\left(\frac{t}{1-t}\right)^b}, \\ G(t) &:= 1 - e^{-\lambda\left(\frac{t}{1-t}\right)^a}. \end{aligned} \tag{6}$$

We define a new distribution in the domain $(0, 1)$ by the CDF

$$F(t) := G(H(t)). \tag{7}$$

We shall call it a H-corrected Kies distribution. We name G the original distribution and H the correcting distribution.

Remark 1. Note that this superposition is possible since the Kies CDF is an increasing from zero to one function in the interval $(0, 1)$.

Proposition 2.3. *The CDF* (7) *can be written as*

$$F(t) = 1 - e^{-\lambda\left(e^{k\left(\frac{t}{1-t}\right)^b} - 1\right)^a}. \tag{8}$$

Proof. We have

$$F(t) = 1 - e^{-\lambda\left(\frac{H(t)}{1-H(t)}\right)^a} = 1 - e^{-\lambda\left(\frac{1}{\overline{H}(t)} - 1\right)^a} = 1 - e^{-\lambda\left(e^{k\left(\frac{t}{1-t}\right)^b} - 1\right)^a}, \tag{9}$$

since

$$\overline{H}(t) = e^{-k\left(\frac{t}{1-t}\right)^b}. \tag{10}$$

□

As a corollary of Definition 2.2 we can establish the quantile function.

Corollary 2.4. *The quantile function of a H-corrected Kies distributed random variable can be derived trough the formula*

$$Q_F(t) = Q_H(Q_G(t)), \tag{11}$$

where $Q_H(t)$ and $Q_G(t)$ are the quantile functions of the original Kies distributions (equation (2)*) $H(t)$ and $G(t)$, respectively.*

Differentiating equation (8) we obtain for the PDF

$$f(t) = ab\lambda k\, e^{-\lambda\left(e^{k\left(\frac{t}{1-t}\right)^b} - 1\right)^a} \left(e^{k\left(\frac{t}{1-t}\right)^b} - 1\right)^{a-1} e^{k\left(\frac{t}{1-t}\right)^b} \frac{t^{b-1}}{(1-t)^{b+1}}. \tag{12}$$

More informative is another form of PDF (12) presented in the following proposition.

Proposition 2.5. *The PDF of the H-corrected Kies distribution* (12) *can be written alternatively as*

$$f(t) = g(H(t))\, h(t). \tag{13}$$

Proof. The prof is an immediate consequence from superposition (7). □

Remark 2. Formula (13) means that the PDF of the H-corrected Kies distribution is the initial PDF weighted by the corresponding correction's PDF.

First we have to derive the PDF value of the H-corrected Kies distribution at the left domain endpoint $t = 0$ (obviously, the value at the right one, $t = 1$, is zero). Analogously to the original Kies distributions, one can expect that $0 < f(0) < \infty$ when $a = b = 1$. This is true, but the values $a = b = 1$ are far from exhausting the cases in which the left endpoint of the PDF is finite and nonzero. The proposition below characterizes the PDF's behavior near the zero.

Proposition 2.6. *The left value $f(0)$ of the H-corrected Kies PDF* (13) *is:*

1. $f(0) = 0$ *when* $ab > 1$*;*

2. $f(0) = \lambda k^a$ *when* $ab = 1$*;*

3. $f(0) = \infty$ *when* $ab < 1$.

Proof. We shall use form (12) of the PDF. We can see that the PDF near the zero depends only on the term

$$L := ab\lambda k \left(e^{kt^b} - 1\right)^{a-1} t^{b-1}. \tag{14}$$

Expanding the exponent in the Taylor series, we transform equation (14) to

$$\begin{aligned} L &= ab\lambda k \left(\left(\sum_{n=1}^{\infty} \frac{k^n t^{nb}}{n!} \right) t^{\frac{b-1}{a-1}} \right)^{a-1} \\ &= ab\lambda k \left(\sum_{n=1}^{\infty} \frac{k^n t^{nb+\frac{b-1}{a-1}}}{n!} \right)^{a-1}. \end{aligned} \tag{15}$$

Suppose first that $a < 1$. If b is such that

$$b + \frac{b-1}{a-1} < 0, \tag{16}$$

then at least one term of the sum above, namely the first one, tends to infinity for $t \to 0$. Therefore $L \to 0$, since $a < 1$. Note that inequality (16) is equivalent to $ab > 1$. If the inequality (16) is opposite in sign, then all terms of the sum tend to zero, and therefore $L \to \infty$ (note again $a < 1$). If we have equality in (16), or equivalently $ab = 1$, then the first term tends to k and the rest tend to zero. Hence $L \to ab\lambda k k^{a-1} = \lambda k^a$.

Assume now that $a > 1$. If inequality (16) holds, equivalently to $ab < 1$, then $L \to \infty$, since the fist term of the sum tends to infinity and $a > 1$. If $ab = 1$, then only the first term is nonzero – its limit is k – and hence $L \to \lambda k^a$. If $ab > 1$ (oppositely to inequality (16)), we conclude that $L \to \infty$, since the sum tends to infinity and the power is positive.

Finally, if $a = 1$, then the desired result holds because formula (14) turns to $L = b\lambda k t^{b-1}$. □

The shape of the PDF of the H-corrected Kies distribution is a consequence of Propositions 2.1 and 2.6. Obviously, it has to be more various than the PDF of the original Kies distribution. Various examples are presented in Figure 1. In each of all six subfigures we vary the coefficients λ and a for the original distribution G as $\lambda \in \{0.5, 1, 1.5\}$ and $a \in \{0.5, 2\}$. Namely, the original Kies distribution G is colored by blue. The rest of the plotted PDFs are produced by the following parameters for the correcting distribution H: $k \in \{1, 2\}$ and $b \in \{0.5, 1, 2\}$.

The following proposition for the expectations of the corrected Kies random variables holds.

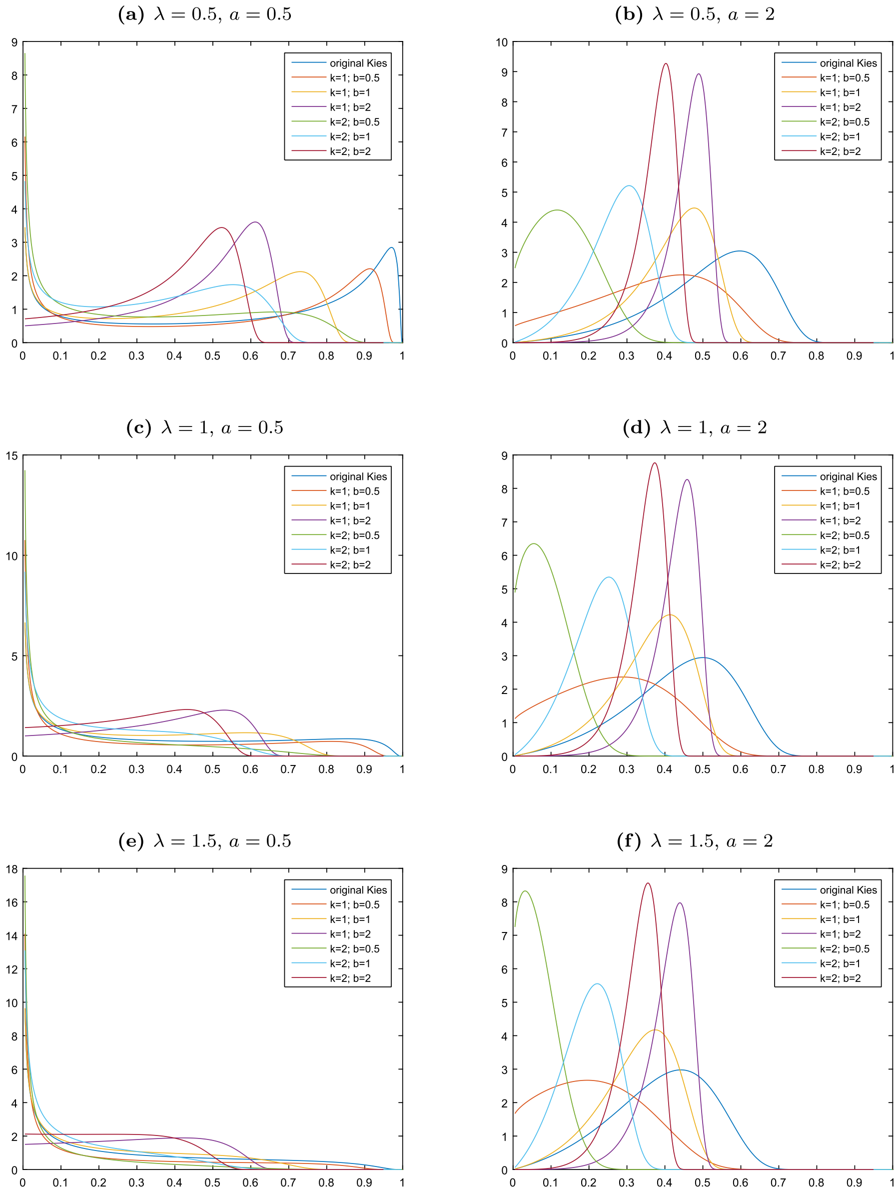


Fig. 1. PDFs of the corrected Kies distributions

Proposition 2.7. *Let ξ_F be an H-corrected Kies distributed random variable with original distribution G, ξ_G be an original Kies distributed random variable, and $\beta(\cdot)$ be a real valued function. The expectation of the random variable $\beta(\xi_F)$ is equal to the expectation of the random variable $\beta(Q_H(\xi_G))$. Written formalized that is*

$$E\left[\beta\left(\xi_F\right)\right] = E\left[\beta\left(Q_H\left(\xi_G\right)\right)\right]. \tag{17}$$

Proof. Using the form of PDF (13) and changing the variables as $x = H(t)$ (equivalently, $t = Q_H(x)$) we derive

$$\begin{aligned} E[\beta(\xi_F)] &= \int_0^1 \beta(t)\, g(H(t))\, dH(t) \\ &= \int_0^1 \beta(Q_H(x))\, g(x)\, dx \\ &= E[\beta(Q_H(\xi_G))]. \end{aligned} \tag{18}$$

$\square$

The following corollaries hold.

Corollary 2.8. *The random variables ξ_F and $Q_H(\xi_G)$ are identically distributed under the assumptions of Proposition 2.7.*

Corollary 2.9. *The H-corrected Kies distributed random variable ξ_F has finite moments and they can be presented as*

$$\mu_n := E\left[\xi_F^n\right] = E\left[(Q_H(\xi_G))^n\right] \tag{19}$$

for $n = 1, 2, \ldots$.

Proof. We can obtain the moments integrating by parts as

$$E\left[\xi_F^n\right] = \int_0^1 t^n dF(t) = 1 - n \int_0^1 t^{n-1} F(t)\, dt \tag{20}$$

and hence they are finite. Formula (19) is an immediate consequence of equation (17). $\square$

Let us consider now the mean residual life function (MRLF, hereafter) of an H-corrected Kies distribution. Usually it is defined as the conditional expectation

$$m_F(t) := E[\xi_F - t \,|\xi_F > t]. \tag{21}$$

We shall use an alternative presentation stated in [7]:

$$m_F(t) := \frac{1}{\overline{F}(t)} \int_t^1 \overline{F}(s)\, ds. \tag{22}$$

The following proposition for the MRLF stands.

Proposition 2.10. *The MRLF of an H-corrected Kies distributed random variable ξ_F can be written as*

$$m_F(t) = \frac{E\left[Q_H(\xi_G)\, I_{\xi_G > H(t)}\right]}{\overline{F}(t)} - t. \tag{23}$$

Proof. Let us consider first the integral in formula (22). Changing the variables as $x = H(s)$ (equivalently to $s = Q_H(x)$) and integrating by parts we derive

$$\begin{aligned}\int_t^1 \overline{F}(s)\,ds &= \int_t^1 \overline{G}(H(s))\,ds = \int_{H(t)}^1 \overline{G}(x)\,dQ_H(x) \\ &= \overline{G}(x)\,Q_H(x)\Big|_{H(t)}^1 + \int_{H(t)}^1 g(x)\,Q_H(x)\,dx \\ &= -\overline{F}(t)\,t + E\left[Q_H(\xi_G)\,I_{\xi_G > H(t)}\right].\end{aligned} \tag{24}$$

We obtain the desired result combining equations (22) and (24). □

Remark 3. A simple validation of Proposition 2.10 can be seen for $t = 0$. Then formula (21) leads to $m(0) = E[\xi_F]$ and thus formulas (17) and (23) coincide when $\beta(\cdot)$ is the identity function.

3 Tail behavior

Let us consider three measures arising from the risk management – *VaR*, *AVaR*, and expectile based VaR; we shall use the notation *EX* for the last one. By its original definition, the *VaR* at level α of a random variable is just the opposite of the quantile function $VaR(\alpha) = -Q(\alpha)$. Since the domain of the Kies family is the interval (0, 1) we shall think $VaR(\alpha) := Q(\alpha)$. As its name shows, *AVaR* is an average *VaR* in some sense – it is defined as

$$AVaR(\alpha) := \frac{1}{\alpha}\int_0^\alpha VaR(u)du. \tag{25}$$

Also, we consider the right tail behavior by defining the following term

$$\overline{AVaR}(\alpha) := \frac{1}{1-\alpha}\int_\alpha^1 VaR(u)du. \tag{26}$$

The expectile function is related to the quantiles in the following way. The α-quantile of the random variable ξ can be viewed as the lower solution of the optimal problem

$$Q(\alpha) = \operatorname*{arg\,min}_{x\in\mathbb{R}}\left\{E\left[\alpha(\xi - x)^+ + (1-\alpha)(\xi - x)^-\right]\right\}, \tag{27}$$

where z^+ and z^- are notations for $\max(z, 0)$ and $\max(-z, 0)$, respectively. For more details, see, for example, [10]. Analogously, the expectile is defined in [19] as the solution of the following quadratic problem

$$EX(\alpha) := \operatorname*{arg\,min}_{x\in\mathbb{R}}\left\{E\left[\alpha\left((\xi - x)^+\right)^2 + (1-\alpha)\left((\xi - x)^-\right)^2\right]\right\}. \tag{28}$$

Note that the expectiles are well defined when the random variable has a finite second moment. For the corrected Kies distributions this is true due to Corollary 2.9. It can be easily proven that expectile (28) is the solution of the following equation w.r.t. the variable x

$$\alpha E\left[(\xi - x)^{+}\right] = (1-\alpha)\, E\left[(\xi - x)^{-}\right]. \tag{29}$$

We derive *AVaR*s and the expectile based *VaR* in the following two propositions.

Proposition 3.1. *We have the following double presentations for* $AVaR(\alpha)$ *and* $\overline{AVaR}(\alpha)$*:*

$$\begin{aligned} AVaR(\alpha) &= \frac{\mu_1}{\alpha} - \frac{1-\alpha}{\alpha}\left[Q_F(\alpha) + m_F(Q_F(\alpha))\right] \\ &= \frac{E\left[Q_H(\xi_G)\, I_{\xi_G < Q_G(\alpha)}\right]}{\alpha} \\ \overline{AVaR}(\alpha) &= Q_F(\alpha) + m_F(Q_F(\alpha)) \\ &= \frac{E\left[Q_H(\xi_G)\, I_{\xi_G > H Q_G(\alpha)}\right]}{1-\alpha}, \end{aligned} \tag{30}$$

where μ_1 *is the first moment given in Corollary 2.9 and* $m_F(\cdot)$ *is the MRLF.*

Proof. We shall use the following relation between the truncated expecations and the MRLF, the proof of which can be found in [7],

$$E\left[(\xi_F - y)^{+}\right] = m_F(y)\, \overline{F}(y). \tag{31}$$

Having in mind equations $x^{-} = x^{+} - x$ and (31), and changing the variables as $s = Q_F(t) \Leftrightarrow t = F(s)$, we derive for the first statement of equation (30)

$$\begin{aligned} AVaR(\alpha) &= \frac{1}{\alpha}\int_0^{\alpha} Q_F(t)\, dt = \frac{1}{\alpha}\int_0^{Q_F(\alpha)} s f(s)\, ds = \frac{E\left[\xi_F I_{\xi < Q_F(\alpha)}\right]}{\alpha} \\ &= \frac{Q_F(\alpha)\, P(\xi_F < Q_F(\alpha))}{\alpha} - \frac{E\left[(\xi_F - Q_F(\alpha))^{-}\right]}{\alpha} \\ &= Q_F(\alpha) - \frac{E\left[(\xi_F - Q_F(\alpha))^{+}\right]}{\alpha} + \frac{E\left[\xi_F - Q_F(\alpha)\right]}{\alpha} \\ &= \frac{\mu_1}{\alpha} - \frac{1-\alpha}{\alpha}\left[Q_F(\alpha) + m_F(Q_F(\alpha))\right]. \end{aligned} \tag{32}$$

To derive the second form of the *AVaR*, we use equations (11), (19), and (23) (for the quantile function, the moment and the MRLF, respectively) and obtain

$$\begin{aligned} AVaR(\alpha) &= \frac{\mu_1 - (1-\alpha)\left[Q_F(\alpha) + m_F(Q_F(\alpha))\right]}{\alpha} \\ &= \frac{E\left[Q_H(\xi_G)\right] - (1-\alpha)\left[Q_F(\alpha) + \frac{E\left[Q_H(\xi_G) I_{\xi_G > H(Q_F(\alpha))}\right]}{\overline{F}(Q_F(\alpha))} - Q_F(\alpha)\right]}{\alpha} \\ &= \frac{E\left[Q_H(\xi_G)\, I_{\xi_G \le Q_G(\alpha)}\right]}{\alpha}. \end{aligned} \tag{33}$$

Let us turn to the right tail term $\overline{AVaR}$. Analogously as above, we obtain

$$\begin{aligned}
\overline{AVaR}(\alpha) &= \frac{1}{1-\alpha}\int_{\alpha}^{1} Q_F(t)\,dt = \frac{1}{1-\alpha}\int_{Q_F(\alpha)}^{1} sf(s)\,ds = \frac{E\left[\xi_F I_{\xi > Q_F(\alpha)}\right]}{1-\alpha} \\
&= \frac{Q_F(\alpha)\,P(\xi_F > Q_F(\alpha))}{1-\alpha} + \frac{E\left[(\xi_F - Q_F(\alpha))^+\right]}{1-\alpha} \\
&= Q_F(\alpha) + \frac{m_F(Q_F(\alpha))\,\overline{F}(Q_F(\alpha)}{1-\alpha} \\
&= Q_F(\alpha) + m_F(Q_F(\alpha))\,.
\end{aligned} \tag{34}$$

Writing equation (23) for $t = Q_F(\alpha)$, we see that

$$Q_F(\alpha) + m_F(Q_F(\alpha)) = \frac{E\left[Q_H(\xi_G)\,I_{\xi_G > Q_G(\alpha)}\right]}{1-\alpha}, \tag{35}$$

which leads to the second form of $\overline{AVaR}$ □

Remark 4. Note that the second forms of *AVaR* and $\overline{AVaR}$ can be obtained directly (without using the first form) changing the variables as $t = Q_G(u) \Leftrightarrow u = G(t)$ in the integral

$$\int Q_F(u)\,du = \int Q_H(Q_G(u))\,du. \tag{36}$$

Next we discuss the expectile based VaR. It can be obtained through both of equations presented in the following proposition.

Proposition 3.2. *The α-expectile based VaR, $EX(\alpha)$, is the solution of the following equivalent equations (w.r.t. the variable t)*

$$\begin{aligned}
&(1-2\alpha)\,m_F(t)\,\overline{F}(t) + (1-\alpha)(t-\mu_1) = 0 \\
&t\left(\overline{G}(H(t)) - \alpha\right) - E\left[Q_H(\xi_G)\left(1-\alpha-(1-2\alpha)\,I_{\xi_G > H(t)}\right)\right] = 0,
\end{aligned} \tag{37}$$

where μ_1 is the first moment given in Corollary 2.9 and $m_F(\cdot)$ is the MRLF.

Proof. Using the formula $x^- = x^+ - x$ and equation (29) which determines the expectile we derive

$$\alpha E\left[(\xi_F - t)^+\right] = (1-\alpha)\,E\left[(\xi_F - t)^+ - (\xi_F - t)\right]. \tag{38}$$

Replacing the truncated expectation from formula (31) we obtain the first equation in (37). It remains to replace the expectation and the MRLF from equations (17) and (23) to derive the second part of (37). □

4 Hausdorff distance and saturation

Let us consider the max-norm in $\mathbb{R}^2$, i.e. if A and B are the points $A = (t_A, x_A)$ and $B = (t_B, x_B)$, then $\|A - B\| := \max\{|t_A - t_B|, |x_A - x_B|\}$. We define the Hausdorff distance, also known as a H-distance, in a sense of [23].

Definition 4.1. The Hausdorff distance $d(g, h)$ between two curves g and h in $\mathbb{R}^2$ is

$$d(g, h) := \max\left\{\sup_{A\in g}\inf_{B\in h}\|A - B\|, \sup_{B\in h}\inf_{A\in g}\|A - B\|\right\}. \tag{39}$$

Remark 5. Roughly said, the Hausdorff distance is the highest optimal path between the curves.

We can define now the *saturation* of a distribution.

Definition 4.2. Let $F(\cdot)$ be the CDF of a distribution with a left-finite domain $[a, b)$, $-\infty < a < b \le \infty$. Its saturation is the Hausdorff distance between the completed graph of $F(\cdot)$ and the curve consisting of two lines – one vertical between the points $(a, 0)$ and $(a, 1)$ and another horizontal between $(a, 1)$ and $(b, F(b))$.

Having in mind that the domain of the Kies distribution is the interval $(0, 1)$, we can prove the following corollary for its saturation.

Corollary 4.3. *The saturation of the Kies CDF, $F(\cdot)$, is the unique solution of the equation*

$$F(d) = 1 - d. \tag{40}$$

Proof. The proof is an immediate corollary of Definitions 4.1 and 4.2. Note that equation (40) has a unique root because the function $F(\cdot)$ is increasing and continuous. $\square$

We shall prove now the following formula of a semiclosed form for the saturation of CDF (8).

Theorem 4.4. *Let y be a positive parameter and the function $\gamma(y)$ be defined as*

$$\gamma(y) := y\left[e^{\lambda(e^y-1)^a} - 1\right]^b. \tag{41}$$

Suppose that $k = \gamma(y)$ for some value of y. Then the H-corrected Kies distribution's saturation is

$$d(y) = e^{-\lambda(e^y-1)^a}. \tag{42}$$

Note that the function $\gamma(y)$ is strictly increasing in the interval $(0, \infty)$ and hence it is invertible. Therefore the saturation can be expressed as a function of λ, k, a, and b as

$$d(\lambda, k, a, b) = e^{-\lambda\left(e^{\gamma^{-1}(\lambda,k,a,b)}-1\right)^a}. \tag{43}$$

Proof. Applying Corollary 4.3 to CDF (8) we see that the saturation d satisfies the equation

$$\mu(d) := \lambda\left(e^{k\left(\frac{d}{1-d}\right)^b} - 1\right)^a + \ln(d) = 0 \tag{44}$$

in the interval $(0, 1)$. Note that the solution exists and is unique, because the function $\mu(d)$ is continuous, increasing, $\mu(0) = -\infty$, and $\mu(1) = +\infty$. Let us change the variables as

$$z = \frac{1}{k}e^{k\left(\frac{d}{1-d}\right)^b} \Leftrightarrow d = \frac{(\ln(kz))^{\frac{1}{b}}}{(\ln(kz))^{\frac{1}{b}} + k^{\frac{1}{b}}}. \tag{45}$$

Thus the function $\mu(d)$ defined as formula (44) turns to

$$\mu(d) = \lambda(zk-1)^a + \ln(d). \tag{46}$$

Another change we need is $y = \ln(kz)$ or, equivalently, $z = \frac{e^y}{k}$. Thus

$$d = \frac{y^{\frac{1}{b}}}{y^{\frac{1}{b}} + k^{\frac{1}{b}}} \tag{47}$$

and therefore the function $\mu(d)$ can be rewritten w.r.t. the variable y as

$$\mu(y) = \ln\left(e^{\lambda(e^y-1)^a} \frac{y^{\frac{1}{b}}}{y^{\frac{1}{b}} + k^{\frac{1}{b}}}\right). \tag{48}$$

Therefore the equation $\mu(y) = 0$ turns to

$$e^{\lambda(e^y-1)^a} \frac{y^{\frac{1}{b}}}{y^{\frac{1}{b}} + k^{\frac{1}{b}}} = 1 \tag{49}$$

or, equivalently,

$$k = y\left[e^{\lambda(e^y-1)^a} - 1\right]^b. \tag{50}$$

Substituting k from equation (50) into formula (47), we obtain

$$d = \frac{y^{\frac{1}{b}}}{y^{\frac{1}{b}} + y^{\frac{1}{b}}\left[e^{\lambda(e^y-1)^a} - 1\right]} = e^{-\lambda(e^y-1)^a}. \tag{51}$$

We finish the proof combining equations (50) and (51). □

The behavior of the corrected Kies CDFs can be seen in Figures 2a–2d together with the saturation $\overline{d}$; it is presented by the red points. The red lines form squares which in fact confirms Corollary 4.3. The used quadruplet for the parameters are $(\lambda, k, a, b) \in \{(2, 2, 5, 1), (2, 2, 1, 0.5), (2, 1, 2, 1), (0.5, 1, 0.5, 0.5)\}$.

Next we discuss a useful in practice interval approximation of the Hausdorff saturation $\overline{d}$. Let us consider first the case $\lambda = a = b = 1$. The saturation $\overline{d}$ has to satisfy equation (44) which now can be written as

$$\mu(d) = e^{k\frac{d}{1-d}} - 1 + \ln(d) = 0. \tag{52}$$

The function $\mu(d)$ can be approximated very well for small ds by the function

$$\mu_1(d) = e^{kd} - 1 + \ln(d). \tag{53}$$

Taking the exponent in the Taylor series we see that the function

$$\mu_1(d) = kd + \ln d + \sum_{n=2}^{\infty} \frac{(kd)^n}{n!}. \tag{54}$$

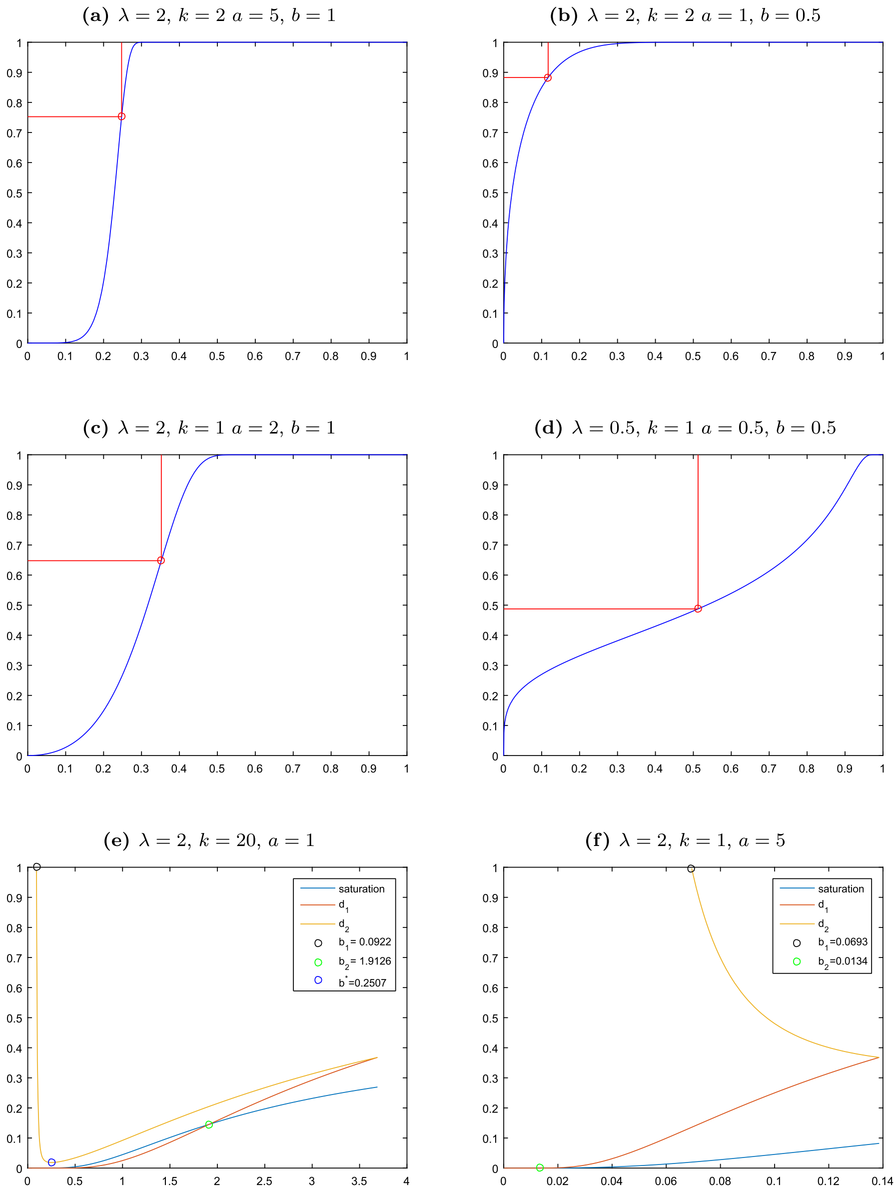


Fig. 2. CDFs of the corrected Kies distributions with the Hausdorff saturation

can be approximated as $\mathcal{O}\left(d^2\right)$ for small enough values of d by the function

$$\mu_2\left(d\right) = kd + \ln d. \tag{55}$$

Let d_1 and d_2 be defined as $d_1 := \frac{1}{k}$ and $d_2 := \frac{\ln k}{k}$. We shall check when $\mu_2\left(d_1\right) < 0 < \mu_2\left(d_2\right)$. Obviously the first inequality holds when $k > e$. Assuming that this restriction holds, we see that $\mu_2(d_2) = \ln \ln k > 0$ and hence the second inequality holds, too. Thus we conclude that the function $\mu_2(d)$ has a unique root in the interval $(0, 1)$ and it belongs to the subinterval (d_1, d_2) when $k > e$, since this function is strictly increasing.

In the next proposition we discuss the general case assuming that $\lambda k^a > 1$.

Proposition 4.5. *Suppose that* $\lambda k^a > 1$. *Let the parameter* b *be such that* $b < \overline{b}$, *where* $\overline{b}$ *is*[2]

$$\overline{b} := \frac{\ln(\lambda k^a)}{a}. \tag{56}$$

Then the function $\mu_2(d)$ *defined as*

$$\mu_2(d) := kd^b - \left(-\frac{\ln d}{\lambda}\right)^{\frac{1}{a}} \tag{57}$$

has a unique root in the interval $(0, 1)$. *Moreover, the root belongs to the subinterval* (d_1, d_2) *where*

$$\begin{aligned} d_1 &:= \left(\frac{1}{\lambda k^a}\right)^{\frac{1}{ab}}, \\ d_2 &:= \left(\frac{\ln(\lambda k^a)}{ab\lambda k^a}\right)^{\frac{1}{ab}}. \end{aligned} \tag{58}$$

Note that $d_1 < d_2$ *due to the condition* $b < \overline{b}$.

Proof. Let us consider first function (57) in the particular case $\lambda = a = 1$. Thus we have

$$\begin{aligned} k &> e^b, \\ \mu_2(d) &= kd^b + \ln d, \\ d_1 &= \left(\frac{1}{k}\right)^{\frac{1}{b}}, \\ d_2 &= \left(\frac{\ln k}{bk}\right)^{\frac{1}{b}}. \end{aligned} \tag{59}$$

Obviously, the function $\mu_2(d)$ is increasing and $\mu_2(d_1) < 0$ due to $k > e^b$. We have for $\mu_2(d_2)$:

$$\mu_2(d_2) = \frac{\ln k}{b} + \frac{1}{b}\left[\ln\left(\frac{\ln k}{b}\right) - \ln k\right] = \frac{1}{b}\ln\left(\frac{\ln k}{b}\right) > 0. \tag{60}$$

The last inequality is true again due to $k > e^b$.

[2] Note that this inequality is equivalent to $k > e$ when $\lambda = a = b = 1$.

Let us remove the restriction $\lambda = a = 1$. Let the function $\overline{\mu}_2(d; k, b) := kd^b + \ln d$ be defined as before – see the second line of (59). Note that we mark the dependence on the variables k and b. We can easily check that the equation $\mu_2(d) = 0$ is equivalent to $\overline{\mu}_2(d; K, B) = 0$ for $K = \lambda k^a$ and $B = ab$ and thus we can use the derived above result. This way the values of K and B lead to formulas (56) and (58). $\square$

Let us return to the saturation of the corrected Kies distribution. We have shown above that it is the solution of equation (44) which is equivalent to

$$e^{k\left(\frac{d}{1-d}\right)^b} - 1 - \left(-\frac{\ln d}{\lambda}\right)^{\frac{1}{a}} = 0. \tag{61}$$

Analogously to the case $a = b = \lambda = 1$, formula (54), we can see that after the Taylor expansion of the exponent, the left hand-side of equation (61) can be approximated by the function $\mu_2(d)$ near zero as $\mathcal{O}\left(d^{2b}\right)$. Hence, its root can be used as an approximation of the corrected Kies distribution's saturation when it is small enough. On the other hand, the function $\mu_2(d)$ is a lower approximation of $\mu(d)$ and therefore $\mu_2(d) < \mu(d)$. Thus the saturation is below the root of the function $\mu_2(d)$ and hence $\overline{d} < d_2$. The question stands, when $d_2 < 1$. Let us define b_1 as

$$b_1 := \frac{\ln(\lambda k^a)}{a\lambda k^a} = \frac{\overline{b}}{\lambda k^a}. \tag{62}$$

Note that $b_1 < \overline{b}$. We shall show that if $b < \overline{b}$, then $d_2 < 1$ only when $b > b_1$. Using again the notations $K = \lambda k^a$ and $B = ab$ and having in mind formula (58) we see that $d_2 < 1$ when $1 > \frac{\ln K}{BK}$ which is equivalent to $b > b_1$. Note that $K > 1$.

On the contrary, d_1 is not always below the saturation $\overline{d}$. It turns out that there exists a value, say b_2, dependent on the other parameters, such that $d_1 < \overline{d}$ for $b < b_2$, and vice versa. To see this, we consider function (44). Obviously, it is increasing in the distribution domain $(0, 1)$. Also, d_1 increases w.r.t. the parameter b because $\lambda k^a > 1$. Therefore $\overline{\mu}(b) := \mu(d_1(b))$ is an increasing function, too; note that $d_1(b) < 1$. Having in mind $\overline{\mu}(0) = -\infty$, $\overline{\mu}(+\infty) = +\infty$, and $\mu\left(\overline{d}(b)\right) = 0$, we conclude that indeed $d_1(b) < \overline{d}(b)$ for $b < b_2$, where b_2 is the unique solution in the interval $(0, \infty)$ of the equation

$$\overline{\mu}(b) = \lambda\left(e^{\lambda^{-\frac{1}{a}}\left(1-\lambda^{-\frac{1}{ab}}k^{-\frac{1}{b}}\right)^{-b}} - 1\right)^a - \frac{\ln\lambda}{ab} - \frac{\ln k}{b}. \tag{63}$$

We shall show that $b_2 < \overline{b}$, too. Let us mark the dependence on b in the terms d_1 and d_2. We can easily check that $d_1\left(\overline{b}\right) = d_2\left(\overline{b}\right)$ and hence $\overline{d} < d_1\left(\overline{b}\right) = d_2\left(\overline{b}\right) < 1$ (because $\overline{d} < d_2\left(\overline{b}\right)$ and $d_1\left(\overline{b}\right) < 1$). Therefore $\overline{\mu}\left(\overline{b}\right) = \mu\left(d_1\left(\overline{b}\right)\right) > \mu\left(\overline{d}\right) = 0$, since $\mu(b)$ is an increasing function. Thus we see that $b_2 < \overline{b}$.

We can formulate these results in the following proposition.

Proposition 4.6. *Suppose that $b < \overline{b}$, where $\overline{b}$ is given in formula (56). Then if $b < b_2$, where b_2 is the solution of equation (63), then $d_1 < \overline{d} < d_2$. If in addition $b > b_1$ for b_1 given in equation (62), then $d_2 < 1$. Having in mind that $b_1 \vee b_2 < \overline{b}$, we can formulate the following statements:*

- *If* $b_1 < b_2$, *then*
  1. $d_1 < \overline{d} < d_2$ *and* $d_2 > 1$ *when* $b < b_1$*;*
  2. $d_1 < \overline{d} < d_2 < 1$ *when* $b \in (b_1, b_2)$*;*
  3. $\overline{d} < d_1 < d_2 < 1$ *when* $b \in \left(b_2, \overline{b}\right)$*;*
- *If* $b_1 > b_2$, *then*
  1. $d_1 < \overline{d} < d_2$ *and* $d_2 > 1$ *when* $b < b_2$*;*
  2. $\overline{d} < d_1 < d_2$ *and* $d_2 > 1$ *when* $b \in (b_2, b_1)$*;*
  3. $\overline{d} < d_1 < d_2 < 1$ *when* $b \in \left(b_1, \overline{b}\right)$*;*

As we can see from definition (58), d_1 can be viewed as an increasing function w.r.t. the parameter b. Let us consider the second value d_2. It can be written as $d_2 = \alpha(\beta) = (\beta c)^{\beta}$, where $\beta = \frac{1}{ab}$ and $c = \frac{\ln k}{k}$. The function $\alpha(\beta)$ decreases in the interval $\beta \in \left(0, \frac{1}{ec}\right)$ and increases for $\beta > \frac{1}{ec}$ because its derivative can be written as $\alpha'(\beta) = \alpha(\beta)(\ln \beta c + 1)$. Thus, we conclude that d_2, considered as a function of the parameter b, $d_2(b)$, decreases for $b < b^*$ and increases otherwise, where

$$b^* := e\frac{\ln(\lambda k^a)}{a\lambda k^a} = eb_1. \tag{64}$$

Some calculus shows that $b^* < \overline{b}$ when $\lambda > \frac{e}{k^a}$, and $b^* > \overline{b}$ otherwise. Note that $b_1 < b^*$.

The interval approximations of the saturation are presented in Figures 2e and 2f. The values of $\overline{d}$, d_1, and d_2 considered as functions of the parameter b are colored in blue, red, and orange, respectively. The parameters for the first figure are $\lambda = 2$, $a = 1$, and $k = 20$. The related important values for the parameter b are $b_1 = 0.0922$, $b_2 = 1.9126$, $b^* = 0.2507$, and $\overline{b} = 3.6889$. We mark b_1, b_2, and b_3 by black, green, and blue points, respectively. In this case $b_1 < b^* < b_2$ and thus the first case of Proposition 4.6 holds. Also $b^* < \overline{b}$, since $\lambda > \frac{e}{k^a}$. We can see in Figure 2e that the interval (d_1, d_2) is a good evaluation for the saturation $\overline{d}$ when $b \in (b^*, b_2)$. Otherwise, if $b > b_2$, then the limitation is $\overline{d} < d_1$.

We choose parameters $\lambda = 2$, $k = 1$, and $a = 5$ for Figure 2f. Now the important values are $b_1 = 0.0693$, $b_2 = 0.0134$, $b^* = 0.1884$, and $\overline{b} = 0.1386$. Note that $b^* > \overline{b}$ because $\lambda < \frac{e}{k^a}$. Also, $b_1 > b_2$ and thus the second case of Proposition 4.6 is actual. We have that $\overline{d} > d_1$ for $b < b_2$, but both values are very close. Otherwise $\overline{d} < d_1$ when $b \in \left(b_2, \overline{b}\right)$.

Let us mention that the relation $\overline{d} < d_2$ still holds if we remove the restriction $b < \overline{b}$. In this case we have $\overline{d} < d_2 < d_1 < 1$.

5 Calibration

The defined corrected Kies distributions depend on four parameters λ, k, a, and b. The maximum likelihood estimator can be obtained in a closed form – for original Kies distributions, see [13]. Unfortunately, it turns out that this method does not work efficiently either in terms of speed or precision. For this we construct a least square errors (LSqE) type algorithm. It falls in the large class of generalized methods of moments (GMM), since it is based on curve fitting to a histogram. Hence we can use the existing results for the GMM; we refer to [8]. These methods produce consistent and asymptotically normal estimators for the Kies distributions since the whole Kies family exhibits finite moments.

Let us have n observations $t_1, t_2, \ldots, t_n$. First we calculate the empirical PDF at m (= 50) bins as

$$l_i^{emp} := \frac{mN_i}{n\left(\left(\max\left\{t_i\right\} - \min\left\{t_i\right\}\right)\right.}. \tag{65}$$

Then we derive the PDF values of the corrected Kies distribution with parameters (λ, k, a, b) in the centers of the bins, say $l_i^{Kies}(\lambda, k, a, b)$, via formula (12). The usual LSqE criterion for minimization is

$$L(\lambda, k, a, b) = \sum_{i=1}^{m} \left(l_i^{emp} - l_i^{Kies}(\lambda, k, a, b)\right)^2. \tag{66}$$

We introduce a little logarithmic modification to minimize the impact of the extremely large values of the PDF. We make this because the PDF is infinitely large at the zero for some values of the parameters. Thus we define the cost function as

$$L(\lambda, k, a, b) := \sum_{i=1}^{m} \left|\ln\left(l_i^{emp} + \epsilon\right) - \ln\left(l_i^{Kies}(\lambda, k, a, b) + \epsilon\right)\right|. \tag{67}$$

We have to minimize the corresponding criterion – (66) or (67) – over all possible parameters $\{\lambda, k, a, b\}$. The additional constant ϵ is introduced, because some empirical values may be equal to zero. We set this constant to be $\epsilon = 10^{-5}$. Also, we can use criterion (66) if the empirical PDF seems to be finite at its left endpoint. We provide some experiments to validate this algorithm. We generate n corrected Kies distributed random numbers as $t_i = Q_F(r_i)$, where $Q_F(\cdot)$ is the quantile function given in equation (11), and r_i are $(0, 1)$-uniformly distributed random numbers. Our choice of n is among $n = 1\,000$, $n = 10\,000$, $n = 100\,000$, and $n = 1\,000\,000$. We fix the coefficients λ and k to one, $\lambda = k = 1$, and vary a and b among 0.5, 1, and 2. We report in Table 1 the results which are returned by our calibration algorithm. The fits can be seen in Figure 3. It turns out that this simple LSqE algorithm is quite fast and accurate.

6 An application

We investigate now the behavior of the S&P500 index. It is one of the most used indicators in the financial markets and provides an important information for the world economy. We use daily observations for the period between January 2, 1980 and July 01, 2022 – totally 10717 ones. We derive the so-called log-returns, denoted by r_i, via the equation

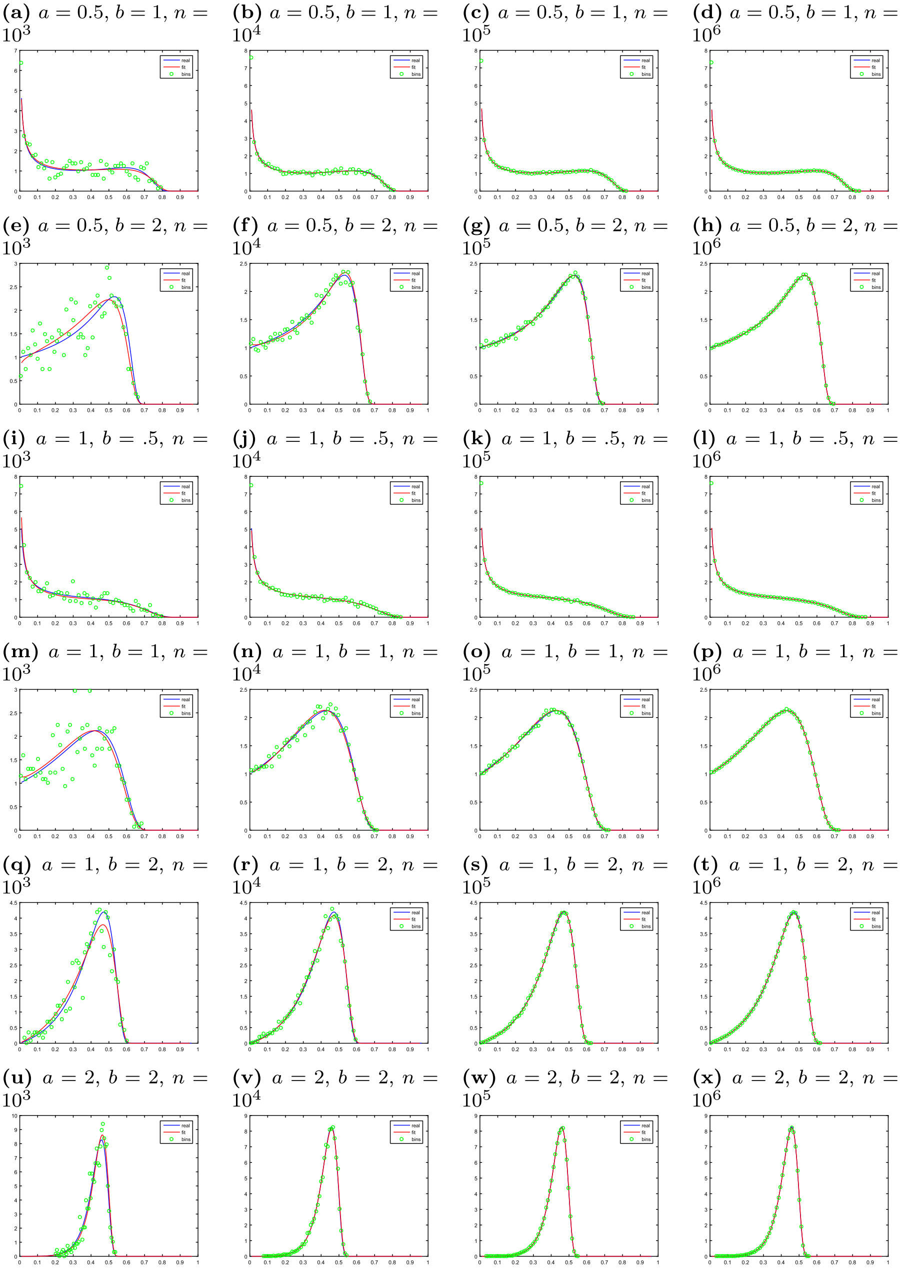
$$r_i := \ln\left(\frac{S_{i+1}}{S_i}\right) \quad \text{for } i = 1, 2, \ldots, 10716, \tag{68}$$

Table 1. The fitted parameters of the corrected Kies distribution

parameter	real	$n = 1\,000$	$n = 10\,000$	$n = 100\,000$	$n = 1\,000\,000$
λ	1	1.1918	0.8809	0.9801	1.0315
k	1	0.8076	1.1230	1.0310	0.9670
a	0.5	0.5637	0.5541	0.5045	0.4923
b	1	0.9379	0.8609	0.9744	1.0258
λ	1	1.1224	0.9164	0.9273	1.0096
k	1	0.9450	1.0817	1.0653	0.9882
a	0.5	0.5814	0.5031	0.5451	0.4969
b	2	1.8257	1.9312	1.7986	2.0200
λ	1	0.9224	0.7287	0.9904	1.0395
k	1	1.1869	1.1270	1.0063	0.9797
a	1	0.7860	1.1613	1.0126	0.9855
b	0.5	0.5761	0.4087	0.4915	0.5096
λ	1	0.7780	0.6540	0.9013	0.9728
k	1	1.1416	1.1456	1.0248	1.0121
a	1	1.1356	1.2856	1.1234	1.0173
b	1	0.8307	0.7231	0.8761	0.9783
λ	1	1.3803	1.4454	0.8623	1.1066
k	1	0.8388	0.7850	1.0578	0.9413
a	1	0.8738	0.9411	1.0997	0.9822
b	2	2.2277	2.2455	1.7768	2.0618
λ	1	0.5789	1.4045	1.0190	1.1136
k	1	1.1381	0.8774	0.9984	0.9617
a	2	2.0803	2.0887	1.9625	2.0007
b	2	1.8879	1.9627	2.0357	2.0085

where S_i are the observed S&P500 values. The log-returns are presented in Figure 4a. It can be seen that there are periods of calm trading as well as high-volatility periods. This is a well observed phenomenon at all financial markets – the so-called volatility clustering. The highest downward peak happens at October 19, 1987 (1971th observation) – the Black Monday. The S&P500 index loses more than twenty percents – this is the highest one-day loss ever. We are interested in the length of the periods between the shocks. We derive them obtaining the dates at which the index falls by more than two percents – there are 357 such dates – and then we calculate the lengths of the periods between these days. The longest such period contains 950 days – between May 19, 2003 and February 26, 2007. We mark these days with red points in Figure 4. We may view the derived lengths as survival times and we examine their distribution. We divide all observations by 1000 to fit the Kies domain, because the maximal value is 950. We calibrate the parameters of four distributions – corrected Kies and its ancestors, namely, exponential, Weibull, and original Kies. We use the following parametrization:

$$\begin{aligned} f_{\text{exponential}} &:= \frac{1}{\lambda} e^{-\frac{x}{\lambda}}, \\ f_{\text{Weibull}} &:= \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k}. \end{aligned} \tag{69}$$



(a) $a = 0.5$, $b = 1$, $n = 10^3$ (b) $a = 0.5$, $b = 1$, $n = 10^4$ (c) $a = 0.5$, $b = 1$, $n = 10^5$ (d) $a = 0.5$, $b = 1$, $n = 10^6$

(e) $a = 0.5$, $b = 2$, $n = 10^3$ (f) $a = 0.5$, $b = 2$, $n = 10^4$ (g) $a = 0.5$, $b = 2$, $n = 10^5$ (h) $a = 0.5$, $b = 2$, $n = 10^6$

(i) $a = 1$, $b = .5$, $n = 10^3$ (j) $a = 1$, $b = .5$, $n = 10^4$ (k) $a = 1$, $b = .5$, $n = 10^5$ (l) $a = 1$, $b = .5$, $n = 10^6$

(m) $a = 1$, $b = 1$, $n = 10^3$ (n) $a = 1$, $b = 1$, $n = 10^4$ (o) $a = 1$, $b = 1$, $n = 10^5$ (p) $a = 1$, $b = 1$, $n = 10^6$

(q) $a = 1$, $b = 2$, $n = 10^3$ (r) $a = 1$, $b = 2$, $n = 10^4$ (s) $a = 1$, $b = 2$, $n = 10^5$ (t) $a = 1$, $b = 2$, $n = 10^6$

(u) $a = 2$, $b = 2$, $n = 10^3$ (v) $a = 2$, $b = 2$, $n = 10^4$ (w) $a = 2$, $b = 2$, $n = 10^5$ (x) $a = 2$, $b = 2$, $n = 10^6$

Fig. 3. PDFs of the corrected Kies distributions

The derived parameters are reported in the first part of Table 2. Immediately after them we provide the results which are returned by the LSqE algorithm described in Section 5. The constant ϵ in cost function (67) is chosen to be $\epsilon = 0.01$. It turns out

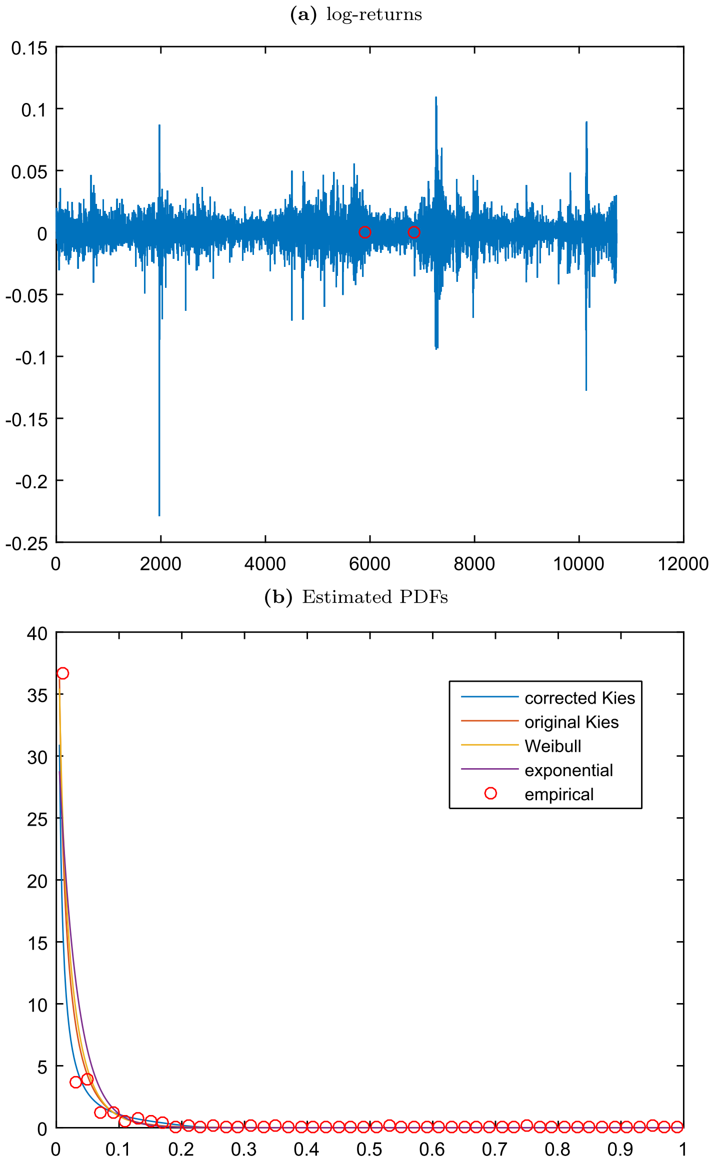


Fig. 4. S&P500 log-returns and the related estimations

that the corrected Kies distribution is significantly closer to the real observations – its error is 23.1820. The value of this error for the original Kies distribution is 25.3491, whereas for the exponential and Weibull distributions it is 26.6652 and 29.4037, respectively.

Having in mind Propositions 2.1 and 2.6 (third statements) we conclude that the initial value of PDF for both Kies style distributions is the infinity, because $b =$

$0.7120 < 1$ for the original distribution and $ab = 0.4509 < 1$ for the corrected one. Hence, a lot of mass is in the left part of the domain. This fact confirms the mentioned above financial phenomenon of volatility clustering. This is true for the Weibull distribution, too, since its parameter k is less than one; $k = 0.8327 < 1$. Also, the initial value of the exponential PDF is relatively large; it is 34.1236.

Additionally, the shape of the calibrated distributions means that the right tails are important. In fact, they present the probabilities of large calm periods in the markets. We compare the results which the four distributions generate for the tail measures *VaR*, *AVaR*, and *EX* with the empirical ones – see again Table 2. The levels we have chosen are 0.9, 0.925, 0.95, and 0.975; the values of *VaR*, *AVaR*, and *EX* are derived via equations (11), (30), and (37). Note that here the meaning of these measures is quite different from their traditional use in finance. We can see again that the corrected Kies distribution produces more realistic values in a comparison with the other distributions.

Here is the place to mention another purpose for which we can use these distributions. The available historical data generates relatively small number of observations for the dates with shocks. Note that the lower empirical values for all tail measures reported in Table 2 can be explained namely by the lack of enough observations. Therefore we can use the theoretical distributions as a tool to fulfill the missing information. In this light the closest tail behavior of the corrected Kies distribution has an additional importance.

Table 2. Fits to the S&P500 data

parameter	corrected Kies	original Kies	Weibull	exponential	empirical
λ	3.1091	-	0.0228	0.0293	-
k	55.0876	15.7857	0.8327	-	-
a	0.2086	-	-	-	-
b	2.1617	0.7120	-	-	-
LSqE	23.1820	25.3491	29.4037	26.6652	-
VaR	corrected Kies	original Kies	Weibull	exponential	empirical
0.9	0.0710	0.0627	0.0622	0.0675	0.0690
0.925	0.0877	0.0732	0.0716	0.0759	0.0920
0.95	0.1106	0.0883	0.0853	0.0878	0.1300
0.975	0.1448	0.1149	0.1095	0.1081	0.1800
$\overline{AVaR}$	corrected Kies	original Kies	Weibull	exponential	empirical
0.9	0.1200	0.1004	0.0969	0.0968	0.1872
0.925	0.1337	0.1112	0.1069	0.1052	0.2221
0.95	0.1513	0.1268	0.1214	0.1171	0.2799
0.975	0.1763	0.1535	0.1468	0.1374	0.4154
expectile	corrected Kies	original Kies	Weibull	exponential	empirical
0.9	0.0666	0.0564	0.0556	0.0591	0.0753
0.925	0.0745	0.0624	0.0612	0.0643	0.0856
0.95	0.0857	0.0711	0.0694	0.0718	0.1009
0.975	0.1044	0.0864	0.0838	0.0849	0.1274

A Proof of Proposition 2.1

The value of PDF (3) at left endpoint of the distribution domain, $h(0)$, can be obtain directly from equation (3). To continue, let us examine the derivative of PDF (3). It can be presented as

$$h'(t) = ke^{-k\eta(t)}\left[\eta''(t) - k\left(\eta'(t)\right)^2\right], \tag{70}$$

where

$$\eta(t) := \left(\frac{t}{1-t}\right)^b. \tag{71}$$

Derivative of function (71) is

$$\eta'(t) = b\frac{t^{b-1}}{(1-t)^{b+1}}. \tag{72}$$

Let us consider first the case $b = 1$ and therefore

$$\eta'(t) = \frac{1}{(1-t)^2}. \tag{73}$$

The second derivative of function (71) is

$$\eta''(t) = 2\frac{1}{(1-t)^3}. \tag{74}$$

Having in mind formulas (73) and (74), we see that derivative (70) is positive when $t < t_2 = 1 - \frac{k}{2}$, and vice versa. Hence, if $k \geq 2$, then derivative (70) is always negative in the distribution domain and therefore the PDF is a decreasing function. Otherwise, if $k < 2$, then $0 < t_2 < 1$, and hence the PDF increases for $t \in (0, t_2)$ and decreases for $t \in (t_2, 1)$.

Suppose now that $b \neq 1$. The second derivative of function (71) now is

$$\eta''(t) = b\frac{t^{b-2}}{(1-t)^{b+2}}(2t + b - 1). \tag{75}$$

Taking in attention formulas (72) and (75), we conclude that PDF's derivative (70) is positive when $\alpha(t) < 0$, and vice versa, where function $\alpha(t)$ is given in equation (4).

Let us consider the case $b > 1$. The derivative $\alpha'(t)$, given in equation (5), is an increasing function with negative left endpoint and positive right endpoint, $\alpha'(0) = -2$ and $\alpha'(1) = +\infty$. Thus it has a unique root in the interval $(0, 1)$, which we denote by t_2. Hence, function (4) decreases for $t \in (0, t_2)$, has a minimum for $t = t_2$ and increases to $\alpha(1) = +\infty$. The shape of the PDF follows, since the left endpoint of the function $\alpha(t)$ is negative, $\alpha(0) = -(b-1) < 0$.

Finally, suppose that $b < 1$. We derive the second derivative of function (4) as

$$\alpha''(t) = kb^2\frac{t^{b-2}}{(1-t)^{b+2}}(2t + b - 1). \tag{76}$$

Therefore $\alpha''(t) < 0$ for $t \in (0, \bar{t})$, $\bar{t} = \frac{1-b}{2}$, and $\alpha''(t) > 0$ for $t \in (\bar{t}, 1)$. Hence, the derivative $\alpha'(t)$ decreases when $t \in (0, \bar{t})$ and increases when $t \in (\bar{t}, 1)$. If $\alpha'(\bar{t}) \geq 0$, then the derivative $\alpha'(t)$ is positive in the whole domain and thus $\alpha(t)$ is an increasing function. Hence, the PDF decreases in the distribution domain, since $\alpha(0) = 1 - b > 0$.

Suppose that $\alpha'(\bar{t}) < 0$. Having in mind $\alpha'(0) = \alpha'(1) = +\infty$, we conclude that the derivative $\alpha'(t)$ has two roots, $\bar{t}_1$ and $\bar{t}_2$. Also, $\alpha'(t) < 0$ for $t \in (t_1, t_2)$ and $\alpha'(t) > 0$ outside the interval. Therefore the function $\alpha(t)$ starts from the positive value $\alpha(0) = 1 - b$, increases to a local maximum for $t = \bar{t}_1$, decreases to a local minimum for $t = \bar{t}_2$ and increases to infinity when $t = 1$. Hence, if $\alpha(\bar{t}_2) \geq 0$, then $\alpha(t)$ is positive in the whole domain and therefore the PDF is a decreasing function. Otherwise, if $\alpha(\bar{t}_2) < 0$, then the function $\alpha(t)$ has two roots, t_1 and t_2, and it is negative between them and positive outside. This finishes the proof.

Acknowledgement

The authors would like to express sincere gratitude to the editor Prof. Yuliya Mishura and to the anonymous reviewers for the helpful and constructive comments which substantially improve the quality of this paper.

Funding

This research has been partially supported by Grant No BG05M2OP001-1.001-0003, financed by the Science and Education for Smart Growth Operational Program (2014–2020) and co-financed by the European Union through the European structural and Investment funds. The first author was supported also by the project KP-06-N32/8 with the Bulgarian National Science Fund.

References

[1] Afify, A.Z., Gemeay, A.M., Alfaer, N.M., Cordeiro, G.M., Hafez, E.H.: Power-modified Kies-exponential distribution: Properties, classical and Bayesian inference with an application to engineering data. Entropy **24**(7), 883 (2022). MR4467767. https://doi.org/10.3390/e24070883

[2] Al-Babtain, A.A., Shakhatreh, M.K., Nassar, M., Afify, A.Z.: A new modified Kies family: Properties, estimation under complete and type-II censored samples, and engineering applications. Mathematics **8**(8), 1345 (2020). MR4199201. https://doi.org/10.3934/math.2021176

[3] Alsubie, A.: Properties and applications of the modified Kies–Lomax distribution with estimation methods. J. Math., **2021**(2), 1–18 (2021). MR4346604. https://doi.org/10.1155/2021/1944864

[4] Berger, M.-H., Jeulin, D.: Statistical analysis of the failure stresses of ceramic fibres: Dependence of the weibull parameters on the gauge length, diameter variation and fluctuation of defect density. J. Mater. Sci. **38**(13), 2913–2923 (2003). https://doi.org/10.1023/A:1024405123420

[5] Bhatti, F.A., Ahmad, M.: On a new family of kies burr iii distribution: Development, properties, characterizations, and applications. Sci. Iran. **27**(5), 2555–2571 (2020). ISSN 1026-3098. . URL http://scientiairanica.sharif.edu/article_21382.html

[6] Emam, W., Tashkandy, Y.: The weibull claim model: Bivariate extension, bayesian, and maximum likelihood estimations. Math. Probl. Eng., **2022**(1), 1–10 (2022)

[7] Gupta, R.C., Bradley, D.M.: Representing the mean residual life in terms of the failure rate. Math. Comput. Model. **37**(12–13), 1271–1280 (2003). MR1996036. https://doi.org/10.1016/S0895-7177(03)90038-0

[8] Hansen, L.P.: Large sample properties of generalized method of moments estimators. Econometrica: Journal of the econometric society, **50**(4) 1029–1054 (1982). MR0666123. https://doi.org/10.2307/1912775

[9] Kies, J.A.: The strength of glass performance. In: Naval Research Lab Report, Washington, D.C., vol. 5093 (1958)

[10] Koenker, R.W.: Quantile regression. Cambridge University Press, (2005). ISBN 9780521845731. MR2268657. https://doi.org/10.1017/CBO9780511754098

[11] Kumar, C.S., Dharmaja, S.H.S.: On some properties of Kies distribution. Metron **72**(1), 97–122 (2014). MR3176964. https://doi.org/10.1007/s40300-013-0018-8

[12] Kumar, C.S., Dharmaja, S.H.S.: The exponentiated reduced Kies distribution: Properties and applications. Commun. Stat., Theory Methods **46**(17), 8778–8790 (2017). MR3680792. https://doi.org/10.1080/03610926.2016.1193199

[13] Kumar, C.S., Dharmaja, S.H.S.: On modified Kies distribution and its applications. J. Stat. Res. **51**(1), 41–60 (2017). MR3702285. https://doi.org/10.47302/jsr.2017510103

[14] Kyurkchiev, N., Zaevski, T., Iliev, A., Rahnev, A.: A modified three–parameter Kies cumulative distribution function in the light of reaction network analysis. Int. J. Differ. Equ. Appl. **21**(2), 1–17 (2022)

[15] Lai, C.-D.: Generalized Weibull distributions. In: Generalized Weibull Distributions, pp. 23–75. Springer, (2014). MR3115122. https://doi.org/10.1007/978-3-642-39106-4_2

[16] Matsushita, S., Hagiwara, K., Shiota, T., Shimada, H., Kuramoto, K., Toyokura, Y.: Lifetime data analysis of disease and aging by the weibull probability distribution. J. Clin. Epidemiol. **45**(10), 1165–1175 (1992). https://www.sciencedirect.com/science/article/pii/089543569290157I. https://doi.org/10.1016/0895-4356(92)90157-I

[17] McCool, J.I.: Using the Weibull distribution: reliability, modeling, and inference, vol. 950. John Wiley & Sons, (2012). MR3014584. https://doi.org/10.1002/9781118351994

[18] Prabhakar Murthy, D.N., Xie, M., Jiang, R.: Weibull models. John Wiley & Sons, (2004). MR2013269

[19] Newey, W.K., Powell, J.L.: Asymmetric least squares estimation and testing. Econometrica: Journal of the Econometric Society **55**(4), 819–847 (1987). MR0906565. https://doi.org/10.2307/1911031

[20] Rinne, H.: The Weibull distribution: a handbook. Chapman and Hall/CRC, (2008). MR2477856

[21] Sanku, D.E.Y., Nassarn, M., Kumar, D.: Moments and estimation of reduced Kies distribution based on progressive type-II right censored order statistics. Hacet. J. Math. Stat. **48**(1), 332–350 (2019). MR3976180. https://doi.org/10.15672/hjms.2018.611

[22] Seguro, J.V., Lambert, T.W.: Modern estimation of the parameters of the weibull wind speed distribution for wind energy analysis. J. Wind Eng. Ind. Aerodyn. **85**(1),

75–84 (2000). https://www.sciencedirect.com/science/article/pii/S0167610599001221. https://doi.org/10.1016/S0167-6105(99)00122-1

[23] Sendov, B.: Hausdorff approximations, vol. 50. Springer, (1990). MR1078632. https://doi.org/10.1007/978-94-009-0673-0

[24] Shafiq, A., Lone, S.A., Naz Sindhu, T., El Khatib, Y., Al-Mdallal, Q.M., Muhammad, T.: A new modified Kies Fréchet distribution: Applications of mortality rate of Covid-19. Results Phys. **28**, 104638 (2021). https://www.sciencedirect.com/science/article/pii/S2211379721007294. https://doi.org/10.1016/j.rinp.2021.104638

[25] Sobhi, M.M.A.: The modified Kies–Fréchet distribution: properties, inference and application. AIMS Math. **6**, 4691–4714 (2021). MR4220431. https://doi.org/10.3934/math.2021276

[26] Soulimani, A., Benjillali, M., Chergui, H., da Costa, D.B.: Multihop weibull-fading communications: Performance analysis framework and applications. J. Franklin Inst. **358**(15), 8012–8044 (2021). https://www.sciencedirect.com/science/article/pii/S0016003221004701. MR4319388. https://doi.org/10.1016/j.jfranklin.2021.08.004

[27] Weibull, W.: A statistical distribution function of wide applicability. J. Appl. Mech. **18**(3), 293–297 (1951). https://doi.org/10.1115/1.4010337

[28] Wilks, D.S.: Rainfall intensity, the weibull distribution, and estimation of daily surface runoff. J. Appl. Meteorol. Climatol. **28**(1), 52–58 (1989). https://doi.org/10.1175/1520-0450(1989)028<0052:RITWDA>2.0.CO;2

[29] Yazhou, J., Molin, W., Zhixin, J.: Probability distribution of machining center failures. Reliab. Eng. Syst. Saf. **50**(1), 121–125 (1995). https://www.sciencedirect.com/science/article/pii/095183209500070I. https://doi.org/10.1016/0951-8320(95)00070-I

[30] Zaevski, T.S., Kyurkchiev, N.: Some notes on the four-parameters Kies distribution. Comptes rendus de l'Académie bulgare des Sciences **75**(10), 1403–1409 (2022). MR4504780