STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **13**, January 2025, pp 27–46.
Published online in International Academic Press (www.IAPress.org)

An Extended Discrete Model for Actuarial Data and Value at Risk Analysis: Properties, Applications and Risk Analysis under Financial Automobile Claims Data

Mohamed Ibrahim [1,2,*], Nadeem Shafique Butt[3], Abdullah H. Al-Nefaie [1], G. G. Hamedani[4], Haitham M. Yousof[5] and Aya Shehata Mahmoud[5]

[1] *Department of Quantitative Methods, college of Business, King Faisal University, Al Ahsa 31982, Saudi Arabia*
[2] *Department of Applied, Mathematical and Actuarial Statistics, Faculty of Commerce, Damietta University, Damietta, Egypt*
[3] *Department of Family and Community Medicine, King Abdul Aziz University, Jeddah, Kingdom of Saudi Arabia*
[4] *Department of Mathematical and Statistical Sciences, Marquette University, 1313 W. Wisconsin Ave., Milwaukee, WI 53233, USA*
[5] *Department of Statistics, Mathematics and Insurance, Benha University, Egypt*

Abstract This paper deals with a new discrete distribution with high flexibility. We have studied many of its mathematical and statistical properties, and we have neglected many other properties due to the narrow scope of the paper. Additionally, we have presented a comprehensive analysis of actuarial risks. A good set of actuarial risk indicators that are used in financial analysis and measurement and evaluation of financial risks. Five discrete data sets have been relied upon in conducting the financial analysis and risk assessment. Necessary comments have been provided on the results of the paper, and a set of necessary recommendations are provided for insurance companies to avoid the occurrence of unexpected large losses. All these financial analyses have been conducted in light of a discrete probability distribution.

Keywords Discrete Model; Maximum Expected Loss; Inflated Claims Data; Over Dispersed Data; Automobile Claims Data; Risk Assessment; Value at Risk

AMS 2010 subject classifications 62E15, 62J05

DOI: 10.19139/soic-2310-5070-2147

1. Introduction

In insurance and risk analysis, examining claims frequencies is crucial for effectively evaluating and managing financial risks. Specifically, analyzing automobile claims data poses distinct difficulties due to inflated and over-dispersed frequencies. Factors like demographic attributes, vehicle types, and geographic regions significantly impact both the frequency and severity of claims (see [32]). The goal of this paper is to use the discrete expanded Burr-Hatke (DEBH) distribution (see [41]) to analyze automobile claims frequencies that exhibit inflation and over-dispersion. Discrete models provide a more adaptable and detailed approach compared to traditional continuous models, especially for handling count data with excessive zeros and greater variability than what is typically predicted by standard statistical distributions (see [36], [4] and [28]). According to [4], the focus on a discrete modeling framework stems from the inadequacies of traditional continuous distributions in accurately capturing the unique characteristics of automobile claims data. By utilizing discrete models, we aim to overcome the limitations of normality and homogeneity assumptions, offering a more accurate representation of the underlying claims processes. This paper contributes to the fields of statistics and risk analysis by introducing a new methodology

[*] Correspondence to: Mohamed Ibrahim (Email:miahmed@kfu.edu.sa). Department of Quantitative Methods, college of Business, King Faisal University, Al Ahsa 31982, Saudi Arabia

ISSN 2310-5070 (online) ISSN 2311-004X (print)

specifically designed to address the excess zeros and variability found in automobile claims frequencies. Our approach integrates advanced statistical techniques to develop a specialized discrete claims model, effectively capturing the complex patterns and relationships within the data, thereby enhancing the accuracy and reliability of risk assessments in the insurance industry.

Recently, [41] introduced an innovative generalized discrete distribution known as the DEBH distribution, which includes the discrete Burr-Hatke distribution. They thoroughly examined its properties, revealing that the distribution's probability mass function can exhibit a range of features such as right skewness with different shapes, bimodality and uniformity. Additionally, the hazard rate function associated with this distribution may display monotonically decreasing, monotonically increasing or constant patterns. In their research, [41] conducted numerical analyses to detail important statistical measures, including mean, variance, skewness, kurtosis, and the index of dispersion. The distribution's flexibility makes it useful for modeling both under-dispersed and over-dispersed count data. The study also included characterizations of the distribution through the conditional expectation of specific functions of the random variable and the hazard rate function. Furthermore, the authors investigated both Bayesian and non-Bayesian estimation techniques, comparing their performance through numerical simulations. The DEBH model was applied to real-world datasets, such as dental carries and kidney cyst counts, illustrating its practical utility. In statistical modeling, selecting the right probability distribution is crucial for accurately representing real-world phenomena. This is especially important in ithe nsurance analytics, where precise assessment and prediction of automobile insurance claims are critical. Given the increasing complexity of the insurance data and the need for accurate risk evaluation, developing new distributions that address the specific complexities of such data is essential (see [33]).

The DEBH distribution shows considerable potential in the realm of automobile insurance claims. Automobile claim data frequently present complexities such as over-dispersion, right skewness and multimodality, which require adaptable distribution models to accurately capture these characteristics. Utilizing the flexibility of the DEBH distribution allows researchers and practitioners to develop more precise and reliable models for evaluating claim frequency and severity, thus improving risk management strategies within the insurance sector (for more useful details, see [29], [12], [37], [5] and [2]. Other useful continuous model can be covered to discrete models, see [13] (for new Lindley extension), [3] (for novel XGamma extension) and [35], [14] (novel compound reciprocal Rayleigh extension), [17] (Burr XII model) and [11] (for new lomax extensions)).

In this context, the paper aims to explore the application of the DEBH distribution for modeling automobile claim data. By conducting empirical analyses and case studies, we intend to illustrate how effectively the DEBH distribution captures the complexities inherent in automobile insurance data, thus improving the accuracy and reliability of risk assessment models. The study utilized five different datasets of automobile insurance claims (see [30], [31] and [38]). The data were analyzed statistically, and the performance of the DEBH distribution was compared with several other competitive distributions in this field. In the statistical literature, there are many important, flexible and applicable continuous probability distributions that deserve attention and conversion from continuous distributions to discrete distributions. The process of converting to discrete distributions will undoubtedly serve the field of statistical and mathematical modeling of discrete numerical data. For more of these previously mentioned distributions, see [39], [22], [23], [15], [16], [6] and [34].

In the following sections, we will provide a comprehensive analysis of the DEBH distribution, covering its mathematical formulation, key statistical properties, and techniques for parameter estimation. We will then illustrate its practical application in modeling automobile claim data through detailed insurance case studies and numerical simulations. These examples will underscore the DEBH distribution's potential to revolutionize statistical modeling practices in the insurance industry. By highlighting the importance of the DEBH distribution and its application in modeling automobile claim data, this research seeks to advance statistical methodologies in insurance analytics, with the goal of improving the robustness and accuracy of decision-making processes in the sector. Additionally, we will consider various risk indicators to analyze automobile insurance claims, including:

1. Value at Risk (VaR$[q;Y]$): This metric measures the potential loss at a given quantile, providing an estimate of the maximum loss expected over a specified time period with certain level of confidence.
2. Tail Value at Risk (TVaR$[q;Y]$): Also known as Conditional Value at Risk (CVaR), this indicator assesses the average loss exceeding the VaR, offering insights into the risk of extreme losses.
3. Tail Variance (TV$[q;Y]$): This measures the variance of losses beyond the quantile q, helping to understand the variability of extreme losses.
4. Tail Mean Variance (TMV$[q;Y]$): This provides the variance of the mean of losses beyond the quantile q, contributing to a deeper understanding of the distribution of extreme outcomes.
5. Expected Loss (EL$[q;Y]$): This indicator calculates the expected value of losses at a specific quantile, reflecting the average loss anticipated in the tail of the distribution.

In this context, the DEBH distribution will be evaluated based on these risk indicators, analyzing how well it performs in terms of risk calculations and its behavior in modeling extreme values.

2. The model and its main properties

In recent years, there has been increasing interest in the discretization of continuous probability distributions. This deals with a three-parameter discrete distribution that encompasses the Burr-Hatke distribution, as described by El-Morshedy et al. [8]. Therefore, this new distribution can be regarded as an extension of the Burr-Hatke distribution. In statistical research, numerous discrete versions of continuous distributions have been developed, examined, and utilized for modeling count data. Examples include the Poisson-Lindley distribution (PLi) by Sankaran [26], the discrete Weibull distribution (DW) by Nakagawa and Osaki [24], the discrete half-normal distribution by Kemp [18], the discrete Rayleigh distribution (DR) by Roy [25], the discrete Pareto distribution (DPa) by Krishna and Pundir [19], the generalized geometric distribution (GGc) by Gomez-Déniz [9], the discrete Lindley distribution (DLi) by Gomez-Déniz and Calderin-Ojeda [10], the discrete linear failure rate distribution (DLFR) by Kumar et al. [20], and the exponentiated discrete Lindley distribution (EDLi) by El-Morshedy et al.[7]. Recently, a new family of discrete distributions based on the Rayleigh distribution, called the discrete Rayleigh G (DRG) family, has been introduced. This family includes various new discrete sub-models (see Aboraya et al. [1]). Yousof et al. [40] defined and studied a new continuous family of probability distributions based on the Burr-Hatke (BH) distribution. A RV Y is said to have the expanded Burr-Hatke (EBH) distribution if its cumulative distribution function (CDF) is given by

$$G_{a,b}\left(y\right)=1-\frac{1}{y+1}\exp\left(-ay^{b}\right)|y>0,\text{ and }a,b>0$$

For $b=1$, the EBH distribution reduces to one parameter BH distribution first introduced by Maniu and Voda [21]. Then, the CDF of the discrete expanded Burr-Hatke (DEBH) model can be expressed as

$$F_{p,a,b}\left(y\right)=1-\frac{1}{y+2}\,p^{a(y+1)^{b}}\quad|0<p<1\text{ and }y\in\mathbb{N}^{*}=\mathbb{N}\cup\left\{0\right\}, \tag{1}$$

where $\mathbb{N}=\{1,2,\dots\}$. For $b=1$, the DEBH distribution reduces to one parameter DEBH distribution as introduced by El-Morshedy et al. [8]. The corresponding reliability function (RF) due to Steutel and van Harn [27] can be written as

$$\overline{F}_{p,a,b}\left(y\right)=S_{p,a,b}\left(y\right)=\frac{1}{y+2}p^{a(y+1)^{b}}\quad|0<p<1\text{ and }y\in\mathbb{N}^{*}. \tag{2}$$

The probability mass function (PMF) of the DEBH distribution corresponding to (1) and (2) can be expressed as

$$P_{p,a,b}\left(y\right)=\overline{F}_{p,a,b}\left(y-1\right)-\overline{F}_{p,a,b}\left(y\right)\quad|_{\left(0<p<1\text{ and }y\in\mathbb{N}^{*}\right)}, \tag{3}$$

that is

$$P_{p,a,b}\left(y\right)=\frac{1}{y+1}p^{ay^{b}}-\frac{1}{y+2}\,p^{a(y+1)^{b}}\quad|0<p<1\text{ and }y\in\mathbb{N}^{*}. \tag{4}$$

As y becomes large, the denominator $y+1$ is approximately y, so

$$\frac{1}{y+1} \approx \frac{1}{y}.$$

Therefore, the first term $\frac{1}{y+1}p^{ay^b}$ asymptotically behaves as $\frac{1}{y}p^{ay^b}$. Since $0<p<1$, p^{ay^b} will tend to 0 as y^b grows, but it will do so very slowly, depending on the values of p, a, b. As y becomes large, $y+2$ is approximately y and $(y+1)^b$ can be expanded using the binomial theorem

$$(y+1)^b = y^b\left(\frac{1}{y}+1\right)^b \approx y^b\left(\frac{b}{y}+1\right).$$

Thus

$$\frac{1}{y+2}\,p^{a(y+1)^b} \approx \frac{1}{y}\,p^{a\left(y^b+by^{b-1}\right)},$$

which simplifies to

$$\frac{1}{y}\,p^{ay^b}p^{aby^{b-1}}.$$

The PMF $P_{p,a,b}\left(y\right)$ of the DEBH distribution asymptotically decays as

$$P_{p,a,b}\left(y\right) \approx \frac{1}{y}\left[p^{ay^b} - p^{ay^b}p^{aby^{b-1}}\right] \approx \frac{1}{y}\,p^{ay^b}\left[-\ln\left(p\right)\right]aby^{b-1}.$$

The model in (4) can be considered as a new generalization of the model of [42].This illustrates the flexibility of the proposed two-parameter DEBH distribution and the importance of the parameter b in this regard. The hazard rate function (HRF) can be written as

$$H_{p,a,b}\left(y\right) = \frac{1}{\overline{F}_{p,a,b}\left(y-1\right)}P_{p,a,b}\left(y\right).$$

Then,

$$H_{p,a,b}\left(y\right) = \frac{\left(y+2\right)p^{ay^b}}{\left(y+1\right)p^{a(y+1)^b}} - 1|0<p<1 \text{ and } y\in\mathbb{N}^*. \tag{5}$$

This can be further broken down as

$$H_{p,a,b}\left(y\right) = \frac{\left(y+2\right)}{\left(y+1\right)}p^{a\left[y^b-(y+1)^b\right]} - 1|0<p<1 \text{ and } y\in\mathbb{N}^*.$$

As $y\to\infty$, $\frac{(y+2)}{(y+1)}=1$, $\left(y+1\right)^b=y^b$, implying that $y^b-\left(y+1\right)^b$ becomes increasingly negative. Therefore, for large y we have

$$p^{a\left[y^b-(y+1)^b\right]} \to 0 \Rightarrow H_{p,a,b}\left(y\right) \to -1.$$

As y increases, the HRF approaches -1, indicating a decreasing probability of occurrence as y grows large. For small y, the term $\frac{(y+2)}{(y+1)}$ is slightly greater than 1, and the exponent $a\left[y^b-\left(y+1\right)^b\right]$ is less negative. Hence, the HRF might be small but positive for smaller values of y. Specifically, we have

$$H_{p,a,b}\left(1\right) = \frac{3}{2}p^{a\left(1^b-2^b\right)} - 1.$$

Given $0<p<1$, and $a\left(1^b-2^b\right)$ being negative, the HRF for $y=1$ will also be less than 1, indicating a low probability of occurrence at the first step. Now, let Y be a random variable following the DEBH distribution. Then, the probability generating function (PGF) of Y is given by

$$p(s) = \mathbb{E}(s^y) = 1 + (s-1)\sum_{y=1}^{\infty} s^{y-1}\frac{1}{y+1}p^{ay^b} \quad |0 < s < 1. \tag{6}$$

The r^{th} ordinary moments of Y is given by

$$\mathbb{E}(Y^r) = \sum_{y=1}^{\infty} \left[y^r - (y-1)^r\right]\frac{1}{y+1}p^{ay^b}. \tag{7}$$

Therefore, the mean and variance of the DEBH distribution do not have analytical forms. We can ,however, expressed them as the following series expressions:

$$\mathbb{E}\left(Y\right) = \sum_{y=1}^{\infty} \frac{1}{y+1}p^{ay^b}, \tag{8}$$

and

$$\mathbb{V}\left(Y\right) = \sum_{y=1}^{\infty}(2y-1)\frac{1}{y+1}p^{ay^b} - \left(\sum_{y=1}^{\infty}\frac{1}{y+1}p^{ay^b}\right)^2. \tag{9}$$

Based on (8) and (9), the index of dispersion is

$$\mathbb{D}\left(Y\right) = \frac{1}{\mathbb{E}\left(Y\right)}V\left(Y\right).$$

Similarly, we can express the first four moments of Y, allowing to define the following skewness and kurtosis measures:

$$\mathbb{S}\left(Y\right) = \frac{1}{\left[V\left(Y\right)\right]^{3/2}}\left\{\mathbb{E}(Y^3) - 3\mathbb{E}\left(Y\right)\mathbb{E}(Y^2) + 2[\mathbb{E}\left(Y\right)]^3\right\}.$$

and

$$\varsigma\left(Y\right) = \frac{1}{\left[V\left(Y\right)\right]^2}\left\{\mathbb{E}(Y^4) - 4\mathbb{E}\left(Y\right)\mathbb{E}(Y^2) + 6[\mathbb{E}\left(Y\right)]^2\mathbb{E}(Y^2) - 3[\mathbb{E}\left(Y\right)]^4\right\}.$$

All these measures can be determined numerically with the help of any mathematical software. Now, let us establish some general relations regarding the order statistics of the DEBH distribution. The CDF of the k^{th} order statistic from the DEBH distribution is given by

$$F_{p,a,b,k:m}\left(y\right) = \sum_{\varsigma=k}^{m}\binom{m}{\varsigma}\left[F_{p,a,b}\left(y\right)\right]^{\varsigma}\left[\overline{F}_{p,a,b}\left(y\right)\right]^{m-\varsigma}.$$

Applying the binomial formula to $\left[F_{p,a,b}\left(y\right)\right]^{\varsigma} = \left[1 - \overline{F}_{p,a,b}\left(y\right)\right]^{\varsigma}$, we have

$$F_{p,a,b,k:m}\left(y\right) = \sum_{\varsigma=k}^{m}\sum_{l=0}^{\varsigma}\binom{m}{\varsigma}\binom{\varsigma}{l}(-1)^l\left[\overline{F}_{p,a,b}\left(y\right)\right]^{l+m-\varsigma}.$$

Then,

$$F_{p,a,b,k:m}\left(y\right) = \sum_{\varsigma=k}^{m}\sum_{l=0}^{\varsigma}\binom{m}{\varsigma}\binom{\varsigma}{l}(-1)^l\frac{p^{(l+m-\varsigma)a(y+1)^b}}{(y+2)^{l+m-\varsigma}}, \quad |0 < p < 1 \text{ and } y \in \gtrdot.$$

Also, the corresponding PMF is obtained as

$$P_{p,a,b,k:m}\left(y\right) = F_{p,a,b,k:m}\left(y\right) - F_{p,a,b,k:m}\left(y-1\right).$$

Then

$$P_{p,a,b,k:m}\left(y\right)=\sum_{\varsigma=k}^{m}\sum_{l=0}^{\varsigma}\binom{m}{\varsigma}\binom{\varsigma}{l}(-1)^{l}\left[\begin{array}{c}\frac{1}{(y+2)^{l+m-\varsigma}}p^{(l+m-\varsigma)a(y+1)^{b}}\\-\frac{1}{(y+1)^{l+m-\varsigma}}p^{(l+m-\varsigma)ay^{b}}\end{array}\right].$$

From this form of the PMF, several measures and functions can be derived, as done for the former DEBH distribution.

The entropy $H\left(Y\right)$ of the DEBH distribution is a measure of uncertainty and can be computed as:

$$H\left(Y\right)=-\sum_{y=0}^{+\theta}P_{p,a,b}\left(y\right)\log P_{p,a,b}\left(y\right).$$

Let $M_m=\max(Y_1,Y_2,\ldots,Y_m)$. Then, the distribution function of M_m is

$$P(M_m\leq y)=\left[1-\frac{1}{y+2}\,p^{a(y+1)^{b}}\right]^{m}.$$

Analyze the behavior of $\left[1-\frac{1}{y+2}\,p^{a(y+1)^{b}}\right]^{m}$ as $m\to+\infty$ and determine the limiting distribution. For modeling extreme values, compute the return level and return period. Given a return period T, the return level y_T is such that $P(M_m\leq y_T)=1-1/T$. Solving for y_T provides insights into extreme value predictions. For a given extreme value, the return period T is $T=1/\left[1-F_{p,a,b}\left(y\right)\right]$. The tail index ξ characterizes the heaviness of the tails and can be estimated from the data. Using methods like the Hill estimator or the Pickands estimator, estimate the tail index to understand the extremal behavior of the DEBH distribution.

3. Risk indicator derivations

The VaR at a specified confidence level a is the q-quantile of the distribution. For a discrete distribution represented by the PMF, VaR can be computed by finding the smallest value v such that

$$\Pr(y\leq v)\geq q.$$

For the PMF, the VaR at confidence level a can be found numerically by solving:

$$\Sigma_{y\leq v}\;P_{P,\beta}(y)=\Sigma_{y\leq v}\left[\frac{1}{y+1}P^{ay^{b}}-\frac{1}{y+2}P^{a(y+1)^{\beta}}\right]\geq q,$$

here, $v=$VaR$[q;Y]$ is our VaR. The Tail Value at Risk at a specified confidence level $-q$ (or Conditional Value at Risk) represents the expected loss beyond the VaR. It is computed as the conditional expectation of losses exceeding the VaR. TVaR at confidence level a can be calculated as:

$$\text{TVaR}\left[q;Y\right]=\frac{1}{1-q}\Sigma_{y\leq\text{VaR}}\left(y-\text{TVaR}\left[q;Y\right]\right)\left[\frac{1}{y+1}P^{ay^{b}}-\frac{1}{y+2}P^{a(y+1)^{\beta}}\right].$$

Tail-variance at a specific quantile q represents the variance of losses beyond the q-quantile. It is calculated as:

$$\text{TV}\left[q;Y\right]=\Sigma_{y\leq q}\left(y-q\right)^{2}\times\left[\frac{1}{y+1}P^{ay^{b}}-\frac{1}{y+2}P^{a(y+1)^{\beta}}\right].$$

The tail-mean-variance at a quantile $-q$ is the mean of squared losses beyond the $-q$-quantile. It can be computed as:

$$\text{TMV}\left[q;Y\right]=\frac{1}{1-q}\Sigma_{y\leq q}\left(y-q\right)^{2}\times\left[\frac{1}{y+1}P^{ay^{b}}-\frac{1}{y+2}P^{a(y+1)^{\beta}}\right].$$

The expected-loss at a quantile q is the expected value of losses beyond the q-quantile. It is given by:

$$\mathrm{EL}\left[q;Y\right]=\Sigma_{y\leq q}\left(y-q\right)\times\left[\frac{1}{y+1}P^{ay^{b}}-\frac{1}{y+2}P^{a(y+1)^{\beta}}\right],$$

where, the VaR$[q;Y]$ provides a threshold value representing the maximum loss that might occur with a given confidence level. It is a crucial risk metric for setting risk limits and capital requirements. The TVaR$[q;Y]$ quantifies the average severity of losses beyond the VaR, providing deeper insights into potential extreme losses and tail risk. the TV$[q;Y]$ measures the spread or variability of losses in the tail of the distribution beyond a specified quantile, offering information on the risk of extreme outcomes. The TMV$[q;Y]$ provides the average squared deviation of losses beyond a quantile, indicating the average variability of extreme losses. Finally, the EL$[q;Y]$ estimates the average amount of loss expected beyond a certain quantile, aiding in risk assessment and scenario planning (for more details see [2]).

4. Describing automobile claims data

In this Section, we undertake a thorough examination of automobile insurance claim frequencies across different countries. This analysis is guided by the work of [31], who offer an extensive evaluation of this topic through the presentation of five distinct datasets. These datasets, which are also cited in the study by [38], provide a valuable basis for our exploration. Table 1 presents a detailed summary of these datasets, illustrating the occurrence of inflated over-dispersion in automobile insurance claims. This phenomenon, characterized by greater variability and higher-than-expected frequencies of claims, is a significant factor in understanding the complexities of insurance data across different regions.

Additionally, Figure 1 provides a comprehensive visualization of the automobile claims frequencies across five different countries. It includes three types of plots: box plots, scatter plots, and Q-Q (Quantile-Quantile) plots, each offering unique insights into the distribution and characteristics of the data sets. The box plots display the distribution of automobile claims frequencies for each country. Each plot presents a box-and-whisker diagram showing the median, quartiles, and potential outliers for the claims frequencies. The central box represents the interquartile range (IQR) between the first and third quartiles, with a line inside the box indicating the median frequency. The whiskers extend to 1.5 times the IQR from the quartiles, and any points outside this range are considered potential outliers. The scatter plots illustrate the relationship between the number of claims and the frequency of claims for each country.

Each plot is separated by country, allowing for a detailed examination of how the frequency of claims varies with the number of claims. Data points are plotted on the x-axis (number of claims) against the y-axis (frequency of claims), with different colors representing different countries. The Q-Q plots compare the quantiles of the claim frequencies against a theoretical normal distribution. Each Q-Q plot assesses how well the claim frequencies follow a normal distribution by plotting the quantiles of the data against the quantiles of a normal distribution. A straight line in the Q-Q plot indicates that the data follows a normal distribution closely. These visual representations offer a clear comparison of the distributions and variability within the data, highlighting the presence of inflated over-dispersion. By analyzing these plots, we can better understand how claim frequencies differ across countries and the implications for modeling and risk assessment in automobile insurance.

The average number of claims is relatively low at 0.15514, indicating that most policyholders had very few claims. Both the median and mode are zero, showing that a large portion of the data is concentrated around no claims. This is supported by the quartiles, all of which are zero. The data exhibits a high skewness of 3.153594, signifying a strong right skew. This means there are a few instances of higher claims, but most data points are clustered around zero. In Switzerland, the mean number of claims is also low at 0.1317373, with the median and mode at zero, and quartiles indicating that most claims are zero. The skewness is slightly lower than in the previous case at 2.980456, but still shows a right-skewed distribution. This suggests that while most claims are zero, there are some higher claims that affect the distribution. Belgium has a slightly higher mean claim frequency of 0.2143537. Although the median and mode remain zero, and quartiles show that most data points are zero, the skewness is the highest among the datasets at 3.4812. This indicates a pronounced right skew, with a few high claims having a significant impact on the distribution. Zaire exhibits the lowest mean claim frequency at 0.0865, indicating fewer claims on average. The median and mode are zero, and quartiles reinforce this concentration around zero. The skewness is the highest at 5.31602, reflecting an even more pronounced right skew, with very few high claims compared to the rest of the data. Germany's mean claim frequency is similar to Switzerland and Great Britain at 0.1442198. The median and mode are zero, and quartiles suggest that most claims are zero. The skewness is slightly lower than Belgium's at 3.230354, but still indicates a notable right skew.

Table 1: Automobile claim data.

data set	**Country**	**Year**	**Automobile claims frequencies**
			0, 1, 2, 3, 4, 5, 6, 7
I	Switzerland	1961	103704, 14075, 1766, 255, 45, 6, 2, 0
II	Great-Britain	1968	370412, 46545, 3935, 317, 28, 3, 0, 0
III	Belgium	1958	7840, 1317, 239, 42, 14, 4, 4, 1
IV	Zaire	1974	3719, 232, 38, 7, 3, 1, 0, 0
V	Germany	1960	20592, 2651, 297, 41, 7, 0, 1, 0

Table 1 provides detailed insights into automobile insurance claim data from various countries and years. Here's an expanded analysis of the data entries:

1. Switzerland (1961): The data for Switzerland in 1961 reveals a predominant number of claims in the ”0 claims” category, totaling 103,704 instances. This distribution shows a pronounced right skew, with a substantial drop in the number of claims as we progress to higher claim categories. This pattern, where fewer claims are observed in higher categories, is typical in insurance claim data, reflecting a common trend of many policyholders experiencing no claims while a few incur higher numbers of claims.
2. Great Britain (1968): The dataset from Great Britain in 1968 exhibits a similar trend to Switzerland, with a very high frequency of claims in the ”0 claims” category, numbering 370,412. As with the Swiss data, there is a noticeable decrease in the number of claims in the higher categories, indicative of a right-skewed distribution. Notably, there are no reported claims in categories 6 and 7, which points to a relatively infrequent occurrence of extreme claims.
3. Belgium (1958): The Belgian data from 1958 shows a lower overall number of automobile claims compared to Switzerland and Great Britain. This could suggest either fewer insurance policies or differences in insurance practices. The distribution remains right-skewed, with most claims clustered in the lower categories. Although there are a few claims reported in categories 6 and 7, these are relatively rare, indicating that extreme claims are uncommon.
4. Zaire (1974): The dataset for Zaire in 1974 reflects a significant drop in the number of automobile claims relative to the previous entries. This reduction could be attributed to lower insurance coverage or different insurance practices prevalent in Zaire at the time. The distribution is strongly right-skewed, with the majority of claims in the ”0 claims” category. There are very few claims in the higher categories, suggesting that moderate to severe claims are infrequent.
5. Germany (1960): The German data from 1960 also follows the right-skewed pattern seen in the other datasets, with the majority of claims falling into the lower categories. There is a clear decline in the number of claims as the claim categories increase, although this decline is less sharp compared to some of the other countries' data.

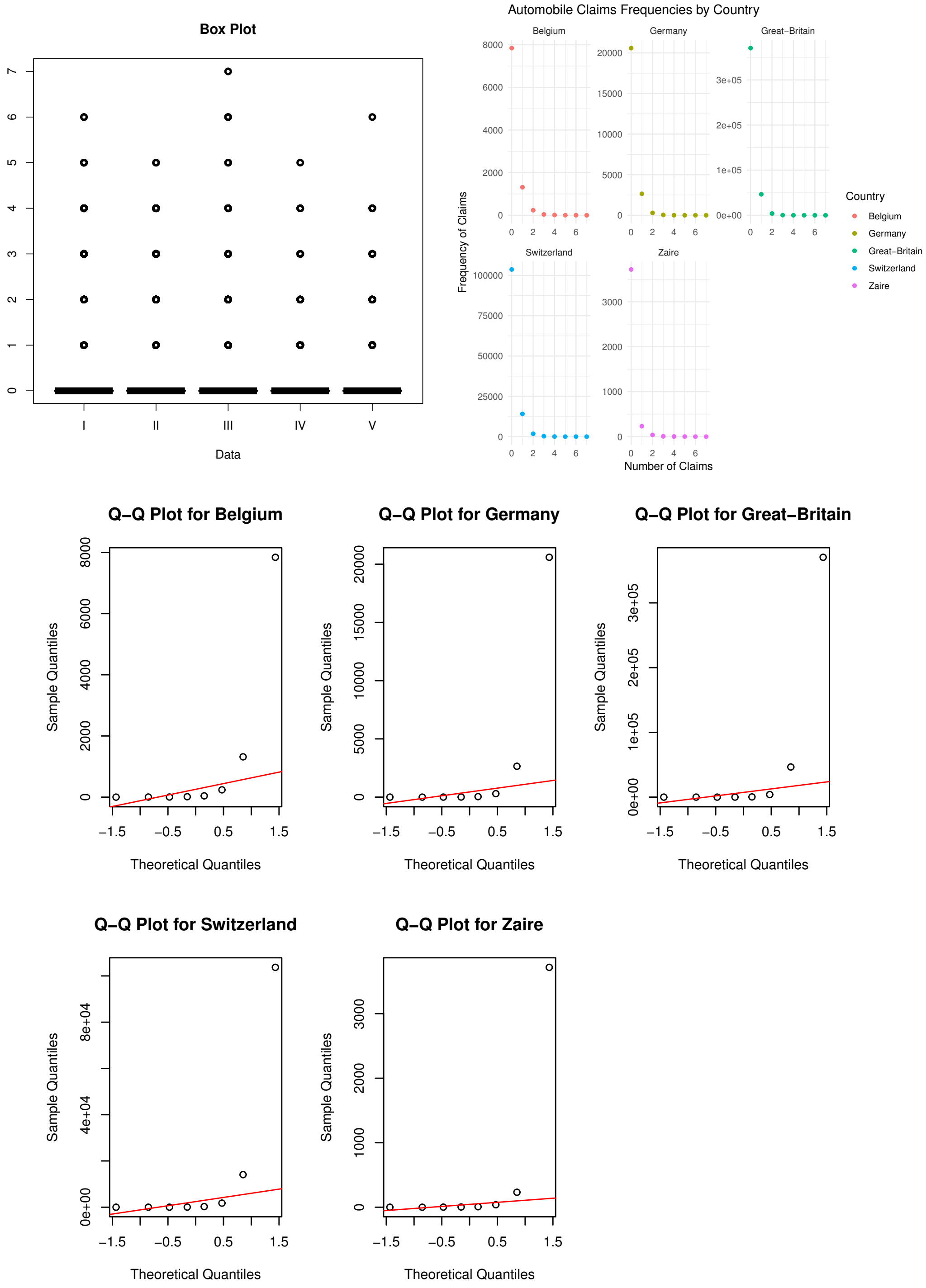

Figure 1. The box plots, scatter plots and Q-Q plots for the five data sets.

5. Data modeling

In this section, we delve into the challenges of modeling automobile claims data, which is vital for actuarial science and insurance analytics. Accurate claims data modeling is essential for insurers to effectively assess risk, determine suitable premiums, and ensure financial stability. Our analysis seeks to meet these needs by applying established claims distributions from actuarial literature. We have two main objectives: first, to use recognized claims distributions to model the automobile claims data, and second, to conduct a thorough comparison and evaluation of these distributions. This analysis is important as it enhances actuaries' and insurance professionals' understanding and predictive capabilities, providing valuable insights into automobile claims data behavior. To achieve our objectives, we will evaluate the distributions using three key criteria: the negative log-likelihood (NLL), the Akaike Information Criterion (AIC), and the Bayesian Information Criterion (BIC). These metrics serve as quantitative measures of each distribution's goodness-of-fit, helping us determine how well each model captures the underlying data patterns and characteristics. By comparing these criteria, our aim is to identify the distribution that best fits the automobile claims data and to highlight the strengths and weaknesses of each model. This information is crucial for practitioners to make informed decisions regarding risk management, pricing strategies, and resource allocation in the insurance industry. Table 2 below presents the results for the NLL, AIC, and BIC across five different models (BDLD, DLD, Poisson, DP, DEBH) applied to five datasets.

Table 2: Results of NLL, AIC and BIC for all automobile claims data.

data set	Criteria	BDLD	DLD	Poisson	DP	DEBH
I	NLL	54659.100	54659.614	55108.455	56351.011	54616.705
	AIC	109320.201	109321.227	110218.910	112704.021	109239.411
	BIC	109329.895	109330.921	110228.604	112713.715	109268.493
II	NLL	171198.407	171196.166	171373.176	178321.718	171141.352
	AIC	342398.813	342394.333	342748.352	356645.437	342288.704
	BIC	342409.764	342405.283	342759.303	356656.388	342321.557
III	NLL	5377.784	5377.510	5490.780	5486.714	5347.532
	AIC	10757.57	10757.02	10983.56	10975.43	10701.064
	BIC	10764.72	10764.18	10990.72	10982.58	10722.529
IV	NLL	1217.358	1217.698	1246.077	1186.498	1183.380
	AIC	2436.717	2437.397	2494.154	2374.997	2372.760
	BIC	2443.011	2443.691	2500.448	2381.291	2391.643
V	NLL	10228.342	10228.453	10297.843	10551.846	10223.857
	AIC	20458.684	20458.906	20597.686	21105.693	20453.714
	BIC	20466.752	20466.975	20605.755	21113.761	20477.920

Table 2 illustrates that for data set I, the DEBH model exhibits the lowest values across all criteria: a negative log-likelihood (NLL) of 54616.705, an Akaike Information Criterion (AIC) of 109239.411, and a Bayesian Information Criterion (BIC) of 109268.493. These results indicate that the DEBH model provides the best fit for this dataset, making it the preferred choice. In data set II, the DEBH model again leads with the lowest NLL value of 171141.352, the lowest AIC of 342288.704, and the lowest BIC of 342321.557. This confirms that the DEBH model is the best option for this dataset as well. For data set III, the DEBH model continues to perform the best, with the lowest NLL value of 5347.532, the lowest AIC of 10701.064, and the lowest BIC of 10722.529. This consistent performance across all metrics makes the DEBH model the top choice for this dataset. In data set IV, the DEBH model achieves the lowest NLL of 1183.380, the lowest AIC of 2372.760, and the lowest BIC of 2391.643, reinforcing its status as the best model for this data set. Similarly, for data set V, the DEBH model has the lowest NLL value of 10223.857, the lowest AIC of 20453.714, and the lowest BIC of 20477.920, confirming that it is the optimal model based on all three metrics.

The DEBH model consistently outperforms other models across all datasets when assessed using NLL, AIC, and BIC criteria. It achieves the lowest values for each of these metrics, reflecting its superior fit, efficiency, and balance between model complexity and goodness-of-fit. For all datasets presented in Table 2, the DEBH model is the preferred choice due to its best overall performance according to these statistical measures. Its reliable performance across various datasets indicates that it is robust and dependable for modeling automobile insurance claims data. Selecting the DEBH model allows practitioners to obtain a more accurate and dependable representation of claims data, enhancing risk assessment and decision-making processes within the insurance sector.

6. Automobile claims data and risk analysis

In the fields of risk management and financial analysis, assessing risk at various confidence levels is essential for understanding how adverse events might impact an organization's financial stability. The analysis of risk indicators such as VaR$[q;Y]$, TVaR$[q;Y]$, TV$[q;Y]$, TMV$[q;Y]$, and EL$[q;Y]$ offers crucial insights into the nature of loss distributions and the behavior of tail risks. VaR$[q;Y]$ (Value at Risk) estimates the maximum potential loss that is unlikely to be exceeded with a given probability (confidence level). Higher VaR$[q;Y]$ values at elevated confidence levels reflect a more conservative risk management strategy, indicating a lower tolerance for potential losses. TVaR$[q;Y]$ (Tail Value at Risk) represents the average loss that exceeds the VaR$[q;Y]$ threshold. It measures the expected shortfall beyond the maximum loss estimate, providing a sense of the average severity of losses in the tail of the distribution. TV$[q;Y]$ (Tail Variance) assesses the variability of the distribution's tail at a specified quantile level, highlighting the potential magnitude of losses in extreme scenarios that surpass the VaR$[q;Y]$. TMV$[q;Y]$ (Tail Mean Variance) captures the mean of losses in the tail of the distribution beyond the quantile level q, offering insights into the average severity of extreme losses. EL$[q;Y]$ reflects the expected value of the loss distribution beyond the quantile q, helping to evaluate the average impact of extreme events on the financial position of an organization.

Tables 3, 4, 5, 6, and 7 display various risk indicators for data sets I through V, respectively, at different quantile levels (70%, 75%, 80%, 85%, 90%, 95%, 99%). These tables offer a comprehensive view of risk assessment using the DEBH distribution, highlighting potential losses, tail behavior, and expected outcomes across different confidence levels. The consistent patterns observed across quantiles (70%, 90%, 99%) suggest that the DEBH distribution provides a stable and reliable framework for modeling and predicting risk. This stability is crucial for risk managers, analysts, and decision-makers, as it improves their ability to accurately quantify and manage risks in various scenarios.

Additionally, we present a series of visual aids to further substantiate our findings. Each figure provides a detailed graphical representation of the risk indicators across various confidence levels for different data sets. Figure 2 illustrates the risk indicators for data set I, highlighting how these indicators vary with different confidence levels. Figure 3 showcases the risk indicators for data set II, offering insights into the behavior of these metrics across varying levels of confidence. Figure 4 presents the risk indicators for data set III, depicting the changes in risk measures with different confidence thresholds. Figure 5 displays the risk indicators for data set IV, providing a visual overview of how these indicators fluctuate with confidence levels. Figure 6 reveals the risk indicators for data set V, illustrating the patterns and trends in risk metrics across various confidence levels.These figures collectively offer a comprehensive view of the risk indicators' patterns and behavior under different confidence levels. By visualizing the data in this manner, the figures not only validate the robustness and effectiveness of the DEBH distribution in modeling automobile insurance claims but also enhance our understanding of how risk indicators perform across a range of confidence levels. The clarity provided by these plots underscores the reliability of the

DEBH model in capturing and analyzing the complexities of insurance data, thereby reinforcing the credibility of our results and methodology.

Table 3: Risk analysis for data set I under the DEBH model.

CL	VaR$[q;Y]$	TVaR$[q;Y]$	TV$[q;Y]$	TMV$[q;Y]$	EL$[q;Y]$
70%	2	0.12884	0.2708	0.26424	-1.10418
75%	2	0.1546	0.32098	0.31509	-1.30458
80%	2	0.19325	0.39375	0.39013	-1.5541
85%	3	0.0496	0.15619	0.1277	-2.08553
90%	3	0.0744	0.23245	0.19062	-2.63333
95%	4	0.02506	0.10411	0.07712	-3.71485
99%	7	0.00037	0.00268	0.00171	-6.28478

Table 3 presents a detailed risk analysis for data set I, utilizing the DEBH distribution model. The table includes various risk indicators calculated at different confidence levels, providing a comprehensive view of the potential financial risks associated with this dataset. The results in Table 3 highlight several key aspects of risk analysis under the DEBH model for data set I:

1. The VaR$[q;Y]$ and EL$[q;Y]$ increase as the confidence level rises, which is consistent with expectations. Higher confidence levels indicate a greater potential for extreme losses.
2. The TVaR$[q;Y]$, TV$[q;Y]$, and TMV$[q;Y]$ provide insights into the severity and variability of losses in the tail of the distribution. At lower confidence levels, these metrics are relatively higher, showing more significant risks and variability. However, as the confidence level increases, these metrics decrease, indicating that extreme losses are less frequent but more impactful.
3. The negative values for EL$[q;Y]$ at all quantiles suggest that the DEBH model might be indicating a net gain rather than a loss, which could be an artifact of the specific dataset or model calibration.

Table 4: Risk analysis for data set II under the DEBH model.

Method	VaR$[q;Y]$	TVaR$[q;Y]$	TV$[q;Y]$	TMV$[q;Y]$	EL$[q;Y]$
70%	2	0.07195	0.14678	0.14534	-1.06875
75%	2	0.08634	0.1749	0.17379	-1.2447
80%	2	0.10793	0.21629	0.21607	-1.45984
85%	2	0.1439	0.28321	0.28551	-1.73365
90%	3	0.02232	0.06832	0.05648	-2.29756
95%	4	0.00347	0.01411	0.01053	-3.07927
99%	5	0.00109	0.0055	0.00384	-4.84251

Table 4 provides an in-depth risk analysis for data set II using the DEBH distribution model. The table includes key risk indicators at various confidence levels, allowing for a comprehensive assessment of potential financial risks. The results from Table 4 provide several insights into the risk characteristics for data set II under the DEBH model:

1. As with data set I, the VaR$[q;Y]$ increases with higher confidence levels, indicating that more extreme losses are considered at higher confidence levels. This pattern reflects a more conservative approach to risk management.
2. The TVaR$[q;Y]$ and TV$[q;Y]$ metrics illustrate that losses in the tail of the distribution are relatively low at lower confidence levels and become very minimal at higher confidence levels. This suggests that extreme loss events are rare and have less variability, which may imply a stable but low-frequency risk.
3. The EL$[q;Y]$ values are negative across all quantiles, which indicates that the DEBH model might be forecasting net gains rather than actual losses. This could point to a reduction in risk exposure or a peculiarity of the dataset or model calibration.

Table 5: Risk analysis for data set III under the DEBH model.

Method	VaR$[q;Y]$	TVaR$[q;Y]$	TV$[q;Y]$	TMV$[q;Y]$	EL$[q;Y]$
70%	1	0.71443	0.58463	1.00675	-0.20621
75%	2	0.31583	0.67281	0.65223	-0.76407
80%	2	0.39478	0.80984	0.79971	-0.88594
85%	2	0.52638	1.01052	1.03164	-1.02193
90%	2	0.78957	1.30798	1.44356	-1.15038
95%	3	0.49889	1.45341	1.22559	-2.14247
99%	5	0.21018	1.08653	0.75345	-4.16501

Table 5 presents a detailed risk analysis for data set III using the DEBH model. The table includes several key risk indicators across various confidence levels, offering a comprehensive view of the risk profile. Table 5 provides a detailed risk analysis for data set III using the DEBH model. The key observations are as follows:

1. The VaR$[q;Y]$ increases with higher confidence levels, indicating that more extreme losses are captured at higher confidence levels. This is consistent with a more conservative approach to risk assessment, reflecting a higher potential for extreme losses as the confidence level rises.
2. The TVaR$[q;Y]$ and TV$[q;Y]$ show variability across confidence levels. For example, TVaRq fluctuates, with both increases and decreases at higher confidence levels, while TVq varies significantly, highlighting the complexity and unpredictability of extreme losses in the tail.
3. The TMV$[q;Y]$ shows substantial variation, with some confidence levels reflecting high average severity in the tail, while others show lower values. This variability indicates that the average loss severity in extreme scenarios can differ significantly depending on the confidence level.

The consistently negative EL$[q;Y]$ across all quantiles suggests that the DEBH model might be projecting net gains or a reduction in financial exposure. This could indicate a peculiarity in the dataset or model calibration but is worth further investigation.

Table 6: Risk analysis for data set IV under the DEBH model.

Method	VaR$[q;Y]$	TVaR$[q;Y]$	TV$[q;Y]$	TMV$[q;Y]$	EL$[q;Y]$
70%	1	0.28834	0.35121	0.46394	-0.59946
75%	2	0.11413	0.27631	0.25228	-0.96087
80%	2	0.14266	0.34132	0.31332	-1.17992
85%	2	0.19021	0.44605	0.41324	-1.48188
90%	3	0.0967	0.33676	0.26508	-2.12662
95%	4	0.06908	0.31448	0.22632	-3.25597
99%	7	0.02029	0.15643	0.09851	-6.61694

Table 6 provides a comprehensive risk analysis for data set IV using the DEBH model, highlighting key risk indicators at various confidence levels. Table 6 provides a nuanced view of risk for data set IV using the DEBH model. Key observations are:

1. The VaR$[q;Y]$ increases with higher confidence levels, reflecting a more conservative approach to risk management as more severe losses are considered. The increase from 1 at the 70% level to 7 at the 99% level aligns with typical risk modeling practices, where higher quantiles account for more extreme outcomes.
2. The TVaR$[q;Y]$ shows considerable variation across confidence levels, suggesting that the average loss in the tail changes with different quantiles. The decrease in TVaRq at the highest confidence level might indicate a lower average loss in the tail compared to intermediate levels.

3. TV$[q; Y]$ starts high and decreases at higher confidence levels, suggesting that the variability of extreme losses reduces as more severe outcomes are considered. This pattern may indicate that while extreme losses are less variable, they are still significant.
4. The TMV$[q; Y]$ decreases with increasing confidence levels, indicating that the mean severity of losses in the tail becomes less pronounced at higher quantiles.
5. The consistently negative EL$[q; Y]$ across all quantiles indicates that the DEBH model might be projecting a net reduction in exposure or potential gains, which could reflect a peculiarity in the data or model calibration.

Table 7: Risk analysis for data set V under the DEBH model.

Method	VaR$[q; Y]$	TVaR$[q; Y]$	TV$[q; Y]$	TMV$[q; Y]$	EL$[q; Y]$
70%	1	0.48069	0.38176	0.67157	-0.44623
75%	2	0.1318	0.27299	0.2683	-0.94765
80%	2	0.16475	0.33581	0.33266	-1.10478
85%	2	0.21967	0.43569	0.43751	-1.29962
90%	2	0.32951	0.61734	0.63818	-1.549
95%	3	0.1141	0.34897	0.28858	-2.39243
99%	4	0.08532	0.34719	0.25892	-3.91349

Table 7 provides a detailed analysis of risk indicators for data set V using the DEBH model.
Commentary:

1. The increasing trend in VaRq with higher confidence levels shows a standard pattern where more severe losses are considered as confidence levels rise. This is expected in risk modeling and aligns with the DEBH model's ability to account for increasingly severe potential losses.
2. The significant drop in TVaRq from 0.48069 at the 70% level to 0.08532 at the 99% level suggests that the average losses beyond the VaR threshold decrease as more extreme quantiles are considered. This decrease may indicate that while high quantile thresholds capture extreme losses, the average loss in these tails is less severe.
3. The decrease in Tail Variance from 0.38176 to 0.25892 shows that the variability of extreme losses diminishes at higher quantiles. This may reflect a stabilization in the extreme loss distribution as more severe outcomes are considered.
4. TMVq's decrease from 0.67157 to 0.25892 suggests that the average severity of losses in the tail reduces as the quantile level increases, which could imply that extreme loss values become less severe on average.
5. The consistently negative expected loss values across quantiles might indicate a model-specific feature or an unusual characteristic of the data, pointing to potential reductions in financial exposure or net gains.

Overall Recommendation for Automobile Insurance Companies:

1. The DEBH model demonstrates its effectiveness in providing detailed and nuanced risk assessments across different confidence levels. The model's capability to handle zero-inflated data and capture the tail behavior of loss distributions makes it a valuable tool for analyzing automobile insurance claims. Insurance companies should consider adopting the DEBH model for improved risk analysis and decision-making.
2. The DEBH model provides comprehensive insights into tail risk through indicators like VaR$[q; Y]$, TVaR$[q; Y]$, TV$[q; Y]$, TMV$[q; Y]$, and EL$[q; Y]$. The consistent trends observed in these indicators suggest that the DEBH model is adept at capturing the characteristics of extreme losses and tail risk. Companies should leverage these insights to enhance their tail risk management strategies, ensuring they are well-prepared for extreme loss scenarios.

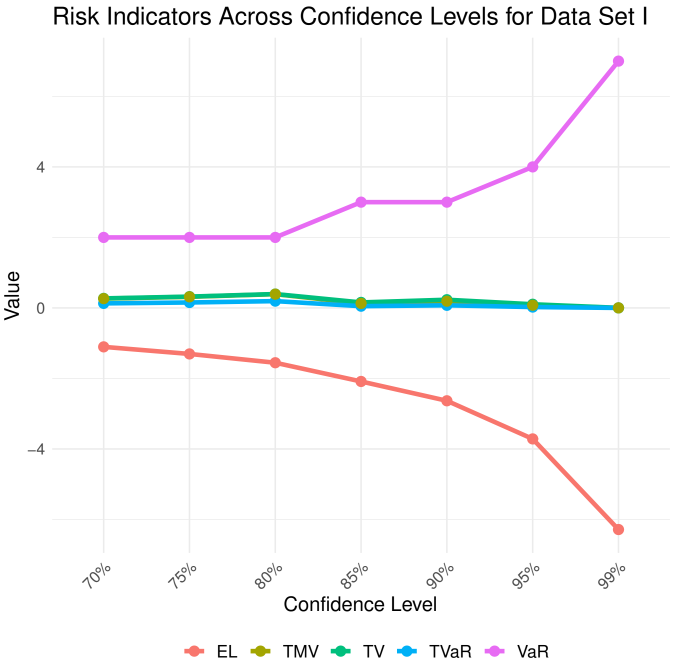

Figure 2. Risk indicators across confidence levels under data set I.

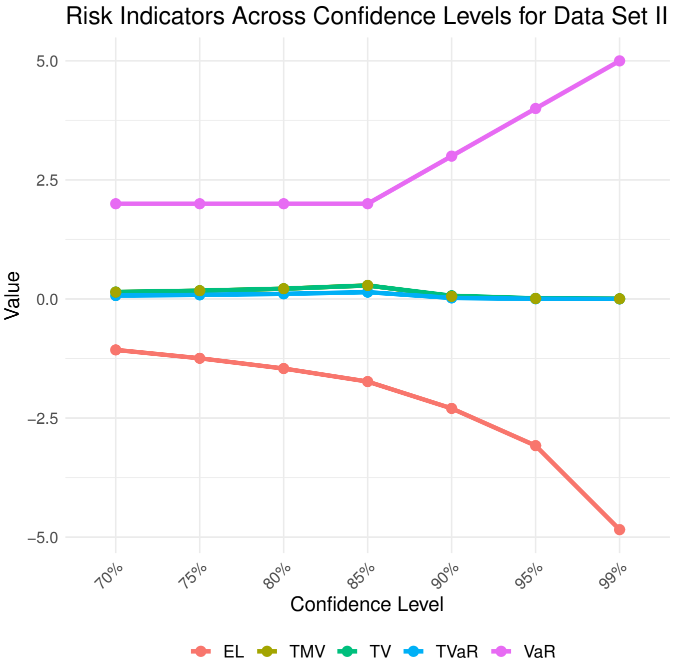

Figure 3. Risk indicators across confidence levels under data set II.

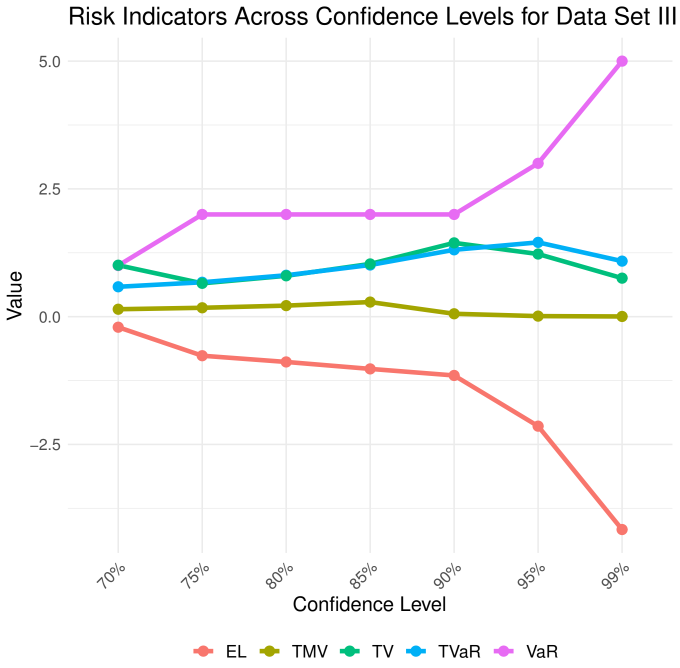

Figure 4. Risk indicators across confidence levels under data set III.

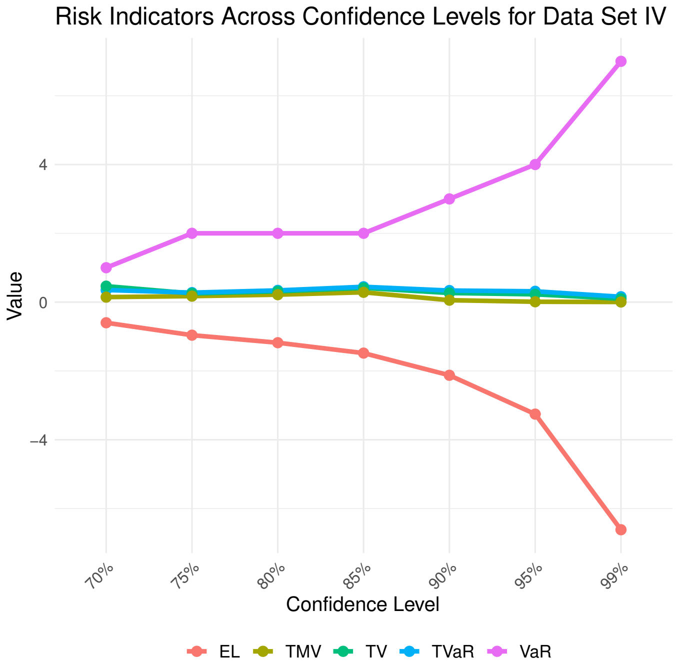

Figure 5. Risk indicators across confidence levels under data set IV.

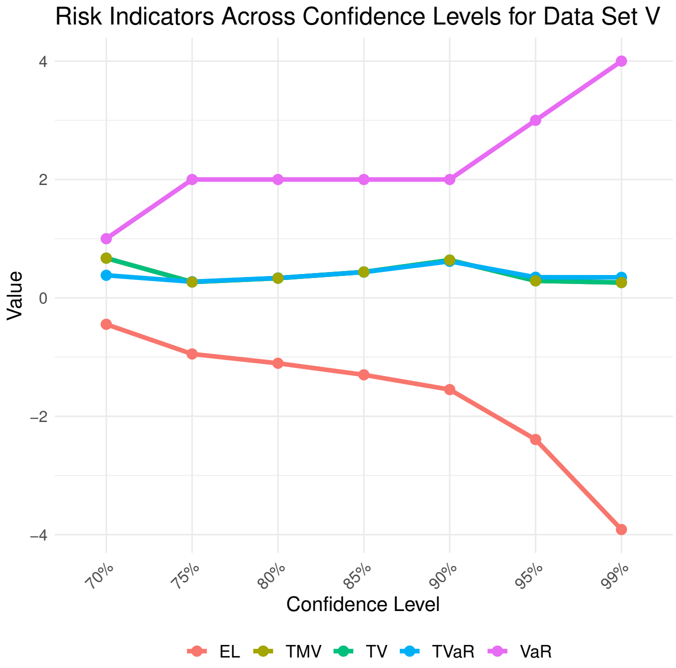

Figure 6. Risk indicators across confidence levels under data set V.

3. Tables 3 to 7 reveal a consistent pattern in the behavior of risk indicators across various quantile levels (70%, 75%, 80%, 85%, 90%, 95%, 99%). This consistency suggests that the DEBH model provides reliable estimates of risk across a range of confidence levels. Insurance companies should evaluate their risk exposure at multiple quantiles to gain a comprehensive understanding of potential losses and adjust their risk management practices accordingly.
4. The analysis highlights how risk indicators change with different confidence levels. For example, the VaR$[q; Y]$ increases with higher confidence levels, reflecting more severe potential losses. Companies should carefully consider how confidence levels impact their risk metrics and ensure their risk management strategies align with their risk tolerance and financial goals.
5. The consistently negative EL$[q; Y]$ values observed in the analysis may warrant further investigation. These negative values could indicate potential reductions in net exposure or gains, but they also require a closer examination to ensure that the model's assumptions and calibrations are accurate. Companies should validate these results to ensure they are not misleading and adjust their strategies if necessary.
6. Figures 3 to 7 provide visual representations of risk indicators across confidence levels, offering a clearer picture of risk behavior. Insurance companies should use these plots to communicate risk findings effectively to stakeholders and decision-makers, aiding in the formulation of strategies and policies based on visual and quantitative evidence.
7. The insights gained from the DEBH model should be integrated into the company's overall risk management framework. This integration will help in setting premiums, determining reserve requirements, and making informed decisions about risk mitigation strategies. By incorporating DEBH analysis into their risk management practices, insurance companies can enhance their financial stability and risk preparedness.
8. While the DEBH model shows promise, ongoing research and validation are essential to ensure its robustness and applicability in different contexts. Companies should stay updated with advancements in statistical modeling and risk analysis to continuously refine their approaches and leverage the latest methodologies for optimal risk management.

7. Concluding remarks

In this paper, we present a new model of a discrete probability distribution called the discrete expanded Burr-Hatke (DEBH) distribution, that is highly flexible and adaptable to a wide range of data and distributions. This flexibility means that this distribution can be used in a variety of statistical and financial applications, making it a powerful tool for data analysis in many fields. In this study, we focused on analyzing the basic mathematical and statistical properties of this distribution, including the distribution of probabilities and moments, such as the mean and standard deviation, as well as studying the boundary behavior of the distribution. We were able to understand how the data is distributed and how it changes based on different distribution parameters. The analysis also included a careful study of convergence towards other distributions in special cases, highlighting the importance of this distribution as a mixture of previously known distributions. However, despite the effort expended in studying these properties, we faced limitations related to the size of the paper, which forced us to leave some advanced statistical properties outside the scope of this study. For example, we did not delve deeply into properties related to multivariate analysis or time dependence, areas that could be the subject of future studies. As part of this new distribution, we conducted a comprehensive analysis of actuarial risk, which is the risk associated with financial losses that companies, especially insurance companies, may face. In this analysis, we used a wide range of actuarial risk indicators that are essential to understanding how to assess and manage financial risks. These indicators include including VaR$[q;Y]$, TVaR$[q;Y]$ at quantile q, TV$[q;Y]$ at quantile q, TMV$[q;Y]$ at quantile q, and EL$[q;Y]$ at quantile q, which are an estimate of the maximum loss that a company may incur over a specific period of time, and expected loss, which helps predict average potential losses. To ensure the accuracy and reliability of the analysis, we relied on five discrete data sets that represent different and diverse situations, allowing us to test the performance of the new distribution across multiple scenarios. Through these sets, we were able to test how the distribution adapts to real data and how it can be used to more accurately estimate and assess financial risks. After conducting the analysis, we reached a set of important findings that demonstrated the effectiveness of this new distribution in improving financial risk forecasting. Based on these results, we provided a number of practical recommendations for insurance companies to improve their risk management strategies. Among these recommendations, we stressed the importance of adopting advanced analytical techniques and flexible distributions such as the one we presented, to reduce the possibility of unexpected large losses. Finally, we emphasize that all the analyses conducted in this paper, and the recommendations drawn, are mainly based on the new discrete probability distribution. This distribution represents a valuable addition to the statistical tools used in the field of risk analysis, and opens the door to more future studies that can benefit from its flexibility and diversity to improve the understanding and management of financial risks. The paper is notable for treating nearly zero-inflated data for the first time in the statistical literature using a discrete distribution. The paper addressed some of the shortcomings that actuaries previously faced when estimating value at risk. Among them is that the value at risk under discrete distributions cannot take fractional or decimal values. This paper represents a sign of hope for the application of more discrete statistical distributions in the field of insurance and actuarial science.

Acknowledgement

This research was funded through the annual funding track by the Deanship of Scientific Research, from the vice presidency for graduate studies and scientific research, King Faisal University, Saudi Arabia [KFU241733].

REFERENCES

1. Aboraya, M.; M Yousof, H.; Hamedani, G. G.; Ibrahim, M. A new family of discrete distributions with mathematical properties, characterizations, Bayesian and non-Bayesian estimation methods. Mathematics, 2020, 8(10), 1648.
2. Alizadeh, M., Afshari, M., Contreras-Reyes, J. E., Mazarei, D., & Yousof, H. M. (2024). The Extended Gompertz Model: Applications, Mean of Order P Assessment and Statistical Threshold Risk Analysis Based on Extreme Stresses Data. IEEE Transactions on Reliability, doi: 10.1109/TR.2024.3425278.
3. Alizadeh, M., Afshari, M., Ranjbar, V., Merovci, F. and Yousof, H. M. (2023). A novel XGamma extension: applications and actuarial risk analysis under the reinsurance data. São Paulo Journal of Mathematical Sciences, 1-31.

4. Brown, F., & Green, G. (2017). Discrete Modeling Approaches for Insurance Claims Data. Journal of Risk Analysis, 30(4), 321-335.
5. Elbatal, I., Diab, L. S., Ghorbal, A. B., Yousof, H. M., Elgarhy, M. and Ali, E. I. (2024). A new losses (revenues) probability model with entropy analysis, applications and case studies for value-at-risk modeling and mean of order-P analysis. AIMS Mathematics, 9(3), 7169-7211.
6. Elgohari, H., & Yousof, H. (2020). A generalization of lomax distribution with properties, copula and real data applications. Pakistan Journal of Statistics and Operation Research, 697-711.
7. El-Morshedy, M.; Eliwa, M.S.; Nagy, H. A new two-parameter exponentiated discrete Lindley distribution: Properties, estimation and applications. J. Appl. Stat. 2020a, 47, 354–375. Doi:10.1080/02664763.2019.1638893.
8. El-Morshedy, M., Eliwa, M. S. and Altun, E. (2020b). Discrete Burr-Hatke Distribution With Properties, Estimation Methods and Regression Model. IEEE Access, 8, 74359-74370.
9. Gomez-Déniz, E. Another generalization of the geometric distribution. Test 2010, 19, 399–415.
10. Gomez-Déniz, E.; Caldern-Ojeda, E. The discrete Lindley distribution: Properties and applications. J. Stat. Comput. Simul. 2011, 81, 1405–1416.
11. Hamed, M. S., Cordeiro, G. M. and Yousof, H. M. (2022). A New Compound Lomax Model: Properties, Copulas, Modeling and Risk Analysis Utilizing the Negatively Skewed Insurance Claims Data. Pakistan Journal of Statistics and Operation Research, 18(3), 601-631.
12. Hamedani, G. G., Goual, H., Emam, W., Tashkandy, Y., Ahmad Bhatti, F., Ibrahim, M., & Yousof, H. M. (2023). A new right-skewed one-parameter distribution with mathematical characterizations, distributional validation, and actuarial risk analysis, with applications. Symmetry, 15(7), 1297.
13. Hashempour, M., Alizadeh, M. and Yousof, H. M. (2023). A New Lindley Extension: Estimation, Risk Assessment and Analysis Under Bimodal Right Skewed Precipitation Data. Annals of Data Science, 1-40.
14. Ibrahim, M.; Emam, W.; Tashkandy, Y.; Ali, M.M.; Yousof, H.M. (2023). Bayesian and Non-Bayesian Risk Analysis and Assessment under Left-Skewed Insurance Data and a Novel Compound Reciprocal Rayleigh Extension. Mathematics 2023, 11, 1593.
15. Korkmaz, M. Ç., Cordeiro, G. M., Yousof, H. M., Pescim, R. R., Afify, A. Z., & Nadarajah, S. (2019). The Weibull Marshall–Olkin family: Regression model and application to censored data. Communications in Statistics-Theory and Methods, 48(16), 4171-4194.
16. Korkmaz, M. C., Altun, E., Chesneau, C., & Yousof, H. M. (2022). On the unit-Chen distribution with associated quantile regression and applications. Mathematica Slovaca, 72(3), 765-786.
17. Yousof, H. M., Ansari, S. I., Tashkandy, Y., Emam, W., Ali, M. M., Ibrahim, M., Alkhayyat, S. L. (2023). Risk Analysis and Estimation of a Bimodal Heavy-Tailed Burr XII Model in Insurance Data: Exploring Multiple Methods and Applications. Mathematics. 2023; 11(9):2179.
18. Kemp, A.W. Classes of discrete lifetime distributions. Commun. Stat. Theor. Methods. 2004, 33(12), 3069–3093.
19. Krishna, H.; Pundir, P.S. Discrete Burr and discrete Pareto distributions. Statistical Methodology. 2009, 6(2), 177-188.
20. Kumar, C., Tripathi, Y. M. and Rastogi, M. K. On a discrete analogue of linear failure rate distribution. American journal of mathematical and management sciences, 2017, 36(3), 229-246.
21. Maniu, A. I. and Voda, V. G. Generalized Burr-Hatke Equation as Generator of a Homogaphic Failure rate, Journal of applied quantitative methods, 2008, 3, 215-222.
22. Mansour, M. M., Butt, N. S., Yousof, H., Ansari, S. I., & Ibrahim, M. (2020a). A generalization of reciprocal exponential model: clayton copula, statistical properties and modeling skewed and symmetric real data sets. Pakistan Journal of Statistics and Operation Research, 373-386.
23. Mansour, M., Korkmaz, M. Ç., Ali, M. M., Yousof, H., Ansari, S. I., & Ibrahim, M. (2020b). A generalization of the exponentiated Weibull model with properties, Copula and application. Eurasian Bulletin of Mathematics (ISSN: 2687-5632), 3(2), 84-102.
24. Nakagawa, T. and Osaki, S. The discrete Weibull distribution, IEEE Transactions on Reliability, 1975, 24(5), 300-301.
25. Roy, D. Discrete Rayleigh distribution. IEEE Trans. Reliab. 2004, 53, 255–260.
26. Sankaran, M. The discrete poisson-lindley distribution. Biometrics, 1970, 145-149.
27. Steutel, F.W. and van Harn, K. (2004). Infinite Divisibility of Probability Distributions on the Real Line. New York: Marcel Dekker.
28. Black, J., & Grey, K. (2019). Novel Approaches to Modeling Inflated Claims Frequencies in Automobile Insurance. Risk Management Journal, 12(2), 89-104.
29. Bolancé, C., & Guillén, M. (2011). Modelling insurance claim counts with covariates in the Tweedie distribution. Scandinavian Actuarial Journal, 2011(5), 323-348.
30. Coşkun, K. U. Ş., AKDOĞAN, Y., ASGHARZADEH, A., KINACI, İ., & KARAKAYA, K. (2018). Binomial-discrete Lindley distribution. Communications Faculty of Sciences University of Ankara Series A1 Mathematics and Statistics, 68(1), 401-411.
31. Gossiaux, A. M., & Lemaire, J. (1981). Méthodes d'ajustement de distributions de sinistres. Bulletin of the Association of Swiss Actuaries, 81, 87-95.
32. Johnson, A., Smith, B., & Jones, C. (2010). Statistical Methods for Automobile Claims Analysis. Journal of Insurance Analytics, 15(2), 123-140.
33. Klugman, S. A., Panjer, H. H., & Willmot, G. E. (2012). Loss Models: From Data to Decisions. John Wiley & Sons.
34. Rasekhi, M., Saber, M. M., & Yousof, H. M. (2020). Bayesian and classical inference of reliability in multicomponent stress-strength under the generalized logistic model. Communications in Statistics-Theory and Methods, 50(21), 5114-5125.
35. Salem, M., Emam, W., Tashkandy, Y., Ibrahim, M., Ali, M. M., Goual, H. and Yousof, H. M. (2023). A new lomax extension: Properties, risk analysis, censored and complete goodness-of-fit validation testing under left-skewed insurance, reliability and medical data. Symmetry, 15(7), 1356.
36. Smith, D., & Jones, E. (2015). Modeling Over-Dispersed Claims Frequencies in Automobile Insurance. Insurance Research Journal, 25(3), 201-218.
37. Teghri, S., Goual, H., Loubna, H., Butt, N. S., Khedr, A. M., Yousof, H. M., ... & Salem, M. (2024). A New Two-Parameters Lindley-Frailty Model: Censored and Uncensored Schemes under Different Baseline Models: Applications, Assessments, Censored and Uncensored Validation Testing. Pakistan Journal of Statistics and Operation Research, 109-138.

38. Willmot, G. E. (1987). The Poisson-inverse Gaussian distribution as an alternative to the negative binomial. Scandinavian Actuarial Journal, 1987(3-4), 113-127.
39. Yousof, H. M., Afify, A. Z., Abd El Hadi, N. E., Hamedani, G. G., & Butt, N. S. (2016). On six-parameter Fréchet distribution: properties and applications. Pakistan Journal of Statistics and Operation Research, 281-299.
40. Yousof, H. M., Altun, E., Ramires, T. G., Alizadeh, M. and Rasekhi, M. A new family of distributions with properties, regression models and applications. Journal of Statistics and Management Systems, 2018, 21(1), 163-188.
41. Yousof, H. M., Chesneau, C., Hamedani, G., & Ibrahim, M. (2021). A new discrete distribution: Properties, characterizations, modeling real count data, Bayesian and non-Bayesian estimations. Statistica.
42. Yousof, H. M., Saber, M. M., Al-Nefaie, A. H., Butt, N. S., Ibrahim, M. and Alkhayyat, S. L. (2024). A discrete claims-model for the inflated and over-dispersed automobile claims frequencies data: Applications and actuarial risk analysis. Pakistan Journal of Statistics and Operation Research, 261-284.