Academic Journal of Mathematical Sciences
ISSN 2616-5805 Vol. 5, Issue 3: 82-90, DOI: 10.25236/AJMS.2024.050312

The Four-Parameter Exponential-Transmuted Exponential Distribution: Properties and Applications

Bin Han

Tianjin University of Commerce, Tianjin, China

Abstract: *In this paper, a new four-parameter exponential-transmuted exponential (ETE) distribution is proposed based on the T-X transformation method, which extends and improves the existing ETE distribution by introducing an additional shape parameter. Several important properties of the new distribution are systematically explored in the paper, including the construction of moments, mother-of-moment function and risk function. In order to verify the applicability of the new distribution in real data, the paper uses the great likelihood estimation method to fit the new distribution to the payout data with heavy-tailed characteristics in the U.S. compensation loss dataset. By comparing with the classical models such as the original three-parameter ETE distribution, the Weibull exponential (WE) distribution and the exponential (E) distribution, the results show that the newly proposed ETE distribution outperforms other models in terms of fitting effect and information criterion, which further validates its robustness and applicability in complex data processing.*

Keywords: *T-X Distribution, ETE Distribution, Heavy Tailed Data, Maximum Likelihood Estimate*

1. Introduction

In statistical modelling and data analysis, classical probability distributions such as normal, t and exponential distributions have long been important tools in research and practice due to their theoretical simplicity and wide range of applications. However, these traditional distributions often face certain limitations when dealing with complex data in the real world. For example, the distribution of asset returns in financial markets tends to be thick-tailed, while environmental data may exhibit asymmetry and heteroskedasticity, which make it difficult for standard distributions to effectively capture the true characteristics of the data. To solve this problem, the T-X family of distributions was created. The T-X distribution introduces greater flexibility by introducing transformations or extensions to the classical distribution, allowing it to accommodate a wider range of data characteristics. In recent years, T-X distributions have been applied to a variety of fields including risk management, engineering reliability, and life sciences. In particular, the exponential-transmuted exponential (ETE) distribution in the family of T-X distributions has received a lot of attention due to its great flexibility and applicability. However, the existing form of T-X distribution still exhibits shortcomings in dealing with some specific data patterns, especially when dealing with extreme asymmetric or multi-peaked distributions, and its flexibility is still limited. In view of this, a new form of ETE distribution is proposed in this paper, which further extends the scope of application of ETE distribution by introducing a new transformation mechanism. The new distribution not only has greater theoretical flexibility, but also shows stronger adaptability when dealing with complex data [1-2].

2. Literature Review

Alzaatreh et al. (2013) proposed a method for generating a family of continuous distributions, called T-X distributions, which set a precedent for later scholars to study families of T-X distributions. Tahir et al. (2016) defined a special lifetime model called the Poisson power-Cauchy based on the T-X distribution and investigated some of its properties with flexible hazard rate shapes such as increasing, decreasing, bathtub and upside-down bathtub. Moolath et al. (2017) investigated a special case of the T-X family of distributions: exponential-transmuted exponential (ETE) distribution, and demonstrated the effectiveness and flexibility of this distribution in modelling and predicting lifetime data by analysing two real lifetime datasets. Aslam et al. (2020) proposed an improved family of T-X distributions and discussed the estimation of the parameters in the framework of classical methods and Bayesian

ISSN 2616-5805 Vol. 5, Issue 3: 82-90, DOI: 10.25236/AJMS.2024.050312

framework, and verified through empirical analyses that the proposed new family of distributions has better flexibility and can be applied to the study of survival analysis and reliability theory. Ahmad et al. (2021) proposed exponential T-X (ETX) family, on the basis of which a new extension of the Weibull model was made to make it more flexible in modelling heavy-tailed data. Baharith and Aljuhani (2021) obtained an alpha power Weibull–exponential distribution based on the T-X distribution and the alpha power transformation approaches, which is highly adaptable for time data and material property data. Haiyue Wang (2021) constructed two families of generalised exponential Weibull distributions and investigated their statistical properties based on two methods, T-X transformation and cubic transformation, using the exponential Weibull distribution as the base distribution. Shah et al. (2022) proposed a member of the T-X family, which contains heavy-tailed distributions and is referred to as "a new exponential-X family of distribution". Kamal et al. (2023) introduced generalised exponential-U family of distributions as a new approach to improve the flexibility of existing classical and modified distributions. Their combination of the T-X family approach with an exponential model yields a new distribution family, which has wider and more extensive applications in the field of prediction and modelling of healthcare phenomena[3-5].

3. Exponential-Transmuted Exponential (ETE) Distribution Based on T-X Distribution

3.1. T-X distribution

Alzaatreh (2013) defines the cumulative distribution function (cdf) of the T-X family as

$$J(x)=\int_a^{W(F(x))} r(t)dt, \tag{1}$$

where $F(x)$ is the distribution function of any random variable $X\in R$, $W(F(x))$ is a function of the distribution function $F(x)$; $r(t)$ is the probability density function of the random variable $T\in[a,b]$. At the same time to meet the following conditions:

(1) $W(F(x))\in[a,b]$,

(2) $W(F(x))$ is absolutely continuous and monotonically non-decreasing,

(3) $W(F(x))\to a$ as $x\to-\infty$ and $W(F(x))\to b$ as $x\to\infty$.

3.2. Exponential-Transmuted Exponential (ETE) Distribution

In particular, in the T-X family of distributions, let $W(F(x))=-\ln[1-F(x)]$, and $F(x)$ is defined by the quadratic rank transmutation map (QRTM) approach of Shaw and Buckley (2009) as

$$F(x)=(1+\lambda)G(x)-\lambda[G(x)]^2,\ |\lambda|\le 1, \tag{2}$$

where $G(x)$ is the baseline distribution. At this point the distribution function of the family of T-X distributions is

$$J(x)=\int_0^{-\ln[\overline{G}(x)[1-\lambda G(x)]]} r(t)dt. \tag{3}$$

Let the distribution functions $G(x)$ and $R(t)$ both be exponentially distributed, i.e., $G(x)=1-e^{-\beta x}$, $x>0$, $\beta>0$. $R(t)=1-e^{-\theta t}$, $t>0$, $\theta>0$. The exponential-transmuted exponential (ETE) distribution can be obtained with the following distribution function

$$J(x)=1-e^{-\theta\beta x}(1-\lambda+\lambda e^{-\beta x})^{\theta},\ x>0,\theta>0,\beta>0,|\lambda|\le 1, \tag{4}$$

its probability density function (pdf) is

ISSN 2616-5805 Vol. 5, Issue 3: 82-90, DOI: 10.25236/AJMS.2024.050312

$$j(x) = \theta\beta e^{-\theta\beta x}\frac{1-\lambda+2\lambda e^{-\beta x}}{(1-\lambda+\lambda e^{-\beta x})^{1-\theta}},\ x>0, \theta>0, \beta>0, |\lambda|\le 1 \quad (5)$$

For the convenience of the later description, we refer to this ETE distribution as the three-parameter ETE distribution. Figure 1 shows the trend of the probability density function image of the ETE distribution with respect to different parameters.

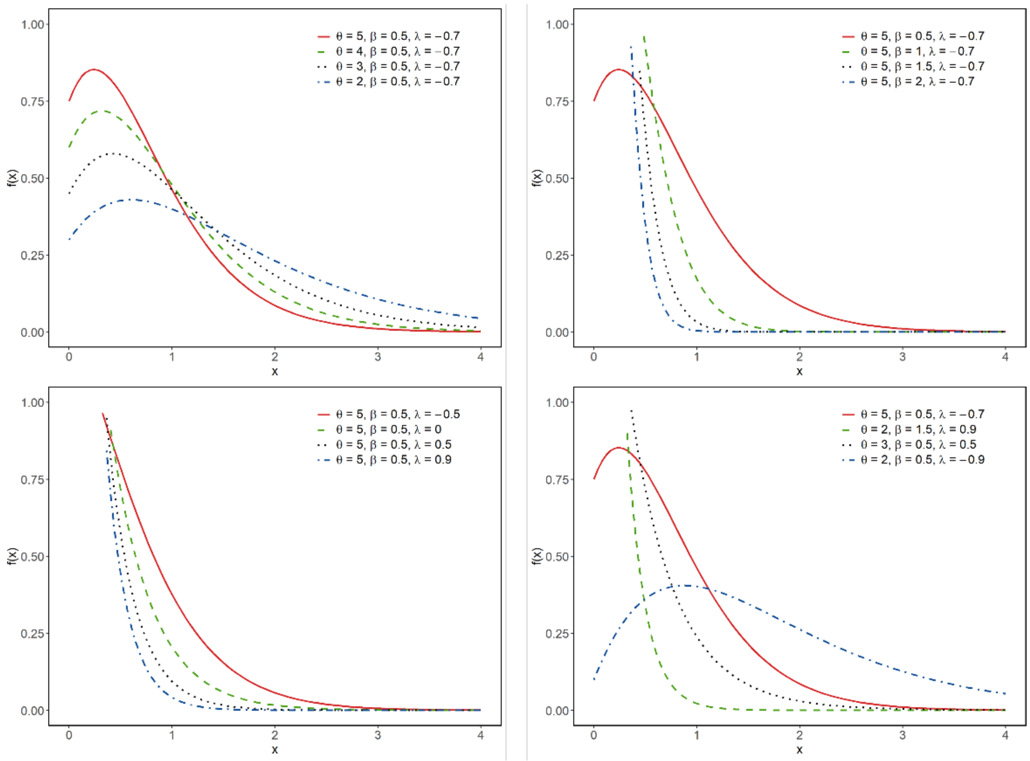

Figure 1: Different plots for the pdf of the three-parameter ETE distribution.

4. Improved ETE Distribution

4.1. Model Construction

The survival function of the distribution function in Equation (2) can be expressed as

$$\overline{F}(x) = 1 - F(x) = \overline{G}(x)[1-\lambda[1-\overline{G}(x)]], \quad (6)$$

In this paper, a shape parameter r is introduced on the basis of this survival function to generate a new distribution, which has greater flexibility and can be effectively adapted to a variety of complex data, the new survival function is given by the following equation

$$\overline{F}(x) = 1 - F(x) = \overline{G}^{r}(x)[1-\lambda[1-\overline{G}(x)]],\ r+\lambda \ge 0. \quad (7)$$

Let $W(F(x)) = -\ln[\overline{F}(x)]$, it is easy to prove that $W(F(x))$ satisfies the three conditions for generating a family of T-X distributions, then then the distribution function of the family of T-X distributions becomes

$$J(x) = \int_0^{-\ln[\overline{G}^{r}(x)[1-\lambda[1-\overline{G}(x)]]]} r(t)dt, \quad (8)$$

Let the distribution functions $G(x)$ and $R(t)$ both be exponentially distributed, i.e., $G(x) = 1 - e^{-\beta x}$, $x>0$, $\beta>0$. $R(t) = 1-e^{-\theta t}$, $t>0$, $\theta>0$. Then the distribution function of the

ISSN 2616-5805 Vol. 5, Issue 3: 82-90, DOI: 10.25236/AJMS.2024.050312

four-parameter ETE distribution can be obtained as

$$J(x) = 1 - e^{-r\theta\beta x}(1 - \lambda + \lambda e^{-\beta x})^{\theta}, \tag{9}$$

where $x \geq 0$, $\theta > 0$, $\beta > 0$, $|\lambda| \leq 1$, $r \geq -\lambda$. Derivation of this distribution function yields the pdf of the four-parameter ETE distribution as

$$j(x) = \theta\beta e^{-r\theta\beta x} \frac{r(1-\lambda) + \lambda e^{-\beta x}(r+1)}{(1 - \lambda + \lambda e^{-\beta x})^{1-\theta}}. \tag{10}$$

Figure 2 shows the image of the probability density function of this distribution for different parameters.

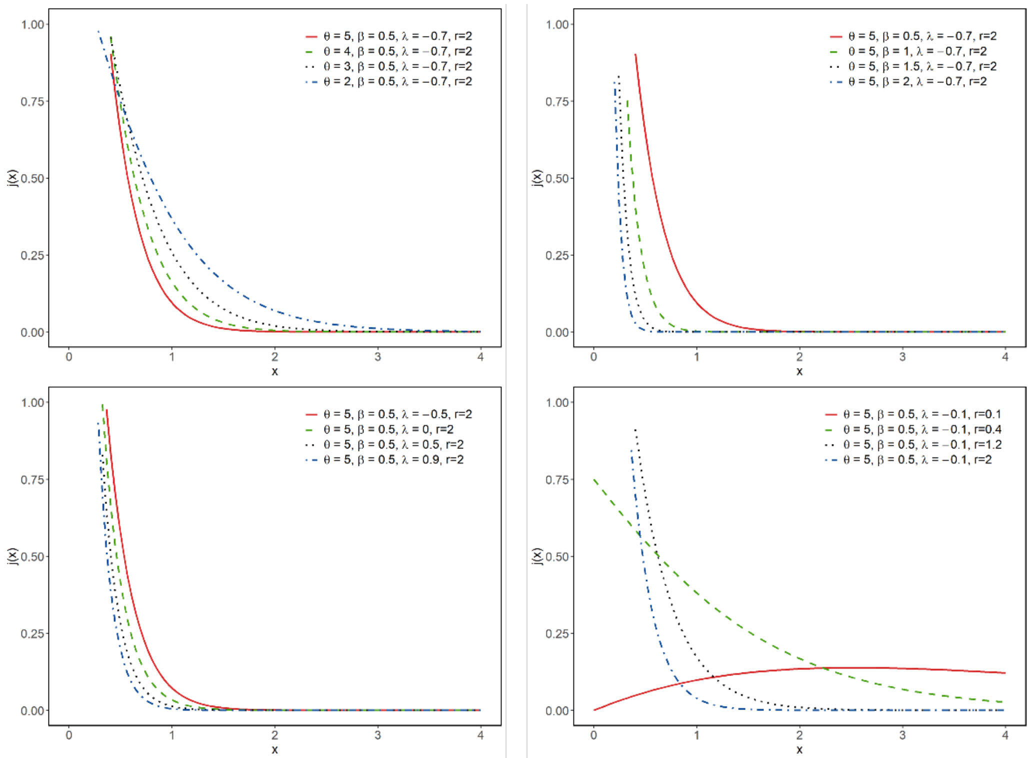

Figure 2: Different plots for the pdf of the four-parameter ETE distribution.

To simplify the derivation of the relevant properties later, the cdf of the new ETE distribution can be obtained with the help of the generalised binomial distribution theorem as

$$J(x) = 1 - \sum_{k=0}^{\infty} S_k(\theta, \lambda) e^{-(r\theta + k)\beta x}, \tag{11}$$

where $S_k(\theta, \lambda) = \sum_{i=k}^{\infty} (-1)^{i+k} \binom{\theta}{i} \binom{i}{k} \lambda^i$, the derivation process is

ISSN 2616-5805 Vol. 5, Issue 3: 82-90, DOI: 10.25236/AJMS.2024.050312

$$\begin{aligned} J(x) &= 1-e^{-r\theta\beta x}(1-\lambda(1-e^{-\beta x}))^{\theta} \\ &= 1-e^{-r\theta\beta x}\sum_{i=0}^{\infty}(-1)^{i}\binom{\theta}{i}\lambda^{i}(1-e^{-\beta x})^{i} \\ &= 1-e^{-r\theta\beta x}\sum_{k=0}^{\infty}\sum_{i=k}^{\infty}(-1)^{i+k}\binom{\theta}{i}\binom{i}{k}\lambda^{i}e^{-k\beta x} \\ &= 1-\sum_{k=0}^{\infty}S_k(\theta,\lambda)e^{-(r\theta+k)\beta x} \end{aligned}$$

4.2. Properties of the New ETE Distribution

4.2.1. Moments

Moments are used in probability distributions to quantify their basic characteristics, including centrality (mean), dispersion (variance), symmetry (skewness) and cuspiness (kurtosis), providing an important basis for statistical analysis and modelling.

Theorem 1: The nth order moments of the new ETE distribution as

$$\begin{aligned} M_X(t) &= E(e^{tX}) \\ &= \int_0^{\infty} e^{tx}\sum_{k=0}^{\infty}S_k(\theta,\lambda)(r\theta+k)\beta e^{-(r\theta+k)\beta x}dx \\ &= \sum_{k=0}^{\infty}S_k(\theta,\lambda)(r\theta+k)\beta\frac{1}{(r\theta+k)\beta-t} \end{aligned} \quad . \tag{12}$$

4.2.2. Moment Generating Function

The moment generating function is an important tool to characterise the distribution of random variables, especially in the study of the distribution of random variables and the sum of independently and identically distributed random variables, the moments function has unique advantages.

Theorem 2: The moment generating function of the new ETE distribution as

$$\begin{aligned} M_X(t) &= E(e^{tX}) \\ &= \int_0^{\infty} e^{tx}\sum_{k=0}^{\infty}S_k(\theta,\lambda)(r\theta+k)\beta e^{-(r\theta+k)\beta x}dx \\ &= \sum_{k=0}^{\infty}S_k(\theta,\lambda)(r\theta+k)\beta\frac{1}{(r\theta+k)\beta-t} \end{aligned} \quad . \tag{13}$$

4.2.3. Hazard Rate Function

The hazard rate function is used in statistical decision theory to assess the performance of a decision rule. It measures the expected value of the actual loss or cost under a given decision rule. With the hazard rate function, the advantages and disadvantages of different decision rules can be compared and the rule that minimises the risk can be selected, thus improving the validity and reliability of the decision.

Theorem 2: The hazard rate function of the new ETE distribution as

$$h(x) = \frac{j(x)}{1-F(x)} = \theta\beta\frac{r(1-\lambda)+\lambda(r+1)e^{-\beta x}}{1-\lambda+\lambda e^{-\beta x}} \quad . \tag{14}$$

Figure 3 illustrates the variation of the risk function with different values of the parameters.

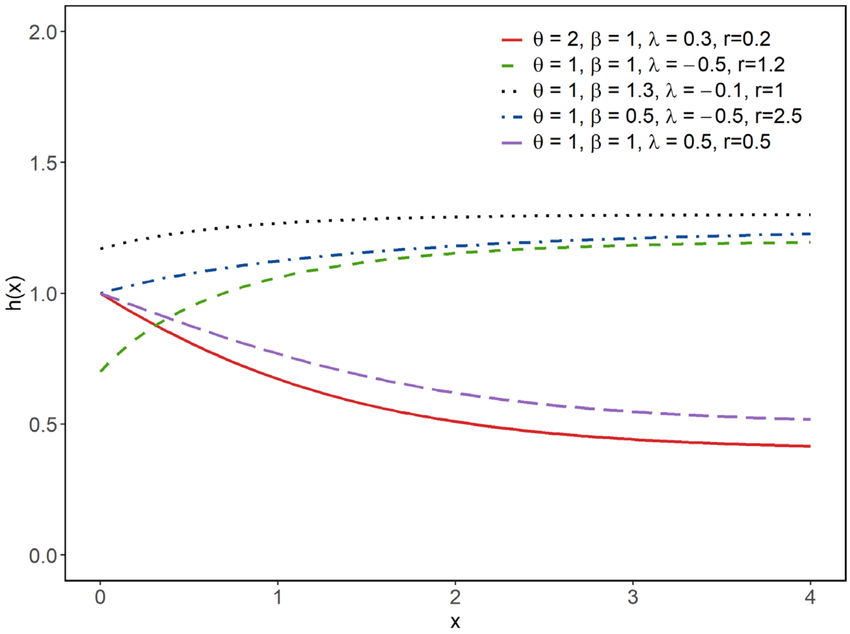

Figure 3: Hazard rate function of the four-parameter ETE distribution for various parameter values

4.2.4. Maximum Likelihood Estimation of the Parameters

The likelihood function is an important tool for estimating model parameters, and by measuring the likelihood of an observation occurring for a given parameter value, it helps us to find the parameter estimates that are most likely to lead to the data, and is used in model evaluation and selection. The likelihood function for the four-parameter ETE distribution is

$$L(x;\theta,\beta,\lambda,r)=(\theta\beta)^n e^{-r\theta\beta\sum_{i=1}^{n}x_i}\prod_{i=1}^{n}(1-\lambda+\lambda e^{-\beta x_i})^{\theta-1}\prod_{i=1}^{n}[r(1-\lambda)+\lambda e^{-\beta x_i}(r+1)] \tag{15}$$

The log likelihood function is

$$ln(L)=n\ln(\theta)+n\ln(\beta)-r\theta\beta\sum_{i=1}^{n}x_i+(\theta-1)\sum_{i=1}^{n}\ln(1-\lambda+\lambda e^{-\beta x_i})$$
$$+\sum_{i=1}^{n}\ln[r(1-\lambda)+\lambda e^{-\beta x_i}(r+1)] \tag{16}$$

Maximising the above equation yields maximum likelihood estimates for θ, β, λ and r with the following derivatives respectively

$$\frac{\partial\ln L}{\partial\theta}=\frac{n}{\theta}-r\beta\sum_{i=1}^{n}x_i+\sum_{i=1}^{n}\ln(1-\lambda+\lambda e^{-\beta x_i}), \tag{17}$$

$$\frac{\partial\ln L}{\partial\beta}=\frac{n}{\beta}-r\theta\sum_{i=1}^{n}x_i-\sum_{i=1}^{n}\frac{(\theta-1)\lambda x_i e^{-\beta x_i}}{1-\lambda+\lambda e^{-\beta x}}-\sum_{i=1}^{n}\frac{\lambda x_i(r+1)e^{-\beta x_i}}{r(1-\lambda)+\lambda e^{-\beta x_i}(r+1)}, \tag{18}$$

$$\frac{\partial\ln L}{\partial\lambda}=(\theta-1)\sum_{i=1}^{n}\frac{e^{-\beta x_i}-1}{1-\lambda+\lambda e^{-\beta x}}-\sum_{i=1}^{n}\frac{(r+1)e^{-\beta x_i}-r}{r(1-\lambda)+\lambda e^{-\beta x_i}(r+1)}, \tag{19}$$

$$\frac{\partial\ln L}{\partial r}=-\theta\beta\sum_{i=1}^{n}x_i+\sum_{i=1}^{n}\frac{1-\lambda+\lambda e^{-\beta x_i}}{r(1-\lambda)+\lambda e^{-\beta x_i}(r+1)}. \tag{20}$$

ISSN 2616-5805 Vol. 5, Issue 3: 82-90, DOI: 10.25236/AJMS.2024.050312

5. Application

In order to evaluate the performance of the four-parameter ETE distribution proposed in this paper, it is compared with the exponential distribution (E), the original ETE distribution and the Weibull exponential (WE) distribution to see how well they fit the heavy-tailed data. The density functions of the exponential (E) distribution and the Weibull exponential distribution (WE) are, respectively:

$$f(x) = \theta e^{-\theta x}, x > 0, \theta > 0 , \tag{21}$$

$$f(x) = \theta\beta\lambda e^{-\lambda x}(e^{-\lambda x} - 1)^{\beta-1} e^{-\alpha(e^{\lambda x}-1)^{\beta}}, x, \theta, \beta, \lambda > 0 . \tag{22}$$

In this paper, we use the payout data from the U.S. Compensation Losses dataset in the R language package, which contains a total of 1,500 pieces of data. The data were reduced by a factor of 1000 for ease of computation, and the parameters of these four distribution models were fitted by maximum likelihood estimation, and the AIC (Akaike Information Criterion), AICC (Amended AIC), BIC (Bayesian Information Criterion), and HQIC (Hannan-Quinn Information Criterion) values were computed separately to fully assess the model fit. The comparison of these criteria will further demonstrate the advantages of the new distribution in complex data processing[6-9].

AIC and AICC focus on the balance between model complexity and goodness-of-fit, with AICC correcting for AIC when the sample is small. bIC has a stricter penalty for model complexity and is suitable for use with large samples, while HQIC falls between AIC and BIC, providing an alternative balance. Smaller values of these metrics indicate that the model performs better in balancing goodness-of-fit and complexity. They are expressed as

$$AIC = -2\ln(L) + 2k , \tag{23}$$

$$AICC = -2\ln(L) + \frac{2kn}{n-k-1} , \tag{24}$$

$$BIC = -2\ln(L) + k\ln(n) , \tag{25}$$

$$HQIC = -2\log L + 2k \cdot \log(\log(n)) . \tag{26}$$

where $\ln(L)$ is the maximum value of the log-likelihood function, k is the number of parameters, and n is the sample size.

Table 1: Descriptive statistics of the loss of compensation dataset

n	Min	Max	Mean	Median	Sd	Kurtosis	Skewness
1500	10	2713595	41208	12000	102747.7	9.15	141.98

The descriptive statistics of this dataset are shown in Table 1, the kurtosis of this dataset is obtained to be 9.15 and skewness to be 141.98, which indicates that the data has obvious heavy-tailed characteristics and extremes, and the heavy-tailed distribution model should be used to describe this dataset. Next, the normal Q-Q plot of this data set is drawn, as shown in Figure 4. If the data conforms to a normal distribution, the sample quartiles should be aligned roughly along the red reference line. It is clear that the data are closer to the theoretical normal distribution in the middle section of the distribution, but significantly deviate from the reference line in the right tail, showing extreme large values much higher than predicted by the normal distribution. This phenomenon indicates that the data has a longer right tail and more outliers in the tail, suggesting that there are more significant extreme values in the sample, again validating the heavy-tailed nature of the data.

ISSN 2616-5805 Vol. 5, Issue 3: 82-90, DOI: 10.25236/AJMS.2024.050312

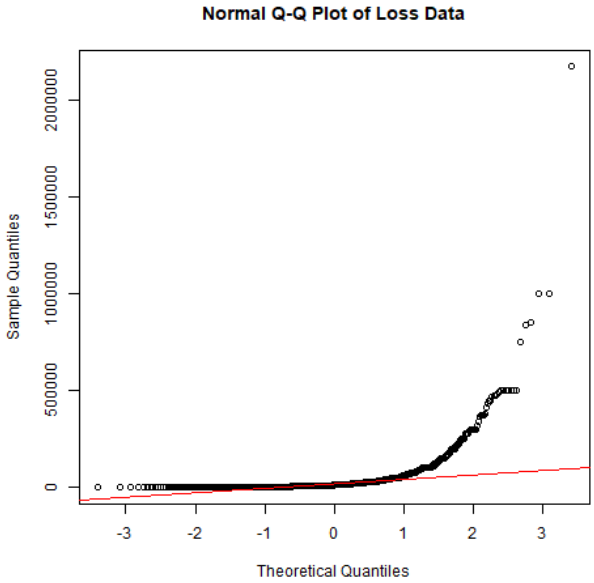

Figure 4: Normal Q-Q plots for the loss of compensation dataset

The results of model parameter estimation are shown in Table 2, and the results of model fitting indexes are shown in Table 3. By comparing the various information criteria (AIC, AICC, BIC, HQIC), Table 1 shows that the values of the newly proposed four-parameter ETE distribution are lower than those of the traditional ETE, WE, and E distributions under all the evaluation criteria, which suggests that it has a better performance and less model complexity in fitting the data.

Table 2: Model parameter estimation results

Model	ML estimates			
	θ	β	λ	r
NEW ETE	48.686	0.037	0.034	0.004
ETE	0.020	0.943	0.999	
WE	73.199	0.626	0.00004	
E	0.024			

Table 3: Results of model fit indicators

Model	$-\ln L$	AIC	AICC	BIC	HQIC
NEW ETE	6579.81	13167.61	13167.64	13188.87	13175.53
ETE	6820.14	13646.27	13646.29	13662.21	13652.21
WE	6660.05	13326.10	13326.11	13342.03	13332.03
E	7077.96	14157.93	14157.93	14163.24	14159.91

6. Conclusion

The empirical evidence shows that the four-parameter ETE distribution performs superiorly with respect to all types of information criteria (AIC, AICC, BIC, HQIC), and the values of all assessment metrics are significantly lower than those of other traditional models. This means that the new model is able to fit the data better when modelling data with heavy-tailed distributions, while maintaining the simplicity of the model and avoiding the problem of overfitting. In particular, when dealing with data with high skewness and kurtosis, the four-parameter ETE distribution demonstrated a greater ability to capture extreme values and data complexity. Future research can explore the potential applications of the model in areas such as risk management, insurance actuarial and financial modelling, and assess its applicability to different types of data to further extend its theoretical and practical impact.

ISSN 2616-5805 Vol. 5, Issue 3: 82-90, DOI: 10.25236/AJMS.2024.050312

References

[1] Alzaatreh A, Lee C, Famoye F. A new method for generating families of continuous distributions[J]. Metron, 2013, 71(1): 63-79.
[2] Tahir M H, Zubair M, Cordeiro G M, et al. The Poisson-X family of distributions[J]. Journal of Statistical Computation and Simulation, 2016, 86(14): 2901-2921.
[3] Aslam M, Asghar Z, Hussain Z, et al. A modified TX family of distributions: classical and Bayesian analysis [J]. Journal of Taibah University for Science, 2020, 14(1): 254-264.
[4] Baharith L A, Aljuhani W H. New method for generating new families of distributions[J]. Symmetry, 2021, 13(4): 726.
[5] Shah Z, Ali A, Hamraz M, et al. A new member of TX family with applications in different sectors[J]. Journal of Mathematics, 2022.
[6] Ahmad Z, Mahmoudi E, Alizadeh M, et al. The exponential TX family of distributions: properties and an application to insurance data[J]. Journal of Mathematics, 2021, 2021: 1-18.
[7] Kamal M, Alsolmi M M, Al Mutairi A, et al. A new distributional approach: estimation, Monte Carlo simulation and applications to the biomedical data sets[J]. Networks and Heterogeneous Media, 2023, 18(4): 1575-1599.
[8] Moolath G B, Jayakumar K. T-transmuted X family of distributions[J]. Statistica, 2017, 77(3): 251-276.
[9] Wang Haiyue. Study on the construction and properties of two-class statistical model based on exponential Weibull distribution [D]. Liaoning Normal University, 2021.