Review Article

Distributional Censored and Uncensored Validation Testing under a Modified Test Statistic with Risk Analysis and Assessment

Yusra Tashkandy, Walid Emam, Gauss M. Cordeiro, M. Masoom Ali, Khaoula Aidi, Haitham M. Yousof, and Mohamed Ibrahim

1 Department of Statistics and Operations Research, Faculty of Science, King Saud University, P.O. Box 2455, Riyadh 11451, Saudi Arabia
2 Universidade Federal de Pernambuco, Departamento de Estatística, Recife, Brazil
3 Department of Mathematical Sciences, Ball State University, Muncie, IN, USA
4 Laboratory of Probability and Statistics LaPS, University Badji Mokhtar, Annaba, Algeria
5 Department of Statistics, Mathematics and Insurance, Benha University, Benha, Egypt
6 Department of Applied, Mathematical and Actuarial Statistics, Faculty of Commerce, Damietta University, Damietta, Egypt

Correspondence should be addressed to Mohamed Ibrahim; mohamed_ibrahim@du.edu.eg

Received 7 January 2023; Revised 28 April 2023; Accepted 26 May 2023; Published 30 June 2023

1. Introduction

A novel continuous distribution will be introduced and studied in this work, but we will approach it from new viewpoints that depart from those often covered by specialists. We will not focus on numerous theoretical findings and algebraic derivations—not because they are unimportant, but rather to give us the chance to highlight more practical features of risk analysis, distributive verification, and their associated applications in both complete and censored data. We will discuss certain theoretical facets of the new distribution, though. But, we will pay particular attention to features that are applicable and useful in the following areas:

(1) A set of frequently used financial indicators, such as value-at-risk (VAR), tail-value-at-risk (TVAR) (also known as conditional tail expectation), conditional-value-at-risk, tail variance (TV), tail mean-variance (TMV), and the mean excess loss (MEL) function, is studied when examining and evaluating the risks that insurance companies face. The maximum likelihood estimation (MLE), the ordinary least squares (OLS), the weighted least squares estimation (WLSE), and the Anderson–Darling estimation (ADE) are all described as estimate strategies for the main key risk indicators (KRIs). These four methodologies are applied in two distinct ways for financial and
actuarial evaluation, including simulation with three confidence levels (CLSs), under different sample sizes, and applications to data from insurance claims.

(2) We present a simulation experiment to compare the performance of the estimators of VAR based on insurance data in order to satisfy the requirements of the actuarial analysis of risks.

(3) The well-known Nikulin–Rao–Robson (NRR) statistic \( Y_a^2 (r - 1) \), which is based on the uncensored maximum likelihood estimators (UMLEs) on initial non-grouped data, is considered under the new Rayleigh generalized gamma (RGG) model in the framework of distributional validation and statistical hypothesis tests for complete data. Three real datasets and a simulation study evaluate the statistic \( Y_a^2 (r - 1) \).

(4) The RGG model considers a modified NRR statistic \( M_a^2 (r) \), which is based on the censored maximum likelihood estimators (CMLEs) on the original non-grouped data. This statistic is used in the framework of distributional validation and statistical hypothesis testing for censored data. The statistic \( M_a^2 (r) \) is evaluated using three real datasets and a thorough simulation analysis.

(5) It is worth noting that risk indicators can be applied in the field of engineering, especially in the field of structural engineering, with the aim of developing mathematical measurement and statistical modelling processes in these fields. In engineering, maximum likelihood estimation can be used to estimate the parameters of a distribution for the failure time of a product or system, which can be used to assess reliability and inform maintenance decisions. Moreover, in engineering, right censored maximum likelihood estimation can be used to estimate the reliability function of a system, such as a machine or a bridge. For example, if the failure time of a machine follows a Weibull distribution, the parameters of the distribution can be estimated using right censored maximum likelihood estimation, and the reliability of the machine can be assessed based on the estimated distribution. Generally, it is important to monitor the system over time and review the risk assessment periodically to ensure that any changes or updates to the design or operating conditions are taken into account. This will help to ensure that the system remains safe and reliable over its lifetime. For more details, see Amini et al. [1] and El-Morshedy et al. [2].

The cumulative distribution function (CDF) of the generalized gamma model [3] can be expressed as

\[
G_\varphi (z) = [1 - (1 + \lambda z) \exp (-\lambda z)]^\theta, \hspace{1cm} (1)
\]

where \( z \geq 0 \), \( \varphi = (\lambda, \theta) \), and \( \lambda, \theta > 0 \), which is flexible enough to accommodate both monotonic and non-monotonic failure rates. Following Yousof et al. [4], the CDF of the RGG model has the following form:

\[
F_\varphi (z) = 1 - \exp \left\{ - \left[ \xi(z)^{-\theta} - 1 \right]^{2} \right\}, \hspace{1cm} (2)
\]

where \( \xi(z) = 1 - (1 + \lambda z) \exp (-\lambda z) \). The probability density function (PDF) corresponding to (2) can be expressed as

\[
f_\varphi (z) = 2\theta \lambda^2 z \exp (-\lambda z) \xi(z)^{2\theta-1} \left[ 1 - \xi(z)^{-\theta} \right]^{-3} \exp \left\{ \left[ \xi(z)^{-\theta} - 1 \right]^{2} \right\}, z > 0. \hspace{1cm} (3)
\]

When comparing probability distributions for applications in insurance, several criteria are commonly considered. These criteria help insurers select the most appropriate distribution to model the claims data accurately. Here are some of the main criteria:1. This criterion assesses how well a probability distribution fits the observed historical claims data. Insurers typically use statistical tests, such as the Kolmogorov-Smirnov test or the chi-square test, to evaluate the goodness of fit. A distribution that closely matches the data is preferred, as it provides more reliable estimates for future claims.2. Skewness measures the asymmetry of a distribution, while kurtosis measures its tail heaviness. In insurance, it is important to consider the skewness and kurtosis of the claims data to capture any non-normal characteristics. Some distributions, such as the lognormal or Pareto distributions, are better suited for modeling skewed and heavy-tailed data commonly observed in insurance claims.3. Probability distributions often have parameters that need to be estimated from the data. Insurers consider the ease and accuracy of parameter estimation methods for a given distribution. Maximum likelihood estimation is a common approach used to estimate parameters, but other methods like the method of moments or Bayesian estimation may also be applicable. A distribution with easily estimable parameters is desirable, as it simplifies the modeling process.4. The interpretability of a probability distribution is crucial for insurance applications. Insurers and actuaries need to understand the underlying assumptions and characteristics of the selected distribution. Distributions like the normal (Gaussian) distribution or the gamma distribution are well-known and have established interpretations, making them popular choices. However, it is important to balance interpretability with the goodness of fit to ensure accurate modeling.5. The tail behavior of a distribution is important for capturing extreme events or catastrophic losses in insurance. Insurers need to assess whether a distribution adequately represents the tail risks inherent in the claims data. Heavy-tailed distributions, such as the Pareto or generalized Pareto distributions, are often considered for modeling extreme events.6. The historical credibility of a probability distribution is based on its track record and applicability to similar insurance portfolios or lines of business. Insurers often rely on industry experience, expert judgment, and historical data from similar risks to assess the suitability of a particular distribution. Distributions that have been successfully used in the past for similar insurance applications may carry more weight in the selection process. It’s important to note that the choice of probability
distribution should be based on a comprehensive analysis of the specific insurance context and the characteristics of the claims data. Insurance companies often employ experienced actuaries and statistical experts to evaluate and compare various distributions, considering these criteria, to make informed decisions about the most appropriate probability distribution for their insurance applications. The test statistic $Y^2_n(r - 1)$ depends on the MLEs on the initial non-grouped real datasets, and the test statistic $Y^2_n(r - 1)$ is of particular importance among all goodness-of-fit tests. The test statistic $Y^2_1(r - 1)$, pioneered by Nikulin [5, 6] and Rao and Robson [7], has a chi-square distribution and recovers information lost during data grouping. However, censoring makes all widely used goodness-of-fit tests worthless and leads to a variety of practical issues. As a result, additional researchers provide a wide range of improvements to the material that was already accessible. Bagdonavicic and Nikulin [8] developed a modified NRR statistic for statistical distributions with right censoring and unknown parameters. Since it recovers all information lost during data regrouping, this variant of the NRR statistic is most likely to fit data from fields where the data are frequently censored, such as in survival analysis, dependability, and others. Following Nikulin [5, 6] and Rao and Robson [7], we will present modified NRR $\chi^2$ goodness-of-fit tests for adjusting the proposed model to full and right censored data.

In case of complete data, the NRR statistic is a well-known alternative to the conventional $\chi^2$ tests. It is founded on variations between two estimates of the chance of falling within grouping intervals. One estimate is based on the empirical distribution function, while the other is based on maximum likelihood calculations employing ungrouped starting data to estimate the unobserved parameters of the tested model. For further information, see Nikulin [5, 6] and Rao and Robson [7].

In general, the development of statistical methods for testing hypotheses and the validity of parametric distributions under censorship is accelerating, although the existence of censorship is considered as a huge challenge. There have been numerous contributions to the field of application in the history of statistical literature for verification tests in the case of controlled data.

The NRR test has been the subject of several research studies in the statistical literature. Because of their rarity, these studies may be tallied; here, we will list the most recent ones. In this work, the RGG distribution is derived in risk analysis and distributional validation, and the uncensored and the right censored scenarios are used to validate a modified $\chi^2$ goodness-of-fit test statistic based on the NRR statistic test ($Y^2_n(r - 1)$) and the modified NRR statistic ($M^2_n(r)$), respectively. The statistic $Y^2_n(r - 1)$ is adopted for testing the null hypothesis $H_0$ according to which a certain complete sample follows the RGG model. The statistic $Y^2_n(r - 1)$ is assessed utilizing a comprehensive simulation study using the Barzilai–Borwein (BB) algorithm for complete data (see Ravi and Gilbert [9]). Then, the statistic $Y^2_n(r - 1)$ is assessed utilizing a comprehensive simulation study using the BB algorithm for censored data. To examine the performance of the test when the sample size is increased, we have relied on the standard mean square error (MSE) in all simulated experiments, accounting for varied sample sizes.

Three uncensored real datasets (uncensored times between failures for repairable items’ data, uncensored reliability data, and uncensored strength data) are used in statistical testing under the statistic $Y^2_n(r - 1)$ for distributional validation. The uncensored real data included in the analysis have a wide reputation in the statistical literature and have had a great deal of analysis. In this work, we will focus on an important aspect of statistical analysis, which is the aspect of hypothesis tests using these data. This most important material is motivated in the introduction, as most of the statistical works did not study and analyze this aspect despite the importance of hypothesis tests in statistical theory. On the other hand, three right censored real datasets are used to assess the statistic $M^2_n(r)$ for distributional validation under the RGG model. The censored real data that have been considered and included in the analysis are the censored bone marrow transplant data, the censored times to infection of kidney dialysis patients data (censored times to infection data), and censored strength of certain type of braided cord data.

The new NRR statistical test showed that using the new model as a stand-in for looking at two right censored datasets is successful. In this context, we shall outline several recent studies that added to or changed the NRR. It is important to note that the browser for statistical literature on this topic (NRR goodness-of-fit test) will not find many new NRR goodness-of-fit extensions and will find few research studies that applied this test because the NRR goodness-of-fit test has specific requirements and strict procedures and demands censored data. As is generally known, obtaining novel censored data to apply to and emphasize the significance of the new test is a difficult duty. In the next few sections, we will discuss few recent research studies that looked at using this test on real data that had been subject to right censoring datasets, along with a description of the conclusions from each study independently.

In this introduction, we do not fail to review some of the limitations of the current study, while we present two basic limitations:

(i) The datasets used must be positive only because the new probability distribution is bound by this constraint and is conditioned by this condition.

(ii) The new modified NRR test introduced in this version can only be applied to data subject to right censored only.

The main novelties of this work can be highlighted as follows:

(i) Employing the new probability distribution in the analysis and evaluation of actuarial risks through a set of actuarial measures. These actuarial measures...
(indicators) have been carefully selected due to their quality in results and popularity in application.

(2) The use of the new probability distribution in modelling processes and statistical hypothesis tests using the NRR test.

(3) Developing the theory of statistical hypothesis testing for controlled data by presenting a modified NRR test and applying it to real data.

(4) Applying the modified NRR test in distributional validations using the new distribution.

(5) Evaluating the performance of several different estimation methods in risk analysis processes. This approach to risk assessment using different estimation methods is considered a recent approach, and many studies are not available.

This paper is distinguished from these competing papers with the following points:

(i) Actuarial application on insurance data.

(ii) Using different estimation methods in evaluating and analyzing risks.

(iii) Combining the original test and the modified test.

(iv) Providing various applications on the original statistical test and the modified test.

2. Risk Indicators

The key risk indicators (KRIs) play a crucial role in understanding actuarial analysis of insurance risks. These indicators provide valuable insights into the level and nature of risks associated with insurance portfolios, allowing actuaries to quantify, measure, and manage these risks effectively. Here are some key points highlighting the importance of risk indicators in actuarial analysis: 1. KRIs help actuaries quantify the level of risk in insurance portfolios. By using various metrics, such as loss ratios, claim frequencies, severity measures, or aggregate reserves, actuaries can assess the potential financial impact of risks. These indicators provide a numerical representation of the risk exposure and assist in decision-making processes related to pricing, reserving, and capital allocation. 2. KRIs enable actuaries to measure and assess the magnitude and likelihood of potential losses. Actuarial models and techniques, such as probability distributions, statistical methods, and simulation studies, are used in conjunction with risk indicators to estimate the probability and severity of adverse events. This helps insurers evaluate the financial implications and potential solvency risks associated with the insurance business. 3. KRIs serve as monitoring tools, allowing actuaries to track changes in risk levels over time. By regularly analyzing and comparing risk indicators against pre-defined thresholds or benchmarks, actuaries can identify emerging risks or deviations from expected risk levels. These indicators act as early warning signals, enabling insurers to take proactive measures to mitigate risks, adjust pricing, or revise underwriting strategies. 4. KRIs assist in segmenting insurance portfolios based on the level of risk. Actuaries can use risk indicators to identify high-risk segments or policyholders and allocate appropriate resources for risk mitigation. This segmentation helps in portfolio management, allowing insurers to optimize their risk exposure, diversify risks, and balance their overall risk portfolio. 5. KRIs play a crucial role in regulatory compliance and financial reporting for insurers. Regulators often require insurers to report and disclose risk-related information to ensure solvency and protect policyholders. Risk indicators, such as risk-based capital (RBC) ratios, economic capital models, or stress testing results, provide insights into the financial stability and resilience of insurance companies, ensuring compliance with regulatory requirements. 6. KRIs support informed decision-making and strategic planning for insurers. Actuaries rely on risk indicators to assess the profitability of insurance products, evaluate potential risks associated with new business ventures, and determine appropriate pricing strategies. These indicators help in setting risk appetite, formulating risk management policies, and aligning business strategies with the risk tolerance and objectives of the insurance company. This indicator is frequently used to calculate the amount of capital needed to deal with such probable negative events. The VAR of the RGG distribution at the 100q% level, say VAR (Z) or πq (Q), is the 100q% quantile (or percentile). Then, we can simply write

\[
VAR(Z) = Pr(Z > Q(U)) = \begin{cases} 
(1 - q)\% \text{ if } q \leq 99\%, \\
(1 - q)\% \text{ if } q > 99\%. 
\end{cases} 
\]

where Q(U) can be derived by inverting (2). For a one-year time when q = 99%, the interpretation is that there is only a very small chance (0.01) that the insurance company will be bankrupted by an adverse outcome over the next year (see Wirch [10] for more details). Generally speaking, if the distribution of gains (or losses) is limited to the normal distribution, it is acknowledged that the number VAR(X) meets all coherence requirements. The datasets for insurance such as the insurance claims and reinsurance revenues are typically skewed to the right or to the left. The normal distribution is not suitable to describe the revenues from reinsurance and insurance claims. The TVAR of Z at the 100q% confidence level is the expected loss given that the loss exceeds the 100q% of its distribution. Then, the TVAR of Z can be expressed as

\[
TVAR(Z) = E(Z \mid Z > \pi_Z(q)) = \frac{1}{1 - F_\Phi(\pi_Z(q))} \int_{\pi_Z(q)}^{\infty} z f_\Phi(z)dz = \frac{1}{q} \int_{\pi_Z(q)}^{\infty} z f_\Phi(z)dz. 
\]
The quantity TVAR ($X$), which gives further details about the tail of the RGG distribution, is therefore the average of all the VAR values mentioned above at the confidence level $q$. Moreover, TVAR ($X$) can also be expressed as TVAR ($Z$) = $e(Z)$ + VAR ($Z$), where $e(Z)$ is the mean excess loss (MEL) function evaluated at the 100$q$th quantile (see Wirch [10]; Acerbi and Tasche [11]; and Tasche [12]). If $e(Z)$ vanishes, then TVAR ($Z$) = VAR ($Z$), and for very small values of $e(Z)$, the value of TVAR ($Z$) will be very close to VAR ($Z$). The TV risk indicator, which Furman and Landsman [13] developed, calculates the loss’s deviation from the average along a tail. Explicit expressions for the TV risk indicator ($TV (Z)$) can then be expressed as

$$TV (Z) = E(Z | Z > \pi_{Z}(q)) - [TVAR (Z)]^2.$$  

(6)

As a statistic for the best portfolio choice, Furman and Landsman [13] developed the TMV risk indicator, which is based on the TV risk indicator. Thus, the measure TMV has the following form:

$$TMV (Z) = TVAR (Z) + \pi TV (Z)|_{0 < \pi < 1}.$$  

(7)

Then, for any RV, TMV ($Z$) > TV ($Z$), and for $\pi = 1$, TMV ($Z$) = TVAR ($Z$). We will use the methodologies that provide numerical solutions to this complex function, and we will use ready-made programs like “R” and “MATHCAD” to facilitate numerical operations. The use of numerical methods has recently become popular for many reasons, the most important of which is the availability of ready-made statistical programs, and the quantum function of the RGG is not known in a certain closed form.

Numerical approaches were applied in this paper’s risk analysis and evaluation procedure (see Section 4), as well as in the issue of distributional validation under the NRR and its updated equivalent version (see Section 7). For more details, see Alanzi et al. [14], Zhou and Gao [15], Hamed et al. [16], and Yousof et al. [17].

3. Risk Analysis Using Different Estimation Methods

3.1. Risk Assessment under Artificial Data. For the purpose of computing the abovementioned KRI$s$, the four estimation methods are discussed in this section: MLE, OLS, WLSE, and ADE. Three CLs ($q$ = 70%, 90%, 99%) and $N = 1,000$ for sample sizes $n = (50, 150, 300, 500)$ are considered. All results are reported in Table 1 (KRI$s$ under artificial data for $n = 50$), Table 2 (KRI$s$ under artificial data for $n = 150$), Table 3 (KRI$s$ under artificial data for $n = 300$), and Table 4 (KRI$s$ under artificial data for $n = 500$). The main goal of the simulations is to assess the effectiveness of the four risk analysis techniques in order to choose the most suitable and effective ones. Tables 1–4 provide the significant conclusions:

(1) The measures VAR ($Z$), TVAR ($Z$), and TMV ($Z$) increase when $q$ increases for all estimation methods.

(2) The measures TV ($Z$) and MEL ($Z$) decrease when $q$ increases for all estimation methods.

(3) VAR ($Z$)$_{MLE} <$ VAR ($Z$)$_{ADE} <$ VAR ($Z$)$_{WLSE} <$ VAR ($Z$)$_{OLSE}$ for most values of $q$.

(4) We can confirm by the numbers of the four tables that all functions are satisfactory and that one approach cannot be clearly recommended over another method. Based on these findings, we are obligated to provide an application based on real data that can select one way over another, determining the best and most appropriate methods. In other words, the simulation study did not help weighting the four methods decisively because they showed similar results in risk assessment. These convergent results assure that all methods have good and acceptable performance in modelling actuarial data and risk assessment.

3.2. Risk Assessment under the Insurance Payment Claims Data. Analyzing historical insurance data of claims using probability distributions is essential for several reasons: 1. Probability distributions provide a mathematical framework to model and analyze the frequency and severity of insurance claims. By fitting historical claims data to appropriate probability distributions, insurers can estimate the likelihood of different claim amounts and frequencies occurring in the future. This information is crucial for assessing the overall risk exposure of the insurance company. 2. Probability distributions help insurers determine appropriate premiums for insurance policies. By understanding the distribution of claim amounts and frequencies, insurers can calculate the expected value of claims and incorporate it into the pricing structure. This ensures that premiums charged to policyholders align with the potential risks faced by the insurer, maintaining a fair and sustainable pricing model. 3. Accurate estimation of potential claim costs is vital for insurers to set aside adequate reserves and plan their financial stability. By analyzing historical claims data using probability distributions, insurers can estimate the potential range of future claim payments and allocate sufficient funds to cover these liabilities. It enables insurers to make informed decisions about capital management, investment strategies, and financial reserves. 4. Probability distributions provide insights into the potential volatility and tail risks associated with insurance claims. Insurers can analyze the shape of the distribution and its parameters to identify extreme events and tail risks that may have a significant impact on the financial health of the company. This knowledge helps insurers manage risks effectively, develop appropriate underwriting guidelines, and implement risk mitigation strategies. 5. Probability distributions allow insurers to simulate various scenarios and evaluate the potential outcomes of different policy design choices. By incorporating historical claims data into probabilistic models, insurers can assess the impact of different policy terms, coverage limits, deductibles, and other factors on claim frequencies and amounts. This information helps insurers make data-driven decisions when designing
### Table 1: KRIs under artificial data for $n = 50$.  

<table>
<thead>
<tr>
<th>Method</th>
<th>$\hat{\lambda}$</th>
<th>$\hat{\theta}$</th>
<th>VaRq ($Z$)</th>
<th>TVaRq ($Z$)</th>
<th>TVq ($Z$)</th>
<th>TMVq ($Z$)</th>
<th>ELq ($Z$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLE</td>
<td>0.50895</td>
<td>0.20142</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70%</td>
<td>0.61704</td>
<td>0.87178</td>
<td>0.04269</td>
<td>0.89313</td>
<td>0.25474</td>
<td></td>
<td></td>
</tr>
<tr>
<td>90%</td>
<td>0.92187</td>
<td>1.1123</td>
<td>0.02643</td>
<td>1.12552</td>
<td>0.19044</td>
<td></td>
<td></td>
</tr>
<tr>
<td>99%</td>
<td>1.34073</td>
<td>1.46944</td>
<td>0.01334</td>
<td>1.47611</td>
<td>0.12871</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLSE</td>
<td>0.43791</td>
<td>0.19792</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70%</td>
<td>0.69477</td>
<td>0.8966</td>
<td>0.05624</td>
<td>1.01472</td>
<td>0.29183</td>
<td></td>
<td></td>
</tr>
<tr>
<td>90%</td>
<td>1.04376</td>
<td>1.2626</td>
<td>0.03497</td>
<td>1.28008</td>
<td>0.21884</td>
<td></td>
<td></td>
</tr>
<tr>
<td>99%</td>
<td>1.5253</td>
<td>1.67363</td>
<td>0.01773</td>
<td>1.6825</td>
<td>0.14834</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WLSE</td>
<td>0.49538</td>
<td>0.20223</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70%</td>
<td>0.63848</td>
<td>0.90104</td>
<td>0.04532</td>
<td>0.9237</td>
<td>0.26256</td>
<td></td>
<td></td>
</tr>
<tr>
<td>90%</td>
<td>0.95271</td>
<td>1.14886</td>
<td>0.02802</td>
<td>1.16287</td>
<td>0.19615</td>
<td></td>
<td></td>
</tr>
<tr>
<td>99%</td>
<td>1.38409</td>
<td>1.51658</td>
<td>0.01413</td>
<td>1.52364</td>
<td>0.13249</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADE</td>
<td>0.47129</td>
<td>0.19390</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70%</td>
<td>0.62177</td>
<td>0.8883</td>
<td>0.04713</td>
<td>0.91186</td>
<td>0.26653</td>
<td></td>
<td></td>
</tr>
<tr>
<td>90%</td>
<td>0.94027</td>
<td>1.14087</td>
<td>0.02947</td>
<td>1.1556</td>
<td>0.2006</td>
<td></td>
<td></td>
</tr>
<tr>
<td>99%</td>
<td>1.38192</td>
<td>1.51837</td>
<td>0.01503</td>
<td>1.52588</td>
<td>0.13645</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 2: KRIs under artificial data for $n = 150$.  

<table>
<thead>
<tr>
<th>Method</th>
<th>$\hat{\lambda}$</th>
<th>$\hat{\theta}$</th>
<th>VaRq ($Z$)</th>
<th>TVaRq ($Z$)</th>
<th>TVq ($Z$)</th>
<th>TMVq ($Z$)</th>
<th>ELq ($Z$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLE</td>
<td>0.50211</td>
<td>0.20039</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70%</td>
<td>0.61969</td>
<td>0.87682</td>
<td>0.04355</td>
<td>0.89859</td>
<td>0.25713</td>
<td></td>
<td></td>
</tr>
<tr>
<td>90%</td>
<td>0.92732</td>
<td>1.11971</td>
<td>0.02699</td>
<td>1.13321</td>
<td>0.19239</td>
<td></td>
<td></td>
</tr>
<tr>
<td>99%</td>
<td>1.35054</td>
<td>1.48069</td>
<td>0.01364</td>
<td>1.48751</td>
<td>0.13015</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLSE</td>
<td>0.43616</td>
<td>0.19788</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70%</td>
<td>0.69733</td>
<td>0.99028</td>
<td>0.05668</td>
<td>1.01862</td>
<td>0.29295</td>
<td></td>
<td></td>
</tr>
<tr>
<td>90%</td>
<td>1.04766</td>
<td>1.26735</td>
<td>0.03525</td>
<td>1.28498</td>
<td>0.21969</td>
<td></td>
<td></td>
</tr>
<tr>
<td>99%</td>
<td>1.53108</td>
<td>1.67999</td>
<td>0.01787</td>
<td>1.68894</td>
<td>0.14892</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WLSE</td>
<td>0.47431</td>
<td>0.20542</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70%</td>
<td>0.68561</td>
<td>0.9633</td>
<td>0.05052</td>
<td>0.98856</td>
<td>0.27769</td>
<td></td>
<td></td>
</tr>
<tr>
<td>90%</td>
<td>1.01813</td>
<td>1.22502</td>
<td>0.03111</td>
<td>1.24058</td>
<td>0.20689</td>
<td></td>
<td></td>
</tr>
<tr>
<td>99%</td>
<td>1.53108</td>
<td>1.67999</td>
<td>0.01787</td>
<td>1.68894</td>
<td>0.14892</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADE</td>
<td>0.46943</td>
<td>0.19354</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70%</td>
<td>0.62209</td>
<td>0.88924</td>
<td>0.04737</td>
<td>0.91293</td>
<td>0.26716</td>
<td></td>
<td></td>
</tr>
<tr>
<td>90%</td>
<td>0.94131</td>
<td>1.14246</td>
<td>0.02964</td>
<td>1.15728</td>
<td>0.20115</td>
<td></td>
<td></td>
</tr>
<tr>
<td>99%</td>
<td>1.38418</td>
<td>1.52105</td>
<td>0.01512</td>
<td>1.52861</td>
<td>0.13686</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 3: KRIs under artificial data for $n = 300$.  

<table>
<thead>
<tr>
<th>Method</th>
<th>$\hat{\lambda}$</th>
<th>$\hat{\theta}$</th>
<th>VaRq ($Z$)</th>
<th>TVaRq ($Z$)</th>
<th>TVq ($Z$)</th>
<th>TMVq ($Z$)</th>
<th>ELq ($Z$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLE</td>
<td>0.50061</td>
<td>0.20021</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70%</td>
<td>0.62056</td>
<td>0.87828</td>
<td>0.04375</td>
<td>0.90015</td>
<td>0.25772</td>
<td></td>
<td></td>
</tr>
<tr>
<td>90%</td>
<td>0.92889</td>
<td>1.12175</td>
<td>0.02713</td>
<td>1.13531</td>
<td>0.19286</td>
<td></td>
<td></td>
</tr>
<tr>
<td>99%</td>
<td>1.35315</td>
<td>1.48363</td>
<td>0.01371</td>
<td>1.49048</td>
<td>0.13048</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLSE</td>
<td>0.43709</td>
<td>0.19783</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70%</td>
<td>0.69554</td>
<td>0.98781</td>
<td>0.05642</td>
<td>1.01602</td>
<td>0.29227</td>
<td></td>
<td></td>
</tr>
<tr>
<td>90%</td>
<td>1.04505</td>
<td>1.26424</td>
<td>0.03509</td>
<td>1.28179</td>
<td>0.21919</td>
<td></td>
<td></td>
</tr>
<tr>
<td>99%</td>
<td>1.52737</td>
<td>1.67595</td>
<td>0.01779</td>
<td>1.68485</td>
<td>0.14859</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WLSE</td>
<td>0.45390</td>
<td>0.20762</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70%</td>
<td>0.72996</td>
<td>1.02258</td>
<td>0.05597</td>
<td>1.05056</td>
<td>0.29262</td>
<td></td>
<td></td>
</tr>
<tr>
<td>90%</td>
<td>1.08048</td>
<td>1.2981</td>
<td>0.03438</td>
<td>1.31529</td>
<td>0.21762</td>
<td></td>
<td></td>
</tr>
<tr>
<td>99%</td>
<td>1.55877</td>
<td>1.70514</td>
<td>0.01722</td>
<td>1.71375</td>
<td>0.14637</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADE</td>
<td>0.47025</td>
<td>0.19342</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70%</td>
<td>0.62032</td>
<td>0.88688</td>
<td>0.04717</td>
<td>0.91046</td>
<td>0.26656</td>
<td></td>
<td></td>
</tr>
<tr>
<td>90%</td>
<td>0.93883</td>
<td>1.13954</td>
<td>0.02951</td>
<td>1.1543</td>
<td>0.20072</td>
<td></td>
<td></td>
</tr>
<tr>
<td>99%</td>
<td>1.38075</td>
<td>1.51734</td>
<td>0.01506</td>
<td>1.52487</td>
<td>0.13659</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
insurance policies and developing risk management strategies. Overall, analyzing historical insurance data of claims using probability distributions enables insurers to gain insights into the nature of risks they face, make informed decisions, and effectively manage their operations. It supports pricing, reserving, risk assessment, and underwriting activities, contributing to the financial stability and success of insurance companies. In this work, we will consider and analyze a claims data collected from 2007 to 2013. This work proposes certain KRI quantities for the left-skewed insurance claims data under the EEC distribution, including VAR, TVAR, TV, and TMV [18]. One of the finest techniques for heavy-tailed distributions is based on the t-Hill approach, an upper order statistic modification of the estimator.

Table 4: KRIs under artificial data for \( n = 500 \).

<table>
<thead>
<tr>
<th>Method</th>
<th>( \lambda )</th>
<th>( \theta )</th>
<th>( \text{VaRq} (Z) )</th>
<th>( \text{TVaRq} (Z) )</th>
<th>( \text{TVq} (Z) )</th>
<th>( \text{TMVq} (Z) )</th>
<th>( \text{ELq} (Z) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLE</td>
<td>0.50059</td>
<td>0.20008</td>
<td>0.61986</td>
<td>0.87745</td>
<td>0.04372</td>
<td>0.89931</td>
<td>0.25759</td>
</tr>
<tr>
<td>70%</td>
<td></td>
<td></td>
<td>0.92803</td>
<td>1.12082</td>
<td>0.02711</td>
<td>1.13437</td>
<td>0.19279</td>
</tr>
<tr>
<td>99%</td>
<td></td>
<td></td>
<td>1.35213</td>
<td>1.48258</td>
<td>0.0137</td>
<td>1.48943</td>
<td>0.13045</td>
</tr>
<tr>
<td>OLSE</td>
<td>0.43711</td>
<td>0.19782</td>
<td>0.69542</td>
<td>0.98767</td>
<td>0.05641</td>
<td>1.01587</td>
<td>0.29224</td>
</tr>
<tr>
<td>70%</td>
<td></td>
<td></td>
<td>1.04491</td>
<td>1.26408</td>
<td>0.03508</td>
<td>1.28162</td>
<td>0.21917</td>
</tr>
<tr>
<td>99%</td>
<td></td>
<td></td>
<td>1.52718</td>
<td>1.67576</td>
<td>0.01779</td>
<td>1.68465</td>
<td>0.14588</td>
</tr>
<tr>
<td>WLSE</td>
<td>0.43681</td>
<td>0.20882</td>
<td>0.76618</td>
<td>1.07161</td>
<td>0.06091</td>
<td>1.10207</td>
<td>0.30544</td>
</tr>
<tr>
<td>70%</td>
<td></td>
<td></td>
<td>1.13212</td>
<td>1.35905</td>
<td>0.03736</td>
<td>1.37773</td>
<td>0.22693</td>
</tr>
<tr>
<td>99%</td>
<td></td>
<td></td>
<td>1.63081</td>
<td>1.78331</td>
<td>0.01868</td>
<td>1.79265</td>
<td>0.1525</td>
</tr>
<tr>
<td>ADE</td>
<td>0.47043</td>
<td>0.19333</td>
<td>0.61954</td>
<td>0.88589</td>
<td>0.0471</td>
<td>0.90944</td>
<td>0.26635</td>
</tr>
<tr>
<td>70%</td>
<td></td>
<td></td>
<td>0.93779</td>
<td>1.13836</td>
<td>0.02947</td>
<td>1.1531</td>
<td>0.20058</td>
</tr>
<tr>
<td>99%</td>
<td></td>
<td></td>
<td>1.37941</td>
<td>1.51591</td>
<td>0.01504</td>
<td>1.52343</td>
<td>0.1365</td>
</tr>
</tbody>
</table>

Table 5 reports the KRIs under the insurance claims data and MLE method for the RGG \( (\Phi_1) \) and GG \( (\Phi_2) \) models, where \( \Phi_1 = (0.00015, 0.28234) \) and \( \Phi_2 = (0.00082, 1.22414) \). Table 6 gives the KRIs under the insurance claims data and OLSE method for the RGG \( (\Phi_1) \) and GG \( (\Phi_2) \) models, where \( \Phi_1 = (0.000102, 0.22171) \) and \( \Phi_2 = (0.00065, 0.91821) \). Table 7 provides the KRIs under the insurance claims data and WLSE method for the RGG \( (\Phi_1) \) and GG \( (\Phi_2) \) models, where \( \Phi_1 = (0.00013, 0.24966) \) and \( \Phi_2 = (0.00015, 0.28234) \). Table 8 presents the KRIs under the insurance claims data and WLSE method for the RGG \( (\Phi_1) \) and GG \( (\Phi_2) \) models, where \( \Phi_1 = (0.00013, 0.25097) \) and \( \Phi_2 = (0.00073, 1.06384) \). Based on these tables, the following results can be highlighted:

(1) For all risk assessment methods:

\[ \text{VAR}(Z_{\lambda=0.3}) < \text{VAR}(Z_{\lambda=0.25}) < \ldots < \text{VAR}(Z_{\lambda=0.1}) < \text{VAR}(Z_{\lambda=0.01}). \]  

(2) For all risk assessment methods:

\[ \text{TVAR}(Z_{\lambda=0.3}) < \text{TVAR}(Z_{\lambda=0.25}) < \ldots < \text{TVAR}(Z_{\lambda=0.1}) < \text{TVAR}(Z_{\lambda=0.01}). \]  

(3) For all risk assessment methods:

\[ \text{TV}(Z_{\lambda=0.3}) > \text{TV}(Z_{\lambda=0.25}) > \ldots > \text{TV}(Z_{\lambda=0.1}) > \text{TV}(Z_{\lambda=0.01}). \]  

(4) For all risk assessment methods:

\[ \text{TMV}(Z_{\lambda=0.3}) > \text{TMV}(Z_{\lambda=0.25}) > \ldots > \text{TMV}(Z_{\lambda=0.1}) > \text{TMV}(Z_{\lambda=0.01}). \]
For the RGG model and based on Table 5, it is noted that the VAR (Z) increased to be 10050.30113\% which means that the RGG model is better than the GG model. The same result can be conducted for all other risk indicators.

### Table 6: KRIs under the insurance claims data and OLSE method for the RGG and GG models.

<table>
<thead>
<tr>
<th>RGG</th>
<th>(\lambda)</th>
<th>(\theta)</th>
<th>VAR (Z)</th>
<th>TVaRq (Z)</th>
<th>TV (Z)</th>
<th>TMV (Z)</th>
<th>MEL (Z)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RGG</td>
<td>0.000102</td>
<td>0.2217</td>
<td>3632.99105</td>
<td>4996.48683</td>
<td>1205793.50393</td>
<td>607893.23879</td>
<td>1363.49578</td>
</tr>
<tr>
<td>GG</td>
<td>0.00065</td>
<td>0.9182</td>
<td>3594.53397</td>
<td>5576.34069</td>
<td>3539252.01829</td>
<td>1775202.34984</td>
<td>1981.86072</td>
</tr>
<tr>
<td>GG</td>
<td>0.00013</td>
<td>0.24966</td>
<td>3526.87158</td>
<td>4718.69291</td>
<td>897465.10391</td>
<td>453451.24486</td>
<td>1191.82133</td>
</tr>
</tbody>
</table>

For the RGG model and based on Table 6, it is noted that the \(\text{VAR}(Z)\) can be 3632.99105\% and then 3811.62160\% and increased to be 7484.76402\% which means that the RGG model is better than the GG model. The same result can be conducted for all other risk indicators.

### Table 7: KRIs under the insurance claims data and WLS method for the RGG and GG models.

<table>
<thead>
<tr>
<th>RGG</th>
<th>(\lambda)</th>
<th>(\theta)</th>
<th>VAR(Z)</th>
<th>TVaRq(Z)</th>
<th>TV(Z)</th>
<th>TMV(Z)</th>
<th>MEL(Z)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RGG</td>
<td>0.00013</td>
<td>0.24966</td>
<td>3526.87158</td>
<td>4718.69291</td>
<td>897465.10391</td>
<td>453451.24486</td>
<td>1191.82133</td>
</tr>
<tr>
<td>GG</td>
<td>0.00015</td>
<td>0.28234</td>
<td>3496.50391</td>
<td>4558.57622</td>
<td>699524.80147</td>
<td>354320.97695</td>
<td>1062.07231</td>
</tr>
</tbody>
</table>
(5) For all risk assessment methods:

\[ \text{MEL}(Z | \eta = 0.3) > \text{MEL}(Z | \eta = 0.25) > \ldots > \text{MEL}(Z | \eta = 0.1) > \text{MEL}(Z | \eta = 0.01). \]

(6) For the RGG model: Nearly for all \( q \) values, the OLSE method is recommended since it provides the most acceptable risk exposure analysis; then, the MLE method is recommended as a second one. However, the other two methods perform well. For the GG model: Nearly for all \( q \) values, the OLSE method is recommended since it provides the most acceptable risk exposure analysis; then, the MLE method is recommended as a second one. However, the other two methods perform well.

(7) For all \( q \) values and under all risk methods: The RGG model is better than the GG model. It is worth noting that the distributions have the same number of parameters, but the new distribution is the best in the process of modelling insurance claims reimbursement data and assessing actuarial risk. We hope that the proposed distribution will gain a great deal of interest from actuaries and practitioners in future actuarial and applied studies.

### 4. Distributional Validity

#### 4.1. Distributional Validity under the UMLE Method

Here, the UMLE method is used to estimate the RGG distribution’s parameters. Let \( z_1, \ldots, z_n \) be random samples distributed according to the RGG model. The uncensored likelihood function is obtained by

\[ L(\Phi) = \prod_{i=1}^{n} f_\Phi(z_i). \]

Then, the uncensored log-likelihood function reduces to

\[ l(\Phi) = n \ln(2\theta) + 2n \ln(\lambda) + \sum_{i=1}^{n} \ln(z_i) - \lambda \sum_{i=1}^{n} z_i + (2\theta - 1) \sum_{i=1}^{n} \ln \xi_i - \sum_{i=1}^{n} b_i^2, \]

\[ \text{Table 7: Continued.} \]

<table>
<thead>
<tr>
<th>RGG</th>
<th>( \lambda )</th>
<th>( \vartheta )</th>
<th>VAR(Z)</th>
<th>TVaRq(Z)</th>
<th>TV(Z)</th>
<th>TMV(Z)</th>
<th>MEL(Z)</th>
</tr>
</thead>
<tbody>
<tr>
<td>90%</td>
<td>6431.29618</td>
<td>6919.75575</td>
<td>188726.99863</td>
<td>101283.25506</td>
<td>488.45957</td>
<td></td>
<td></td>
</tr>
<tr>
<td>95%</td>
<td>6431.29618</td>
<td>6919.75575</td>
<td>188726.99863</td>
<td>101283.25506</td>
<td>488.45957</td>
<td></td>
<td></td>
</tr>
<tr>
<td>99%</td>
<td>6431.29618</td>
<td>6919.75575</td>
<td>188726.99863</td>
<td>101283.25506</td>
<td>488.45957</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For the RGG model and based on Table 7, it is noted that the \( \text{VAR}(Z) \) can be 5352.87158\( \eta = 99\% \) and then 3811.62126\( \eta = 95\% \) and increased to 6849.20863\( \eta = 99\% \). For the GG model and based on Table 7, it is noted that the \( \text{VAR}(Z) \) can be 3496.50391\( \eta = 99\% \) and then 3753.40308\( \eta = 95\% \) and increased to 6431.29618\( \eta = 99\% \), which means that the RGG model is better than the GG model. The same result can be noticed for all other risk indicators.

\[ \text{Table 8: KRIs under the insurance claims data and ADE method for the RGG and GG models.} \]

<table>
<thead>
<tr>
<th>RGG</th>
<th>( \lambda )</th>
<th>( \vartheta )</th>
<th>VAR(Z)</th>
<th>TVaRq(Z)</th>
<th>TV(Z)</th>
<th>TMV(Z)</th>
<th>MEL(Z)</th>
</tr>
</thead>
<tbody>
<tr>
<td>70%</td>
<td>0.00013</td>
<td>0.25097</td>
<td>3540.62976</td>
<td>4731.51551</td>
<td>895287.06144</td>
<td>452375.06423</td>
<td>1190.88575</td>
</tr>
<tr>
<td>75%</td>
<td>0.00013</td>
<td>0.25097</td>
<td>3540.62976</td>
<td>4731.51551</td>
<td>895287.06144</td>
<td>452375.06423</td>
<td>1190.88575</td>
</tr>
<tr>
<td>80%</td>
<td>0.00013</td>
<td>0.25097</td>
<td>3540.62976</td>
<td>4731.51551</td>
<td>895287.06144</td>
<td>452375.06423</td>
<td>1190.88575</td>
</tr>
<tr>
<td>85%</td>
<td>0.00013</td>
<td>0.25097</td>
<td>3540.62976</td>
<td>4731.51551</td>
<td>895287.06144</td>
<td>452375.06423</td>
<td>1190.88575</td>
</tr>
<tr>
<td>90%</td>
<td>0.00013</td>
<td>0.25097</td>
<td>3540.62976</td>
<td>4731.51551</td>
<td>895287.06144</td>
<td>452375.06423</td>
<td>1190.88575</td>
</tr>
<tr>
<td>99%</td>
<td>0.00013</td>
<td>0.25097</td>
<td>3540.62976</td>
<td>4731.51551</td>
<td>895287.06144</td>
<td>452375.06423</td>
<td>1190.88575</td>
</tr>
</tbody>
</table>

For the RGG model and based on Table 7, it is noted that the \( \text{VAR}(Z) \) can be 3540.62976\( \eta = 99\% \) and then 3825.31795\( \eta = 95\% \) and increased to 6859.01979\( \eta = 99\% \). For the GG model and based on Table 7, it is noted that the \( \text{VAR}(Z) \) can be 3458.40688\( \eta = 99\% \) and then 3808.11267\( \eta = 95\% \) and increased to 6631.80035\( \eta = 99\% \).
where $\xi_i = \xi(z_i) = 1 - (1 + \lambda z_i) \exp(-\lambda z_i), b_i = \xi_i^\theta/1 - \xi_i^\theta, \omega_i = 1 - \exp(-b_i^2)$. The MLEs can be obtained by solving the following equation:

$$
\frac{\partial l(\Phi)}{\partial \lambda} = \frac{2n}{\lambda} - \sum_{i=1}^{n} z_i + \lambda (2\theta - 1) \sum_{i=1}^{n} \frac{z_i \exp(-\lambda z_i)}{\xi_i} - 2\lambda \theta \sum_{i=1}^{n} \frac{z_i \exp(-\lambda z_i) b_i^2}{\xi_i (1 - \xi_i^\theta)},
$$

$$
\frac{\partial l(\Phi)}{\partial \theta} = \sum_{i=1}^{n} \frac{n \ln(\omega_i)}{\theta} - 2 \sum_{i=1}^{n} b_i^2 \ln \xi_i (1 + b_i).
$$

### 4.2. Distributional Validity Utilizing the Right CMLE Method.

Let $z = (z_1, \ldots, z_n)^T$ be a right censored data sample with fixed censoring time $r$ from the RGG distribution with parameter vector $\Phi$. Each $z_i$ can be written as $z_i = (z_i, \pi_i)$. The right censored log-likelihood function $L_n(\Phi)$ has the form

$$
L_n(\Phi) = \sum_{i=1}^{n} \pi_i \ln f_{\Phi}(z) + \sum_{i=1}^{n} (1 - \pi_i) \ln b_{\Phi}(z),
$$

where $b_{\Phi}(z) = 1 - F_\Phi(z)$ is the survival function of the RGG model, and based on (15), we have

$$
\frac{\partial L_{n,\pi_i}(\Phi)}{\partial \lambda} = \sum_{i=1}^{n} \pi_i \left[ \frac{2}{\lambda} - z_i + \frac{\lambda (2\theta - 1) z_i \exp(-\lambda z_i)}{\xi_i} - \frac{2\lambda \theta x_i^2 \exp(-\lambda z_i) b_i^2}{\xi_i (1 - \xi_i^\theta)} \right],
$$

$$
- \sum_{i=1}^{n} \frac{2\lambda \theta x_i^2 \exp(-\lambda z_i) b_i^2 \exp(-b_i^2)}{\xi_i (1 - \xi_i^\theta) (1 - \omega_i)},
$$

$$
\frac{\partial L_{n,\pi_i}(\Phi)}{\partial \theta} = \sum_{i=1}^{n} \pi_i \left[ \frac{1}{\theta} - 2 \ln(\omega_i) - 2 \xi_i \ln \xi_i (1 + b_i) \right] - \sum_{i=1}^{n} \frac{2\lambda \theta x_i^2 \ln \xi_i (1 + b_i)}{1 - \omega_i}.
$$

Similar to the complete data scenario, we employ numerical techniques like the Newton–Raphson method, the Monte Carlo method, or the BB-solve package to compute the MLEs.

### 5. Testing Procedures

#### 5.1. Testing Procedures from the Statistic $Y_n^2(r - 1)$.

To testing the null hypothesis $H_0$ according to which a sample $Z_1, Z_2, \ldots, Z_n$ belongs to (2), where

$$
H_0 = \text{Pr}(Z_i \leq z) = F_{\Phi}(z),
$$

consider $r$ equiprobable grouping intervals $I_1, I_2, \ldots, I_r$ where $I_j = [b_{j-1}, b_j]; I_j \cap I_j = \emptyset$ and $\cup_{j=1}^{r} I_j = \mathbb{R}^1$ such as

$$
P_j = \int_{b_j}^{b_{j+1}} f_\Phi(z) dz = \frac{1}{r}, j = 1, \ldots, r,
$$

and $b_j = F^{-1}(j/r), j = 1, \ldots, r$. If $v = (v_1, \ldots, v_r)^T$ denotes the number of observed grouping $z_i$ into these intervals $I_j$ and

$$
T_n(\Phi) = \left( \frac{v_1 - n P_1(\Phi)}{\sqrt{n P_1(\Phi)}} \frac{v_2 - n P_2(\Phi)}{\sqrt{n P_2(\Phi)}} \cdots \frac{v_r - n P_r(\Phi)}{\sqrt{n P_r(\Phi)}} \right)^T,
$$

the NRR statistic $Y_n^2(r - 1)$ due to Nikulin [5] and Rao and Robson [7] is defined by

$$
Y_n^2(r - 1) = Y_n^2(r - 1; \Phi) = T_n(\Phi)^T I_n(\Phi)^{-1} I_n(\Phi).
$$

Here, $I(\Phi)$ and $J(\Phi)$ are the estimated information matrices on non-grouped and grouped data, respectively, and $\Phi$ is the vector of the MLEs on initial data. The elements of the vector $I(\Phi) = (I_k(\Phi))_{k=1}^{22}$ are

$$
I_k(\Phi) = \sum_{i=1}^{n} \frac{\partial}{\partial \Phi_k} P_j(\Phi),
$$

$$
J(\Phi) = B(\Phi)^T B(\Phi),
$$
where

\[
B(\Phi) = \left( \frac{1}{\sqrt{P_l}} \frac{\partial}{\partial \Phi} P_l \right)_{r \times d, k=1, \ldots, b_l=1, \ldots, r} ,
\]

and \( b \) is the number of the model parameters. The distribution of \( Y_n^2(r-1; \Phi) \) is \( \chi^2_{r-1} \). To construct the test statistic \( Y_n^2(r-1) \) corresponding to the RGG with a parameter vector \( \Phi = (\lambda, \theta)^T \), we first calculate the MLEs \( \hat{\Phi} = (\hat{\lambda}, \hat{\theta})^T \) and the limit intervals \( b_j \). Secondly, the derivatives \( \frac{\partial}{\partial \Phi} P_l(\hat{\Phi}) \) are derived as

\[
\frac{\partial}{\partial \lambda} P_l(\hat{\Phi}) = 2\lambda 
\]

\[
\frac{\partial}{\partial \theta} P_l(\hat{\Phi}) = 
\]

Finally, we obtain the statistic \( Y_n^2(r-1; \hat{\Phi}) \) which allows to verify if data belong to the RGG distribution.

5.2. Testing Procedures for the \( M^2_\Phi(r) \) Test Statistic with Right Censorship. The NRR statistic described above was adjustment by Bagdonavičius and Nikulin [8]. Generally, the NRR statistic is established based on the vector \( \Lambda_j = 1/\sqrt{n}(O_j(z) - e_j(z)) \) where \( O_j(z) \) and \( e_j(z) \) are the observed number of failures and expected number of failures to fall into the grouping intervals \( I_j \), and the statistic \( M^2_\Phi(r) \) is defined by

\[
M^2_\Phi(r) = \Lambda^T \Sigma^{-1} \Lambda,
\]

where \( \Sigma^{-1} \) refers to the generalized inverse of the well-known covariance matrix \( \Sigma \). For facilitating the calculation process, this novel NRR statistical test can be expressed as follows:

\[
M^2_\Phi(r) = \sum_{j=1}^{r} \frac{1}{O_j(z)} \left( O_j(z) - e_j(z) \right)^2 + \Omega,
\]

with the quadratic form \( \Omega \) obtained as

\[
\Omega = W^T G^{-1} W, \quad \tilde{A}_j = \frac{O_j(z)}{n}, \quad O_j(z) = \sum_{i \in I_j} \pi_i \tilde{G} = [\tilde{g}_{ij}]_{x \times 1},
\]

\[
\tilde{g}_{ij} = \frac{1}{n} \sum_{j=1}^{r} C_{ij} \tilde{G} \tilde{A}_j^{-1}, \quad C_{ij} = \frac{1}{n} \sum_{i \in I_j} \pi_i \frac{\partial}{\partial \Phi} \ln r_{\Phi}(z_i),
\]

\[
\tilde{W}_j = \sum_{i=1}^{r} C_{ij} \tilde{A}_j^{-1} \Lambda_{ij} \quad \text{for } i, j = 1, \ldots, \rho,
\]

where \( r_{\Phi}(z_i) = f_{\Phi}(z_i)/b_{\Phi}(z_i) \) is the hazard rate function of the RGG model. Under the null hypothesis \( H_0 \), the limit distribution of the statistic \( M^2_{\Phi}(r) \) is a chi-square with \( r = \text{rank}(\Sigma) \) degrees of freedom. For more details on modified chi-square tests, one can see the book of Voinov et al. [19]. For testing the null hypothesis that a right censored sample is described by the RGG distribution, we develop \( M^2_{\Phi}(r) \) corresponding to this distribution. To this end, we have to compute the MLEs \( \hat{\Phi} = (\hat{\lambda}, \hat{\theta})^T \) on initial data (see Section 3), the estimated information matrix \( \tilde{I}_{ij} \) which can be deduced from the score functions and the estimated limit intervals \( \tilde{b}_j(z) \). To apply this test statistic, the expected failure times \( e_j(z) \) to fall into the grouping intervals \( I_j \) must be the same for any \( j \), so the estimated interval limits \( \tilde{b}_j(z) \) are equal to

\[
\tilde{b}_j(z) = H^{-1}\left( \frac{1}{n-i+1} \left( E_j(z) + \sum_{i=1}^{r} \frac{H_{\Phi}(z_i)}{e_j(z)} \right) \right),
\]

where \( \tilde{b}_j = \max(Z_{(w)_r}, E_{r,Z} = \sum_{i=1}^{n} H_{\Phi}(z_i) \) and \( E_j(z) = -j/r - \sum_{i=1}^{n} \ln(1 - \omega_i) \) where \( H_{\Phi}(z_i) = -\ln \left[ b_{\Phi}^{-1}(z_i) \right] \) is the cumulative rate function of the RGG model. So, the numbers \( e_j(z) \) and \( O_j(z) \) can be obtained. Then, we can derive (and then calculate) the components of the estimated matrix \( \tilde{K} \) as follows:

\[
\bar{K}_{1,j\Phi} = \frac{1}{n} \sum_{i \in I_j} \pi_i \left[ \frac{2}{\lambda} - z_i + \frac{\lambda(2\theta - 1)z_i \exp(-\lambda z_i)}{\xi_i} - \frac{2\lambda \theta z_i^2 \exp(-\lambda z_i) b_i^2}{\xi_i(1 - \xi_i)} \right],
\]

\[
\bar{K}_{2,j\Phi} = \frac{1}{n} \sum_{i \in I_j} \pi_i \left[ \frac{2\lambda \theta z_i^2 \exp(-\lambda z_i) b_i^2 \exp(-b_i^2)}{\xi_i(1 - \xi_i)(1 - \omega_i)} \right].
\]
and the estimated matrix \( \hat{W} \) is derived from the matrix \( \hat{K} \).
Therefore, the test statistic can be obtained easily:

\[
M_n^2(r) = M_n^2(r; \hat{\Phi}) = \sum_{j=1}^{r} \left( \frac{O_j(z) - e_j(z)}{O_j(z)} \right)^2 \\
+ \hat{W}^T \left[ \hat{S}_y - \sum_{j=1}^{r} \hat{K}_{ij} \hat{K}_{ij}^{-1} \right]^{-1} \hat{W}.
\]

(30)

6. Assessing \( Y_n^2(r - 1) \) and \( M_n^2(r) \) with Some Applications

We performed a significant investigation using numerical simulations in this section to demonstrate the flexibility and effectiveness of the tests suggested in this work. We then used actual data from reliability and survival analysis to run these tests.

6.1. Simulating \( Y_n^2(r - 1) \) under the UMLE Method. For simulating \( M_n^2(r) \) under the UMLE, the data were simulated \( N = 10,000 \) times under the sample sizes \( n_1 = 25, n_2 = 50, n_3 = 130, n_4 = 350, n_5 = 500, \) and \( n_6 = 1000 \). Using the BB algorithm and the R software, the MLEs and their mean square errors (MSEs) are calculated and presented in Table 9. For testing the null hypothesis \( H_0 \) according to which the data follow the RGG distribution, we calculate the \( Y_n^2(r - 1) \) statistical test, and then it is compared with the different empirical levels of rejection of the null hypothesis \( H_0 \) when \( Y_n^2(r - 1) > \chi^2_n(r - 1) \) where empirical levels of rejection of the null hypothesis are \( \alpha_1 = 0.01, \alpha_2 = 0.05, \) and \( \alpha_3 = 0.10 \). Table 10 gives the theoretical risk and empirical risk for complete case. The levels simulated for the statistic \( Y_n^2(r - 1) \) agree with those corresponding to the theoretical levels of the chi-square distribution with \( (r - 1) \) degrees of freedom, and it is noticed after accounting for simulation errors. In light of this, we can state that the test suggested in this study can appropriately adapt the data obtained from a RGG model.

6.2. Simulating \( M_n^2(r) \) under the CMLE Method. For simulating the \( M_n^2(r) \) under the uncensored maximum likelihood method, the data were simulated \( N = 10,000 \) times under the sample sizes \( n_1 = 25, n_2 = 50, n_3 = 130, n_4 = 350, n_5 = 500, \) and \( n_6 = 1000 \). Using the BB algorithm and the R software, the MLEs and their mean square errors (MSEs) are calculated and presented in Table 11. For testing the null hypothesis \( H_0 \) according to which the data follow the RGG distribution, we calculate the \( Y_n^2(r - 1) \) statistical test, and then it is compared with the different empirical levels of rejection of the null hypothesis \( H_0 \) when \( M_n^2(r) > \chi^2_n(r) \) where empirical levels of rejection of the null hypothesis are \( \alpha_1 = 0.01, \alpha_2 = 0.05, \) and \( \alpha_3 = 0.10 \). Table 12 gives the theoretical risk and empirical risk for censored case. The levels simulated for the statistic \( Y_n^2(r - 1) \) agree with those corresponding to the theoretical levels of the chi-square distribution with \( (r) \) degrees of freedom, and it is noticed after accounting for simulation errors. In light of this, we can state that the test suggested in this study can appropriately adapt the data obtained from a RGG model.

7. Data Analysis

7.1. Real Applications for Uncensored Data

7.1.1. Times between Failures for Repairable Items Data. The first dataset is given by Murthy et al. [21]. These data have had a great deal of analysis and study, as many researchers and scholars have modelled and analyzed them and drawn many conclusions about them. The data refer to the time between failures for repairable items (see Ibrahim et al. [22] for more details). The data are as follows: 1.43, 0.11, 0.71, 0.77, 2.63, 1.49, 3.46, 2.46, 0.59, 0.74, 1.23, 0.94, 4.36, 0.40, 1.74, 4.73, 2.23, 0.45, 0.70, 1.06, 1.46, 0.30, 1.82, 2.37, 0.63, 1.23, 1.24, 1.97, 1.86, 1.17. Using R software and BB algorithms, we have \( \hat{\lambda} = 2.1362 \) and \( \hat{\theta} = 4.5236 \). Then, taking, for example, 6 intervals, then \( r = 6 \), and calculate the Fisher information matrix (FIMx) on the initial data, and we have

\[
I(\hat{\Phi}) = \begin{pmatrix} 7.6235 & 12.562 \\ 12.562 & 4.6325 \end{pmatrix}.
\]

Then, by calculating the NRR test statistic \( Y_{0.05}^2(5) \), we have \( Y_{0.05}^2(5) = 7.3265 \). Since

\[
Y_{0.05}^2(5) = 7.3265 < \chi^2_{0.05}(5) = 11.0705,
\]

we can accept the null hypothesis that the times between failures for repairable items data follow the RGG distribution.

7.1.2. Reliability Data. The second dataset includes the reliability data given by Cabarbaye et al. [23]. These data have undergone extensive examination and investigation, as several researchers and academics have modelled, examined, and concluded on them. The data are as follows: 0, 313, 360, 231, 286, 340, 212, 287, 243, 170, 141, 150, 593, 328, 234, 206, 108, 134, 231, 218, 281, 192, 457, 269, 201, 181, 277, 479, 272, 223, 272, 163, 370, 217, 182, 202, 451, 303. Using R software and BB algorithms, we have \( \hat{\lambda} = 1.3256 \) and \( \hat{\theta} = 9.6235 \). Then, taking, for example, 7 intervals, then \( r = 7 \), and calculate the FIMx on the initial data, and we have

\[
I(\hat{\Phi}) = \begin{pmatrix} 10.2368 & 15.6237 \\ 15.6237 & 9.5236 \end{pmatrix}.
\]

Then, by calculating the NRR test statistic \( Y_{0.05}^2(6) \), we have \( Y_{0.05}^2(6) = 8.6325 \). Since
Table 9: MLEs and their MSEs for the complete case.

<table>
<thead>
<tr>
<th>$\lambda \mid (\lambda_0 = 0.9)$</th>
<th>25</th>
<th>50</th>
<th>130</th>
<th>350</th>
<th>500</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>0.9274</td>
<td>0.9253</td>
<td>0.9222</td>
<td>0.9185</td>
<td>0.9043</td>
<td>0.9005</td>
</tr>
<tr>
<td>MSE</td>
<td>0.0069</td>
<td>0.0057</td>
<td>0.0042</td>
<td>0.0034</td>
<td>0.0021</td>
<td>0.0012</td>
</tr>
<tr>
<td>$\hat{\theta} \mid (\theta_0 = 2)$</td>
<td>1.9710</td>
<td>1.9775</td>
<td>1.9823</td>
<td>1.9875</td>
<td>1.9902</td>
<td>1.9978</td>
</tr>
<tr>
<td>MSE</td>
<td>0.0075</td>
<td>0.0064</td>
<td>0.0053</td>
<td>0.0046</td>
<td>0.0035</td>
<td>0.0022</td>
</tr>
</tbody>
</table>

Table 10: Theoretical risk and empirical risk for complete case.

<table>
<thead>
<tr>
<th>$\alpha_{i}, 1, 2, 3$</th>
<th>25</th>
<th>50</th>
<th>130</th>
<th>350</th>
<th>500</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_1 = 0.01$</td>
<td>0.0056</td>
<td>0.0064</td>
<td>0.0072</td>
<td>0.0082</td>
<td>0.0089</td>
<td>0.0095</td>
</tr>
<tr>
<td>$\alpha_2 = 0.05$</td>
<td>0.0282</td>
<td>0.0298</td>
<td>0.0334</td>
<td>0.0358</td>
<td>0.0398</td>
<td>0.0454</td>
</tr>
<tr>
<td>$\alpha_3 = 0.10$</td>
<td>0.0887</td>
<td>0.0899</td>
<td>0.0943</td>
<td>0.0963</td>
<td>0.0987</td>
<td>0.0999</td>
</tr>
</tbody>
</table>

Table 11: MLEs and their MSEs for the censored case.

<table>
<thead>
<tr>
<th>$\lambda \mid (\lambda_0 = 0.5)$</th>
<th>25</th>
<th>50</th>
<th>130</th>
<th>350</th>
<th>500</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>0.4743</td>
<td>0.4821</td>
<td>0.4879</td>
<td>0.4902</td>
<td>0.4956</td>
<td>0.4989</td>
</tr>
<tr>
<td>MSE</td>
<td>0.0066</td>
<td>0.0051</td>
<td>0.0046</td>
<td>0.0038</td>
<td>0.0023</td>
<td>0.0014</td>
</tr>
<tr>
<td>$\hat{\theta} \mid (\theta_0 = 1.5)$</td>
<td>1.5134</td>
<td>1.5106</td>
<td>1.5093</td>
<td>1.5053</td>
<td>1.5024</td>
<td>1.5012</td>
</tr>
<tr>
<td>MSE</td>
<td>0.0056</td>
<td>0.0049</td>
<td>0.0032</td>
<td>0.0026</td>
<td>0.0018</td>
<td>0.0009</td>
</tr>
</tbody>
</table>

Table 12: Theoretical risk and empirical risk for censored case.

<table>
<thead>
<tr>
<th>$\alpha_{i}, 1, 2, 3$</th>
<th>$n_i = 25$</th>
<th>$n_i = 50$</th>
<th>$n_i = 130$</th>
<th>$n_i = 350$</th>
<th>$n_i = 500$</th>
<th>$n_i = 1000$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_1 = 0.01$</td>
<td>0.0042</td>
<td>0.0058</td>
<td>0.0067</td>
<td>0.0072</td>
<td>0.0087</td>
<td>0.0099</td>
</tr>
<tr>
<td>$\alpha_2 = 0.05$</td>
<td>0.0422</td>
<td>0.0446</td>
<td>0.0453</td>
<td>0.0467</td>
<td>0.0488</td>
<td>0.0492</td>
</tr>
<tr>
<td>$\alpha_3 = 0.10$</td>
<td>0.0954</td>
<td>0.0961</td>
<td>0.0972</td>
<td>0.0986</td>
<td>0.0995</td>
<td>0.1010</td>
</tr>
</tbody>
</table>

\[
\chi^2_{0.05}(6) = 8.6325 < \chi^2_{0.05}(6) = 12.59159, \quad (34)
\]

we can accept the null hypothesis that the reliability data follow the RGG distribution.

7.1.3. Strength of Glass Fiber Data. The third dataset includes the strength of glass fiber data given by Smith and Naylor [24]. Additionally, as many researchers and academics have modelled, examined, and drawn several inferences from the strengths of fiber glass data, they have received considerable attention in statistical modelling. The data are as follows: 1.014, 1.081, 1.082, 1.185, 1.223, 1.248, 1.267, 1.271, 1.272, 1.275, 1.276, 1.286, 1.288, 1.292, 1.304, 1.306, 1.355, 1.361, 1.364, 1.409, 1.426, 1.459, 1.46, 1.476, 1.481, 1.484, 1.501, 1.506, 1.524, 1.526, 1.535, 1.541, 1.568, 1.579, 1.581, 1.591, 1.602, 1.666, 1.67, 1.684, 1.691, 1.704, 1.731, 1.735, 1.747, 1.748, 1.757, 1.800, 1.806, 1.867, 1.876, 1.878, 1.91, 1.916, 1.972, 2.012, 2.456, 2.592, 3.197, 4.121. Using R software and BB algorithms, we have $\bar{\lambda} = 2.6351$ and $\bar{\theta} = 7.3262$. Then, taking, for example, 7 intervals, then $r = 7$, and calculate the FIMx on the initial data, and we have

\[
I(\hat{\Phi}) = \begin{pmatrix} 4.7395 & 9.4185 \\ 9.4185 & 8.6134 \end{pmatrix}. \quad (35)
\]

Then, by calculating the NRR test statistic $Y^2_{0.05}(6)$, we have $Y^2_{0.05}(6) = 7.0362$. Since

\[
Y^2_{0.05}(6) = 7.0362 < \chi^2_{0.05}(6) = 12.59159, \quad (36)
\]

we can accept the null hypothesis that the glass fiber data follow the RGG distribution.

7.2. Real Applications for Censored Data

7.2.1. Times to Infection of Kidney Dialysis Patients. Consider data of times to infection of kidney dialysis patients (see Klein and Moeschberger [25]). Infection times: 1.5, 3.5, 4.5, 4.5, 5.5, 8.5, 8.5, 9.5, 10.5, 11.5, 15.5, 16.5, 18.5, 23.5, 26.5. Censored observations: 2.5, 2.5, 3.5, 3.5, 4.5, 5.5, 6.5, 6.5, 7.5, 7.5, 7.5, 8.5, 9.5, 10.5, 11.5, 12.5, 12.5, 13.5, 14.5, 14.5, 21.5, 21.5, 22.5, 22.5, 25.5, 27.5. Using R software and BB algorithms, we have $\hat{\lambda} = 1.02367$ and $\hat{\theta} = 3.4185$. Table 5 gives the values of $\bar{\lambda}_J, e_J(z), O_J(z), R_{1,\Phi}, R_{2,\Phi} \Phi$ under $r = 5$. Table 13 gives the values of $\bar{\theta}_J, e_J(z), O_J(z), R_{1,\Phi}, R_{2,\Phi} \Phi \Phi$ for data of times to infection of kidney dialysis patients.

Then, by calculating the modified NRR test statistic $M^2_{0.05}(5)$, we have $M^2_{0.05}(5) = 7.6324$. Since
we can accept the null hypothesis that the data of times to infection of kidney dialysis patients follow the RGG distribution.

7.2.2. The Bone Marrow Transplant Data. Consider data of times to infection of kidney dialysis patients for 38 patients (see Klein and Moeschberger [25]). Time to death: 1, 86, 107, 110, 122, 156, 162, 172, 243, 262, 262, 269, 276, 371, 417, 418, 466, 487, 526, 716, 781, 1111, 1182, 1199, 1279, 1377, 1433, 1496. Censored observations: 194, 226, 350, 530, 996, 1167, 1330, 1330, 1462, 1602, 2081. Using R software and BB algorithms, we have \( \hat{\lambda} = 3.1253 \) and \( \hat{\beta} = 3.4152 \). Table 6 gives the values of \( \hat{b}_j(z), e_j(z), O_j(z), \hat{K}_{1j}, \hat{K}_{2j}, \) under \( r = 4 \). Table 14 lists values of \( \hat{b}_j(z), e_j(z), O_j(z), \hat{K}_{1j}, \hat{K}_{2j}, \) for the bone marrow transplant data.

Then, by calculating the modified NRR test statistic \( M_{0.05}^2(4) \), we have \( M_{0.05}^2(4) = 6.2345 \). Since \( M_{0.05}^2(4) = 6.2345 < \chi^2_{0.05}(4) = 9.4877 \), we can accept the null hypothesis that the bone marrow transplant data follow the RGG distribution.

7.2.3. Strength of Certain Type of Braided Cord Data. We apply the findings from this analysis to real data derived from reliable sources (Crowder et al. [26]) for the third dataset. The forces of 48 pieces of cord that had withstood for a certain amount of time were examined as part of an experiment to learn more about the strength of a specific type of braided cord after the weather. Strength: 41.7, 43.9, 49.9, 50.1, 50.8, 51.9, 52.1, 52.3, 52.4, 52.6, 52.7, 53.1, 53.6, 53.9, 53.9, 54.1, 54.6, 54.8, 54.8, 55.1, 55.4, 55.9, 56, 56.1, 56.5, 56.9, 57.1, 57.1, 57.3, 57.7, 57.8, 58.1, 58.9, 59.1, 59.6, 60.4, 60.7. Censored observations: 26.8, 29.6, 33.4, 35, 36.3, 50.1, 50.8, 51.9, 52.1, 52.3, 52.4, 52.6, 52.7, 53.1, 53.6, 53.9, 53.9, 54.1, 54.6, 54.8, 54.8, 55.1, 55.4, 55.9, 56, 56.1, 56.5, 56.9, 57.1, 57.1, 57.3, 57.7, 57.8, 58.1, 58.9, 59.1, 59.6, 60.4, 60.7. Using R software and BB algorithms, we have \( \hat{\lambda} = 3.1253 \) and \( \hat{\beta} = 3.4152 \). Table 7 gives the values of \( \hat{b}_j(z), e_j(z), O_j(z), \hat{K}_{1j}, \hat{K}_{2j}, \) under \( r = 4 \). Table 15 gives the values of \( \hat{b}_j(z), e_j(z), O_j(z), \hat{K}_{1j}, \hat{K}_{2j}, \) for strength of certain type of braided cord data.

Then, by calculating the modified NRR test statistic \( M_{0.05}^2(4) \), we have \( M_{0.05}^2(4) = 5.9637 \). Since \( M_{0.05}^2(4) = 5.9637 < \chi^2_{0.05}(4) = 9.4877 \), we can accept the null hypothesis that the strength of certain type of braided cord data follow the RGG distribution.

8. Concluding Remarks

A novel continuous probability distribution called the Rayleigh generalized gamma (RGG) distribution is introduced and studied in this work, but we will approach it from fresh angles that depart from those often covered by scholars. In order to highlight more practical aspects in the areas of risk assessment and analysis, distributive verification, and its related practical applications on complete data and censored data, we chose to ignore many theoretical results and algebraic derivations, and this is not to say that they are not important. However, by presenting and discussing some novel characterizations based on some related theories, such as characterizations based on the two truncated moments, characterizations in terms of the hazard function, and characterizations based on the basis of the

| Table 13: Values of \( \hat{b}_j(z), e_j(z), O_j(z), \hat{K}_{1j}, \hat{K}_{2j} \) for times to infection of kidney dialysis patients. |
|\( b_j(z) \) | 5.4 | 8.4 | 12.9 | 18.4 | 27.5 |
|\( O_j(z) \) | 10 | 8 | 11 | 5 | 9 |
|\( \hat{K}_{1j} \) | 0.9356 | 1.0956 | 1.4263 | 1.2364 | 1.4326 |
|\( \hat{K}_{2j} \) | 1.3256 | 0.9423 | 0.8412 | 0.7136 | 0.8264 |
|\( e_j(z) \) | 5.5134 | 5.5134 | 5.5134 | 5.5134 | 5.5134 |

| Table 14: Values of \( \hat{b}_j(z), e_j(z), O_j(z), \hat{K}_{1j}, \hat{K}_{2j} \) for the bone marrow transplant data. |
|\( b_j(z) \) | 255 | 490 | 1190 | 2081 |
|\( O_j(z) \) | 11 | 10 | 8 | 9 |
|\( \hat{K}_{1j} \) | 0.9265 | 0.8135 | 0.6135 | 0.7366 |
|\( \hat{K}_{2j} \) | 1.0235 | 1.0465 | 0.8813 | 0.9436 |
|\( e_j(z) \) | 4.6523 | 4.6523 | 4.6523 | 4.6523 |

| Table 15: Values of \( \hat{b}_j(z), e_j(z), O_j(z), \hat{K}_{1j}, \hat{K}_{2j} \) for strength of certain type of braided cord data. |
|\( b_j(z) \) | 50 | 54 | 58 | 60.7 |
|\( O_j(z) \) | 11 | 14 | 16 | 7 |
|\( \hat{K}_{1j} \) | 1.0932 | 1.2388 | 1.4362 | 0.9843 |
|\( \hat{K}_{2j} \) | 1.9846 | 2.4196 | 3.5241 | 1.1635 |
|\( e_j(z) \) | 4.2153 | 4.2153 | 4.2153 | 4.2153 |
conditional expectation of a function of the random variable, we were able to cover some theoretical aspects of the RGG distribution. By analyzing a collection of commonly used financial indicators, such as the value-at-risk (VAR), tail-value-at-risk (TVAR), tail variance (TV), tail mean-variance (TMV), and mean excess loss (MEL) function, it is possible to analyze and evaluate the risks that insurance firms face. The maximum likelihood estimation approach, the ordinary least squares method, the weighted least squares estimation method, and the Anderson–Darling estimation method are all described as estimate strategies for the major important risk indicators. These four methods were used and applied for the actuarial evaluation, and a comparison is presented for determining the best method under a simulation study (for artificial assessment) and under an application to insurance claims data. The simulation is performed under three degrees of confidence, considering various sample sizes. With regard to the application to insurance claims data, the following results can be highlighted:

(1) For all risk assessment methods:

\[ \text{VAR}(Z_{(\tau=0.3)}) < \ldots < \text{VAR}(Z_{(\tau=0.01)}), \quad \text{TVAR}(Z_{(\tau=0.3)}) < \ldots < \text{TVAR}(Z_{(\tau=0.01)}), \]
\[ \text{TV}(Z_{(\tau=0.3)}) > \ldots > \text{TV}(Z_{(\tau=0.01)}), \quad \text{TMV}(Z_{(\tau=0.3)}) > \ldots > \text{TMV}(Z_{(\tau=0.01)}), \]
\[ \text{MEL}(Z_{(\tau=0.3)}) > \ldots > \text{MEL}(Z_{(\tau=0.01)}). \]

(2) Under the RGG model and the MLE method: The VAR \((Z)\) is monotonically increasing starts with 3496.50391 and ends with 3496.50391, the TVaR \((Z)\) in monotonically increasing starts with 4558.53397 and ends with 4558.53397. Under the GG model and the MLE method: The VAR \((Z)\) is monotonically increasing starts with 3278.09435 and ends with 3278.09435, the TVaR \((Z)\) in monotonically increasing starts with 859.59031 and ends with 859.59031.

(3) Under the RGG model and the OLSE method: The VAR \((Z)\) is monotonically increasing starts with 3632.99105 and ends with 7484.76402, the TVaR \((Z)\) in monotonically increasing starts with 3632.99105 and ends with 3632.99105, the TVaR \((Z)\) in monotonically increasing starts with 4718.69291 and ends with 7414.04414. Under the GG model and the OLSE method: The VAR \((Z)\) is monotonically increasing starts with 4718.69291 and ends with 7414.04414, the TVaR \((Z)\) in monotonically increasing starts with 4558.53397 and ends with 4558.53397, the TVaR \((Z)\) in monotonically increasing starts with 3632.99105 and ends with 3632.99105.

(4) Under the RGG model and the WLSE method: The VAR \((Z)\) is monotonically increasing starts with 3526.87158 and ends with 7484.20863, the TVaR \((Z)\) in monotonically increasing starts with 3526.87158 and ends with 3526.87158, the TVaR \((Z)\) in monotonically increasing starts with 4718.69291 and ends with 7414.04414. Under the GG model and the WLSE method: The VAR \((Z)\) is monotonically increasing starts with 3526.87158 and ends with 3526.87158, the TVaR \((Z)\) in monotonically increasing starts with 4718.69291 and ends with 7414.04414, the TVaR \((Z)\) in monotonically increasing starts with 7318.62976 and ends with 7318.62976, the TVaR \((Z)\) in monotonically increasing starts with 3526.87158 and ends with 3526.87158.

(5) Under the RGG model and the AE method: The VAR \((Z)\) is monotonically increasing starts with 3540.62976 and ends with 6859.01979, the TVaR \((Z)\) in monotonically increasing starts with 3540.62976 and ends with 3540.62976, the TVaR \((Z)\) in monotonically increasing starts with 7318.62976 and ends with 7318.62976, the TVaR \((Z)\) in monotonically increasing starts with 7318.62976 and ends with 7318.62976.

(6) For the RGG model: nearly for all \(q\) values, the OLSE method is recommended since it provides the most acceptable risk exposure analysis; then, the MLE method is recommended as a second one. For the GG model: nearly for all \(q\) values, the OLSE method is recommended since it provides the most acceptable risk exposure analysis; then, the MLE method is recommended as a second one.

(7) For all \(q\) in comparing various \(q\) values and risk methods, it has been found that the RGG (Risk Gamma-Gamma) model outperforms the GG (Gamma-Gamma) model. The RGG distribution demonstrates superior performance when used to model insurance claims reimbursement data and assess actuarial risk, even though both probability distributions have an equal number of parameters. Given this observation, it is expected that actuaries and practitioners will show significant interest in adopting the RGG distribution for future actuarial and applied research endeavors.

(8) For all risk estimation methods, TV \((Z)\), TMV \((Z)\), and MEL \((Z)\) are monotonically decreasing.

In the framework of distributional validation and statistical hypothesis tests for the complete data, the well-known Nikulin–Rao–Robson statistic \((Y^2_\alpha (r-1))\), which is based on the uncensored maximum likelihood estimators on initial non-grouped data, is considered under the RGG model. The \(Y^2_\alpha (r-1)\) statistic is assessed under a simulation study and under three real datasets as well, and the following results can be highlighted:

(i) For the uncensored times between failures for repairable items data: \(M^2_{0.05} (5) = 7.6324 < x^2_{0.05} (5) = 11.0705\); therefore, we can accept the null hypothesis that the times between failures for repairable items data follow the RGG distribution.

(ii) For the uncensored reliability data: \(Y^2_{0.05} (6) = 8.6325 < x^2_{0.05} (6) = 12.59159\); therefore, we can...
accept the null hypothesis that the reliability data follow the RGG distribution. 

(iii) For the uncensored strength data: $Y_{0.05}^2(6) = 7.0362 < \chi^2_{0.05}(6) = 12.59159$; therefore, we can accept the null hypothesis that the strength data follow the RGG distribution.

In the framework of distributional validation and statistical hypothesis tests for the censored data, a modified NRR statistic ($M^2(r)$), which is based on the censored maximum likelihood estimators on initial non-grouped data, is considered under the RGG model. The $M^2(r)$ statistic is assessed under comprehensive simulation study and under three real datasets, and the following results can be highlighted:

(i) For the censored times to infection of kidney dialysis patients data, $M^2_{0.05}(5) = 7.6324 < \chi^2_{0.05}(5) = 11.0705$; therefore, we can accept the null hypothesis that the data of times to infection of kidney dialysis patients follow the RGG distribution.

(ii) For the censored bone marrow transplant data, $M^2_{0.05}(4) = 6.2345 < \chi^2_{0.05}(4) = 9.4877$; therefore, we can accept the null hypothesis that the bone marrow transplant data follow the RGG distribution.

(iii) For the censored strength of certain type of braided cord data, $M^2_{0.05}(4) = 5.9637 < \chi^2_{0.05}(4) = 9.4877$; therefore, we can accept the null hypothesis that the strength of certain type of braided cord data follow the RGG distribution.

### Appendix

```r
#p represents parameters
library(BB)
library(nleqslv)
g1 <- function(p){n = 30; }
3b6 - (1 + P[1] * z) * exp(-P[1] * z) 
bt - (3b6^theta)/(1 - 3b6^theta)
wj - 1 - exp(-b^2) 

dd j - rep((NA, length(p)))
dd2 = sum(delta * (1/P[2] + 2 * ln(w) - 2 * 3b6^2 + ln(3b6) * (1 - b)) + sum((2 * P[1] * 3b6^2 * ln(3b6) + (1 + b) * exp(2061(-b^2)))/(1 - w)))))
dd)

# We can change it

gg (p0)
BBsolve (par = p0, fn = gg)
BBsolve (par = p0, fn = gg)$par
```

### Data Availability

The dataset used to support the findings of this study is available from the corresponding author upon request.

### Conflicts of Interest

The authors declare that they have no conflicts of interest.

### Acknowledgments

This study was funded by Researchers Supporting Project (RSP2023R488), King Saud University, Riyadh, Saudi Arabia.

### References


