

Research Article

Assessing the Performance of the Discrete Generalised Pareto Distribution in Modelling Non-Life Insurance Claims

S. K.-B. Dzidzornu  and R. Minkah 

*Department of Statistics and Actuarial Science, School of Physical and Mathematical Sciences,
College of Basic and Applied Sciences, University of Ghana, Accra, Ghana*

Correspondence should be addressed to R. Minkah; rminkah@ug.edu.gh

Received 23 February 2021; Accepted 21 May 2021; Published 12 June 2021

Academic Editor: Hyungjun Cho

Copyright © 2021 S. K.-B. Dzidzornu and R. Minkah. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The generalised Pareto distribution (GPD) offers a family of probability spaces which support threshold exceedances and is thus suitable for modelling high-end actuarial risks. Nonetheless, its distributional continuity presents a critical limitation in characterising data of discrete forms. Discretising the GPD, therefore, yields a derived distribution which accommodates the count data while maintaining the essential tail modelling properties of the GPD. In this paper, we model non-life insurance claims under the three-parameter discrete generalised Pareto (DGP) distribution. Data for the study on reported and settled claims, spanning the period 2012–2016, were obtained from the National Insurance Commission, Ghana. The maximum likelihood estimation (MLE) principle was adopted in fitting the DGP to yearly and aggregated data. The estimation involved two steps. First, we propose a modification to the μ and $(\mu + 1)$ frequency method in the literature. The proposal provides an alternative routine for generating initial estimators for MLE, in cases of varied count intervals, as is a characteristic of the claim data under study. Second, a bootstrap algorithm is implemented to obtain standard errors of estimators of the DGP parameters. The performance of the DGP is compared to the negative binomial distribution in modelling the claim data using the Akaike and Bayesian information criteria. The results show that the DGP is appropriate for modelling the count of non-life insurance claims and provides a better fit to the regulatory claim data considered.

1. Introduction

Non-life or general insurance involves the provision of financial loss protection against risks on interests other than life, such as buildings, vehicles, machinery, and equipment. Conditioned on periodic payments or one-off advance of a predetermined amount, called premium, non-life policies are designed to provide coverage against the occurrence of the insured probabilistic events for individuals, private organisations, and public institutions. The payments effected in response to occurrences of such events are termed as insurance claims (Wuthrich [1]). The non-life insurance claim process is characterised by two quantities: claim frequency or count and claim severity or size. As noted by Renshaw [2] and Özgürel [3], the underlying expectations of claim frequency and severity, quantified as a product, are foremost considerations in the computation of pure or risk premiums.

The main objective of this paper is to illustrate that the discrete generalised Pareto (DGP) distribution can be employed to model the counts of non-life insurance claims, collated by an insurance regulatory authority from a licensed class of insurers. In the absence of suitable actuarial models, non-life insurers largely encounter difficulties in conducting evidence-based assessment of risks insured, often resulting in the miscalculation of premiums and inability to settle claims when due. In response, developing probability models that describe claim frequencies offers a distributional framework for evaluating risks to facilitate premium setting and liquidity reserving by non-life insurance service providers.

A random variable X that follows a DGP with shape (α), scale (λ), and location (μ) parameters is denoted by $X \sim \text{DGP}(\alpha, \lambda, \mu)$. This is a parametric model obtained by discretising the continuous generalised Pareto distribution, introduced by Pickands [4], and is particularly noted for tail

modelling properties. The DGP assumes varying forms based on the omission of one or two of its parameters. For instance, if the location parameter $\mu = 0$, the DGP transforms into the discrete Lomax distribution, $DLo(\alpha, \lambda)$.

A fair amount of research has demonstrated the application of probability models to the study of non-life insurance claims. Among the relevant literature, selected studies have principally explored the subject with reference to standard probability distributions from the outlook of randomised spatial effects (Gschlößl and Czado [5]), collective risk simulation (Pacáková [6]), and incorporation of covariates (Renshaw [2]). In this paper, we attempt to contribute to bridging the research gap by providing a method of fitting the DGP to the non-life claims data from the National Insurance Commission in Ghana. Furthermore, similar to the work of Prieto et al. [7], we assess the performance of the DGP by comparing it with the negative binomial distribution.

This study contributes to the field of claims modelling in threefold. First, it proposes a count-based data structure for the analysis of non-life insurance claim frequencies to enhance the precision of statistical models. Second, a modification is made to the μ and $\mu + 1$ frequency method of Prieto et al. [7], to obtain initial estimators of the DGP for the claims data characterised by varying discrete observational intervals. Third, the algorithm implemented under the estimation of the parameters of the DGP offers a resource to researchers on performance analysis in future statistical and/or actuarial work.

The rest of this paper is organised as follows. In Section 2, the methodology is presented including the maximum likelihood estimation of the parameters of the distributions and model selection criteria. Section 3 presents the data and the arrangements needed to put the data into a form necessary for the model fitting. Lastly, in Section 4, we present concluding remarks.

2. Methodology

This section presents the systematic approach followed to model the reported and settled claim datasets. Specifically, the section entails the description of the probability distributions, parameter estimation, and model selection criteria.

The parameter estimation method, maximum likelihood estimation (MLE) technique, is used in fitting the models. Consider the case where a random variable is available from a population with a known probability distribution except for its parameter $\theta \in \mathbb{R}^d$. The maximum likelihood principle suggests that the criterion for making the selection should be the probability (or likelihood) with which a particular distribution can produce the given sample. The value of θ for that distribution is the maximum likelihood estimate of the unknown parameter, $\hat{\theta}$.

Suppose $\mathbf{x} = \{x_1, x_2, \dots, x_n\}'$ is an independent random sample of size n from a distribution with dependence on one or more unknown parameters $\theta = \{\theta_1, \theta_2, \dots, \theta_d\}'$. Let $f(x_i; \theta)$ be the probability density (or mass) function of x_i , with θ restricted to a given parameter space $\Omega \in \mathbb{R}^d$. The likelihood function of the sample is given by

$$L(\theta; \mathbf{x}) = \prod_{i=1}^n f(x_i). \quad (1)$$

The maximum likelihood estimator, $\hat{\theta}$, of θ is the solution to the equation

$$\frac{\partial}{\partial \theta} L(\theta; \mathbf{x}) = 0. \quad (2)$$

Usually, $L(\theta; \mathbf{x})$ may involve exponentials, and hence, $\ln L(\theta; \mathbf{x})$ is maximised. Since the logarithm of a function increases or decreases with the function, the maximiser of $L(\theta; \mathbf{x})$ also maximises $\ln L(\theta; \mathbf{x})$.

2.1. Negative Binomial Distribution. The negative binomial is a discrete probability distribution which characterises the number of successes in a sequence of independent and identically distributed Bernoulli trials before a specified number of failures (denoted r) occur. Suppose a sequence of Bernoulli trials is observed. By definition, a turn of each trial yields two possible outcomes, success or failure, with respective probabilities of occurrence denoted by p and $1 - p$. Also, the trials are independent, and p remains constant for each trial. If X represents the number of trials, or failures, prior to the r -th success, then X follows a negative binomial distribution with probability mass function:

$$P(X = x) = \binom{x-1}{r-1} (1-p)^{x-r} p^r, \quad (3)$$

for $x = r, r+1, r+2, \dots$

The geometric distribution is a special case of the negative binomial, where the Bernoulli trial discontinues at the first failure, $r = 1$. Since the negative binomial may be represented by alternative distributional parametrisations, three factors inform distinctions: starting point of the support, whether at $x = 0$ or $x = r$; definition of p , whether it represents the probability of success or failure; and interpretation of r , whether it denotes the number of successes or failures (DeGroot and Schervish [8]).

Given N independent and identically distributed claim count observations, (k_1, \dots, k_N) , the likelihood function can be expressed as

$$L(r, p) = \prod_{i=1}^N P(X = k_i; r, p). \quad (4)$$

Substituting (3) into (4) and taking the logarithm results in the log-likelihood function given by

$$\begin{aligned} \ell(r, p) = & \sum_{i=1}^N \ln(\Gamma(k_i + o|o_r)) - \sum_{i=1}^N \ln(k_i!) - N \ln(\Gamma(r)) \\ & + \sum_{i=1}^N k_i \ln(p) + Nr \ln(1 - p). \end{aligned} \quad (5)$$

To maximise equation (5), the partial derivative with respect to r and p is set to zero:

$$\frac{\partial l(r, p)}{\partial p} = \left[\sum_{i=1}^N k_i \frac{1}{p} \right] - Nr \frac{1}{1-p} = 0, \quad (6)$$

$$\frac{\partial l(r, p)}{\partial r} = \left[\sum_{i=1}^N \psi(k_i + r) \right] - N\psi(r) + N \ln(1-p) = 0. \quad (7)$$

Here, the digamma function $\psi(k) = \Gamma'(k)/\Gamma(k)$. Furthermore, solving for p in equation (4) produces

$$\hat{p} = \frac{\sum_{i=1}^N k_i}{Nr + \sum_{i=1}^N k_i}. \quad (8)$$

Finally, substituting p in (7) yields

$$\frac{\partial l(r, p)}{\partial r} = \left[\sum_{i=1}^N \psi(k_i + r) \right] - N\psi(r) + N \ln\left(\frac{r}{r + \sum_{i=1}^N k_i/N}\right) = 0. \quad (9)$$

The form of (9) suggests that a closed-form solution for r may not be obtained analytically. Therefore, numerical methods can be used in order to obtain estimators of r and p . We adopt an alternative appealing reparametrisation, $X \sim \text{NegativeBinomial}(r, m/(m+r))$, where m and r represent the mean and dispersion (or shape) parameters, respectively (see, e.g., Piegorsch [9]). For example, in R Core Team [10], the function `fitdistr` in the MASS package provides a routine for estimating the parameters of the negative binomial distribution with reference to the alternative reparametrisation.

2.2. Discrete Generalised Pareto Distribution. The DGP arises from a discretisation of the continuous generalised Pareto distribution. To provide a basis for the discussion on the discrete generalised Pareto, the distribution functions of Pareto type I and generalised Pareto are given as

$$D(x) = 1 - \left(\frac{b}{x}\right)^a, \quad x \geq b, \quad (10)$$

$$G(x) = 1 - \left[1 + \frac{\alpha(x-\mu)}{\lambda}\right]^{(-1/\alpha)}, \quad (11)$$

$$1 + \alpha(x-\mu)/\lambda > 0, x > \mu, \lambda > 0,$$

respectively. The generalised Pareto distribution is noted for its ability to model tails of distribution functions (see, e.g., [11, 12]). Upon discretisation of the generalised Pareto distribution, the resultant DGP inherits the prior continuous properties in forms adapted to the discrete probability space.

From the stated generalised Pareto distribution in (11), the probability mass function of the DGP can be formally deduced. First, consider the cumulative distribution function of the DGP expressed as

$$F(x) = P(X \leq x) = 1 - [1 + \lambda(x - \mu + 1)]^{-\alpha}, \quad (12)$$

$$x = \mu, \mu + 1, \dots,$$

where $\alpha, \lambda, \mu > 0$ and $F(x) = 0$ if $x < \mu$.

Also, Krishna and Pundir [13] addressed the discretisation of a continuous model by observing unit groupings on the failure time axis. The authors reasoned that, for a continuous failure time X , with survival function $S(x) = P[X > x]$ and time groupings of intervals $dX = \lfloor X \rfloor$, the discrete observed variable, dX , would have the probability mass function

$$P(dX = x) = P(x \leq X < x + 1) = S(x) - S(x + 1), \quad (13)$$

$$x = 0, 1, 2, \dots$$

Next, consider a standard generalisation for the survival function from Xekalaki [14]:

$$S(x) = P(X \geq x) = 1 - F(x - 1). \quad (14)$$

Then, employing equations (12) and (14), the survival function of the DGP is given:

$$S(x) = [1 + \lambda(x - \mu)]^{-\alpha}, \quad x = \mu, \mu + 1, \dots \quad (15)$$

Finally, evaluating equations (13) and (15) simultaneously results in the DGP

$$f(x) = [1 + \lambda(x - \mu)]^{-\alpha} - [1 + \lambda(x - \mu + 1)]^{-\alpha}, \quad (16)$$

$$x = \mu, \mu + 1, \dots$$

Suppose x_1, \dots, x_n is a sample of size n from a DGP. The parameters α and λ are estimated on the assumption that μ is known since $\hat{\mu} = x_{\min} \leq x_i, \forall i$. Adopting the μ and $(\mu + 1)$ frequency method of Prieto et al. [7], the initial values $(\alpha_0, \lambda_0, \mu_0)$ can be obtained and used as seed estimators in the subsequent maximum likelihood operation. Thus, the relative frequencies of $x = \mu$ and $x = (\mu + 1)$, respectively, denoted by \hat{f}_μ and $\hat{f}_{\mu+1}$, are calculated from the sample data. Analogously, α and λ are determined by substituting $x = \mu$ and $x = (\mu + 1)$ into the DGP probability mass function in (16) and equating the expressions to their respective \hat{f}_μ and $\hat{f}_{\mu+1}$ values.

However, the μ and $(\mu + 1)$ frequency method assumes that the count data used are observed in increasing steps of 1. However, in real-life situations, such as the data presented in Section 3.1, they may exhibit variation of intervals other than 1. Therefore, applying the method strictly on the count data results in generating several $\mu + 1 = 0$ and hence leading to $\hat{f}_{\mu+1} = 0$.

In this regard, proceeding with the computations with zero relative frequencies will result in a loss of essential frequency information in the dataset. As a result, we provide a modification of the method as μ and $(\mu + \epsilon)$ frequency methods, where $\epsilon > 0$ and $(\mu + \epsilon)$ is the smallest observation larger than the minimum, μ . Therefore, the estimators of α and λ are obtained by solving the resulting expressions,

$$\hat{f}_\mu = 1 - [1 + \lambda]^{-\alpha}, \quad (17)$$

and

$$\hat{f}_{\mu+\varepsilon} = [1 + \lambda]^{-\alpha} - [1 + 2\lambda]^{-\alpha}, \quad (18)$$

simultaneously. The expression in (19) results after α is eliminated from equations (17) and (18):

$$\frac{\ln(1 + 2l)}{\ln(1 + l)} = \frac{\ln(1 - \hat{f}_\mu - \hat{f}_{\mu+\varepsilon})}{\ln(1 - \hat{f}_\mu)}. \quad (19)$$

Following from this, (20) is obtained by appropriate substitutions into (19):

$$\hat{\alpha} = \frac{\ln(1 - \hat{f}_\mu)}{\ln(1 + \hat{\lambda})}. \quad (20)$$

Next, the maximum likelihood estimation method is employed to obtain estimators of the parameters of the discrete generalised Pareto. The log-likelihood function is constructed as

$$\begin{aligned} \ln \ell(\lambda, \alpha) &= \sum_{i=1}^n \ln f(x_i) \\ &= \sum_{i=1}^n \ln [(1 + \lambda(x_i - \mu))^{-\alpha} - (1 + \lambda(x_i - \mu + 1))^{-\alpha}], \end{aligned} \quad (21)$$

where $f(x_i)$ refers to the probability mass function specified in (16). Partial derivatives of (21) are taken with respect to α and λ and set to zero to obtain normal equations:

$$\frac{\partial \ln \ell}{\partial \alpha} = \sum_{i=1}^n \frac{\ln[1 + \lambda(x_i - \mu + 1)]}{[1 + \lambda(x_i - \mu + 1)]^\alpha [1 + \lambda(x_i - \mu)]^{-\alpha} - 1} - \sum_{i=1}^n \frac{\ln[1 + \lambda(x_i - \mu)]}{1 - [1 + \lambda(x_i - \mu)]^\alpha [1 + \lambda(x_i - \mu + 1)]^{-\alpha}} = 0, \quad (22)$$

$$\begin{aligned} \frac{\partial \ln \ell}{\partial \lambda} &= \sum_{i=1}^n \frac{\alpha(x_i - \mu + 1)}{[1 + \lambda(x_i - \mu + 1)]^{\alpha+1} [1 + \lambda(x_i - \mu)]^{-\alpha} - [1 + \lambda(x_i - \mu + 1)]} \\ &\quad - \sum_{i=1}^n \frac{\alpha(x_i - \mu)}{[1 + \lambda(x_i - \mu)] - [1 + \lambda(x_i - \mu)]^{\alpha+1} [1 + \lambda(x_i - \mu + 1)]^{-\alpha}} = 0. \end{aligned} \quad (23)$$

To proceed with the estimation of the parameters of the DGP, an algorithm was implemented in R to perform the following operations:

A1. Specify log-likelihood function (21) based on DGP probability mass function (16). The log-likelihood function is set to return a negation of the log-likelihood value since the R *optim* function is a minimiser. In effect, minimising the negated log-likelihood function at the initial estimates produces the equivalent of maximising the log-likelihood function.

A2. Optimise the log-likelihood functions in (22) and (23) at the seed values by simulated annealing (SANN), a variant of the B elisle [15] technique.

A3. Extract the estimated parameters, α and λ , from the output generated in A2, and compute the standard errors of the estimators using bootstrap resampling of Efron and Tibshirani [16].

2.3. Model Selection Criteria. The Akaike [17] and Schwarz [18] information criteria, denoted by AIC and BIC, respectively, form the basis for selecting the suitable model. The AIC and BIC are stated as follows:

$$\begin{aligned} \text{AIC} &= -2 \ln L + 2d, \\ \text{BIC} &= -2 \ln L + d \ln(n). \end{aligned} \quad (24)$$

While d and n represent the number of parameters and size of the sample, respectively, $\ln L$ specifies the log-

likelihood of the model evaluated at the maximum likelihood estimates. Thus, L is the maximum value of the likelihood function associated with the model.

In comparison with AIC, BIC addresses the issue of overfitting with a factor, $\ln(n)$, thereby placing a higher penalty for model complexity (Dziak et al. [19]). In statistical decision-making, a candidate model with minimum AIC and/or BIC values is selected.

3. Data and Model Fitting

3.1. Data. The study employs secondary data on non-life claims from the National Insurance Commission (NIC) of Ghana for the five-year period, 2012 to 2016. The historical data cover insurance claims of 29 non-life service providers. For each fiscal year, the dataset indicates the total number of claims administered under each of the five business classes of non-life insurance in Ghana. The classes are Fire, Burglary, and Property Damage; Accident; Marine and Aviation; Motor; and General Liability. The claim data are organised into three categories: incurred but not reported (IBNR), reported but not settled (RBNS), and settled but outstanding (SEBO), each bearing the standard actuarial definitions.

However, since IBNR is necessarily an estimate, the study focuses on RBNS and SEBO, hereinafter referred to as reported and settled claims, respectively. Overall, the data consist of 3,878,355 non-life insurance policies, generating 39,563 reported claims, of which 5,210 claims were settled within the period.

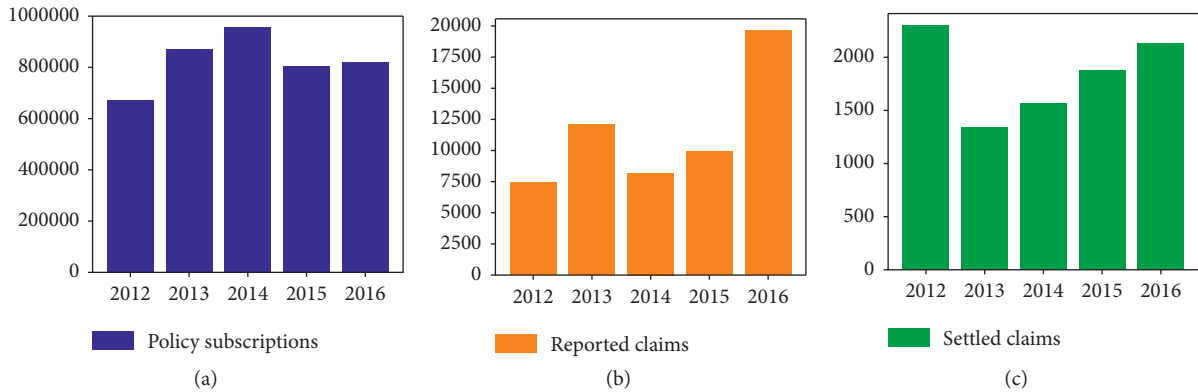


FIGURE 1: Claim data: number of non-life policies, reported claims, and settled claims recorded in the period 2012–2016.

Figure 1 provides an overview of the annual aggregates for policy subscriptions, reported claims, and settled claims. Although policy subscriptions have seen a decrease from 2014 onwards, the number of reported and settled claims has been increasing within the period. This observation shows some evidence of potential liquidity challenges for the non-life insurers if the trend persists into the future.

Following Prieto et al. [7], the dataset is organised into a structure which enables the fit of the discrete distributions to the frequency of occurrence of reported and settled claims. Tables 1 and 2 present descriptive statistics on the reported and settled claim datasets, respectively. For each year of the period considered, the skewness indicates the extent of symmetry and shows that there is positive skewness for the distribution of the count of claims. Also, among the reported claims, the fiscal year 2016 recorded some unusually large values culminating in its large kurtosis value. Similar results can be found in 2013 for the settled claim data.

In addition, Tables 3 and 4 record the individual observations of reported and settled claim counts with corresponding frequencies. For instance, in 2012, there were two records of reported claim count of 19 and six records of cases where no claims were settled.

It should be noted that the respective columns for the count frequencies for both reported and settled claims sum up to 29. Thus, each of them tallies with the total number of non-life insurers from whom records are collated by the National Insurance Commission. Lastly, nonsettlement of reported claims, among other reasons, may result from the eligibility of reported interest, proximity of the cause of the insured event, and noncompliance with coverage provisions of the insurance policy.

3.2. *Model Fitting and Discussion of Results.* This section presents the outcomes of the model fitting methods discussed in Section 2 on the claim data from the preceding section.

The parameter estimates are obtained through the maximum likelihood method. The maximum likelihood estimation of the negative binomial and DGP parameters is performed in R. The negative binomial parameters are

TABLE 1: Descriptive statistics of reported claims.

Year	Mean	Std. dev.	Skewness	Kurtosis	Min	Max
2012	255.86	329.95	1.28	0.71	0	1187
2013	414.9	620.11	1.58	1.12	0	2053
2014	281.9	350.03	1.57	1.47	0	1335
2015	341.72	503.77	2.12	4.07	0	2151
2016	676.59	1839.98	4.34	18.81	2	9902

TABLE 2: Descriptive statistics of settled claims.

Year	Mean	Std. dev.	Skewness	Kurtosis	Min	Max
2012	79.21	123.42	1.85	2.34	0	452
2013	46.24	63.22	2.52	7.35	0	307
2014	54.03	65.38	2.24	5.42	0	307
2015	64.31	90.59	1.93	3.03	0	362
2016	73.38	106.75	2.19	4.44	2	452

TABLE 3: Restructured frequency of reported claims.

Number of reported claims	Frequency of counts by year				
	2012	2013	2014	2015	2016
0	6	5	4	4	3
2	0	0	0	0	2
5	0	1	0	0	0
10	1	0	0	0	0
19	2	0	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮
1520	0	0	0	1	0
1533	0	1	0	0	0
2053	0	2	0	0	0
2151	0	0	0	1	1
9902	0	0	0	0	1
Total	29	29	29	29	29

estimated using the *mle* function and their standard arguments in the *fitdistrplus* package. However, to the best of the authors’ knowledge, no statistical package exists for estimating the parameters of the DGP in R. Therefore, the authors wrote an R-function for estimating the parameters of the DGP, using algorithms A1–A3, and it is available upon request. In addition, the selection criteria for model

TABLE 4: Restructured frequency of settled claims.

Number of settled claims	Frequency of counts by year				
	2012	2013	2014	2015	2016
0	6	5	4	4	3
1	1	1	0	0	0
2	0	1	0	1	2
3	1	0	0	0	0
7	1	0	0	0	1
⋮	⋮	⋮	⋮	⋮	⋮
180	1	0	0	0	1
307	0	1	1	1	0
362	1	0	0	1	1
392	1	0	0	0	0
452	1	0	0	0	1
Total	29	29	29	29	29

TABLE 5: Parameter estimates from the negative binomial.

Variable	Estimator	2012	2013	2014	2015	2016
Reported claims	\hat{r}	0.4915	0.6874	1.0370	0.8603	0.4518
	s.e	(0.1234)	(0.1661)	(0.2605)	(0.2121)	(0.1052)
	\hat{m}	296.8624	542.4572	327.0975	396.9273	783.7698
	s.e	(84.7788)	(130.9043)	(64.3623)	(85.7927)	(233.2420)
Settled claims	\hat{r}	0.4225	0.6735	1.1312	0.7453	0.7815
	s.e	(0.1063)	(0.1764)	(0.3002)	(0.1878)	(0.1912)
	\hat{m}	88.3519	51.5716	60.2751	71.7227	81.8596
	s.e	(26.7214)	(12.4032)	(11.2194)	(16.3761)	(18.2495)

TABLE 6: Parameter estimates from DGP.

Variable	Estimator	2012	2013	2014	2015	2016
Reported claims	$\hat{\mu}$	7.3983	4.8507	5.8075	6.2320	2.2437
	s.e	(3.4175)	(3.9254)	(3.8876)	(5.6555)	(2.2000)
	$\hat{\alpha}$	0.0128	0.0044	0.0020	0.0013	0.0102
	s.e	(0.0104)	(0.0011)	(0.0019)	(0.0010)	(0.0017)
	$\hat{\lambda}$	3.3700	2.7590	2.8036	2.4566	4.9025
	s.e	(0.1318)	(0.8976)	(0.0106)	(0.0829)	(2.3947)
Settled claims	$\hat{\mu}$	3.5866	3.3776	4.6485	3.6443	3.0205
	s.e	(1.0333)	(0.1015)	(3.4941)	(0.6387)	(2.5885)
	$\hat{\alpha}$	3.4157	2.9258	2.0675	2.3356	0.1121
	s.e	(0.0532)	(0.0200)	(0.0235)	(0.0064)	(0.0311)
	$\hat{\lambda}$	3.4015	9.4855	3.1532	3.1127	2.8593
	s.e	(0.9538)	(3.0089)	(0.2831)	(1.5265)	(0.9741)

comparison are presented for the individual years and the aggregated claim data for the five-year period.

3.2.1. Parameter Estimation for Yearly Data. Table 5 shows the parameters of the negative binomial distribution estimated using the alternative reparametrisation in Piegorsch [9]. The standard errors are placed in parenthesis.

Also, Table 6 presents estimates from the estimators, $\hat{\mu}$, $\hat{\alpha}$, and $\hat{\lambda}$, representing the estimated DGP location, shape, and scale parameters, respectively. The bootstrap standard errors are placed in parenthesis.

In terms of reported claims' count, Tables 7 and 8 show that the DGP model presents smaller AIC and BIC

values, in comparison with the negative binomial model. The observation is consistent across the fiscal timeframe under consideration. In addition, for the settled claim counts, the DGP is preferred as it exhibits smaller AIC and BIC values throughout the period as shown in Tables 7 and 8. Therefore, the DGP is recommended as it provides a better fit to both classes of the non-life insurance claim data.

3.2.2. Parameter Estimation for Aggregated Data. This section presents the results of fitting the negative binomial and DGP to the aggregated 5-year count data on reported and settled claims. The results of the parameter estimation

TABLE 7: AIC statistics for the negative binomial and DGP.

Distribution	2012	2013	2014	2015	2016
<i>Reported claims</i>					
Negative binomial model	329.0927	366.1941	343.5535	352.7973	372.5812
DGP model	55.9362	55.9999	55.9998	55.9999	55.9972
<i>Settled claims</i>					
Negative binomial model	274.8660	259.0694	269.3573	277.0937	284.2884
DGP model	57.9813	57.9958	57.9476	57.9653	57.8863

TABLE 8: BIC statistics for the negative binomial and DGP.

Distribution	2012	2013	2014	2015	2016
<i>Reported claims</i>					
Negative binomial model	331.5304	368.6319	345.9913	355.2351	375.0189
DGP model	57.1299	57.1936	57.1935	57.1936	57.1909
<i>Settled claims</i>					
Negative binomial model	277.3822	261.5856	271.8735	279.6099	286.8046
DGP model	59.1750	59.1895	59.1413	59.1590	59.0800

TABLE 9: Parameter estimates from the negative binomial.

Variable	Estimator (s.e)	
	\hat{r}	\hat{m}
Reported claims	0.6538 (0.0710)	477.0055 (53.2312)
Settled claims	0.8596 (0.0978)	74.7975 (7.3175)

TABLE 10: Parameter estimates from DGP.

Variable	Estimator (s.e)		
	$\hat{\mu}$	$\hat{\alpha}$	$\hat{\lambda}$
Reported claims	2.3354 (1.9731)	0.0017 (0.0012)	2.8171 (2.0240)
Settled claims	1.9089 (1.6214)	0.0077 (0.0015)	1.9333 (1.4524)

TABLE 11: Information criteria for model fitting.

Model	AIC	BIC
<i>Reported claims</i>		
Negative binomial model	1750.1770	1755.8010
DGP model	249.9887	251.1824
<i>Settled claims</i>		
Negative binomial model	1311.1770	1316.8010
DGP model	251.9658	253.1595

for the negative binomial and DGP are presented in Tables 9 and 10, respectively.

Comparing the negative binomial and DGP, Table 11 shows the AIC and BIC values for the fit of the two probability distributions. It is obvious that the DGP model produces smaller AIC and BIC values for the aggregate reported claim counts. Also, regarding the aggregate settled claim counts, smaller AIC and BIC values are produced by the DGP. Therefore, in alignment with the year-based modelling conclusion, the DGP is recommended as it provides a better fit to both classes of yearly and the aggregated non-life insurance claims data.

4. Conclusion

This study demonstrates that non-life insurance claims can be described by the three-parameter discrete generalised Pareto distribution. Relative to the negative binomial model, the DGP was observed to provide a better fit to the non-life reported and settled claim counts, as evident from the information criterion values under both yearly and aggregated data scenarios.

First, in organising the regulatory claim data for the distributional investigation, the study disaggregated each dataset into the observed count of claims and the corresponding frequencies, thereby providing an explicit count-frequency breakdown for informed probabilistic modelling. Furthermore, in deriving initial estimators (α_0 , λ_0 , and μ_0) to evaluate the DGP parameters, the μ and $(\mu + 1)$ frequencies of Prieto et al. [7] were adapted to μ and $\mu + \epsilon$, with $\epsilon > 0$, where $\mu + \epsilon$ is the smallest count value larger than the sample minimum, μ . The modified frequency routine involving μ and $\mu + \epsilon$ extends the application of the μ and $(\mu + 1)$ frequency method to the real-world count data exhibiting varying observational intervals.

To assess the strict response of the DGP to the claim count data, the study is conducted from a pure distributional perspective in the absence of explanatory variables. In future studies, however, incorporating relevant covariates into the modelling framework may offer additional insights for enhanced performance evaluation of the DGP. The outcomes will complement the present study in contributing towards optimality considerations for the allocation of premium funds by non-life insurance service providers.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] M. V. Wuthrich, “Non-life insurance: mathematics & statistics,” *SSRN Electronic Journal*, 2013.
- [2] A. E. Renshaw, “Modelling the claims process in the presence of covariates,” *ASTIN Bulletin*, vol. 24, no. 2, pp. 265–285, 1994.
- [3] B. Özgürel, “Risk premium in motor vehicle insurance,” *Dokuz Eylül Üniversitesi İktisadi İdari Bilimler Fakültesi Dergisi*, vol. 20, no. 2, pp. 47–61, 2005.
- [4] J. Pickands III, “Statistical inference using extreme order statistics,” *The Annals of Statistics*, vol. 3, no. 1, pp. 119–131, 1975.
- [5] S. Gschlößl and C. Czado, “Spatial modelling of claim frequency and claim size in non-life insurance,” *Scandinavian Actuarial Journal*, vol. 2007, no. 3, pp. 202–225, 2007.
- [6] V. Pacáková, “Modelling and simulation in non-life insurance,” in *Proceedings of the 5th International Conference on Applied Mathematics, Simulation and Modelling*, WSEAS Press, Corfu Island, Greece, July 2011.
- [7] F. Prieto, E. Gómez-Déniz, and J. M. Sarabia, “Modelling road accident blackspots data with the discrete generalized Pareto distribution,” *Accident Analysis & Prevention*, vol. 71, pp. 38–49, 2014.
- [8] M. H. DeGroot and M. J. Schervish, *Probability and Statistics*, London, UK, 2012.
- [9] W. W. Piegorsch, “Maximum likelihood estimation for the negative binomial dispersion parameter,” *Biometrics*, vol. 46, no. 3, pp. 863–867, 1990.
- [10] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2017.
- [11] A. C. Davison and R. L. Smith, “Models for exceedances over high thresholds,” *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 52, no. 3, pp. 393–425, 1990.
- [12] J. Beirlant, Y. Goegebeur, J. Segers, and J. L. Teugels, *Statistics of Extremes: Theory and Applications*, p. 514, Wiley, England, UK, 2004.
- [13] H. Krishna and S. Pundir, “Discrete burr and discrete Pareto distributions,” *Statistical Methodology*, vol. 6, no. 2, pp. 177–188, 2009.
- [14] E. Xekalaki, “Hazard functions and life distributions in discrete time,” *Communications in Statistics-Theory and Methods*, vol. 12, no. 21, pp. 2503–2509, 1983.
- [15] C. J. P. Bélisle, “Convergence theorems for a class of simulated annealing algorithms on \mathbb{R}^d ,” *Journal of Applied Probability*, vol. 29, no. 4, pp. 885–895, 1992.
- [16] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*, Chapman & Hall, London, UK, 1993.
- [17] H. Akaike, “A new look at the statistical model identification,” *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.
- [18] G. Schwarz, “Estimating the dimension of a model,” *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.
- [19] J. J. Dziak, D. L. Coffman, S. T. Lanza, R. Li, and L. S. Jermini, “Sensitivity and specificity of information criteria,” *Briefings in Bioinformatics*, vol. 21, no. 2, pp. 553–565, 2020.