STATISTICS IN TRANSITION new series, December 2025
Vol. 26, No. 4, pp. 89–110, https://doi.org/10.59139/stattrans-2025-041
Received – 23.11.2024; accepted – 18.05.2025

The truncated Schröter recursive algorithm for the computation of aggregate claim amounts

Friday I. Agu[1]

Abstract

This study introduces and evaluates the truncated Schröter recursive algorithm for computing aggregate claim amounts in the insurance sector. The algorithm addresses the limitations in the existing methods by incorporating truncation at 1, which is crucial for an accurate modelling of insurance claims where the events leading to a claim are pivotal. Using the AutoCollision dataset, the study compares the truncated Schröter algorithm with the Panjer and Schröter recursion algorithms, focusing on computational efficiency and accuracy. Furthermore, the descriptive statistics revealed substantial variability and risk factors, such as higher claim severity for business-use vehicles and young drivers aged 17–20. The results demonstrate that the truncated Schröter algorithm substantially reduces the execution time while maintaining high accuracy, thus making it a superior tool for risk management and premium setting.

Key words: insurance claim amounts, aggregate claim distribution, recursive algorithm, insurance risk management, computational efficiency.

1. Introduction

In the insurance domain, company profits depend largely on the premiums collected from policyholders and the claim amounts paid to insured individuals. Unlike in other market sectors, such as manufacturing, determining the appropriate premium for an insurance portfolio is particularly challenging. This complexity arises from the need to account for future uncertainties and ensure sustained and adequate investment. To address this, insurance companies employ models designed to accurately compute aggregate claim amounts within a collective risk framework and estimate the probability that total claims will not exceed a specified threshold. The process begins with an estimation of expected costs to establish a baseline premium. This is then adjusted by adding margins that account for uncertainties, provide a profit buffer, and

[1] Mathematical Institute, Slovak Academy of Sciences, Bratislava, Slovakia & Department of Sensory Information Systems and Technologies, Institute of Informatics, Slovak Academy of Sciences, Bratislava, Slovakia.
E-mail: agu1@uniba.sk. ORCID: https://orcid.org/0000-0002-2367-4732.

reflect potential aggregate claims payable to policyholders (Yartey, 2020). Central to this approach is the distribution of aggregate claim amounts, which is derived from the convolution of claim frequencies and severities. This distribution plays a crucial role in pricing insurance portfolios as it informs the likelihood and magnitude of potential losses. Accurate estimation of aggregate claim amounts is therefore critical for insurance companies as it supports informed decisions about pricing competitiveness, risk margins, and capital allocation. However, a persistent challenge in actuarial mathematics lies in modeling this distribution when discrete, non-negative integer values represent the number of claims and the severity of claims. Accurately capturing this behavior is essential for reliable risk assessment and premium setting.

1.2. Literature review

Historically, before the advent of modern computing, actuaries relied primarily on estimation and approximation techniques that lacked a rigorous theoretical foundation for determining aggregate claim amounts. These methods were limited in accuracy and reliability, making data-driven decision-making in insurance challenging. A widely adopted approach for analyzing the distribution of aggregate claim amounts involves identifying suitable counting distributions defined over the non-negative integers and fitting them separately to the number of claims and claim severities. However, while claim frequencies are inherently discrete, claim severities are typically modeled as continuous random variables and are thus best represented by continuous distributions. Numerous studies, such as those by Hogg and Stuart (2009), Gray and Pitts (2012), Packová and Brebera (2015), Pacáková and Gogola (2013), Jindrová and Pacáková (2016), and Dzidzornu and Minkah (2021), have examined various methods for fitting distributions to insurance claim datasets. Despite their widespread use, these approaches can be unreliable as they often fail to accurately capture the convolution between the number of claims and claim severity, two central components of the aggregate claim distribution. This convolution forms the basis of the aggregate claims model and has been applied extensively in actuarial science to solve various insurance-related problems (Albrecher et al., 2017; Klugman et al., 2019; Mildenhall & Major, 2022). However, computing this convolution presents substantial challenges, primarily due to the absence of a closed-form expression and the associated computational complexity.

To address these issues, alternative computational strategies have been developed, such as the normal power approximation and fast Fourier transform (FFT) techniques (Beard et al., 1977; Cooley & Tukey, 1965; Heckman & Meyers, 1983; Mildenhall, 2024). Although these methods enhance theoretical understanding, they often become computationally intensive and less accurate when applied to large datasets with high claim frequencies and severities. These limitations have motivated the search for more efficient and robust approaches. One such approach is the recursive method, often referred to as the "exact method". Unlike convolution-based techniques, the recursive approach assumes that the number of claims and claim severity distributions are discrete, enabling the computation of aggregate claim amounts through recursive formulas. This method substantially reduces computational burden while maintaining accuracy, particularly in scenarios involving a large number of claims. A foundational contribution in this area was made by Panjer (1981), who introduced the Panjer recursive family of discrete distributions and the corresponding recursion formula for computing aggregate claim amounts. The Panjer recursive formula has spurred extensive research in actuarial science, with notable contributions from Sundt and Vernic (2009), Yartey (2020), Dickson (2016), and Ghinawan et al. (2021). More recently, Tzaninis and Bozikas (2024) extended the Panjer family of claim number distributions by treating the family's parameters as random variables, thereby deriving a more flexible compound distribution. Their formulation assumes that claim sizes are conditionally independent and identically distributed, as well as conditionally independent of the number of claims. In a related development, Fackler (2023) introduced a reparameterization of the Panjer family, enhancing its modeling flexibility.

Although the Panjer recursion effectively models aggregate claim amounts, its applicability is confined to a narrow class of counting distributions that have a fixed, positive probability at zero. To address this constraint, Schröter (1990) proposed the Schröter recursive formula, which accommodates a broader range of counting distributions and more accurately captures the dynamics of aggregate claims. However, this method relies on convolution operations, making it computationally demanding, especially when dealing with high claim frequencies and large claim amounts. Recent advances in computational modeling have substantially broadened the methodologies available for estimating aggregate claim amounts, supplementing, and in some cases outperforming, traditional actuarial approaches. For instance, Qiu (2019) compared classical reserving methods, such as the Chain Ladder and Bornhuetter-Ferguson techniques, with machine learning-based individual claims reserving. The study found that models like generalized linear models, artificial neural networks, random forests, and support vector machines delivered superior performance on simulated datasets rich in claim-level features. However, these advantages diminished when applied to smaller, real-world datasets. Likewise, Hofmann (2022) proposed fast Fourier transform (FFT)-based algorithms as a computationally efficient alternative to the Panjer recursion under arbitrary claim frequency distributions, incorporating exponential tilting to reduce wrap-around effects and better capture distribution tails. Additionally, Gamaleldin et al. (2025) introduced a hybrid CNN-LSTM model that captures both spatial and temporal patterns in insurance claims data, considerably improving volatility forecasting and enabling proactive risk management. While these studies underscore the growing influence of machine learning in enhancing the precision, scalability, and adaptability of aggregate claims modeling, they also highlight a key trade-off: improved predictive performance often comes at the cost of increased computational complexity and resource demands during implementation and model tuning.

The computation of aggregate claim amounts plays an increasingly pivotal role in risk management and the pricing of insurance coverage. Insurance companies are inherently motivated to minimize claim payouts while maximizing premium income, thereby strengthening their ability to manage future uncertainties and withstand catastrophic losses. Within this highly competitive landscape, insurers face the added challenge of dealing with the unpredictable nature of claim occurrences embedded in insurance contracts.

Despite the utility of the Schröter recursive formula, it does not fully capture the dynamics of claim amounts truncated at one. This practice holds significant practical relevance in real-world insurance settings. In many cases, insurers are primarily concerned with the number of events that generate claims, rather than the exact amounts. Once a claim is reported, the minimum observed claim amount is often truncated at one, effectively implying a zero probability for a claim amount of zero. This reflects typical policy structures that include deductibles, where insured individuals are responsible for losses below a certain threshold, and only the excess is reimbursed. Consequently, minor losses below the deductible are frequently unreported, making one the effective lower bound for observed claim amounts. This truncation has a substantial impact on the modeling of risk exposure, influencing both the accuracy of risk assessment and the determination of premium rates. In risk theory, truncated distributions are essential for modeling claim severities and inter-arrival times, providing insurers and actuaries with critical tools to better understand the frequency and magnitude of losses. As such, accurately modeling the number of claims truncated at one is vital for capturing the true nature of insurance liabilities. It requires careful consideration of the underlying distributions that govern both claim frequency and severity, ultimately supporting more precise pricing and effective risk management. To address this gap, the present study introduces and explores the truncated Schröter recursive formula—a mathematical framework designed to improve accuracy in the computation of aggregate claim amounts. The study further assesses the computational efficiency of the proposed algorithm by analyzing its runtime performance, offering insights into its practical applicability for large-scale insurance datasets.

2. The recursive formulas

2.1. The Panjer recursive formula

The Panjer (1981) recursive formula is defined as

$$P_k = \left(a + \frac{b}{k}\right) P_{k-1}, \ k = 1,2,3,\ldots \tag{1}$$

where a and $b$are parameters, P_k denotes the recurrent probability, $P_{k-1}$is the backward recurrent probability, and by definition, $P_k = 0$ for $k < 0$. The counting distributions that satisfied (1) were explored in Panjer (1981). Furthermore, Panjer (1981) obtained the corresponding recursion algorithm for (1) defined as:

$$g(s) = \frac{1}{1-af_0}\sum_{i=1}^{s}\left(a+\frac{bi}{s}\right)f_i g(s-i), \tag{2}$$

and by definition, $f_0 = P(X=0) = 0$ and $g(0) = p_0$, where p_0 denotes the probability mass function of the counting distribution evaluated at zero, that is, the initial probability. For instance, if p_n is the Poisson distribution function from the recursive family defined in (1), then p_n evaluated at zero (p_0) and one (p_1) represents the initial probabilities of no claim and the probability of a claim, respectively.

2.2. The Schröter recursive formula

While the Panjer recursive formula addresses the challenges of the traditional convolution method, it is limited to a few distributions. Hence, Schröter (1990) generalized (1) and obtained the recursive formula expressed as:

$$P_k = \left(a+\frac{b}{k}\right)P_{k-1} + \frac{c}{k}P_{k-2},\ k = 1,2,3,\ldots, \tag{3}$$

where a, b, and c are parameters, P_{k-1} and P_{k-2} are recursive backward probabilities, and $P_k = 0$ for $k < 0$ (by definition). Note that for $c = 0$, (1) becomes a particular case of (3). Additionally, the counting distributions defined by (3) also contain the convolutions of the Poisson distribution and another distribution from (1) (see Schröter, 1990). Furthermore, Schröter (1990) obtained the corresponding recursion algorithm for (3) defined as:

$$g(s) = \frac{1}{1-af_0}\sum_{i=1}^{s}\left[\left(a+\frac{bi}{s}\right)f_i + \frac{ci}{2s}f_i^{2*}\right]g(s-i), \tag{4}$$

where f_i^{2*} has to be evaluated by the convolution formula $f_i^{2*} = \sum_{j=0}^{i} f_j f_{i-j}$ and for $c = 0$, (4) becomes (2). The parameter estimation of (3) has been studied in Agu, Mačutek, and Szűcs (2023).

3. The truncated Schröter recursive formula

In this section, we present the truncated Schröter recursive formula. We defined the truncated Schröter recursive formula as:

$$P_k = \left(a+\frac{b}{k}\right)P_{k-1} + \frac{c}{k}P_{k-2},\ k = 2,3,4,\ldots, \tag{5}$$

where the parameters are defined as in (3) and note that (5) is truncated at 1.

First, let K be a discrete random variable taking non-negative integer values as defined in (5) and using the fact that the probability generating function is defined as:

$$G(s) = \sum_{k=0}^{\infty} s^k P_k,$$

where $s \in [0,1]$such that $G(s) \geq 0$and $P_k$is the recursive probability defined in (5) and $\sum P_k = 1$. Thus, the probability generating function corresponding to (5) is:

$$G(s) = e^{-\frac{c(s-1)}{a}} \left(\frac{1-a}{1-as}\right)^{\frac{a(a+b)+c}{a^2}}, \tag{6}$$

for $|as| \neq 1$.

The derived truncated probability mass function corresponding to (5) is given as:

$$q_n = \frac{e^{\frac{c}{a}}(1-a)^{\frac{a(a+b)+c}{a^2}} \sum_{i=0}^{n} \binom{\frac{a(a+b)+c}{a^2}+i-1}{i} \frac{\left(-\frac{c}{a}\right)^{n-i} a^i}{(n-i)!}}{1-e^{\frac{c}{a}}(1-a)^{\frac{a(a+b)+c}{a^2}}},\ n = 1,2,\dots,\ 0 < a < 1,\ b, c \in \mathbb{R}. \tag{7}$$

Let $r = \frac{a(a+b)+c}{a^2}$, $x = \frac{c}{a}$, and define the generating function for the negative binomial coefficient as:

$$\sum_{k=0}^{\infty} \binom{r+k-1}{k} z^k = (1-z)^{-r},\ |z| < 1.$$

The goal is to express the finite sum in a form that leverages the generating function.

To relate $\sum_{i=0}^{n} \binom{r+i-1}{i} \frac{(-x)^{n-i} a^i}{(n-i)!}$ to the generating function for the negative binomial coefficient above, we differentiate $(1-z)^{-r}$with respect to $z$evaluated at $z = x - a$ $(0 < a < 1)$ to obtain terms that match the structure of our sum. We have that

$$\sum_{i=0}^{n} \binom{r+i-1}{i} \frac{(-x)^{n-i} a^i}{(n-i)!} = \frac{\Gamma(r+n)}{n!\,\Gamma(r)} \left(\frac{a}{a-c+a^2}\right)^{(r+n)}.$$

Hence, (7) can be expressed as

$$q_n = \frac{e^{\frac{c}{a}}(1-a)^r \frac{\Gamma(r+n)}{n!\Gamma(r)} \left(\frac{a}{a-c+a^2}\right)^{(r+n)}}{1-e^{\frac{c}{a}}(1-a)^r},\ n = 1,2,\dots,\ 0 < a < 1,\ b \geq 0, c \in \mathbb{R}.$$

Also, the log-likelihood function corresponding to (7) can be simplified as:

$$\ell(a,b,c|n_1,\dots,n_k) = k\,log\left[e^{\frac{c}{a}}(1-a)^r\right] + \sum_{j=1}^{k} log\left[\frac{\Gamma(r+n)}{n!\,\Gamma(r)} \left(\frac{a}{a-c+a^2}\right)^{(r+n)}\right] - k\,log\left[1-e^{\frac{c}{a}}(1-a)^r\right].$$

3.1. The truncated Schröter algorithm

Let $X_1, X_2, \dots, X_n$ be independent and identically distributed claim severities over the non-negative integers with probability density $f_k = p(X_i = k)$ for $i = 1,2,\dots,n, k = 0,1,2,\dots$, and $f^{k*} = P(X_1 + X_2 + \dots + X_n = k)$ denotes the n-fold convolution of f_k. Additionally, let Nbe a discrete random variable representing the number of claims with a discrete probability mass function defined as $p_n = P(N = n)$, such that X_i are stochastically independent of N, and $S = \sum_{i=1}^{N} X_i$ is the aggregate claim. While the truncated Schröter algorithm is derived under the classical assumption that claim frequency and severity are stochastically independent, it is important to note that this assumption may not fully reflect the complexities of real-world insurance portfolios. In practice, claim frequency and severity may be influenced by common risk factors (e.g. policyholder behavior, geographic or economic conditions), potentially inducing dependence between them. Ignoring this dependence can lead to biased estimates of aggregate risk, particularly in portfolios characterized by frequent and large claims, although the independence assumption facilitates analytical derivation and computational feasibility. For all the severity distributions f^{k*}, we derived the recursive algorithm as:

$$g(s) = \sum_{k=2}^{\infty} P_k f^{k*}(s), \tag{8}$$

where P_k is defined in (5).

The Panjer recursion formula defined in (2) is based on the expression $f^{k*}(s) = \frac{k}{s}\sum_{i=1}^{s} i f_i f_{s-i}^{k-1}$, $s = k = 1,2,3\ldots$, (see Schröter,1990; page 164). We can write this as: $f(s) = \frac{1}{s}\sum_{i=1}^{s} i f_i$.

Thus,

$$g(s) = \sum_{k=2}^{\infty} \left[\left(a + \frac{b}{k}\right) P_{k-1} + \frac{c}{k} P_{k-2}\right] f^{k*}(s). \tag{9}$$

We have that

$$g(s) = a\sum_{k=0}^{\infty} P_k \left(\sum_{i=0}^{s} f_i f_{s-i}^{k-1}\right) + \sum_{k=0}^{\infty} P_k \left(\sum_{i=1}^{s} \frac{bi}{s} f_i f_{s-i}^{k-1}\right) + \gamma,$$

where $\gamma = \sum_{k=0}^{\infty} P_k \left(\sum_{i=1}^{s} \frac{ci}{s} f_i f_{s-i}^{k-1}\right)$.

Note that $\sum_{k=0}^{\infty} P_k = 1$.

Therefore, it follows that

$$g(s) = \frac{1}{1 - af_0} \sum_{i=1}^{s} \left[a + \frac{i}{s}(b + c)\right] f_i g(s - i), \tag{10}$$

for $s \neq 0$ and a, b, and c are the parameters. Additionally, $f_0 = P(S = 0) = 0$ and $g(0) = p_0$ is the initial probability. If $c = 0$ in (10), we obtain (2), and if we define $f^{k*}(s)$ as $f^{k*}(s) = \frac{k}{ts}\sum_{i=1}^{s} i f_i^{t*} f_{s-i}^{(k-t)*}$, $i = 1,2,\ldots,$ for $t \in \{1,2,\ldots,k\}$ in (9), (4) becomes a special case of (10). To execute (10), we treat f_i as the claim frequencies per number of policies.

To ensure numerical stability and convergence of (10), the parameters a, b, and c were estimated via maximum likelihood of (7) using the *nlminb()* optimizer with box constraints: $0 < a < 1$, $b \geq 0$ and $c \in \mathbb{R}$. These constraints prevent instability in the recursion weights and guarantee the validity of the logarithmic expressions in the likelihood function.

Theoretically, unlike (4), the recursion algorithm defined in (10) eliminates the need for any form of convolution.

This study considers the Negative binomial distribution as the count distribution for the number of claims (see Section 4, Table 4).

The probability mass function for the Negative binomial distribution is defined as

$$h(S = s) = \binom{s + r - 1}{s} (1 - p)^s p^r, \ s = 0,1,2\ldots, \ r > 0, p \in [0,1].$$

We have that

$$h(S = 0) = \binom{r - 1}{0} p^r.$$

From (10), we define

$$g(0) = p_0 = h(S = 0) = p^r. \tag{11}$$

3.2. The numerical implementation procedure of the truncated Schröter algorithm

The implementation of the truncated Schröter recursive algorithm involves several computational stages designed to estimate the distribution of aggregate claim amounts.

The procedure can be summarized as follows:

i. **Data Preparation:** Obtain and clean claim count data (from real-world and simulations). Compute the empirical frequency distribution $f_i$and normalize it to ensure $\sum f_i = 1$. Furthermore, the distribution of the data is determined (see Table 4). The parameters in equation (10) are estimated using the maximum likelihood estimation (MLE) method based on the observed data. The log-likelihood function is constructed from the truncated probability mass function of the claim counts. Parameter estimation is carried out using the nlminb() optimizer in R, which is well-suited for bounded, nonlinear optimization problems. This approach ensures numerical stability and facilitates the explicit enforcement of parameter constraints that are critical to the recursive structure of the model. A similar approach is applied by truncating the corresponding probability mass functions of the Panjer and Schröter families (see Panjer, 1981; Schröter, 1990).

ii. **Initialization:** Determine $g(0)$ as in (11) and initialize a numeric vector to store $g(s)$ for $s = 1,2,\ldots,n$.

iii. **Recursive computation:** For each $s = 2,\ldots,n$, compute $g(s)$ using (10).

iv. **Performance evaluation:** Evaluate the sum of $g(s)$ values and record execution time per iteration to assess computational efficiency.

v. **Visualization:** Utilize graphical tools (e.g. bar plots, execution time plots) to display the algorithm's output and benchmark it against the Panjer and standard Schröter methods.

4. Numerical evaluation

In this section, we examine the run-time computational efficiency of the introduced truncated Schröter algorithm using the Automobile UK Collision Claims (AutoCollision) data obtained from https://instruction.bus.wisc.edu/jfrees/jfreesbooks/Regression%20Modeling/BookWebDec2010/data.html.

First, we began by exploring the descriptive statistics of the dataset, analyzing average claim severity and average claim counts across various age groups and vehicle categories to identify patterns and determine how frequently each group files claims. Particular attention was given to combinations of age groups and vehicle use categories associated with high claim severity and frequency, as these represent higher risk factors for insurers and may necessitate adjustments in insurance coverage strategies.

To model the claim count data, we fitted both the Negative Binomial and Generalized Poisson distributions, selected for their ability to handle overdispersion commonly observed in count data. The choice between these distributions was guided by model fit, using the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) to select the model with the lowest values. Furthermore, to implement the truncated algorithm defined in (10), we employed the truncated probability mass function introduced in (7) to obtain numerical estimates of the parameters $\hat{a}$, $\hat{b}$, and $\hat{c}$as $\hat{a} = 0.99070, \hat{b} = 1.29297,$ and $\hat{c} = 0.29330$ using the maximum likelihood estimation method. The computation of aggregate claim amounts using (2), (4), and (10) was performed as defined in Section 4.1.

In this study, all statistical analyses and computations of recursion algorithm run times were performed using RStudio on a Lenovo PC equipped with an 11th Gen Intel(R) Core(TM) i5-1135G7 @ 2.40GHz processor and 8.00 GB of RAM.

Table 1. Descriptive Statistics

Min.	Max.	Mean	Variance	Kurtosis	Skewness
5.00	970.00	279.44	58374.38	4.08	1.25

The descriptive statistics offer a comprehensive summary of the dataset's distribution and central tendency. The minimum and maximum values define the data range, while the mean provides a central value around which the data are distributed. The high variance indicates substantial variability (overdispersion), and the positive skewness and kurtosis indicate a right-skewed distribution with the presence of outliers.

Table 2. Analysis of Average Claim Severity by Vehicle Use

Vehicle Use	Average Claim Severity	Claim Count
Business	395.21	1075
DriveLong	265.26	2710
DriveShort	231.74	3888
Pleasure	213.20	1269

Table 2 shows that vehicles used for business purposes exhibit the highest average claim severity. Although the claim count in this category is relatively low compared to others, each claim carries a substantial financial impact, indicating that business use presents a higher risk of costly claims.

Vehicles used for long-distance driving exhibit a moderate average claim severity, which is substantially lower than that of business use but higher than for short drives and pleasure use. The relatively high claim count indicates that long drives are associated with frequent incidents, though each claim tends to be less severe than those in the business category.

Short drives register the highest claim count but a lower average claim severity. This indicates that while short trips result in more frequent claims, the financial impact of each is comparatively minor. The high frequency highlights a notable number of incidents with less severe consequences per occurrence.

Pleasure use is associated with the lowest average claim severity and a relatively low claim count, indicating that leisure driving poses the least risk. It results in both fewer claims and lower financial losses, making it the lowest-risk category in terms of both frequency and severity in the UK Automobile Collision Claims dataset.

Table 3. Analysis of Average Claim Severity by Age

Age	Average Claim Severity
17–20	391.80
21–24	293.17
25–29	284.84
30–34	279.73
35–39	212.43
40–49	249.99
50–59	251.11
60+	247.68

Table 3 shows that drivers aged 17–20 have the highest average claim severity, indicating that accidents involving the youngest drivers tend to result in greater financial losses and represent a substantial risk to insurers. A notable decrease in average claim severity is observed among drivers aged 21–24, indicating a reduced but still relatively high financial risk as drivers gain minimal experience. The trend of decreasing claim severity continues in the 25–29 age group, reflecting a further decline in financial impact as drivers mature and gain experience. This downward trend persists in the 30–34 age group, with a slight reduction in average claim severity compared to the previous cohort. A substantial drop is observed in the 35–39 age group, indicating a much lower severity of claims and a correspondingly reduced financial risk. Interestingly, the 40–49 age group sees a modest increase in average claim severity compared to the 35–39 group, though it remains lower than that of drivers under 30, indicating a moderate financial risk. Claim severity levels for the 50–59 age group are comparable to those of the 40–49 cohort, pointing to a stable level of financial risk among middle-aged drivers. Finally, drivers aged 60 and over exhibit slightly lower average claim severity than the 50–59 group, indicating a consistent and moderate financial risk, marginally higher than that of the 35–39 group but lower than the younger cohorts.

Table 4. The fitting of Negative Binomial and Generalized Poisson distributions

	Negative Binomial	**Generalized Poisson**
Parameters Estimate	$\hat{p} = 0.00453, \hat{r} = 1.25042$	$\hat{\theta} = 12.78279, \hat{\lambda} = 0.95426$
N-Loglikelihood	-211.9633	-216.1883
AIC	427.9267	436.3767
BIC	430.8581	439.3081

As shown in Table 4, the Negative Binomial distribution yields a higher (i.e. less negative) log-likelihood and the lowest AIC and BIC values, clearly indicating a superior fit to the AutoCollision claim count data compared to the Generalized Poisson distribution. These results indicate that the Negative Binomial model is more appropriate for capturing the underlying data structure. Both AIC and BIC are essential for model selection, as they balance goodness-of-fit with model complexity, thereby mitigating the risk of overfitting, an especially important consideration in actuarial modeling. Beyond information criteria, residual diagnostics further validate this conclusion. A comparative analysis of the Negative Binomial (see **Fig. 5**) and Generalized Poisson models (see **Fig. 6**) reveals that the former produces Pearson and deviance residuals tightly clustered around zero, with minimal dispersion and no extreme outliers. The histogram of Pearson residuals is approximately symmetric and unimodal. In contrast, the Q–Q plot of deviance residuals aligns closely with the theoretical quantile line, indicating that the model assumptions are well met. In contrast, the Generalized Poisson model exhibits more dispersed residuals, noticeable outliers, a skewed residual histogram, and a Q–Q plot that substantially deviates from the reference line, indicating potential model misspecification. Taken together, these statistical and graphical diagnostics confirm that the Negative Binomial model provides a more accurate and reliable representation of the claim count data, establishing it as the preferred modeling choice for this analysis.

4.1. Computation of aggregate claim

In this section, we used the estimates of $\hat{a}$, $\hat{b}$, and $\hat{c}$for (2), (4), and (10) to compute aggregate claim amounts for each recursive algorithm and their computational run time using Claim count data from the AutoCollision data. Based on Table 5, the performance of the three recursion algorithms is analyzed in terms of computational run time and the computed aggregate claim amounts.

The truncated Schröter recursion algorithm demonstrates the fastest run time at 0.051093 seconds and yields the highest aggregate claim sum of 0.005483, indicating that it either captures more aspects of the claim data or incorporates a more comprehensive modeling approach (see **Fig. 1**). This result implies a probability of approximately 0.55% that the total claim amount will not exceed 32 units, and conversely, a 99.45% probability that it will exceed this threshold.

The Panjer recursion algorithm, with a slightly longer runtime of 0.060898 seconds, computes an aggregate claim sum of $ 0.004887. Although still efficient, this result may indicate a more conservative or less data-sensitive approach (see Fig. 2). The corresponding probability that the total claim amount does not exceed $32 is approximately 0.49%, implying a 99.51% chance of exceeding this amount. The standard Schröter recursion algorithm, which has the longest runtime at 0.173438 seconds, produces an aggregate claim sum of $ 0.004930. This outcome implies a balance between sensitivity and comprehensiveness; however, it comes with higher computational demands due to the convolution component involved in the algorithm (see Fig. 3). The probability that the total claim amount will not exceed 32 units is approximately 0.45%. In contrast, the probability that it will exceed 32 units is around 99.55%.

The general interpretation of these results is that the likelihood of the total claim amount being less than or equal to 32 units is very low, with probabilities ranging from approximately 0.45% to 0.55%.

Consequently, the probability that the total claim amount will exceed 32 units is extremely high, ranging from 99.45% to 99.55% for the AutoCollision dataset. These findings indicate that, across all recursion algorithms evaluated, it is almost certain that total claims will surpass 32 units, underscoring the high-risk nature of the claims being modeled.

These results provide valuable insights for effective risk management and premium setting in the insurance sector. The high probability of large aggregate claims indicates that insurers must prepare for substantial payouts. Understanding this risk landscape allows insurers to more accurately assess claim distributions and frequencies, leading to more informed pricing strategies that ensure financial sustainability. Insurers can utilize these insights to allocate adequate reserves for high-expectation claims, thereby reducing the risk of insolvency. Moreover, policy designs can incorporate deductibles, limits, and exclusions that align with the high likelihood of large claims, striking a balance between customer affordability and insurer profitability. These findings also support the development of targeted reinsurance strategies, allowing insurers to transfer a portion of high-risk exposures and minimize the financial impact of large claims.

Table 5. Aggregate claim for the truncated Schröter, Panjer, and Schröter algorithms

s	g(s) The Truncated Schröter	g(s) The Panjer	g(s) The Schröter
1	711000×10^{-3}	711000×10^{-3}	711000×10^{-3}
2	13488×10^{-5}	11953×10^{-5}	11958×10^{-5}
3	78790×10^{-6}	69718×10^{-6}	69763×10^{-6}
4	18245×10^{-6}	16054×10^{-6}	16132×10^{-6}
5	21313×10^{-5}	18881×10^{-5}	18908×10^{-5}
6	58042×10^{-5}	51406×10^{-5}	51450×10^{-5}
7	32214×10^{-5}	28454×10^{-5}	28493×10^{-5}
8	15979×10^{-5}	14070×10^{-5}	14113×10^{-5}
9	48176×10^{-5}	42617×10^{-5}	42727×10^{-5}
10	11819×10^{-4}	10454×10^{-4}	10474×10^{-4}
11	11254×10^{-4}	99323×10^{-5}	99514×10^{-5}
12	49170×10^{-5}	43135×10^{-5}	43306×10^{-5}
13	46463×10^{-5}	40787×10^{-5}	41095×10^{-5}
14	15944×10^{-4}	14065×10^{-4}	14119×10^{-4}
15	13782×10^{-4}	12090×10^{-4}	12149×10^{-4}
16	74915×10^{-5}	65005×10^{-5}	65502×10^{-5}
17	65619×10^{-5}	57003×10^{-5}	57681×10^{-5}
18	18092×10^{-4}	15882×10^{-4}	15999×10^{-4}
19	16315×10^{-4}	14191×10^{-4}	14324×10^{-4}
20	95516×10^{-5}	81572×10^{-5}	82706×10^{-5}
21	11549×10^{-4}	99748×10^{-5}	10108×10^{-4}
22	36393×10^{-4}	31961×10^{-4}	32173×10^{-4}
23	30416×10^{-4}	26473×10^{-4}	26719×10^{-4}
24	17428×10^{-4}	14879×10^{-4}	15094×10^{-4}
25	15023×10^{-4}	12823×10^{-4}	13062×10^{-4}
26	35570×10^{-4}	30989×10^{-4}	31358×10^{-4}
27	27630×10^{-4}	23653×10^{-4}	24074×10^{-4}
28	17519×10^{-4}	14558×10^{-4}	14920×10^{-4}
29	18952×10^{-4}	15950×10^{-4}	16353×10^{-4}

Table 5. Aggregate claim for the truncated Schröter, Panjer, and Schröter algorithms (cont.)

s	g(s) The Truncated Schröter	g(s) The Panjer	g(s) The Schröter
30	30447×10^{-4}	26076×10^{-4}	26676×10^{-4}
31	27511×10^{-4}	23012×10^{-4}	23706×10^{-4}
32	22584×10^{-4}	18434×10^{-4}	19048×10^{-4}
Sum of Probabilities	0.005483	0.004887	0.004535
Execution time in seconds(s)	0.051093	0.060898	0.173438

The observed differences in computational run times and aggregate claim sums across algorithms are attributable to the inherent complexity and structural differences of the recursion methods. The truncated Schröter algorithm, with its three-parameter structure, strikes an efficient balance between model complexity and computational speed, yielding both fast run times and higher aggregate claims. The Panjer recursion algorithm, while simpler with only two parameters, offers efficient computation but may not capture as many underlying data features. In contrast, the Schröter recursion algorithm, which incorporates an additional convolution term, requires more computation time but provides a nuanced perspective on aggregate claim modeling.

4.2. Simulation study

Here, we generate random claim amounts data from the Negative binomial distribution by setting $r = 100$ and $p = 0.05$. We varied the sample size to examine the aggregate claim computational efficiency and run time of the truncated Schröter, Panjer, and Schröter recursion algorithms. Initially, we generated 5000 random numbers from the Negative binomial distribution and fit (7) to the data to obtain the estimate of the parameters $\hat{a}$, $\hat{b}$, and $\hat{c}$as $\hat{a} = 0.9905$, $\hat{b} = 18.3096$, $\hat{c} = -1.2840$, and compute$g(0) = 0.00117$ to implement the algorithms. Tables 6, 7, and 8 present the sample sizes, aggregate claim amounts, and the execution time in seconds for each recursion algorithm.

Fig. 4 illustrates the execution times of the truncated Schröter recursion, Panjer recursion, and Schröter recursion algorithms for varying values of n, highlighting significant differences in computational efficiency as the sample size increases. The truncated Schröter recursion algorithm consistently demonstrates the lowest execution times across all values of n, starting at 0.0006919 s for n = 20 and increasing to 1.5046701 s for n = 5000 (see **Fig. 4**).

Table 6. Efficiency of the truncated Schröter algorithm on simulated claim data

Recursion Algorithm	Sample (n)	sum of g(s)	Execution time (s)
The truncated Schröter algorithm			
	20	2.7080348	0.0006919
	50	1.5211150	0.0036724
	100	1.1730166	0.0188396
	150	1.1747598	0.0335643
	200	1.1296751	0.0586591
	300	0.9208012	0.1158113
	600	0.8504156	0.3374069
	1500	0.6127028	0.8211629
	2000	0.5397614	1.0198436
	3000	0.4756758	1.2040498
	4000	0.4180231	1.4369745
	5000	0.3979199	1.5046701

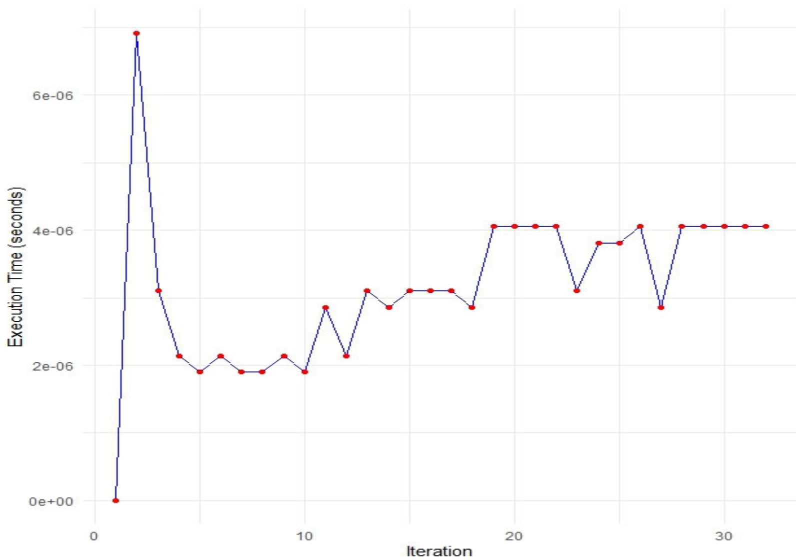

Figure 1. The execution time plot of the truncated Schröter recursion algorithm for each iteration

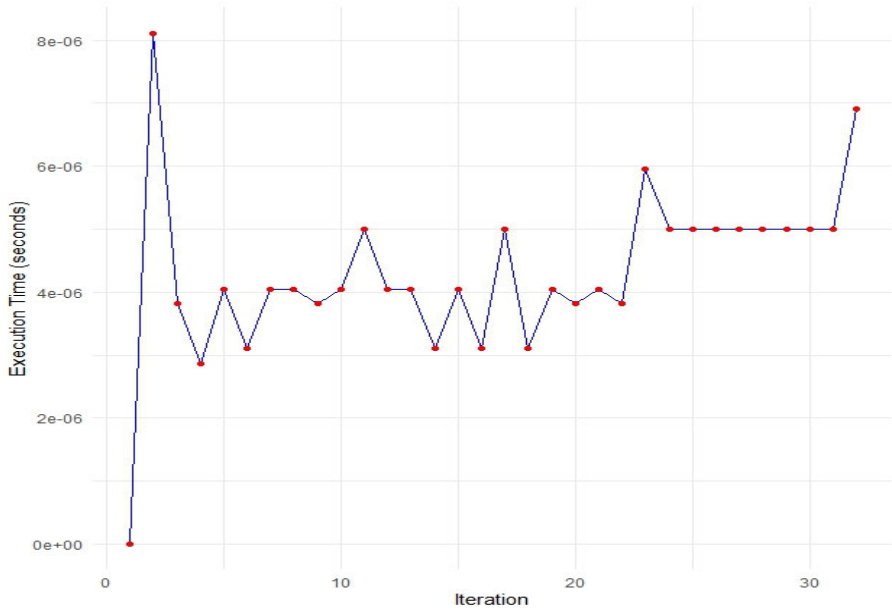

Figure 2. The execution time plot of the Panjer recursion algorithm for each iteration

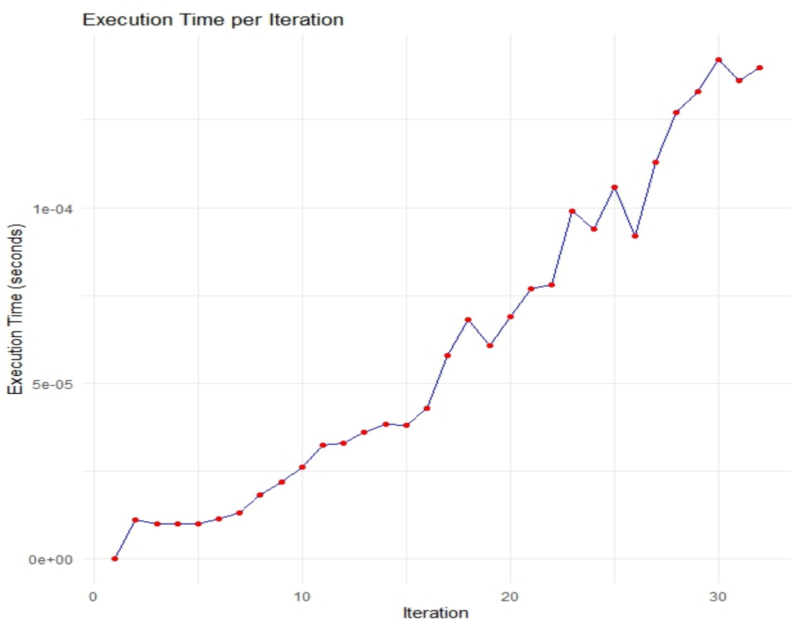

Figure 3. The execution time plot of the Schröter recursion algorithm for each iteration

Table 7. Efficiency of the Panjer algorithm on simulated claim data

Recursion Algorithm	Sample (n)	sum of g(s)	Execution time (s)
The Panjer algorithm			
	20	0.0075441	0.0014127
	50	0.0074026	0.0047266
	100	0.0074002	0.0212255
	150	0.0074619	0.0378468
	200	0.0074672	0.0692441
	300	0.0072761	0.1391730
	600	0.0073219	0.4021811
	1500	0.0073130	1.0212069
	2000	0.0072711	1.0606146
	3000	0.0072833	1.3638492
	4000	0.0072209	1.6406126
	5000	0.0072351	1.7650454

Table 8. Efficiency of the Schröter algorithm on simulated claim data

Recursion Algorithm	Sample (n)	sum of g(s)	Execution time (s)
The Schröter algorithm			
	20	0.0062785	0.0028598
	50	0.0063757	0.0195415
	100	0.0064331	0.0765483
	150	0.0064909	0.1735694
	200	0.0065073	0.2585254
	300	0.0063946	0.6078202
	600	0.0064442	1.8534706
	1500	0.0064868	4.6577935
	2000	0.0064729	5.4729755
	3000	0.0065035	6.8906786
	4000	0.0064749	8.0277340
	5000	0.00649490	8.7507973

This performance indicates that the algorithm is highly efficient and scalable, capable of handling larger datasets with minimal computational burden. The Panjer recursion algorithm also exhibits increasing execution times with larger n, beginning at 0.0014127 s for $n = 20$ and rising to 1.7650454 s for $n = 5000$. While reasonably efficient, it demonstrates less scalability compared to the truncated Schröter algorithm (see **Fig. 4**). In contrast, the Schröter recursion algorithm, which includes a convolution component, f_i^{2*}, shows substantially higher execution times, starting at 0.0028598 s for $n = 20$ and escalating sharply to 8.7507973 s for $n = 5000$ (see **Fig. 4**). This steep increase reflects poor scalability and reduced efficiency, particularly for large sample sizes, making it the least optimal option among the three algorithms evaluated. Overall, the truncated Schröter recursion algorithm emerges as the most efficient and scalable, followed by the Panjer recursion algorithm. The Schröter recursion algorithm, while potentially offering greater modeling flexibility, is substantially less efficient due to its computational complexity.

To assess the consistency of execution times, each sample size was tested across five independent runs. The variation in computational times was negligible, indicating that the execution times were stable and reproducible. However, it is worth noting that minor fluctuations may still be influenced by the operational state of the computing system during execution.

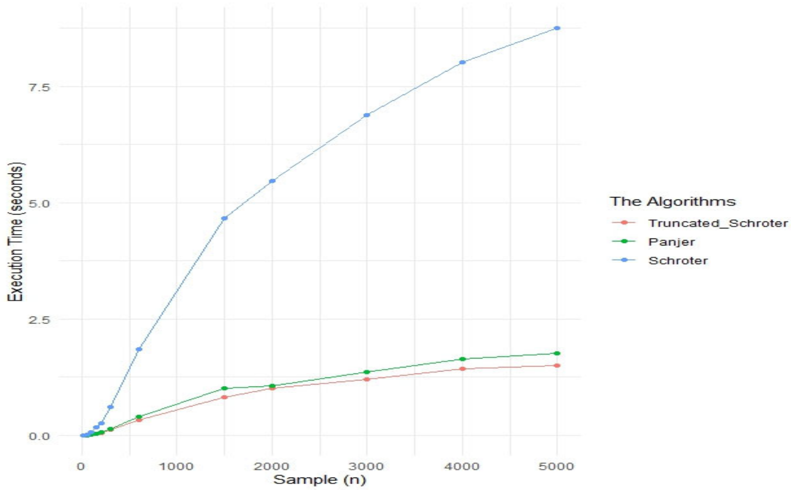

Figure 4. Visual representation of the execution time of the truncated Schröter recursion, Panjer recursion, and Schröter recursion algorithms

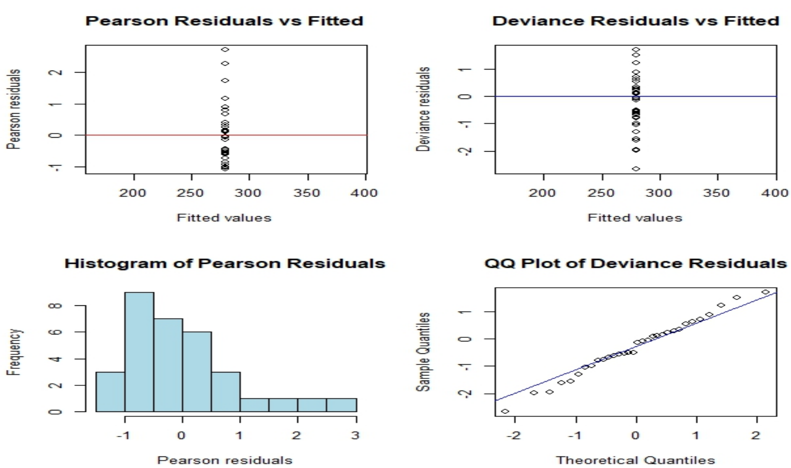

Figure 5. Graphical residual analysis of the fitted Negative Binomial distribution to the claim data

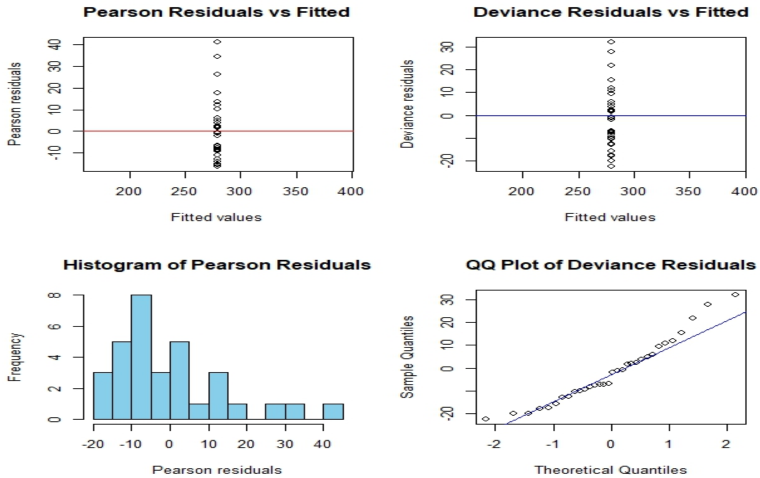

Figure 5. Graphical residual analysis of the fitted Generalized Poisson distribution to the claim data

5. Conclusion

This study investigated the computation of aggregate claim amounts using various recursive algorithms, with a particular focus on the newly introduced truncated Schröter recursion algorithm. The primary objective was to enhance both the accuracy and computational efficiency of aggregate claim estimation, an essential component of effective risk management and premium setting in the insurance industry.

The truncated Schröter recursion algorithm demonstrated superior performance in numerical evaluations and comparative analysis. When applied to the AutoCollision dataset, it consistently delivered the fastest execution times and the highest aggregate claim sums, indicating both computational efficiency and modeling comprehensiveness. For modeling claim count data, the Negative Binomial distribution was favored over the Generalized Poisson distribution due to its ability to accommodate overdispersion, as supported by AIC and BIC selection criteria.

Simulation studies further validated the performance of the truncated Schröter algorithm across varying sample sizes, consistently outperforming the Panjer and standard Schröter recursion algorithms in terms of execution time and scalability, while effectively capturing data variability for more refined analysis. The findings also underscored the critical importance of selecting appropriate counting distributions and recursion methods when modeling aggregate claim amounts.

In conclusion, the truncated Schröter recursion algorithm emerges as a robust and reliable tool for calculating aggregate claim amounts, offering a strong balance between computational speed and modeling accuracy. Its adoption has the potential to improve risk assessment substantially and premium pricing strategies, ultimately benefiting both insurers and policyholders. Future research could explore enhancements to this algorithm, such as incorporating machine learning techniques to optimize parameter estimation based on evolving claim patterns dynamically. Moreover, applying the algorithm to other insurance domains beyond automobile claims could further validate its generalizability and inform domain-specific refinements.

Funding

This study was supported by the Slovak Academy of Sciences Doktogrant: APP0515 (Friday I. Agu).

Acknowledgement

The author expresses his sincere gratitude to his esteemed supervisor, Dr. Jan Macutek, for his invaluable guidance, support, and encouragement throughout this study. His unwavering belief in my abilities and insightful mentorship have shaped my academic and professional development. I am deeply grateful for his continued support and inspiration.

Conflict of interest

There is no conflict of interest for this study.

References

Agu, F. I., Mačutek, J. and Szűcs, G., (2023). A Simple Estimation of Parameters for Discrete Distributions from the Schröter Family. *Statistika: Statistics & Economy Journal*, 103(2).

Albrecher, H., Beirlant, J. and Teugels, J. L., (2017). Reinsurance: actuarial and statistical aspects. *John Wiley & Sons*.

Beard, R. E., Pentikäinen, T. and Pesonen, E., (1977). Risk theory (2nd ed.). *Chapman and Hall*.

Cooley, J. W., Tukey, J. W., (1965). An algorithm for the machine calculation of complex Fourier series. *Mathematics of computation*, 19(90), pp. 297–301.

Dickson, D. C. (2016). Insurance risk and ruin. *Cambridge University Press*.

Dzidzornu, S. B., Minkah, R., (2021). Assessing the Performance of the Discrete Generalised Pareto Distribution in Modelling Non-life Insurance Claims. *Journal of Probability and Statistics*, 2021(1), 5518583.

Fackler, M., (2023). Panjer class revisited: one formula for the distributions of the Panjer (a, b, n) class. *Annals of Actuarial Science*, 17(1), pp. 145–169.

Gamaleldin, W., Attayyib, O., Alnfiai, M. M., Alotaibi, F. A. and Ming, R., (2025). A hybrid model based on CNN-LSTM for assessing the risk of increasing claims in insurance companies. *PeerJ Computer Science*, 11, e2830.

Ghinawan, F., Nurrohmah, S. and Fithriani, I., (2021). Recursive and moment-based approximation of aggregate loss distribution. In *Journal of Physics*: *Conference Series*, Vol. 1725, No. 1, p. 012101. *IOP Publishing*.

Gray, R. J., Pitts, S. M., (2012). Risk modeling in general insurance: From principles to practice. *Cambridge University Press*.

Heckman, P. E., Meyers, G. G., (1983). The calculation of aggregate loss distributions from claim severity and claim count distributions. *In Proceedings of the Casualty Actuarial Society* , Vol. 70, No. 133–134, pp. 49–66. Casualty Actuarial Society.

Hofmann, L., (2022). Approximation Methods for the Total Claim Amount in Collective Risk Modeling/submitted by Hofmann Louisa.

Hogg, R. V., Klugman, S. A., (2009). Loss distributions. *John Wiley & Sons*.

Jindrová, P., Pacáková, V., (2016). Modeling of extreme losses in natural disasters. *International Journal of Mathematical Models and Methods in Applied Sciences*, Vol. 10, issue 2016.

Klugman, S. A., Panjer, H. H. and Willmot, G. E., (2012). Loss models: from data to decisions, Vol. 715. *John Wiley & Sons*.

Mildenhall, S., (2024). Aggregate: fast, accurate, and flexible approximation of compound probability distributions. *Annals of Actuarial Science*, pp. 1–40.

Mildenhall, S. J., Major, J. A., (2022). Pricing insurance risk: Theory and practice. *John Wiley & Sons*.

Pacáková, V., Gogola, J., (2013). Pareto Distribution in Insurance and Reinsurance. In Conference proceedings from 9th International Scientific Conference Financial Management of Firms and Financial Institutions. *VŠB Ostrava*, pp. 298–306.

Packová, V., Brebera, D., (2015). Loss distributions in insurance risk management. *Recent advances in economics and business administration*, pp. 17–22.

Panjer, H. H., (1981). Recursive evaluation of a family of compound distributions. ASTIN Bulletin: *The Journal of the IAA*, 12(1), pp. 22–26.

Qiu, D., (2019). Individual claims reserving: Using machine learning methods (*Doctoral dissertation, Concordia University*).

Schröter, K. J., (1990). On a family of counting distributions and recursions for related compound distributions. *Scandinavian Actuarial Journal*, 1990(2–3), pp. 161–175.

Sundt, B., Vernic, R., (2009). Recursions for convolutions and compound distributions with insurance applications. *Springer Science & Business Media*.

Tzaninis, S. M., Bozikas, A., (2024). Extensions of Panjer's recursion for mixed compound distributions. *arXiv preprint arXiv*:2406.17726.

Yartey, E., (2020). The (a, b, r) class of discrete distributions with applications (*Doctoral dissertation, Laurentian University of Sudbury*).