University of Milan

Faculty of Political, Economic and Social Sciences

A Bayesian Approach to Aggregate Insurance Claim Modeling

Final Project in the Subject Bayesian Analysis

Julia Maria Wdowinska (43288A)
Edoardo Zanone (33927A)

Data Science for Economics
II Year
Master's Degree

May 5, 2025

Contents

1 Introduction

2 Theoretical Background

In insurance modeling, accurately estimating the aggregate value of claims is essential for premium calculation, risk assessment, and maintaining solvency of insurance providers. The aggregate claim value refers to the total amount paid out for claims over a specified period. This quantity is typically modeled using a compound distribution, which separately describes the frequency and severity of claims. The frequency component, representing the number of claims within the period, is often modeled using a Poisson distribution. The severity component, which captures the magnitude of individual claims, is commonly modeled using heavy-tailed distributions such as the Pareto or Gamma distributions.

The traditional framework for modeling such processes is grounded in frequentist statistics. In this approach, the parameters of the underlying distributions—such as the Poisson rate parameter λ, or the shape and scale parameters α and β of the Pareto distribution—are considered fixed but unknown quantities. These parameters are typically estimated from observed data using techniques such as maximum likelihood estimation (MLE).

In contrast, the Bayesian framework treats these parameters as random variables characterized by prior distributions. This allows for the incorporation of prior knowledge or beliefs about the parameters before any data is observed. The prior distribution is updated using Bayes' theorem upon observing data, resulting in a posterior distribution that reflects the updated beliefs. This posterior distribution forms the basis for inference and prediction.

Bayesian methods offer significant advantages in insurance modeling. They are well-suited for incorporating expert knowledge or external information, handling limited or noisy datasets, and accommodating complex hierarchical structures commonly found in insurance data. Moreover, Bayesian inference provides a coherent framework for uncertainty quantification by yielding full posterior distributions over model parameters.

Nevertheless, Bayesian methods also come with certain limitations. The choice of prior distribution can introduce subjectivity, and computational complexity can be significant, especially in models with high dimensionality or non-conjugate priors. Despite these challenges, the flexibility and robustness of the Bayesian framework make it a valuable approach for modeling aggregate insurance claims.

3 Replication of Dudley (2006)

Building on the theoretical background, this project aims to replicate the Bayesian modeling approach proposed by Dudley (2006). The analysis is based on a dataset of automobile insurance claims exceeding 1.5 million, collected over a five-year period. The data, originally reported by Rytgaard (1990), is presented in Table 1.

3.1 Model Specification

To model this dataset within a Bayesian framework, assumptions about the distributions of both the number of claims in year t (N_t) and the amount of the i-th claim in year t ($Y_{i,t}$) were necessary. Claims were assumed to occur randomly and independently at a constant rate over time, so N_t was modeled using a Poisson distribution.

[1] To manage risk exposure, insurers frequently employ reinsurance strategies, which help reduce their financial liability on large claims. Under such arrangements, if a claim amount y exceeds a predetermined threshold d (the retention), the insurer is responsible only for paying up to d, while any excess $y - d$ is covered by the reinsurer.

Table 1: Insurance Claim Amounts Exceeding 1.5 Million (Data from Rytgaard, 1990)

Year	Claim Amounts (in millions)				
1	2.495	2.120	2.095	1.700	1.650
2	1.985	1.810	1.625	–	–
3	3.215	2.105	1.765	1.715	–
4	–	–	–	–	–
5	19.180	1.915	1.790	1.755	–

The threshold of 1.5 million corresponds to the retention level of an excess-of-loss insurance policy[1].

A Pareto distribution was selected for $Y_{i,t}$, as a heavy-tailed loss distribution was needed to account for the fact that individual claim amounts are positive and may include large outliers. That is,

$$\begin{aligned} N_t &\sim \text{Poisson}(\theta), \quad 0 < \theta < \infty, \\ Y_{i,t} &\sim \text{Pareto}(\alpha, \beta), \quad \alpha > 0, \quad 0 < \beta < y. \end{aligned}$$

The Pareto(α, β) distribution with support $[\beta, \infty)$ was particularly suitable in this context, as it was employed to model claim amounts exceeding a specified threshold.

In addition, the following assumptions were made:

- N_t are independently and identically distributed (i.i.d.) across t,
- $Y_{i,t}$ are i.i.d. across both i and t,
- N_t and $Y_{i,t}$ are independent for all i and t.

Under these assumptions, the aggregate claim amount in year t was defined as

$$S_t = Y_{1,t} + Y_{2,t} + \cdots + Y_{N_t,t}.$$

Next, prior distributions for the parameters α, β, and θ were specified. Due to limited prior knowledge about their true values—beyond the assumption that they are strictly positive—vague Gamma priors[2] were chosen:

$$\alpha \sim \text{Gamma}(1, 0.0001), \quad \beta \sim \text{Gamma}(1, 0.0001), \quad \theta \sim \text{Gamma}(1, 0.0001),$$

with the constraint $0 < \beta < \min\{y_{i,t}\}$ to ensure validity of the Pareto distribution.

Finally, the posterior distributions were derived. First, the joint posterior distribution of (α, β) was obtained via Bayes' theorem[3]:

$$\begin{aligned} \pi(\alpha, \beta \mid \mathbf{y}) &\propto \pi(\alpha) \cdot \pi(\beta) \cdot f(\mathbf{y} \mid \alpha, \beta) \\ &\propto 0.0001 \cdot \exp(-0.0001\alpha) \cdot 0.0001 \cdot \exp(-0.0001\beta) \cdot \prod_{i=1}^{n} \frac{\alpha \beta^{\alpha}}{y_i^{\alpha+1}} \\ &\propto \exp(-0.0001\alpha) \cdot \exp(-0.0001\beta) \cdot \alpha^n \cdot \beta^{n\alpha} \left(\prod_{i=1}^{n} y_i \right)^{-(\alpha+1)} \\ &\propto \alpha^n \cdot \exp(-0.0001\alpha) \cdot \left(\prod_{i=1}^{n} y_i \right)^{-\alpha} \cdot \beta^{n\alpha} \cdot \exp(-0.0001\beta) \\ &\propto \alpha^n \cdot \exp\left(- \left(0.0001 + \sum_{i=1}^{n} \ln(y_i) \right) \alpha \right) \cdot \beta^{n\alpha} \cdot \exp(-0.0001\beta) \end{aligned}$$

As a result, the full conditional posterior distributions of α and β were as follows:

$$\begin{aligned} \pi(\alpha \mid \beta, \mathbf{y}) &\propto \alpha^n \cdot \exp\left(- \left(0.0001 - n \ln(\beta) + \sum_{i=1}^{n} \ln(y_i) \right) \alpha \right), \\ \pi(\beta \mid \alpha, \mathbf{y}) &\propto \beta^{n\alpha} \cdot \exp(-0.0001\beta), \end{aligned}$$

[2]Each of these Gamma priors has a variance of 10^8, implying minimal prior influence so that most of the information about the parameters is derived from the dataset. Additionally, the Gamma distribution is conjugate to both the Poisson and Pareto likelihoods, facilitating analytical tractability in Bayesian inference.

[3]Here, assuming that α and β are independent, the joint prior $\pi(\alpha, \beta)$ was computed as $\pi(\alpha) \cdot \pi(\beta)$.

which implied that:

$$\alpha \mid \beta, \mathbf{y} \sim \text{Gamma}\left(n+1, \sum_{i=1}^{n} \ln(y_i) - n\ln(\beta) + 0.0001\right),$$
$$\beta \mid \alpha, \mathbf{y} \sim \text{Gamma}(n\alpha + 1, 0.0001).$$

Similarly, the posterior distribution of θ was obtained via Bayes' theorem:

$$\begin{aligned}
\pi(\theta \mid \mathbf{n}) &\propto \pi(\theta) \cdot f(\mathbf{n} \mid \theta) \\
&\propto \exp(-0.0001\theta) \cdot \prod_{t=1}^{T} (\theta^{n_t} \cdot \exp(-\theta)) \\
&\propto \exp(-0.0001\theta) \cdot \theta^{\sum_{t=1}^{T} n_t} \cdot \exp(-5\theta) \\
&\propto \exp(-5.0001\theta) \cdot \theta^{\sum_{t=1}^{T} n_t}
\end{aligned}$$

which implied that:

$$\theta \mid \mathbf{n} \sim \text{Gamma}\left(\sum_{t=1}^{T} n_t + 1, 5.0001\right)$$

3.2 Sampling Results

Since all three posterior distributions were standard distributions, the Gibbs sampling method was employed to draw realizations from them[4]. This was implemented using the `JAGS` program, which was called from within `R`. Following Dudley (2006), three Markov chains were run in parallel. The initial values of α, β, and θ were chosen to be well-dispersed and are presented in Table 2.

Table 2: Initial Parameter Values

Chain	α	β	θ
1	0.000 01	0.000 01	0.000 01
2	100 000	1	100 000
3	3.076	1.625	3.200

These values are taken from Dudley (2006).

A burn-in of 20,000 iterations was used. The statistics computed from the subsequent 30,000 iterations are presented in Table 3. A comparison with those reported by Dudley (2006) shows a close match, indicating that the model was properly specified and the Gibbs sampler executed correctly.

Table 3: Posterior Statistics

	Mean	**Standard Deviation**	**95% Bayesian Credible Interval**
α	3.076	0.762	(1.762, 4.752)
β	1.591	0.035	(1.498, 1.624)
θ	3.399	0.820	(1.986, 5.185)
$E[Y]$	2.507	1.071	(2.024, 3.637)

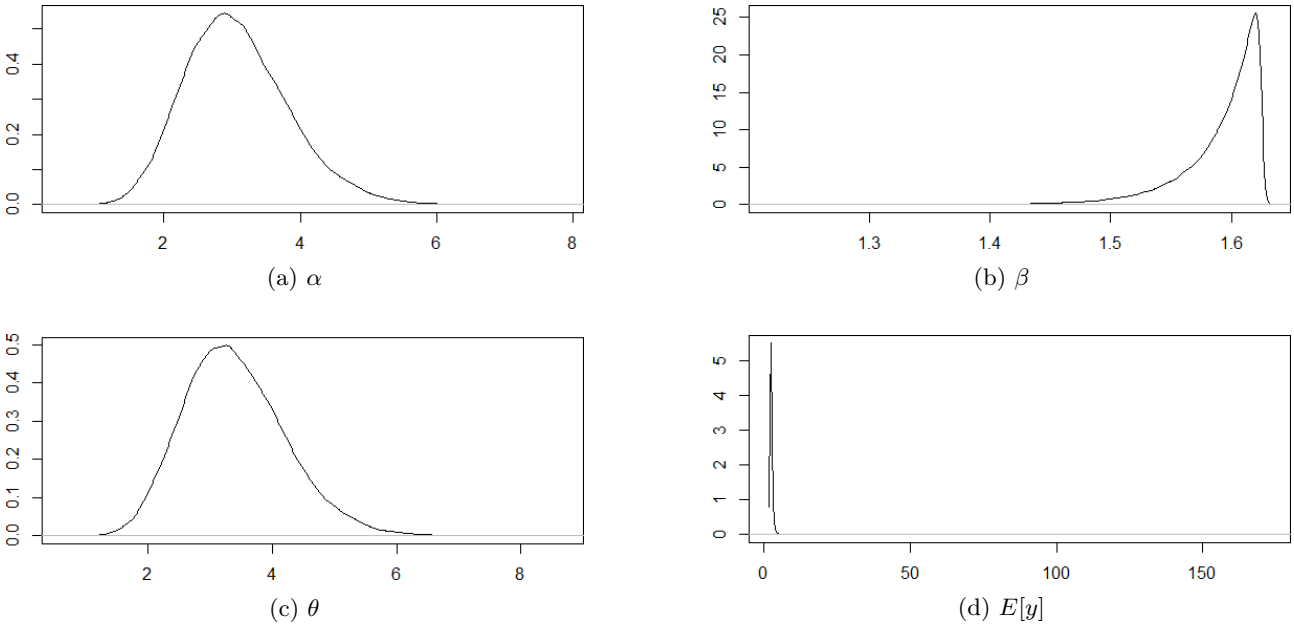
$E[Y]$ was calculated for each simulated set of parameters α and β, and from these values, the mean, standard deviation, and 95% Bayesian credible interval were subsequently computed.

In addition, density plots were generated for each of the parameters and for $E[Y]$, as presented in Figure 1. The resulting densities for the parameters resemble Gamma distributions, with the density of β appropriately truncated at $\min\{y_{i,t}\} = 1.625$. The plot for $E[Y]$ displays a right-skewed distribution that permits very large values, albeit with very low probability—consistent with expectations.

The posterior means of α and β were used as parameters of the Pareto distribution, and the corresponding cumulative distribution function (CDF) was plotted against the empirical cumulative data ($y_{i,t}$). Similarly, the posterior mean of θ was used as the parameter of the Poisson distribution, and its CDF was plotted against the empirical cumulative data (n_t). The Pareto(3.079, 1.592) distribution provides a close fit to the empirical data. The Poisson(3.396) distribution also fits the observed frequencies quite well (see Figures 2 and 3).

[4]Markov Chain Monte Carlo (MCMC) methods, like Gibbs sampling, are used to draw samples from intractable posterior distributions. Gibbs sampling is efficient when full conditional posteriors are in closed form, often with conjugate priors. When posteriors are not in closed form, the Metropolis-Hastings algorithm can be used to generate candidate values from a proposal distribution.

Figure 1: Posterior Densities



(a) α (b) β

(c) θ (d) $E[y]$

Figure 2: Empirical vs. Fitted Pareto CDF

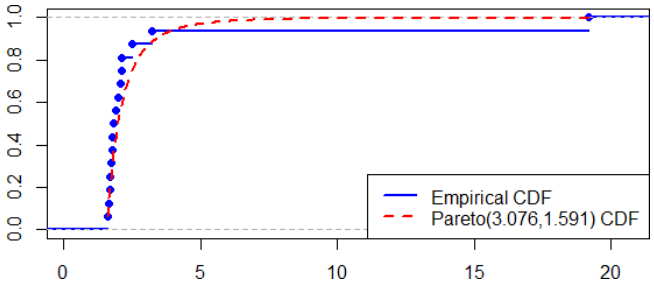
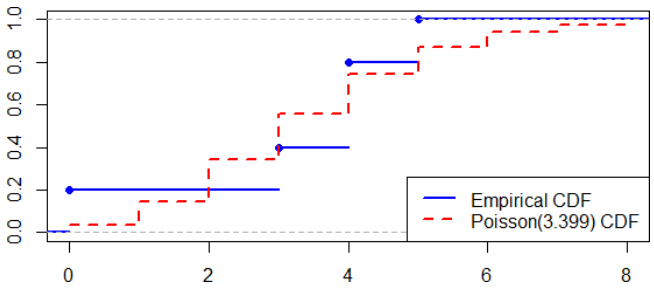


Figure 3: Empirical vs. Fitted Poisson CDF



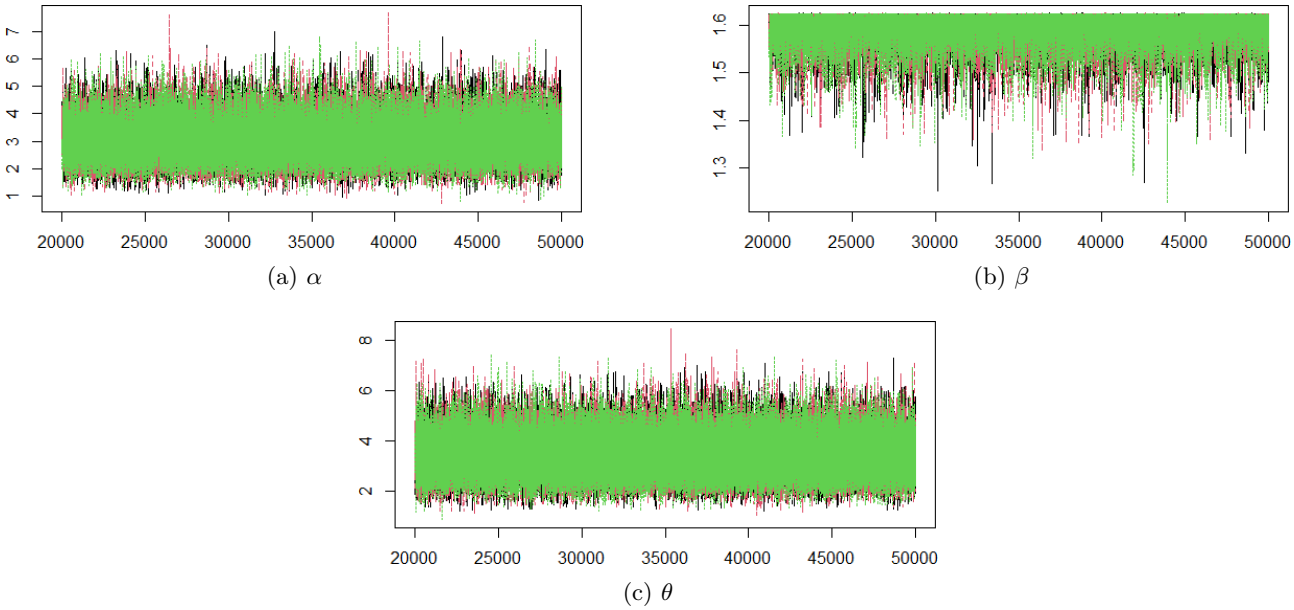
3.3 Convergence Diagnostics

Throughout all computations above, it was assumed that the Markov chains had converged to a stationary distribution after discarding the initial 20,000 iterations, following the approach of Dudley (2006). However, it was essential to verify that convergence had indeed occurred. The first method of assessment involved visual inspection[5]. Figure 4 presents trace plots for all three parameters, showing the sampled values across iterations. These plots indicate good mixing, suggesting that the chains have likely converged.

Convergence was further assessed using the Gelman–Rubin diagnostic (Gelman and Rubin, 1992), applied to the post-burn-in iterations (20,001–50,000). The results, shown in Table 4, indicate that all univariate potential scale reduction factors (PSRFs) have point estimates and upper confidence bounds equal to 1. The multivariate PSRF is also 1. These values again suggest that the Markov chains have likely converged, both individually and jointly.

[5]Visual inspection involves assessing how well a chain explores the parameter space. Poor mixing—where the chain moves slowly or gets stuck—indicates potential convergence issues. Trace plots help identify such problems by showing how sampled values evolve over iterations for each parameter.

Figure 4: Trace Plots



(a) α (b) β

(c) θ

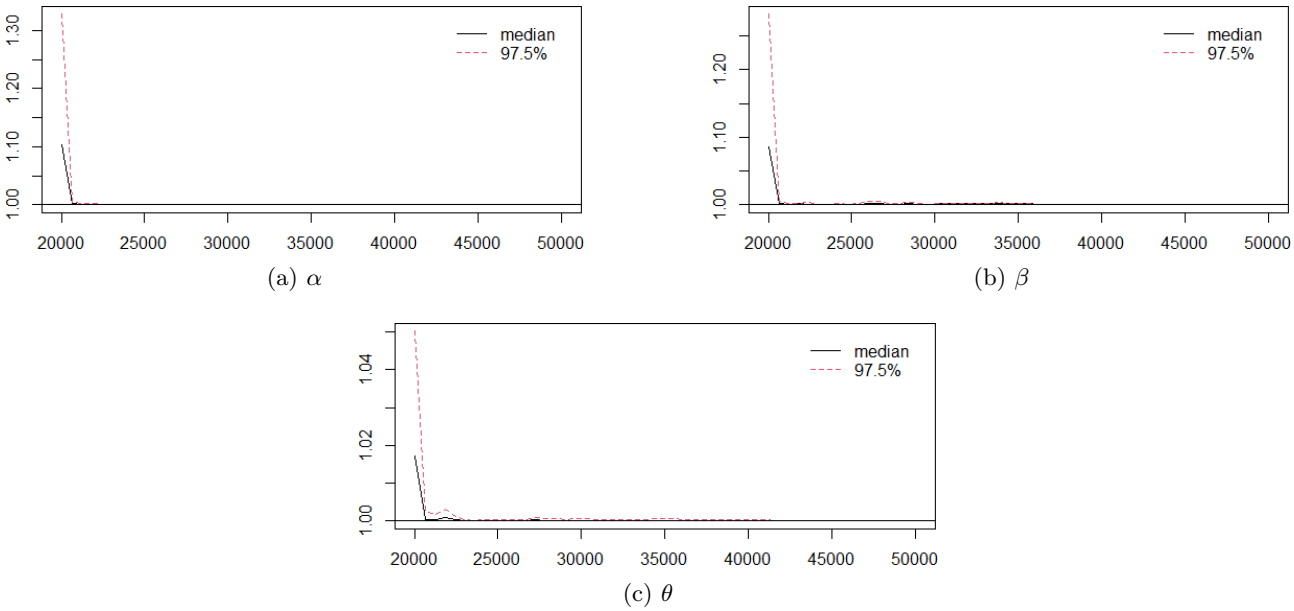
Table 4: Potential Scale Reduction Factors (Gelman–Rubin Diagnostic)

Parameter	**Point Estimate**	**Upper C.I.**
α	1.00	1.00
β	1.00	1.00
θ	1.00	1.00
Multivariate PSRF	1.00	

The diagnostic was applied to iterations 20,001–50,000.

Figure 5 shows how the univariate PSRF point estimates evolve with increasing iterations. Throughout all iterations, all estimates remain below 1.1, which is commonly considered an acceptable threshold for convergence. This further confirms that the chains have likely reached a stable distribution.

Figure 5: PSRF Values (Gelman–Rubin Diagnostic)



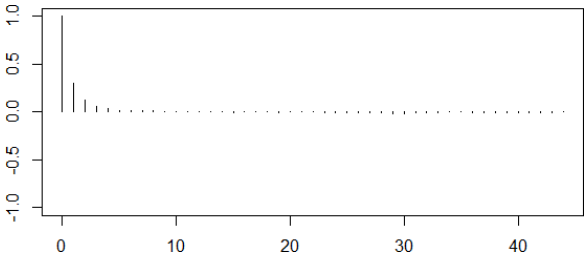
(a) α (b) β

(c) θ

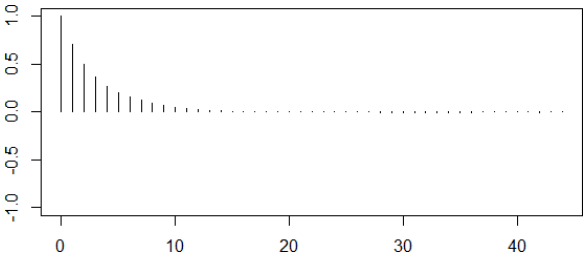
In addition, autocorrelation[6] plots were generated for all three parameters (see Figure 6), and values at lags 1 through 10 are reported in Table 5. Several high autocorrelations were observed, particularly for β, which motivated the use of a thinning interval of 10 iterations, as suggested by Dudley (2006).

[6]The autocorrelation function (ACF) shows the correlation between samples at different lags. High autocorrelation indicates strong dependence between draws, leading to slow mixing. Thinning reduces this dependence by keeping only every kth sample.

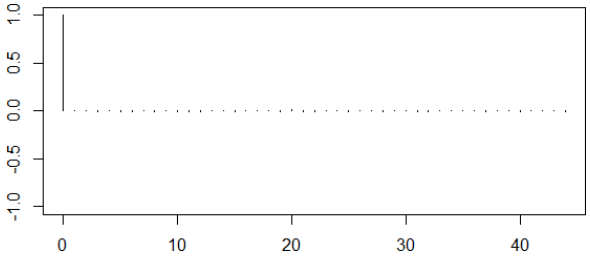
Figure 6: Autocorrelation Plots



(a) α



(b) β



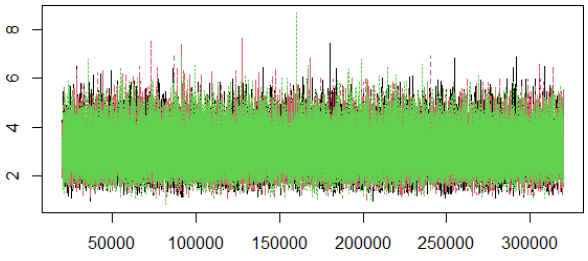
(c) θ

Table 5: Autocorrelations at Lags 1–10

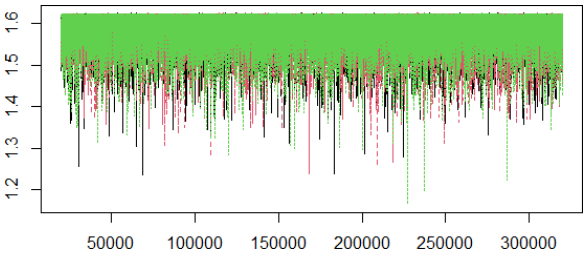
Lag	α	β	θ
1	0.288	0.699	−0.001
2	0.109	0.498	−0.001
3	0.053	0.356	0.003
4	0.035	0.257	−0.007
5	0.019	0.191	0.002
6	0.008	0.144	−0.006
7	0.010	0.104	0.003
8	0.013	0.079	−0.007
9	0.007	0.061	0.001
10	0.006	0.048	0.000

Consequently, the chains were rerun with this thinning. Figure 7 presents the corresponding trace plots, and Figure 8 shows the updated autocorrelation plots. The trace plots indicate that the chains have mixed well, and the autocorrelation plots demonstrate that all autocorrelation values at lags 1, 2, and beyond have become negligible.

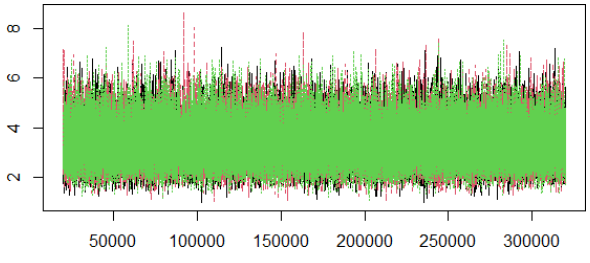
Figure 7: Trace Plots After Thinning



(a) α

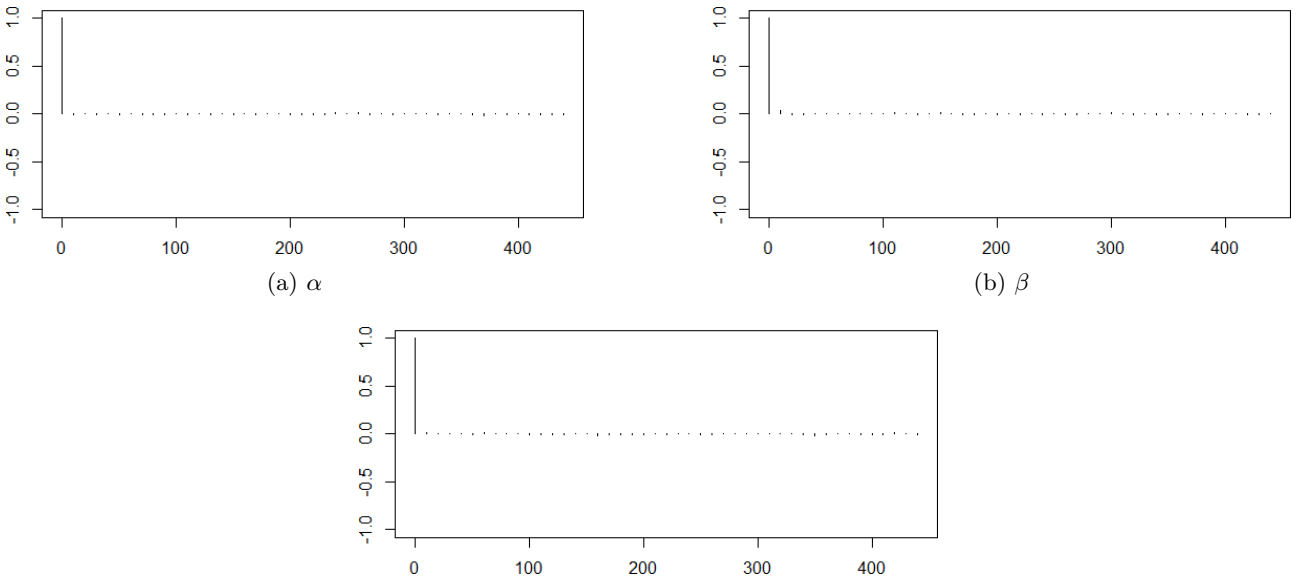


(b) β



(c) θ

Figure 8: Autocorrelation Plots After Thinning



(a) α (b) β

(c) θ

3.4 Predictive Inference

The ultimate objective of the analysis was to predict future values of S_t, which was accomplished using the posterior predictive distribution. For a variable z, this distribution is defined as[7]:

$$\pi(z \mid \mathbf{y}) = \int_{\Theta} f(z \mid \phi)\, \pi(\phi \mid \mathbf{y})\, d\phi$$

where $\pi(\phi \mid \mathbf{y})$ is the posterior distribution of ϕ, and $f(z \mid \phi)$ is the likelihood of z given ϕ.

Let $\mathbf{n}$ denote the observed data, and let N_f represent a future observation of N_t. Then the posterior predictive distribution of N_f is given by:

$$\begin{aligned} p(N_f = n \mid \mathbf{n}) &= \int_0^{\infty} p(N_f = n \mid \theta)\pi(\theta \mid \mathbf{n})\, d\theta \\ &= \mathbb{E}_{\theta|\mathbf{n}}\left[p(N_f = n \mid \theta)\right] \\ &= \mathbb{E}_{\theta|\mathbf{n}}\left[\frac{\theta^n e^{-\theta}}{n!}\right] \end{aligned}$$

Since the integral cannot be evaluated analytically, it is typically approximated using samples from the posterior distribution obtained via MCMC. Specifically, the expectation is approximated as:

$$p(N_f = n \mid \mathbf{n}) \approx \frac{1}{m}\sum_{i=1}^{m} \frac{(\theta^{(i)})^n e^{-\theta^{(i)}}}{n!}$$

where $\theta^{(i)}$ is the i-th sample from the MCMC chain and m is the number of iterations after burn-in and thinning.

Similarly, let $\mathbf{y}$ denote the observed data, and let Y_f represent a future observation of $Y_{i,t}$. Then the posterior predictive cumulative distribution function (CDF) of Y_f is given by:

$$\begin{aligned} p(Y_f \le y \mid \mathbf{y}) &= \int_{\mathbf{u}} p(Y_f \le y \mid \mathbf{u})\, \pi(\mathbf{u} \mid \mathbf{y})\, d\mathbf{u} \\ &= \mathbb{E}_{\mathbf{u}|\mathbf{y}}\left[p(Y_f \le y \mid \alpha, \beta)\right] \end{aligned}$$

Again,

$$p(Y_f \le y \mid \mathbf{y}) \approx \frac{1}{m}\sum_{i=1}^{m}\left(1 - \left(\frac{\beta^{(i)}}{y}\right)^{\alpha^{(i)}}\right)$$

where $\alpha^{(i)}$ and $\beta^{(i)}$ denote the i-th samples from the MCMC chain.

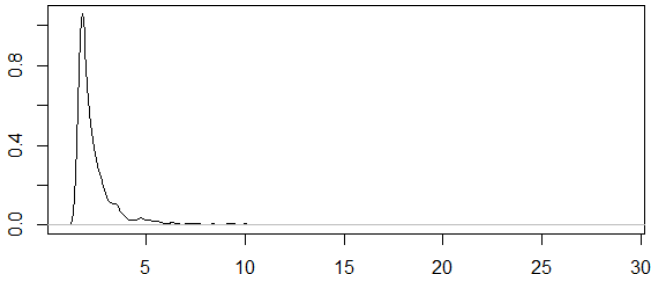
[7]This method accounts for the uncertainty in $\boldsymbol{\phi}$ by integrating over its possible values, weighted by their posterior probabilities.

Table 6 shows the estimated probabilities of $N_f = n$ for $n = 0, \ldots, 14$, which are consistent with the findings of Dudley (2006). Figure 9 displays the estimated predictive probability density function (PDF) of Y_f, obtained using the inverse CDF method[8].

Table 6: Estimates of $p(N_f = n \mid \mathbf{n})$

n	**Probability**
0	0.0453
1	0.1282
2	0.1919
3	0.2023
4	0.1683
5	0.1177
6	0.0718
7	0.0393
8	0.0196
9	0.0091
10	0.0039
11	0.0016
12	0.0006
13	0.0002
14	0.0001

Figure 9: Estimated Predictive PDF of Y_f



The draws of S_f, representing a future observation of S_t, were generated as follows:

1. 1,000 values of N_f were drawn using the inverse CDF method[9].
2. For each simulated value of N_f, the corresponding number of Y_f values was drawn using the same procedure as in the predictive PDF estimation, and these values were then summed to generate a draw of S_f.

Figure 10 presents the histogram of the resulting S_f samples, along with three fitted distributions using moment matching. As observed by Dudley (2006), the Gamma distribution provides the best fit. The fitted Gamma distribution has parameters $\alpha = 1.726$ and $\beta = 0.202$.

Figure 10: Histogram and Fitted Distributions for Predictive S_f

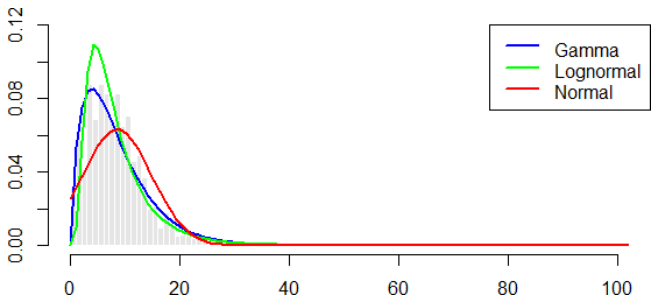


Table 7 reports various percentiles of the simulated S_f values. The maximum value of 116.396 reflects a long tail, consistent with expectations for a heavy-tailed claim size distribution. This indicates that the simulation effectively captured the tail behavior of the predictive distribution of S, which is important since most aggregate claims are moderate, but extreme values can occasionally occur.

[8]Specifically, 1,000 values $U \sim \text{Uniform}(0, 1)$ were generated. For each value, the transformation $\beta^{(i)}(1-U)^{-1/\alpha^{(i)}}$ was applied for each i, and the results were averaged across i. The resulting 1,000 values were then used to approximate the PDF of Y_f via kernel density estimation (KDE).

[9]Once again, 1,000 values $U \sim \text{Uniform}(0, 1)$ were generated. For each value, Poisson samples were drawn for each $\theta^{(i)}$ using the inverse CDF method. The results were averaged across i and rounded to produce 1,000 integer values.

Table 7: Percentiles of Simulated S_f Values

Percentile	Value
Median	7.504
90th Percentile	15.065
95th Percentile	18.092
99th Percentile	25.171
Maximum	116.396

4 Implementation on Alternative Data

To assess the generalizability of the Bayesian approach proposed by Dudley (2006) for insurance claim modeling, the same methodology was applied to a different dataset. The `itamtplcost` dataset comprises 457 motor insurance claims exceeding 500,000 euros, recorded in Italy between 1997 and 2012 (Dutang, 2020).

Table 8 presents a comparison of summary statistics between the Rytgaard (1990) and `itamtplcost` datasets. Claims in the `itamtplcost` dataset are significantly larger, with a mean of 1.015 billion euros compared to 3.058 million euros in the Rytgaard (1990) dataset. The variability is also greater, with a range of 6,637.34 versus 17.56. Regarding the number of claims per year, the `itamtplcost` dataset exhibits a minimum of 5 and a maximum of 40, while the Rytgaard (1990) dataset ranges from 0 to 5. These differences highlight the greater complexity of the `itamtplcost` dataset, providing a valuable opportunity to assess whether the Poisson–Pareto model continues to yield meaningful posterior inference in a more realistic insurance context.

Table 8: Comparison of Summary Statistics Between Datasets

Dataset	Min	1st Qu.	Median	Mean	3rd Qu.	Max
Claim Amounts (in millions)						
Rytgaard (1990)	1.625	1.745	1.863	3.058	2.109	19.180
`itamtplcost`	2.161	627.719	844.011	1015.352	1224.316	6639.500
Claim Counts						
Rytgaard (1990)	0.00	3.00	4.00	3.20	4.00	5.00
`itamtplcost`	5.00	24.25	30.50	28.56	35.25	40.00

4.1 Poisson–Pareto Model

Similarly to the previous case, Gibbs sampling was used to draw realizations from the posterior distributions of α, β, and θ, with three Markov chains run in parallel and initial parameter values chosen to be well-dispersed. A burn-in of 5,000 iterations was initially applied, but the Gelman–Rubin diagnostic (Gelman and Rubin, 1992) suggested extending it to 10,000. Consequently, a burn-in of 10,000 was adopted, and autocorrelation plots were generated. High autocorrelations were observed, prompting the use of a thinning interval of 10. The chains were rerun with this thinning, and the trace plots, along with the updated autocorrelation plots, indicated good mixing and negligible autocorrelation across all lags.

Despite this, all simulated values of α were below 1, indicating that the mean of the Pareto distribution was infinite. This posed a risk of instability in the results and led to the conclusion that the Poisson–Pareto model was not suitable for the `itamtplcost` dataset. Consequently, alternative loss distributions were considered, including the Weibull and lognormal. Of these two, the lognormal distribution was preferred, as it provided a heavier tail for the predictive distribution of S_f. Alternatives to the Poisson distribution were also explored, since the empirical mean and variance of N_t differed substantially—28.563 versus 90.529—violating the Poisson assumption of equality between mean and variance. This contrasts with the Rytgaard (1990) dataset, where the mean and variance were more closely aligned (3.2 vs. 3.7). Both the Poisson and Negative Binomial distributions provided a good fit. Therefore, two models are reported below: the Poisson–Lognormal and the Negative Binomial–Lognormal.

4.2 Poisson–Lognormal Model

The lognormal distribution was parameterized in terms of μ and τ, where $\tau = 1/\sigma^2$. A vague normal prior was assigned to μ, and vague gamma priors were used for both σ and θ. As in previous models, three Markov chains were run in parallel with well-dispersed initial parameter values. For the third chain, initial values were set equal to the maximum likelihood estimates (see Table 9).

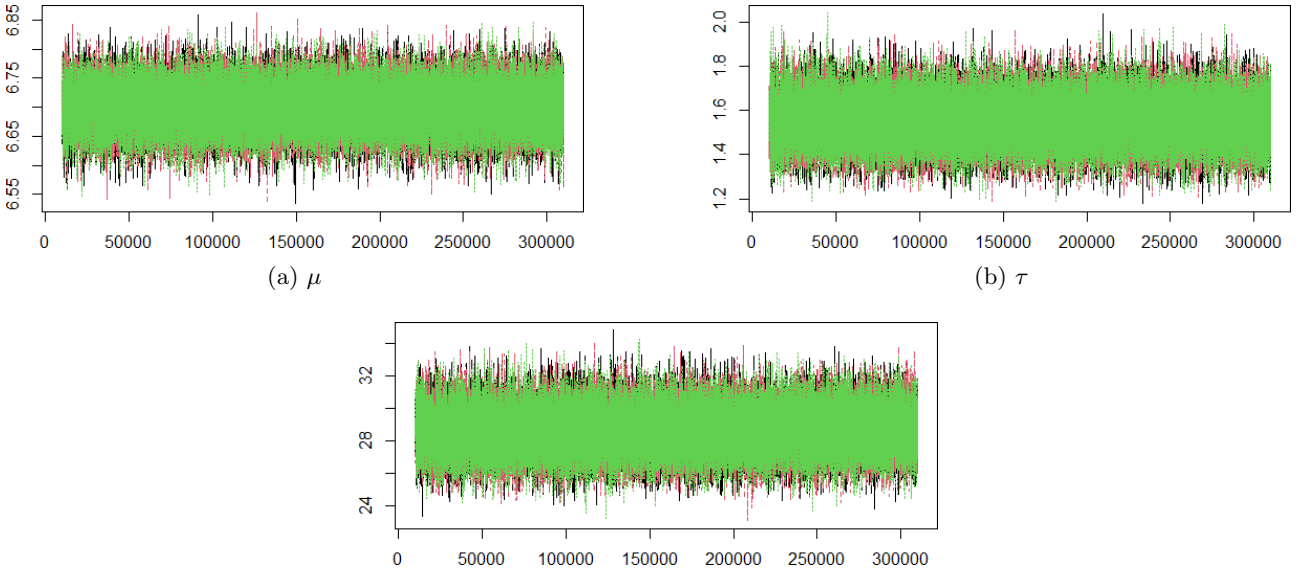
An initial burn-in of 1,000 iterations was applied. However, the Gelman–Rubin diagnostic (1992) suggested that this was insufficient, as convergence had not yet been achieved. The PSRF point estimates stabilized around 1 only after approximately 10,000 iterations, which was then adopted as the revised burn-in. Moderate autocorrelation was observed in the chains, prompting the use of a thinning interval of 10. The chains were rerun with

Table 9: Initial Parameter Values

Chain	μ	σ	θ
1	0.000 01	0.000 01	0.000 01
2	100 000	100 000	100 000
3	6.700	0.802	28.563

this setting, and updated diagnostics showed negligible autocorrelation across all lags. Trace plots of the parameters, shown in Figure 11, support the conclusion of good mixing and convergence.

Figure 11: Trace Plots After Thinning



(a) μ (b) τ

(c) θ

Table 10 reports posterior summary statistics based on 30,000 post-burn-in samples. The posterior means are closely aligned with the maximum likelihood estimates (MLE)[10], as expected under vague priors, since the inference is primarily driven by the data. The sample mean of Y, equal to 1,015.352, lies within the Bayesian credible interval for $\mathbb{E}[Y]$. The posterior mean of $\mathbb{E}[Y]$ is slightly higher, reflecting the influence of several extremely large claim amounts (exceeding 3 billion euros). Compared to the results from the Rytgaard (1990) dataset, the 95% credible intervals are narrower or of comparable width, indicating good estimation precision—likely due to the larger sample size.

Table 10: Posterior Statistics

	Mean	**Standard Deviation**	**95% Bayesian Credible Interval**
μ	6.700	0.038	(6.626, 6.774)
τ	1.552	0.103	(1.358, 1.759)
θ	28.628	1.337	(26.058, 31.313)
$E[Y]$	1124.280	48.873	(1034.404, 1225.412)

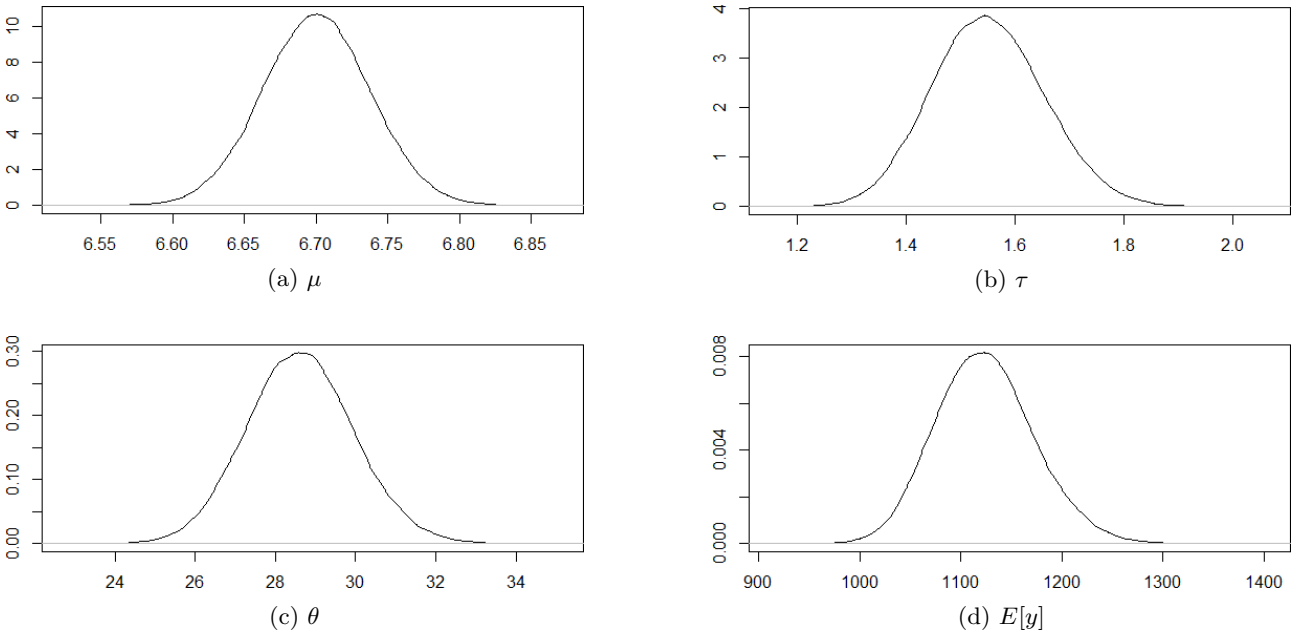
Figure 12 shows the posterior density plots for μ, τ, and θ. The posterior for μ appears approximately normal, while the distributions of τ and θ resemble gamma distributions. The posterior for $\mathbb{E}[Y]$ is right-skewed, though notably less so than in the Rytgaard (1990) dataset.

Figure 13 compares the empirical cumulative distribution of observed claim amounts $y_{i,t}$ with the cumulative distribution function (CDF) of a lognormal distribution parameterized by the posterior means of μ and τ. The Lognormal(6.700, 1.552) distribution closely matches the empirical data. Figure 14 presents a similar comparison for the claim frequency data n_t, evaluated against the CDF of a Poisson distribution with mean equal to the posterior mean of θ. The Poisson(28.628) distribution captures the overall trend of the empirical distribution well, though there is room for improvement.

Figure 15 displays the estimated predictive probability density function (PDF) of a future individual claim amount Y_f, derived via the inverse CDF method, as in the previous section. The resulting distribution is markedly right-skewed, capturing the heavy-tailed nature of claim sizes. While extremely large values are possible, they occur with very low probability.

[10]The MLE for τ is 1.554.

Figure 12: Posterior Densities



(a) μ (b) τ

(c) θ (d) $E[y]$

Figure 13: Empirical vs. Fitted Lognormal CDF

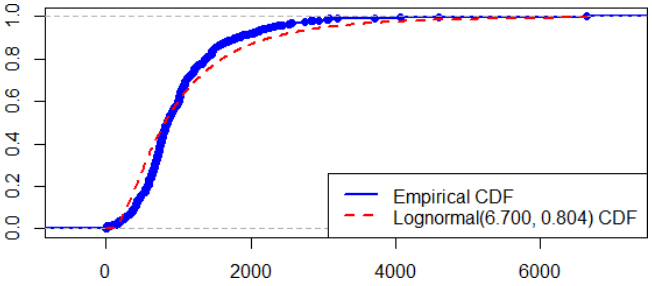


Figure 14: Empirical vs. Fitted Poisson CDF

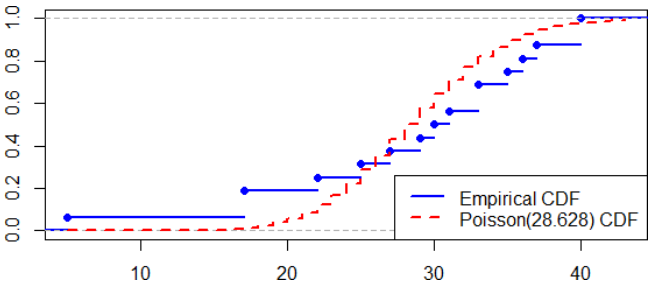
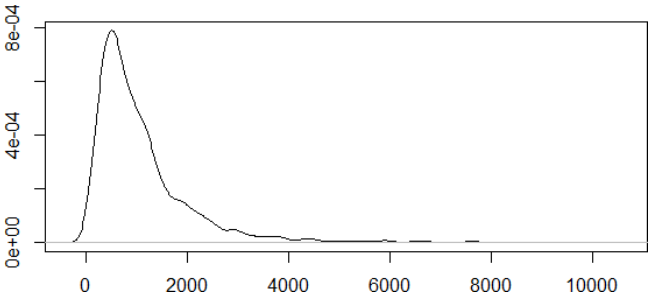


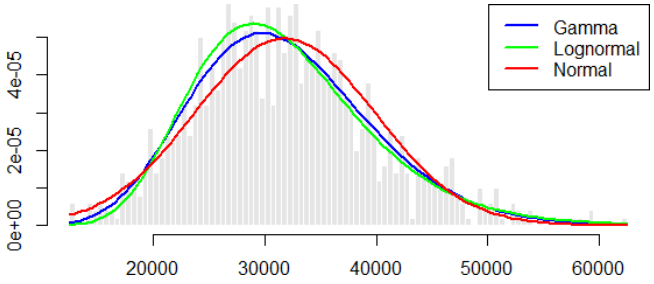
Figure 15: Estimated Predictive PDF of Y_f



Finally, Figure 16 presents the histogram of simulated aggregate claims S_f, obtained using the same predictive simulation approach as in the previous section. Three theoretical distributions were fitted to the simulated values using moment matching. Among these, the Gamma distribution appears to provide the best fit, with esti-

mated parameters $\alpha = 15.6920$ and $\beta = 0.0005$.

Figure 16: Histogram and Fitted Distributions for Predictive S_f



4.3 Negative Binomial–Lognormal Model

References

Dudley, C. (2006). Bayesian Analysis of an Aggregate Claim Model Using Various Loss Distributions. Master's dissertation, Heriot-Watt University, School of Mathematical and Computer Sciences, Actuarial Mathematics & Statistics.

Dutang, C. (2020). CASdatasets: Insurance Datasets. R package version 1.0.

Gelman, A. and Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, 7(4):457–472.

Rytgaard, M. (1990). Estimation in the Pareto Distribution. *ASTIN Bulletin*, 20(2):201–216.