

A Bayesian Model for Estimating Long Tailed Excess of Loss Reinsurance Loss Costs

Abstract: In this paper we will describe a Bayesian model for excess of loss reinsurance pricing which has many advantages over existing methods. The model is currently used in production for multiple lines of business at one of the world's largest reinsurers. This model treats frequency and severity separately. In estimating ultimate frequency, the model analyzes nominal claim count data jointly against uncertain ultimate frequency and development pattern priors, allowing for more careful analysis of sparse claim count information and properly differentiating between triangulated and last diagonal data. The severity model is pragmatic, yet accounts for severity distribution development and weighs the volume of data against prior distributions. The model is programmed in R and Stan, thus eliminating the need for a considerable amount of algebra and calculus and the necessity to use conjugate prior distribution families. We compare this method with the more established Buhlmann-Straub credibility application to excess of loss pricing (for instance in Cockroft), and the more complex model given by Mildenhall, showing numerous advantages of our method.

Keywords: Credibility, Bayesian statistics, Reinsurance, Excess of loss pricing, Pareto distribution, Gamma distribution, Poisson distribution, Weibull distribution, Copulas

1. INTRODUCTION

Cedent claim datasets used in the pricing of excess of loss reinsurance are often limited in size. Individual claim data is provided for claims above a certain threshold and can produce sparse excess claim count triangles for long tailed lines. In some cases, excess triangles might not be available. A difficult problem faced by reinsurance pricing actuaries is how much credibility to assign cedent experience with limited excess claims and how to blend this data with industry excess loss development patterns and exposure rating methods.

We introduce a hierarchical Bayesian model which allows us to specify prior distributions and use Bayes' Theorem to estimate posterior distributions based on the data. The model is programmed in R and Stan. The code, a sample dataset and case studies are provided so users can easily experiment for themselves. The foundation of this model is a standard Poisson-Pareto compound frequency severity distribution with gamma prior distributions. It includes a claim count emergence model that combines a Weibull distribution with gamma marginal priors. Codependence between the two Weibull parameters, which was found in our study, is described using a copula function. The model is used in practice for multiple long-tailed lines of business at a major reinsurance company. The next section describes practical applicabilities and benefits practitioners can enjoy.

The fundamental question on excess loss frequency is whether observing very few claims reported in an immature year is due to low ultimate frequency or slow development. Conversely, does a high observed claim count in an immature year imply high ultimate frequency or just quick development? The answer is that it could be an indication of either situation or a combination of both. Observed claim counts inform our view on the ultimate

frequency and development pattern. How strong of an indicator the observed claim counts are depends on the relative uncertainty in the ultimate frequency compared to the uncertainty in the development pattern. The frequency model compares the observed claim counts to the joint probability of that observation under the uncertain ultimate frequency distribution and uncertain development pattern, allowing the two major assumptions (i.e., prior distributions) to be weighed against each other simultaneously.

To complete the compound frequency severity distribution, we will also explain how to develop the Gamma-Pareto ultimate severity model. We will detail how our method addresses additional factors related to severity such as development, policy limit capping and loss adjustment expenses. Finally, we discuss comparisons to existing methods and Appendix A contains helpful one-page step-by-step guides to implementing the model in practice.

2. FREQUENCY MODEL

In this section we present a Bayesian frequency model for long-tailed, reinsurance excess of loss claim counts. A diagram of the frequency model is provided in Figure 1. Each component will be discussed at length in subsequent sections of this paper.

The quantity of interest in this model is the ultimate expected frequency per exposure, λ . However, we can only observe incremental claim counts in finite time intervals. A key concept is that claim counts are observed over a certain time period and the expected number of claims observed in that time period, given an expected ultimate frequency, is determined by the development pattern. The Poisson formula in the observation level of the diagram is the part of the model that connects the observed claim counts with the prior distributions of the development pattern and the ultimate expected frequency in order to estimate posterior distributions. The relative certainty or uncertainty in each of the two company level components of the diagram (i.e., ultimate expected frequency and development pattern) are simultaneously weighed against each other to derive posterior distributions.

Note that $CDF(t_2) - CDF(t_1)$ is the incremental expected percent of claims reported according to the Weibull development pattern assumption, where CDF refers to the cumulative distribution function. Also note that a hyperparameter in the context of a Bayesian hierarchical model is simply a parameter of a prior distribution in the model. Adding an additional “hyper” suffix simply moves us one level up the model hierarchy of parameters and prior distributions.

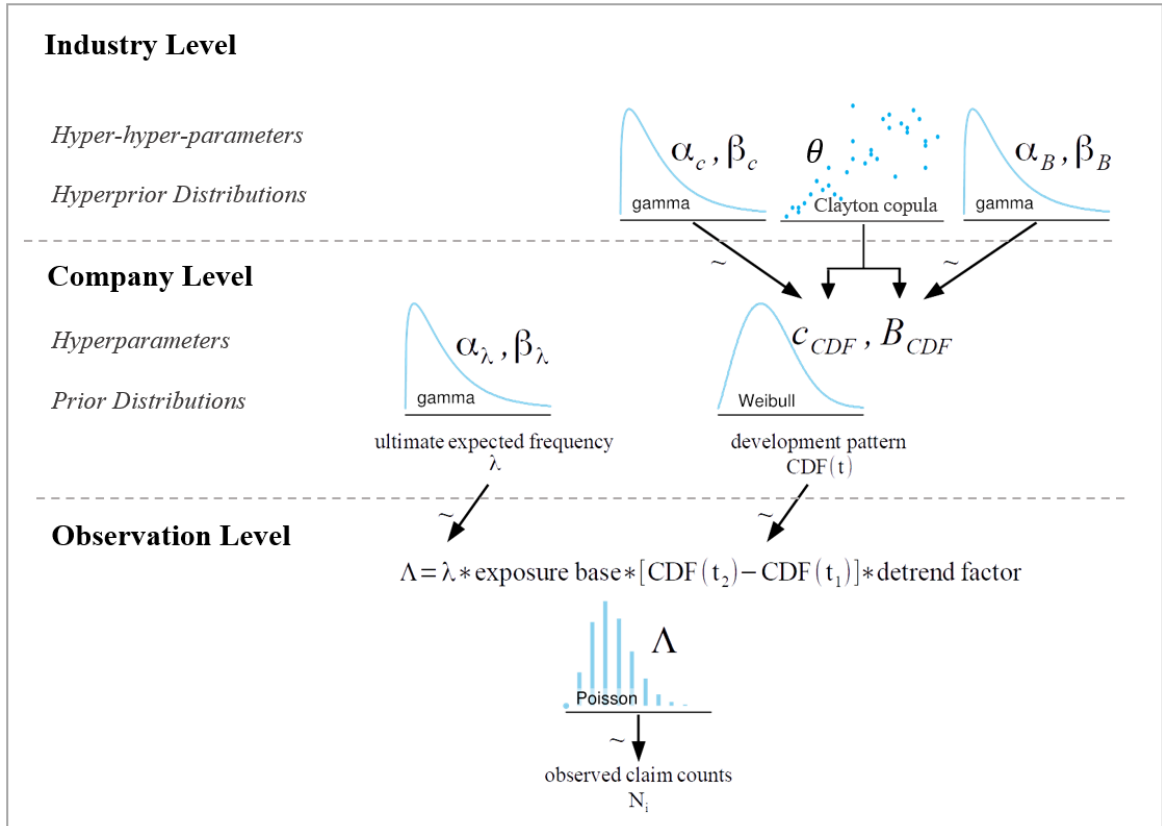


Figure 1. Diagram of Hierarchical Bayesian Frequency Model

The following are several advantages of the frequency model that we feel are important motivations for investing your time and understanding. Each advantage considers development pattern uncertainty or aims for a more consistent and efficient application of assumptions across cedents. We hope it will be useful for readers to know that these are actual benefits the authors have seen from the use of this model in a live reinsurance renewal pricing process, not just theoretical benefits that could potentially materialize. The technical explanation of the model continues in the next section.

Development Pattern Uncertainty

- i. The frequency model incorporates development pattern uncertainty on large loss frequency. A common subjective consideration in reinsurance is the development pattern of the cedent and its relation to an assumed default pattern. Usually, a large amount of data is required in traditional pricing techniques to use a cedent pattern instead of a default pattern. Having an uncertain Weibull development distribution

with prior hyperparameters allows the cedent data to inform the pattern, i.e., update the posterior, to the extent it is credible.

- ii. Frequency and development are evaluated at the same time and the model weighs the relative uncertainty in both. Observed claim counts by year and development period are compared to a distribution of possible observed claim counts based on the prior distributions of ultimate frequency and the incremental percent of claims reported. The claim count data contains information about both the development pattern and the ultimate frequency, and the model attempts to extract and separate that information according to the prior distributions and Bayes' Theorem.
- iii. Data used in the frequency model can be triangulated or in last diagonal form – i.e., a large loss listing as of a particular date. The model has a way of judging the amount of information contained in a triangle. It differentiates credibility between triangulated and non-triangulated data so that credibility assignment increases when a cedent triangle is available.

Consistency

- i. Actuarial judgement is applied in specification of the prior, not in the individual loss selection for each cedent or final loss estimate where a subjective credibility is assigned to experience. When given those priors, every cedent is analyzed on the same basis and thus removes some subjectivity of experience rating and credibility selection. The method of implying credibility is applied consistently among all cedents. Two cedents with same mean and amount of data will get same posterior results. No judgement is used in the selection of credibility, leading to a more consistent blending of exposure rating and experience.
- ii. The model meaningfully and consistently incorporates cedent data with few historical losses and sparse development triangles, including being able to use and appropriately update prior distributions when zero claims are observed. Credibility implied by the model varies by “sparseness” of the data triangle.

Efficiency

- i. R and Stan codes derive posterior ultimate excess frequencies simultaneously for

multiple cedents. This approach allows for increased efficiencies in the reinsurance pricing process. Prior distribution parameters and company priors can be updated during times of the year when a reinsurer's workload is slower and not as hectic.

- ii. Closed form posteriors are unnecessary for this model because everything is programmed using R and Stan. Formulas are nice in theory but in practice there is usually a danger that long formulas in papers have typos, and there is the possibility of translation error between a formula in a paper and recreating that formula in Excel or a programming language. Bayesian MCMC produces samples of the posterior from which all desired values can be estimated without having to deal with algebraic formulas.

2.1 The Observation Level of the Hierarchical Model

We assume observed claims are Poisson distributed. When we observe a claim count in a certain time period, we can then compare that number with the theoretical mean of the Poisson which is a function of both the Weibull distribution of the development pattern and the gamma distribution of the expected ultimate frequency as shown in Figure 1.

We use the following parameterization of the Poisson distribution:

$$p(k; \lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$$

One might ask, why not just develop claims to ultimate using traditional techniques, and then treat those ultimate claim counts as observations of a Poisson distribution to more directly infer the ultimate frequency posterior? The answer is that this does not incorporate pattern uncertainty. A traditional development to ultimate will not capture the greater uncertainty (and weaker information relative to the prior) of the newer ages. We believe the model presented in this paper is a relatively direct way of incorporating pattern uncertainty into a large loss frequency analysis.

2.2 Data Organization and Model Inputs

Our study included 35 cedents providing primary General Liability coverage. The approach outlined in this paper can also be applied to other long-tailed lines of business such as Auto Liability or Workers' Compensation. Our General Liability dataset contained more than 12,500 claims with 3100 claims above \$500,000. All companies had at least 10 accident years

of information with some going back 20 years. There were 10 cedents with triangulated individual claims data and the remainder only provided claims along the last diagonal.

Reporting thresholds in the dataset were not consistent for each company. Less than a handful of cedents provided ground up claims. The other thresholds ranged from \$50,000 to \$500,000 with \$500,000 being the most common. Therefore, our Bayesian model was constructed with threshold of \$500,000.

Table 1 shows how to organize input data for the frequency model. Incremental claim counts are excess of a fixed threshold for all years and all companies. Columns **obs_start** and **obs_end** indicate starting and ending points of an observation age.

company	year	premium	obs_start	obs_end	incr_claims
1	2010	150,000,000	0	11	14
1	2011	153,000,000	0	10	11
1	2012	156,060,000	0	9	12
1	2013	159,181,200	0	8	7
1	2014	162,364,824	0	7	17
1	2015	165,612,120	0	6	9
1	2016	168,924,363	0	5	10
1	2017	172,302,850	0	4	13
1	2018	175,748,907	0	3	10
1	2019	179,263,885	0	2	8
1	2020	182,849,163	0	1	2
2	2010	54,000,000	0	1	10
2	2010	54,000,000	1	2	2
2	2010	54,000,000	2	3	5
2	2010	54,000,000	3	4	1
2	2010	54,000,000	4	5	1
2	2010	54,000,000	5	6	2
2	2010	54,000,000	6	7	0
2	2010	54,000,000	7	8	0
2	2010	54,000,000	8	9	0
2	2010	54,000,000	9	10	0
2	2010	54,000,000	10	10.5	0
⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮

Table 1. Sample Frequency Model Data

Notice that starting observation ages are equal to zero in every year for Company 1. This set up is used when only the last diagonal of data is received. Compare this to Company 2 where data for year 2010 is entered based on a claim count triangle. The input file, and the Bayesian frequency model, is structured to accommodate both triangulated and non-triangulated data.

Regarding the issue of loss trend, the model uses nominal data as input rather than trended data. Under “observed claim counts” in Figure 1, we show the application of a detrend factor. This is an excess frequency trend factor which includes ground up severity trend and the impact on excess frequency from ground up severity trend. Appendix B includes the frequency model Stan code. An assumed detrend factor of 10% is entered in the body of the code.

Since the ultimate expected frequency value refers to the prospective analysis year, we detrend the frequency to an expected amount for each prior accident year. Losses increase in severity over time which implies that the expected ultimate frequency at a fixed nominal threshold should be lower for each previous year as one goes back in time¹. There is no provision or allowance in this model for uncertainty in the trend, such as a prior distribution of annual trend amount, but it could be added relatively easily.

Input data assumes all premium and exposures are brought to current levels. Exposure trend and other premium related trends are accounted for by using onlevel exposure or onlevel premium as data input under the “Premium” column in Table 1. Premium is used in our example as a proxy for exposure, but the most accurate measure of frequency is number of claims per exposure base.

An important note on Company 2 is with regard to the **obs_end** value of 10.5 for year 2010. Section 2.4 will provide details on how the frequency model is able to handle data with partial year development. Company 2 demonstrates a common reinsurance scenario where a 12-month claim count development triangle is provided with a partial last diagonal – e.g., as of mid-year or June 30th.

When working with triangulated data, incremental claim counts are used because there is correlation between cumulative claims at each age. Modeling cumulative claim counts requires additional parameterization of the correlation between ages which is an unnecessary

¹ Another way to recognize the impact of trend would be to detrend the loss threshold for older years. In that case, the expected ultimate frequency would be the same for each year but that frequency would be observed at a different threshold each year. We designed the model with a fixed nominal loss threshold for all years because that is the typical presentation of losses in a reinsurance submission, hence data available to reinsurance actuaries.

Also, a flat trend for all years may not be technically correct depending on the shape of the severity distribution. For many distribution families, there is a leveraging effect where the increase in the probability of a large loss is greater than the increase in the average severity if a fixed factor is applied to all claim sizes.

complication. We work with the simpler assumption that given an ultimate frequency and Weibull development pattern the observed claim counts from disjoint time periods are independent.

2.3 Ultimate Expected Frequency and Parameterizing the Prior

Ultimate expected frequency in the reinsurance layer is the quantity of interest because it determines the reinsurance loss cost. However, it is unobservable. We can only make inferences based on a prior distribution and observed claim counts at finite ages. Ultimate frequency is defined in the model as ultimate expected claim counts excess of a threshold per unit of premium.

An ultimate frequency prior, denoted by $\hat{\lambda}$, is established for each company in the portfolio. The authors of this paper derived priors using an exposure model, see (Clark 2014) page 27-31. However, other methods of determining an a priori are also acceptable. In our study containing 35 cedents, a gamma distribution was a reasonable fit to the set of ultimate expected frequencies indicated by the classical exposure model.

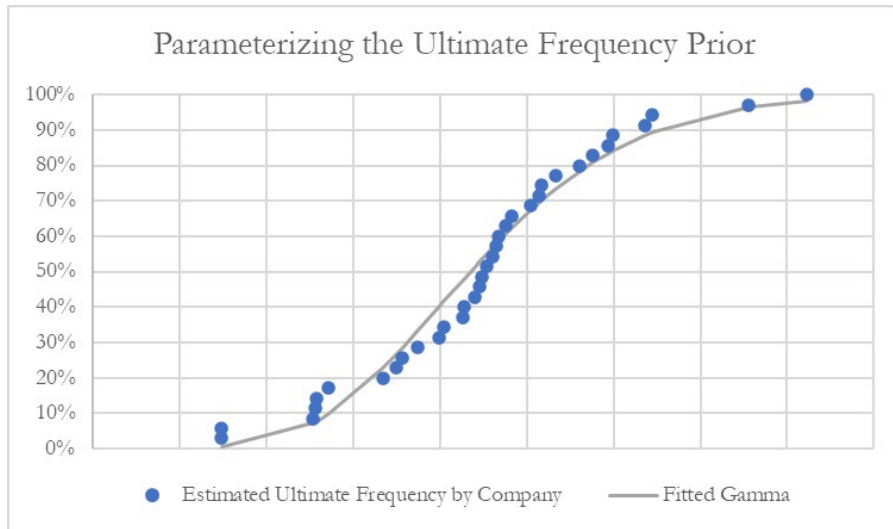


Figure 2. Fitted Gamma to empirical distribution of $\hat{\lambda}_{\text{prior_co}[i]}$

We use the following parameterization of the gamma distribution:

$$f(\lambda; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}$$

The mean of a gamma distribution is α/β . Therefore, we have:

$$\lambda_{prior_co[i]} \sim \Gamma(\beta * \hat{\lambda}_{prior_co[i]}, \beta)$$

A universal value for β is assumed. It represents the variability of each individual cedent's actual frequency from the exposure model estimate, which may be based only on a few variables such as state and product mix without accounting for things specific to the cedent like risk selection or claims handling skill/philosophy/etc.

Theoretically there is a difference between the variability among each company's exposure model frequency, versus the amount a cedent could differ from their own exposure model, but in the absence of a good way of measuring the latter, we use the former as a proxy.

2.4 Claim Count Development

A distribution of excess claim count development patterns is assumed in order to recognize that losses develop differently for each cedent, and that each cedent's development pattern is uncertain. The frequency model responds in a proper way to the data based on the level of certainty represented by the prior distributions of the development pattern.

It is preferable to select an excess claim count development pattern from a family of continuous distributions over discrete distributions. This allows for data with partial year development to be used as inputs into the model. Reinsurance submission data is rarely received as of December 31st. Continuous distributions have the flexibility to model latest diagonals at non-integer values such as 6 months or 18 months.

A Weibull distribution was found to fit excess development patterns reasonably well for a substantial number of cedents. We will use the following notation for a cumulative claim count pattern:

$$F(x) = 1 - e^{-(x/B)^c}; x \geq 0$$

The scale parameter, B, controls the development pattern speed with higher values corresponding to a longer tail. The value of this parameter approximates the number of years until claims reach 63% development to ultimate. To see this, plugging in $x=B$ produces the following cumulative development regardless of the value of c:

$$F(B) = 1 - \frac{1}{e}$$

Figure 3 provides the development pattern under a Weibull distribution for fixed c and varying values for B. Much faster development is observed for lower B values while higher

scale parameter values substantially slow down the development pattern.

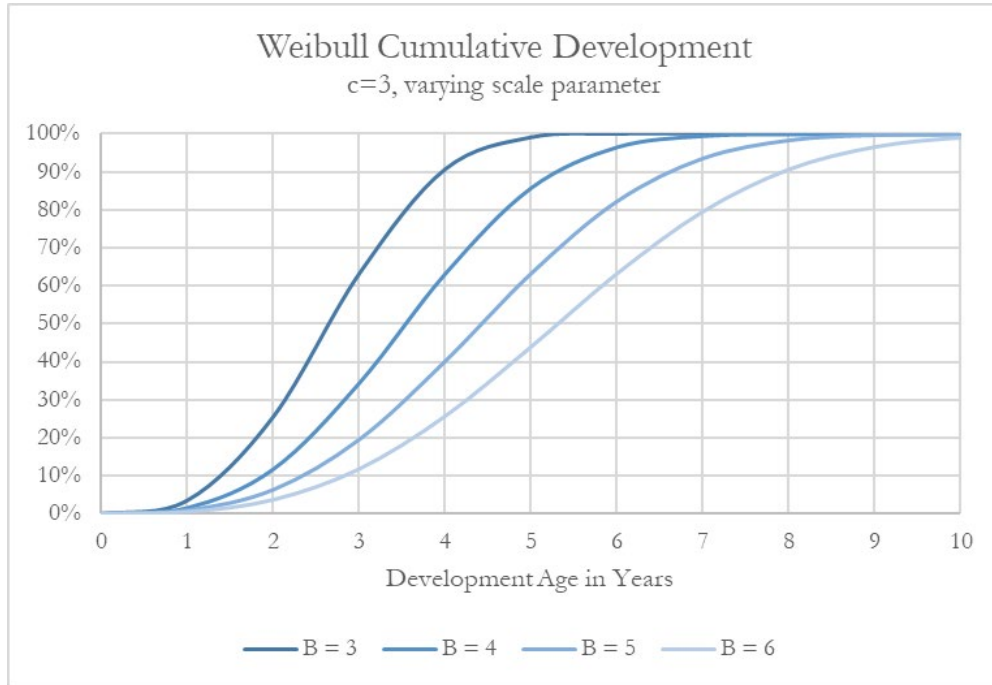


Figure 3. Weibull cumulative development pattern with fixed shape and varying scale parameter

According to Figure 4, the shape parameter, c , also impacts the length of the tail. Lower values increase the tail length while higher values shorten the length assuming similar ranges of B . The distribution also becomes more S-shaped with an increasing c value.

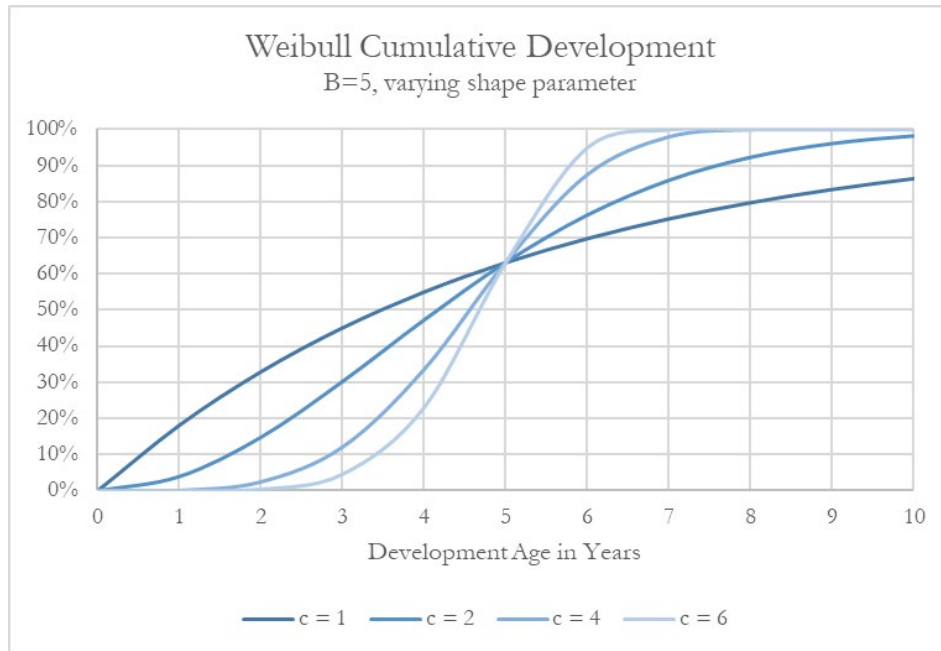


Figure 4. Weibull cumulative development pattern with fixed scale and varying shape parameter

2.5 Parameterizing the Weibull Development Prior

Not all cedents have a full triangle or enough data to estimate a complete development pattern. Our dataset included 10 companies with triangulated individual claims. Of those companies, 8 had enough data with which to estimate a development pattern. This subset of cedents is used to parameterize the development pattern prior distributions. A Weibull distribution is assumed to describe the cumulative excess development pattern for each cedent with sufficient triangulated data. Judgement needs to be used as it typically is in selecting LDFs and the tail factor. Once a pattern is selected for each cedent, a Weibull distribution is fit to that pattern using maximum likelihood or any other suitable fitting algorithm so that a portfolio of Weibull parameters is constructed. Our frequency model included 8 shape and 8 scale parameters, one set of parameters for each cedent having enough data to generate a development pattern.

A distribution fitting exercise performed on the set of Weibull shape parameters and another distribution fitting on the set of scale parameters produces marginal prior distributions for c and B . Gamma prior distributions were a reasonable fit for each parameter:

$$c_{prior} \sim \Gamma(\alpha_{c_shape}, \beta_{c_scale})$$

$$B_{prior} \sim \Gamma(\alpha_{B_shape}, \beta_{B_scale})$$

α_{c_shape} , β_{c_scale} , α_{B_shape} and β_{B_scale} are all fixed values based on the distribution fit.

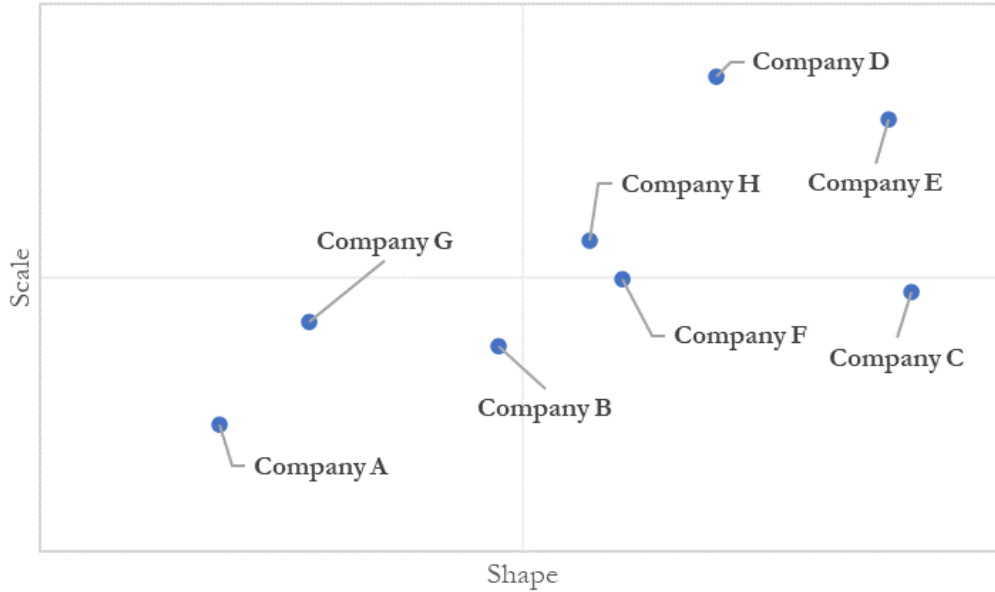


Figure 5. Weibull parameter joint distribution resembles a Clayton copula

Plotting combinations of c and B values for each cedent generates the graph in Figure 5. It appears that the c and B values are not independent but have positive correlation. To keep this relationship in our frequency model, we introduced a copula dependence between each company's c and B values.

Different subsets of data may be modelled with different copula functions. A bivariate Clayton copula is the best fit to our data as determined by the Akaike information criterion (AIC). Figure 6 shows resulting AIC rankings and data points from the fitted joint distribution.

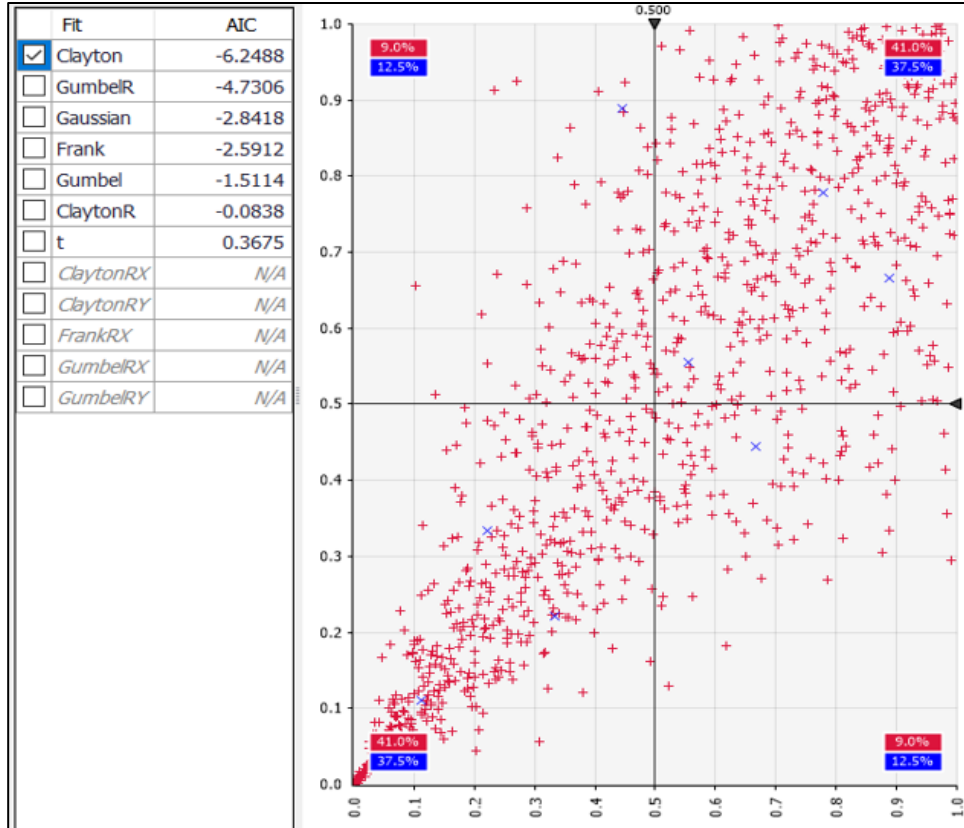


Figure 6. Fitted Clayton copula and Akaike information criterion rankings

We use the following notation for the cumulative density of a Clayton copula:

$$C(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}}$$

Modeling shape and scale dependence assuming a fixed value of θ results in posterior parameter estimates that fit the original data rather well. When plotted on a unit scale in Figure 7, one can clearly see the greater (left) negative tail dependence exhibited by both the observed data points and the Bayesian posteriors.

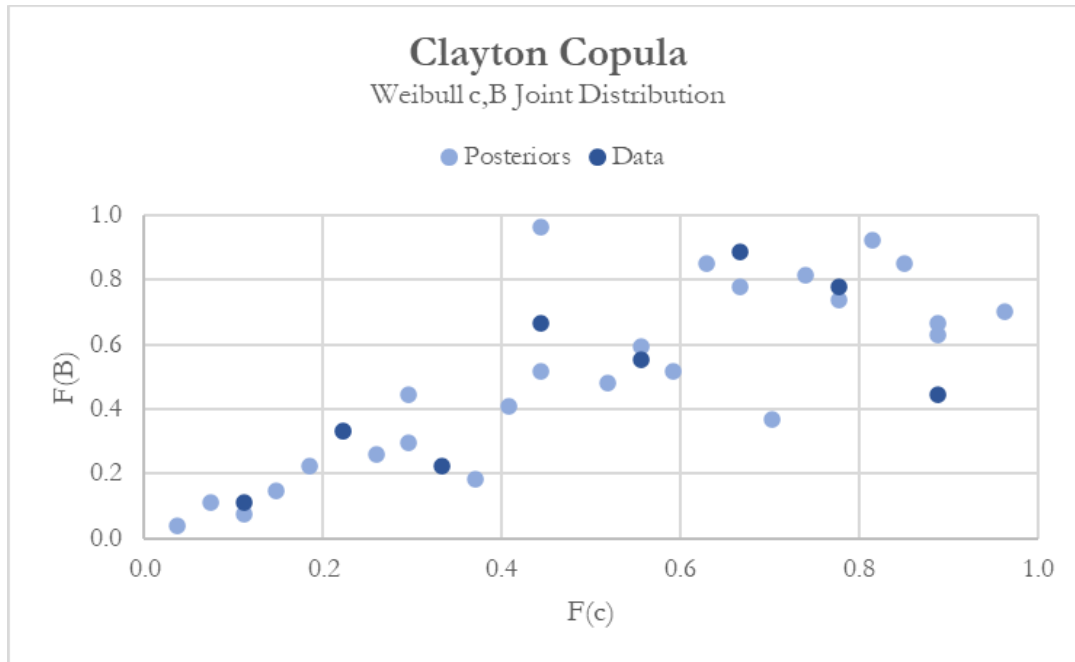


Figure 7. Posterior Weibull parameters plotted against cedent data

2.6 Implied Credibility and the Value of a Triangle

An implied credibility factor, Z_i , can be calculated for each cedent by specifying:

$$\lambda_{posterior}[i] = (1 - Z_i) * \lambda_{prior}[i] + Z_i * \lambda_{empirical}[i]$$

where $\lambda_{empirical}[i]$ is the maximum likelihood estimate of the Poisson rate parameter based only on the i^{th} cedent's data. In practice we found these values by running the frequency model with uninformative priors.

In Figure 8 we plot the implied credibility factor for each cedent against the expected number of claims based on the prior. Using the actual number of claims as a basis for credibility has a problem in that observing zero claims may actually contain significant information, for example if one hundred claims were expected to be observed. The chart also identifies cedents where a full claim count triangle was available as opposed to only the latest diagonal.

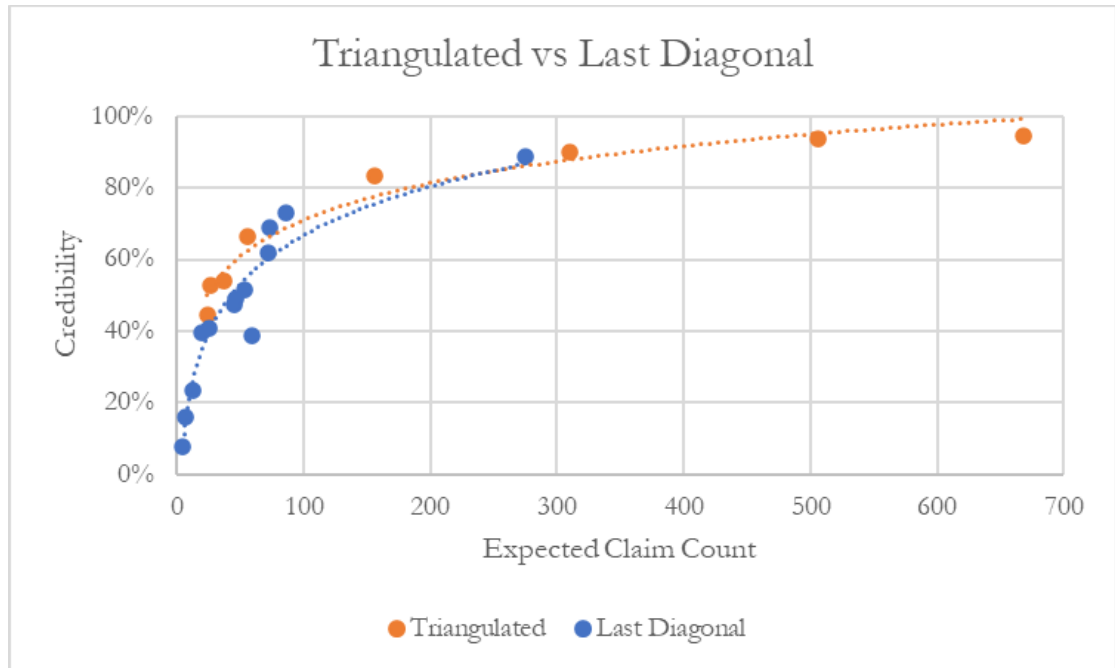


Figure 8. Implied Frequency Credibility Comparison

We can see that the majority of cedents had frequency credibility above 40%. A desirable and intuitive feature of the model is that it assigns higher credibility to cedents with triangulated data than those only providing losses from the last diagonal. No cedents with losses in triangulated form had frequency credibility falling below 40%. Cedents assigned the lowest credibility were all companies without triangles. This result is difficult to capture with traditional actuarial methods and is a distinct advantage of our Bayesian model.

Figure 9 plots the Weibull parameter joint probability densities for a certain company in our study. The graph plots a heatmap of the prior distribution in yellow and two posterior distributions in green and red. The range of parameters for the prior distribution is much wider and follows the gamma marginal distributions and Clayton copula described in Section 2.5. Both posterior distributions are concentrated within the company's prior distribution.

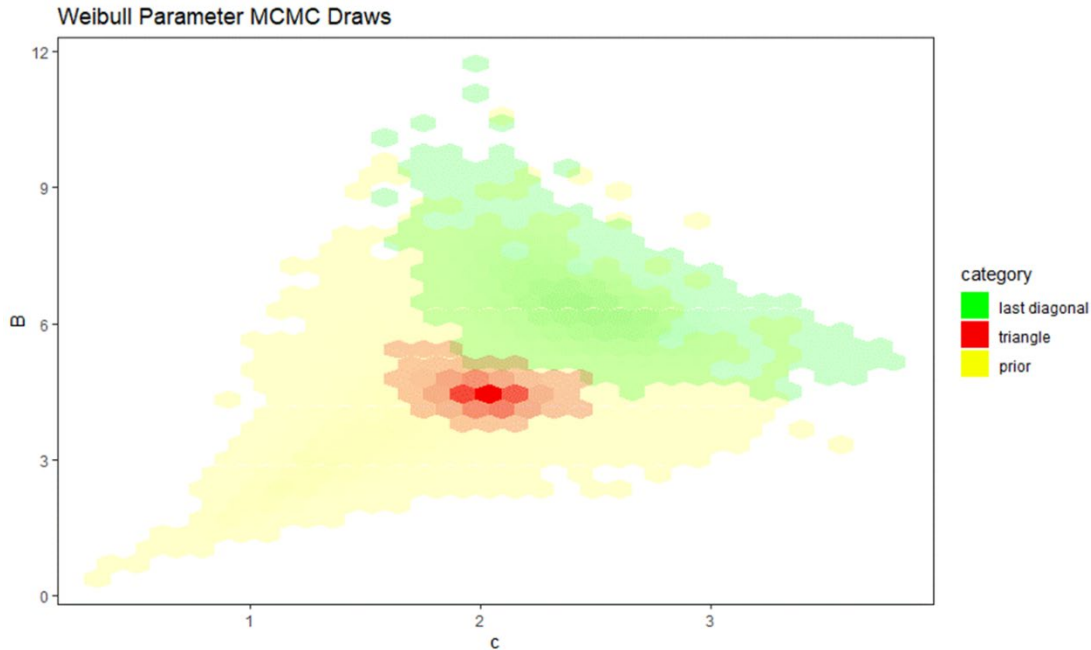


Figure 9. Comparison of Weibull Parameter Joint Distributions

The area in green represents the posterior distribution when only the last diagonal of excess claim count data is entered into the frequency model. What’s happening is that this company has, to the extent we can know, lower ultimate frequency than the mean of the prior distribution. When we observe the latest diagonal of observed claims, it is lower than expected. However, that could also be due to slow development. Without a triangle, the model must consider that possibility according to the prior distributions and Bayes’ Theorem. The result is the posterior distribution in green, which is very wide but populates mostly the “slow” region of the development parameter space (recall that higher values of c and B imply a slower development pattern).

Compare this to the area in red which is the frequency model’s posterior distribution using a full triangle of excess claim counts. Triangulated data produces more concentrated c and B values because of the increased certainty in the development pattern compared to only having the latest diagonal. What’s happening is that the model has much more information with which to reject the slow development hypothesis indicated by the low observed claim counts. Increasing the certainty of the development pattern while holding observed claim counts constant means that the model can be more certain that the ultimate frequency actually is lower than the prior mean, which is what occurred in this case.

Interestingly, the full triangle posterior is concentrated around the center of the prior but looks like an extreme outlier relative to the green distribution. The full triangle posterior can also be viewed as a Bayesian posterior with the green distribution as prior, with the marginal additional information being the claim counts in the interior of the triangle. This is what we mean by the “value of a triangle.” The triangle’s interior claim count data contains significant information about the pattern and ultimate frequency of a company’s large losses, a fact that is often overlooked or only accounted for subjectively in a reinsurance pricing analysis.

2.7 Case Study

A frequency data input file with two mock companies is provided in Appendix B. Only last diagonal data is given for Company 1 which has three excess claims over an 11-year period. Company 2 has 149 excess claims over the same period and data is available in triangulated form. A \$500,000 threshold is used in this example. Parameters for the prior distributions are illustrative and programmed into the R/Stan codes in Appendix B.

Posterior means,² standard deviations, distribution percentiles and MCMC convergence plots from the Bayesian model are presented in Figure 10. Company 2 has a smaller standard deviation around the posterior mean ultimate frequency due to a larger volume of claim count data and the availability of an excess claim count triangle. Using the formula in section 2.6, the implied credibility for Company 2 is 90% compared to 8% for Company 1. Note that implied credibilities are commensurate with the claim count versus credibility plot in Figure 8. However, they do not perfectly line up because prior distribution parameters for this case study are not the same.

Company 2 also has tighter ranges in the Weibull parameter posterior distributions. This is clearly seen in the MCMC convergence plots and evidenced by the relatively smaller standard deviation around posterior c and B means. There is increased certainty in the development pattern for Company 2 when compared to Company 1. A full triangle of excess claim counts provides the model with more confidence that the ultimate frequency for Company 2 is actually lower than the prior mean. The model can reject the hypothesis that lower-than-average observed claim counts are due to slow development.

² Ultimate frequency is calculated per \$10M of premium in sample R code

On the other hand, Company 1 has low observed claim counts and only last diagonal data. The model does not have enough information to reject the slow development hypothesis. Therefore, a wider range of c and B values is contained in the posterior distributions.

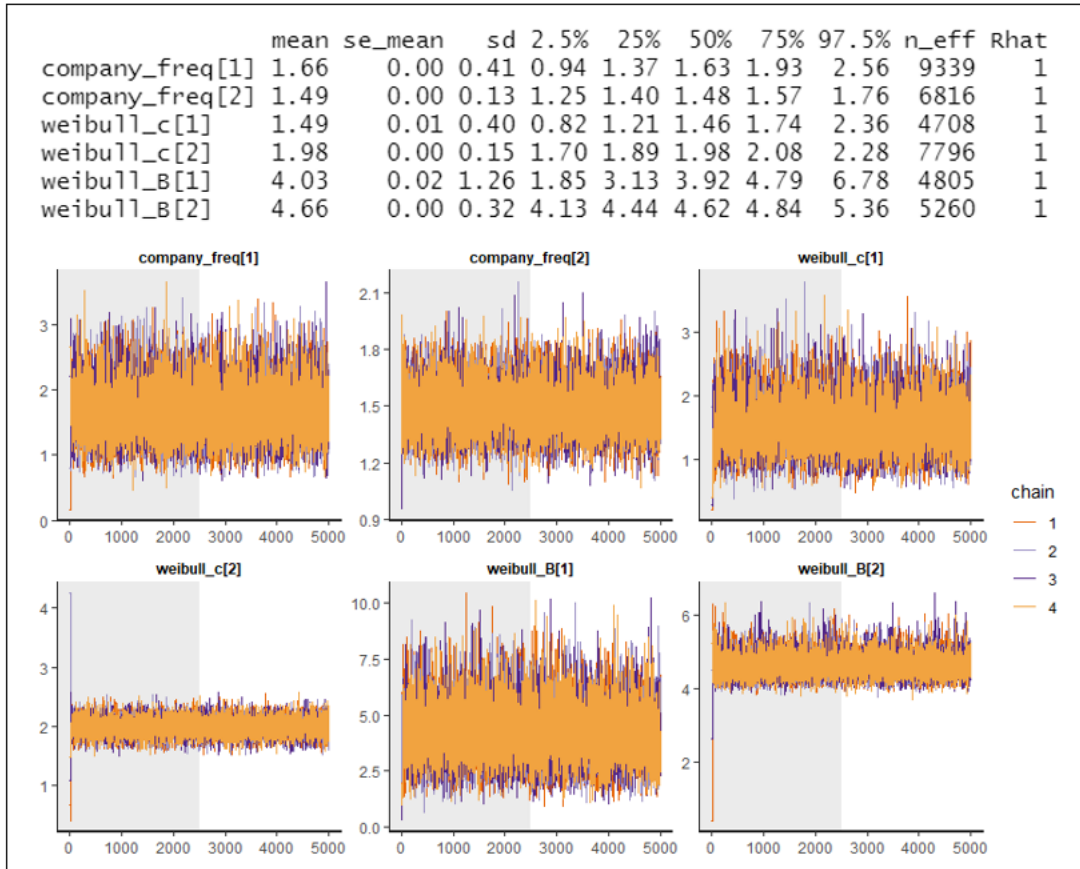


Figure 10. Bayesian frequency model posterior output and MCMC convergence plots

A visual comparison of joint probabilities densities is presented in Figure 11. The graph for Company 1 resembles the yellow prior distribution heatmap from Figure 9. With an implied credibility of only 8%, a wider range of distribution parameters makes sense. Low credibility assignments by the frequency model will yield posterior distributions that are closer to the prior distribution.

In contrast, the graph for Company 2 looks similar to the red area in Figure 9. A full triangle of excess claim counts produces more concentrated posterior joint probability densities. The model has more information with which to reject a larger number of parameter values.

This case study included two sample companies. However, the input files and code are set up to receive a more extensive number of cedents for simultaneous modeling. We encourage users to experiment with their own excess reinsurance portfolios.

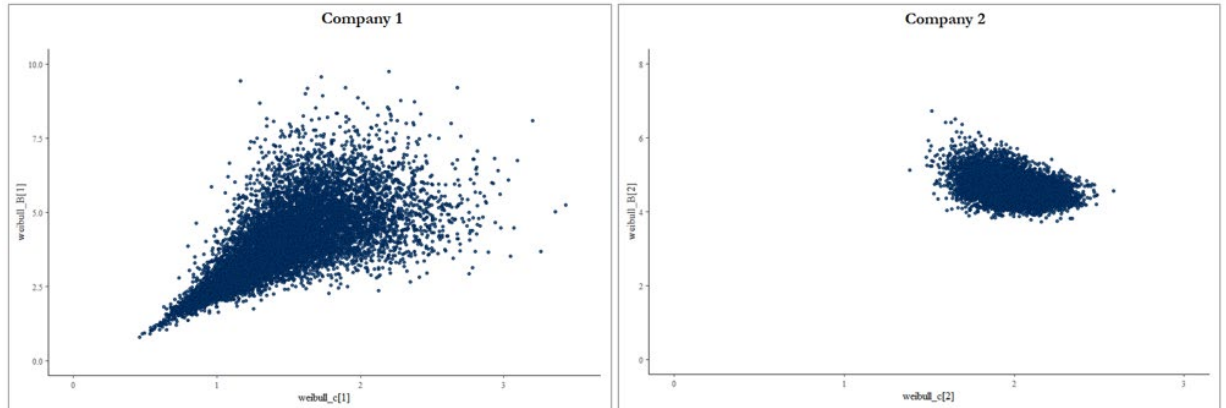


Figure 11. Weibull parameter joint distribution

3. SEVERITY MODEL

We also develop a Bayesian model for claim severity as depicted in Figure 12. We assume claims are Pareto distributed above the threshold. The shape parameter of the Pareto is allowed to differ by company. The Bayesian model weighs observed individual claim amounts against the prior distribution of each company’s shape parameter, and we use stan/R/MCMC to calculate the posterior distribution, in particular the posterior mean, of each company’s Pareto shape parameter.

One additional consideration is made to account for severity development, which is non-trivial in many long-tailed lines of business. We treat this in a simplified way by applying an age factor to the “ultimate” company Pareto α . We stop short of the more complicated model in (McNulty 2017).

We assume there is no trend in the excess severity, as would be consistent with a Pareto distribution. Note that ground up severity trend is accounted for in the trend factors for excess frequency at the chosen threshold.

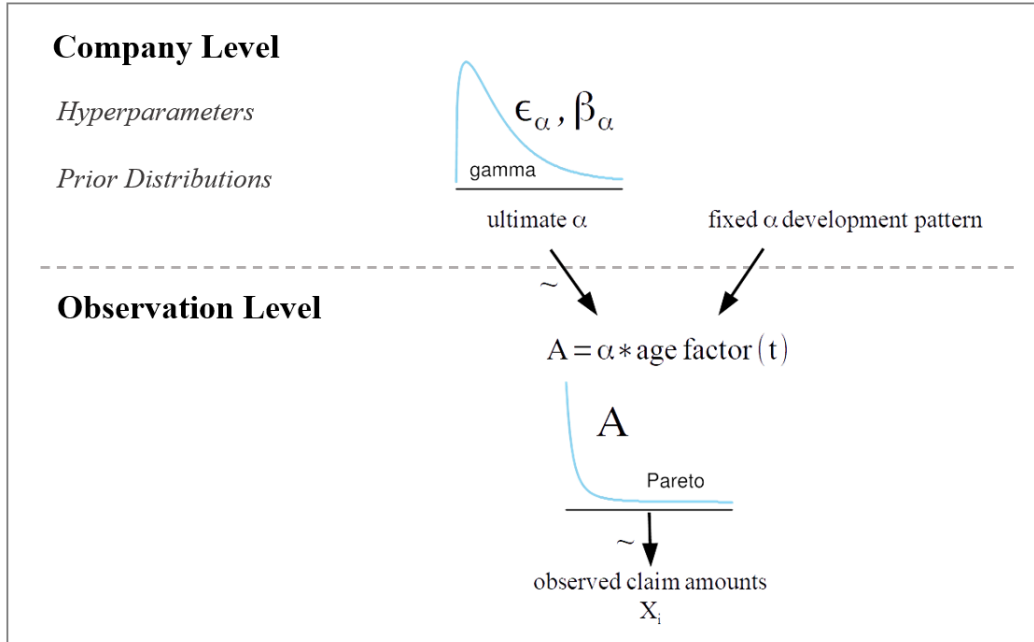


Figure 12. Diagram of Hierarchical Bayesian Severity Model

3.1 Gamma-Pareto Ultimate Severity

We use the following parameterization of the Pareto and gamma distributions with excess threshold T :

$$f(x; \alpha, T) = \left(\frac{\alpha T^\alpha}{x^{\alpha+1}} \right); x > T$$

$$f(x; \epsilon, \beta) \sim \frac{\beta^\epsilon}{\Gamma(\epsilon)} x^{\epsilon-1} e^{-\beta x}$$

Instead of using α, β as parameters of the gamma distribution we are utilizing ϵ, β to avoid confusion with the Pareto α .

3.2 Parameterizing the α Prior

A prior mean for each company's uncertain α Pareto shape parameter, denoted by $\hat{\alpha}$, is established for each company in the portfolio. See Table 2. As with the ultimate frequency prior it is derived using an exposure rating method for this paper, but other methods of determining an a priori are also acceptable. The Pareto fitting exercise to determine each cedent's $\hat{\alpha}$ prior mean is only performed on the portion of the curve between T and \$1M. Cedent data is right censored at policy limits. The most common commercial general liability limit in our data was \$1M.

Similar to what we saw in the previous section on ultimate frequency, the array of best fit $\hat{\alpha}$ prior means was well fit by a gamma distribution. The variance parameter from that fit, β , we then use as a universal value for β in the gamma prior distribution of each company's uncertain Pareto α severity shape parameter. The mean of a gamma distribution is ε/β . Therefore, we have for each company:

$$\alpha_{prior_co[i]} \sim \Gamma(\beta * \hat{\alpha}_{prior_co[i]}, \beta)$$

Cedent	Best fit Pareto α
Company 1	$\hat{\alpha}_{prior_co[1]}$
Company 2	$\hat{\alpha}_{prior_co[2]}$
Company 3	$\hat{\alpha}_{prior_co[3]}$
Company 4	$\hat{\alpha}_{prior_co[4]}$
Company 5	$\hat{\alpha}_{prior_co[5]}$
⋮	⋮
Company N	$\hat{\alpha}_{prior_co[N]}$

Table 2. Pareto α Priors by Cedent

3.3 Pareto α Development

If we observe a certain Pareto α in early ages, the ultimate α might not be the same because excess claim severity, for example in the \$500K excess \$500K layer, develops just as excess claim counts develop. A development pattern needs to be determined for the α parameter. A factor from this pattern is applied to $\alpha_{prior_co[i]}$ values, which are at ultimate, to bring the severity expectation to a level corresponding to the age of each observed claim.

We assume a fixed α development pattern in the severity model. The pattern is obtained by creating an aggregated triangle of claim count data, excess of threshold T . Data is compiled from cedent loss listings that provide individual loss development. A development triangle of Pareto α 's by year and age can be constructed from the excess claim count triangle. This can be done because given the distribution of individual claims in each cell of the triangle, a fitted Pareto shape parameter can be determined by maximum likelihood, moment matching, or fitting to percentiles such as the survival at \$1M.

A cumulative α development pattern is then determined, similar to a selected development pattern from a triangle of link ratios. Unlike for frequency, we allow for no uncertainty in this pattern, or variation between cedents. In reality, there is uncertainty in this pattern. There is likely variation between cedents and there may even be correlation between a cedent's frequency development and their severity development. These considerations would be good additions to a more complex model which we stopped short of in this paper.

Below is a graph of the selected pattern used for our application of the severity model:

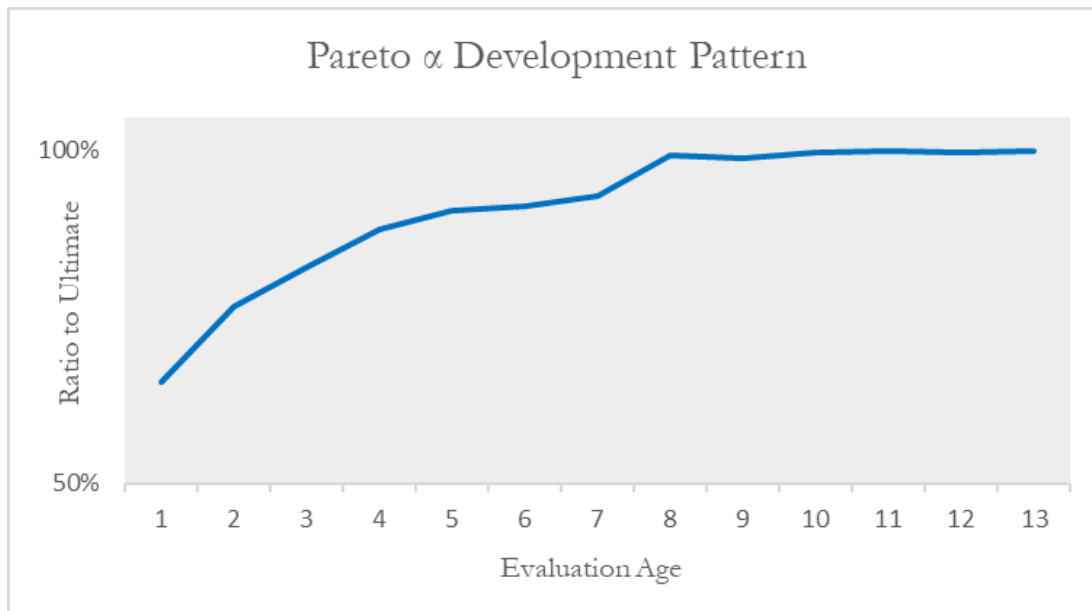


Figure 13. Fixed Pareto α development

3.4 Accounting for Policy Limit Capping

One complication that arises in some liability coverages in the US market is the presence of policy limits. Especially in a reinsurance application to large losses as in this paper, a significant portion of individual loss indemnity amounts excess of \$500,000 will be exactly equal to the policy limit of \$1 million.

This is censoring of the data and fitting distributions when presented with censored data is treated in *Loss Models* (Klugman, Panjer and Willmot 2012). In a traditional maximum likelihood fitting exercise, to incorporate data censored above at T, instead of using the pdf value at T for those points the survival function is used which is the cumulative probability of the observation being greater than or equal to the censored value.

Our model uses Bayesian MCMC instead of maximum likelihood, but the principle is the same and the severity model code (see attached files) uses the complementary cumulative

distribution function for the marginal probability associated with censored data points. The complementary cumulative distribution function is also known as the survival function and is one minus the CDF.

3.5 Estimating Posterior Distributions

The Bayesian model to estimate posterior distributions using the data is more straightforward for severity. Pseudo-code for the severity model would read as follows:

$$\text{Claims size} \sim \text{Pareto}(\Gamma, \alpha_{\text{company } i} * \text{age adjustment})$$

Only the company α is uncertain with its own distribution since the age adjustments are fixed values by age. In the actual model, extra steps are made to implement the policy limit censorship adjustment where the cumulative distribution function (CDF) is used for the marginal probability of censored values versus the probability distribution function (PDF) for uncensored values. See (Panjer et. al. 2012) for more detail on fitting distributions to truncated and censored data.

3.6 Results

The Stan program simulates the posterior distribution of each company's Pareto shape parameter. Although technically the compound Pareto mixture with variable α may be a more complex distribution (McNulty 2021), we simply took the mean of the posterior distribution of the α as a fixed value. This gave us a Pareto distribution for each company's claims excess of T , which is a Bayesian blending of the prior distribution and the data, although not technically the full posterior distribution.

Interestingly we found that there was less “credibility” in the severity data than for frequency. The chart below shows the implied credibility factors, Z_i , such that:

$$\alpha_{\text{posterior}}[i] = (1 - Z_i) * \alpha_{\text{prior}}[i] + Z_i * \alpha_{\text{empirical}}[i]$$

where $\alpha_{\text{empirical}}[i]$ is the maximum likelihood estimate of the Pareto shape parameter based only on the i^{th} cedent's data. In practice we found these values by running the severity model with uninformative priors.

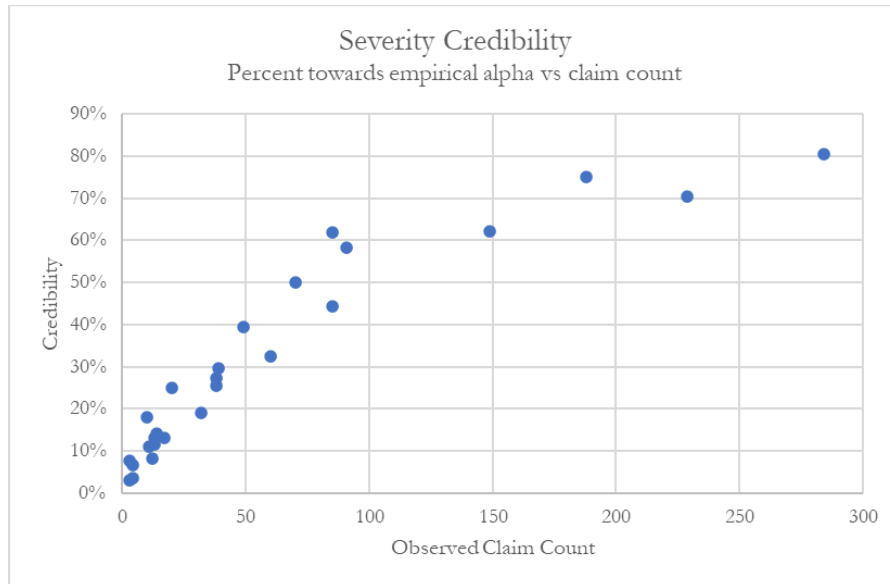


Figure 14. Implied severity credibility for various observed claim counts

We can see in Figure 14 that the majority of cedents had severity credibility below 30%, and even in the case of the cedents with the most claims the credibility was at most 80%.

3.7 Case Study

A severity data input file with two mock companies is provided in Appendix B. The sample dataset was created and perturbed independently from the frequency dataset. Thus, the total number of claims and number of claims per year might not match the frequency input file. However, the information is more than sufficient to serve as a concrete example for the severity model.

Company 1 has five excess claims over a 10-year period. Company 2 has 64 claims over a 14-year period. As with the frequency model case study, a \$500,000 threshold is used in this example. Parameters for the prior distributions are illustrative and programmed into the R/Stan codes in Appendices B.

Posterior means, standard deviations, distribution percentiles and MCMC convergence plots from the Bayesian model are in Figure 15. In the severity model R code shown in Appendix B, the company prior α values are shown as (0.95, 1.05). Company 1, which has barely any data, results in a posterior mean of 0.95 after running the model, i.e., exactly the same as the prior. This is consistent with the expected implied credibility based on Figure 14 of 0% for such a low claim count.

Company 2, on the other hand, has a significant amount of data that would result in an implied credibility of perhaps 30% to 40% weight towards the empirical Pareto α (assuming a

similar credibility profile as Figure 14 and a claim count of 64), and away from the prior α of 1.05 as specified in the code. The posterior mean α as shown in Figure 15 is 1.21, meaning that Company 2 displays lower than average severity in the excess layer, after adjusting for policy limits and severity development.

As with the frequency model case study, input files and R/Stan codes are set up to receive a more extensive number of cedents for simultaneous modeling. The model is set up to handle entire portfolios or large subsets of reinsurance portfolios.

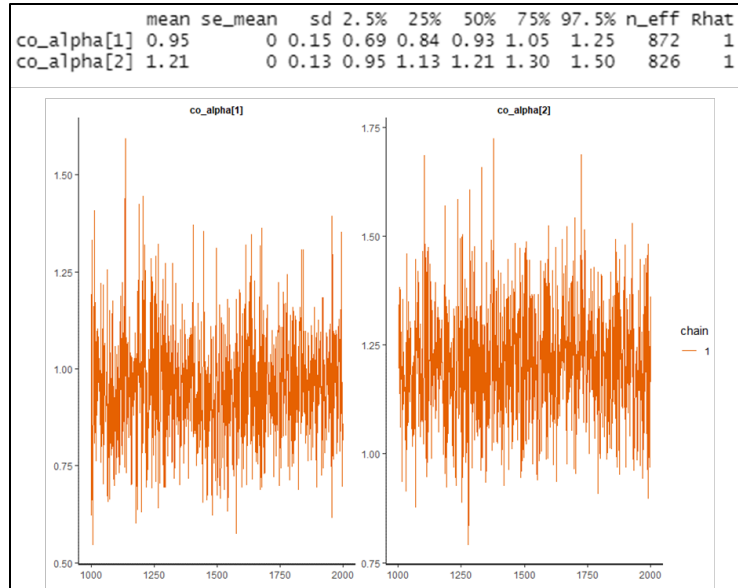


Figure 15. Bayesian severity model posterior output and MCMC convergence plots

3.8 Treatment of ALAE

The model developed to this point in Section 3 has concerned indemnity only. In the US market, claims in long tailed lines of business often incur significant loss adjustment expense, or ALAE, in addition to indemnity. A full model of large casualty losses for reinsurance in the US needs to have a provision for ALAE, which can be treated as included in-addition or pro-rata with loss for determining the reinsured portion of ALAE.

The standard way of modeling ALAE would be to multiply every indemnity loss by a factor, for example 1.2, to gross up to a loss amount with ALAE. This is what the authors did in implementing the method in this paper for use in reinsurance pricing. There is a more realistic, although also more complicated, model for ALAE which fits a separate distribution for ALAE amounts, and then models the bivariate distribution of ALAE and indemnity with a copula. That model would be beyond the scope of this paper but see, e.g., (Micocci and Masala 2009) for more information.

4. COMPARISON TO EXISTING METHODS

The defining features of the model developed in this paper, and the reason why we believe it is novel research, is that we address both problems of severity development and uncertainty in the claim count development in a Bayesian large loss model. We wanted to highlight some other papers that come the closest and compare the differences and similarities between the methods.

4.1 Cockroft/Buhlmann

Cockroft, in his 2004 paper (Cockroft 2004), develops what we would call a classical credibility treatment for excess of loss claims. He derives the formulas for applying Buhlmann best linear approximation to the Bayesian posterior mean in the case of a compound Gamma-Poisson, Gamma-Pareto model.

There are a few elements developed in our paper which we believe are advantages. First, the Cockroft treatment does not allow for claim development, so it is more suitable for short-tailed lines of business. For long-tailed lines, where claim development and uncertainty in the development pattern are keys to estimating the mean, we suggest incorporating the methods from this paper.

Trend is another very important parameter in any reinsurance pricing. Cockroft's paper assumes all data is trended before fitting the model. In practice this would make it difficult to include trend as an uncertain parameter in a Bayesian analysis. We also chose to use fixed trend factors, but since the trend factor is explicitly used in our frequency model with untrended data, we could easily give the trend a prior distribution. This would allow for uncertainty in the posterior and cedent specific differentiation.

On the severity side, Cockroft does not address policy limit capping, whereas we have explicitly incorporated it into our severity model and provided code for practitioners who may have policy limit censored data.

Finally, the Buhlmann method is fundamentally concerned with a best estimate of the mean and develops a best linear approximation. Cockroft goes further by deriving full posterior distributions for some quantities of interest but not all. Since R and Stan are simulation based, it would be easy to define the mean layer loss as a variable and observe the MCMC samples of the full distribution of the posterior. It's also not clear that in the case of reinsurance, with heavily right-skewed distributions that the best linear approximation to the posterior mean is a good approximation.

4.2 Barnett

The authors wanted to highlight (Barnett 2020) as it is one of the only papers we found that addressed the issue of development in a Bayesian large loss or reinsurance excess of loss model. In order to stay practical, the author develops the model in tables and spreadsheets which makes it more usable but probably limits the complexity of the model. For example, the model does not consider uncertainty in the development pattern, which we believe to be a key factor in reinsurance excess of loss estimation.

Note that Bayesian reserving models or models for estimating ground up loss ratios which do incorporate uncertainty in the development pattern have been well developed by, e.g., (Zhang, Dukic and Guszcza 2012) and (Clark 2003).

4.3 Mildenhall

Mildenhall's 2006 paper, "A Multivariate Bayesian Claim Count Development Model With Closed Form Posterior and Predictive Distributions" (Mildenhall 2006) introduced a completely new model, breaking the Buhlmann mold. The model uses Gamma-Poisson ultimate claim counts and a multinomial Dirichlet claim count development distribution.

The greatest achievement of this paper is that it addresses the fundamental question of weighing uncertainty in the ultimate frequency against uncertainty in the development pattern. The author includes a very enlightening chart on page 472 that shows where established reserving methods lie on the spectrum of the relative size of the uncertainty in those two parameters.

In contrast to Mildenhall, we prefer the simpler Weibull development pattern distribution to the multinomial Dirichlet for several reasons. The Weibull is a 2-parameter distribution, and as discussed above there are intuitive visual interpretations of the parameters. This can be beneficial when priors are specified using judgment. In contrast, the Dirichlet has as many parameters as there are development ages (minus one) and so it would be difficult to intuitively judge or visualize the prior. This also leads the total model to have $n + 2$ parameters when there are n time periods as stated in Section 7 of the paper, whereas the model presented in this paper has a total of seven³ parameters independent of how many time periods there are.

³ The seven parameters being 2 Gamma parameters for the prior ultimate frequency, 2 Gamma parameters for the prior of each of the 2 Weibull parameters, one Clayton copula parameter. One could argue that the copula parameter, rather than adding a degree of freedom, reduces the degrees of freedom in the model by constraining the independence of the two Weibull parameters. Viewed this way, the degrees of freedom would be between 5 and 6.

Another advantage of the Weibull is that it is continuous which means that we are easily able to incorporate information from a partial diagonal of a triangle or data at different ages, such as a 12-month triangle with a six-month latest diagonal which is quite common in reinsurance practice. The multinomial Dirichlet is discrete with each point representing one time period which would typically be a year. It's not clear how a practitioner would deal with data that is at different ages.

Lastly, consider the following statement from the Mildenhall paper on page 468:

“Proposition 4 shows the predictive distribution does not depend on the individual observed values b_1, \dots, b_t but only on their sum $b(t) = b_1 + \dots + b_t$. Thus, the GPDM model has a kind of Markov property that the future development depends only on the total number of claims observed to date, and not on how those claims were reported over time.”

The authors of this paper would argue that the Markov property is unintuitive for insurance losses. The times at which various claims have been reported in the past contains valuable information about future development. This was demonstrated in Section 2 which showed that the frequency model can provide a much more concentrated posterior for the development pattern in the case of a cedent who provides a full triangle. If only the latest diagonal or total number of claims reported to date mattered in the model, i.e., the model demonstrated the Markov property, then we would get the same results when giving the full triangle or simply the latest diagonal. We view the absence of the Markov property to be a fundamental advantage of our model.

5. REFERENCES

- [1] Barnett, J. 2020. “A Bayesian Approach to Excess of Loss Pure Premium Rating.” *Variance* 13 (1): 54-79.
- [2] Clark, D. 2003. “LDF Curve-Fitting and Stochastic Reserving: A Maximum Likelihood Approach.” *Casualty Actuarial Society Forum*, Spring, 41-92.
- [3] Clark, D. 2014. “Basics of Reinsurance Pricing.” Casualty Actuarial Society Study Note, 1996, revised 2014.
- [4] Cockroft, M. 2004. “Bayesian Credibility for Excess of Loss Reinsurance Rating.” GIRO Conference.
- [5] McNulty, G. 2021. “The Pareto-Gamma Mixture.” *Casualty Actuarial Society E-Forum*, Spring.
- [6] McNulty, G. 2017. “Severity Curve Fitting for Long-Tailed Lines: An Application of Stochastic Processes and Bayesian Models.” *Variance* 11 (1): 118-132.
- [7] Micocci, M., and G. Masala. 2009. “Loss-ALAE modeling through a copula dependence structure.” *Investment Management and Financial Innovations* 6 (4): 67-80.
- [8] Mildenhall, S. 2006. “A Multivariate Bayesian Claim Count Development Model With Closed Form Posterior and Predictive Distributions.” *Casualty Actuarial Society Forum*, Winter, 451-493.
- [9] Klugman, S., H. Panjer and G. Willmot. 2008. *Loss Models: From Data to Decisions*, 381-388. 3rd ed. New Jersey: Wiley.
- [10] Zhang, Z., V. Dukic, and J. Guszcz. 2012. “A Bayesian Nonlinear Model for Forecasting Insurance Loss Payments,” *Journal of the Royal Statistical Society: Series A* 175 (2): 637-656.

6. APPENDIX A

Frequency Model One-Page Step-by-Step Guide:

1. Parameterize the prior distribution of Poisson λ
 - 1.1. Collect samples of the exposure-based frequency of claims excess of a universal threshold, e.g., \$500k, for several (ideally 10+) cedents
 - 1.2. Fit a gamma distribution to the set of frequencies stated as a ratio to a common exposure measure, such as \$10M premium as used in the paper. The variance parameter of this fit determines our industry/market prior distribution for the Poisson λ of an individual cedent. The mean of the prior λ for an individual cedent can be that recorded from step 1. The overall mean is an alternative.
2. Parameterize the Weibull claim count pattern prior distribution
 - 2.1. Collect a subsample of experience data that has a full triangle of development available
 - 2.2. Use traditional techniques to choose an LDF pattern for each triangle
 - 2.3. Fit a Weibull distribution to each development pattern using percentile matching or a visual match method. Record the best fit Weibull parameters, c and B in our parameterization, for each cedent.
 - 2.4. Fit a gamma distribution to the list of c 's and another gamma to the B 's
 - 2.5. Fit a copula to the percentiles of each (c, B) pair after transforming the c and B values to percentiles via the gamma fit, e.g., $F^{-1}(c)$
3. Run the Bayesian model
 - 3.1. Input the parameters found in steps 1 and 2 into the code for the frequency model
 - 3.2. Input the data on excess claim counts in the tabular format explained in the paper
 - 3.3. Run the code
 - 3.4. Enjoy your posterior excess frequency per exposure distributions which will differ by cedent. Assuming claims will be Poisson distributed with expected claims equal to the mean should be fine, a Poisson mixed with a gamma frequency parameter is technically negative binomially distributed. Although the pattern posteriors are not needed directly after this stage for estimating expected loss to the reinsurance layer, it can be of interest to see the relative certainty in each cedent's pattern and confirm agreement with which cedents are deemed fast or slow.

Severity Model One-Page Step-by-Step Guide:

1. Parameterize the prior distribution of Pareto α
 - 1.1. Collect samples of the exposure-based severity distribution of claims excess of a universal threshold, e.g., \$500k, for several (ideally 10+) cedents
 - 1.2. Fit a Pareto to each of the collected severity distributions and record the best fit Pareto α parameter. A good ad hoc method would be to match the survival at \$1M using the Pareto CDF.
 - 1.3. Fit a gamma distribution to the set of α 's. The variance parameter of this fit determines our industry/market prior distribution for the Pareto α of an individual cedent. The mean of the prior gamma for an individual cedent can be the actual fitted α from step 2 if desired. The overall mean is an alternative.
2. Create a deterministic Pareto α development pattern
 - 2.1. Using available triangulated excess claim amount data, calculate a best fit α for each cell of the triangle. This can be somewhat complicated by the presence of policy limited data. A simplification is just determining the percent of claims that exceed \$1M given they are already over \$500k. The implied α can be computed by inverting the CDF of the Pareto distribution with threshold \$500k.
 - 2.2. Using traditional actuarial techniques, determine a set of development factors between successive ages of the triangle, for example a straight average of link ratios in the triangle, or perhaps a recent 5-year average.
 - 2.3. Convert the selected development factors into a pattern representing percent of ultimate. Note that we are developing Pareto α 's, not total claims or incurred amounts. The age 1 numbers may develop downward in some applications.
3. Run the Bayesian model
 - 3.1. Input the parameters found in steps 1 and 2 into the code for the severity model
 - 3.2. Input the data on excess claim amounts in the format provided. Note only latest diagonal is needed for severity
 - 3.3. Run the code
 - 3.4. Enjoy your posterior Pareto α distributions which will differ by cedent. Note that the full posterior distribution is not a Pareto with α equal to the mean of the α 's posterior distribution, although the authors use this simplification in practice. See the Pareto-Gamma Mixture paper.

7. APPENDIX B

The following sample input files are used for the frequency and severity models, respectively, and called from the R code:

freq data.csv

company	year	premium	obs_start	obs_end	incr_claims
1	2010	3000000	0	11	1
1	2011	3000000	0	10	0
1	2012	2000000	0	9	0
1	2013	2000000	0	8	0
1	2014	2000000	0	7	1
1	2015	2000000	0	6	1
1	2016	2000000	0	5	0
1	2017	2000000	0	4	0
1	2018	2000000	0	3	0
1	2019	4000000	0	2	0
1	2020	4000000	0	1	0
2	2010	350000000	0	1	1
2	2010	350000000	1	2	4
2	2010	350000000	2	3	6
2	2010	350000000	3	4	2
2	2010	350000000	4	5	3
2	2010	350000000	5	6	3
2	2010	350000000	6	7	0
2	2010	350000000	7	8	2
2	2010	350000000	8	9	0
2	2010	350000000	9	10	1
2	2010	350000000	10	11	0
2	2011	330000000	0	1	2
2	2011	330000000	1	2	4
2	2011	330000000	2	3	3
2	2011	330000000	3	4	3
2	2011	330000000	4	5	7
2	2011	330000000	5	6	0
2	2011	330000000	6	7	4
2	2011	330000000	7	8	1
2	2011	330000000	8	9	0
2	2011	330000000	9	10	0
2	2012	320000000	0	1	1
2	2012	320000000	1	2	4
2	2012	320000000	2	3	3
2	2012	320000000	3	4	3
2	2012	320000000	4	5	3

2	2012	320000000	5	6	3
2	2012	320000000	6	7	1
2	2012	320000000	7	8	0
2	2012	320000000	8	9	0
2	2013	320000000	0	1	1
2	2013	320000000	1	2	5
2	2013	320000000	2	3	2
2	2013	320000000	3	4	1
2	2013	320000000	4	5	6
2	2013	320000000	5	6	3
2	2013	320000000	6	7	1
2	2013	320000000	7	8	0
2	2014	300000000	0	1	0
2	2014	300000000	1	2	2
2	2014	300000000	2	3	4
2	2014	300000000	3	4	6
2	2014	300000000	4	5	4
2	2014	300000000	5	6	4
2	2014	300000000	6	7	0
2	2015	300000000	0	1	1
2	2015	300000000	1	2	3
2	2015	300000000	2	3	6
2	2015	300000000	3	4	4
2	2015	300000000	4	5	3
2	2015	300000000	5	6	0
2	2016	310000000	0	1	0
2	2016	310000000	1	2	5
2	2016	310000000	2	3	4
2	2016	310000000	3	4	6
2	2016	310000000	4	5	0
2	2017	310000000	0	1	0
2	2017	310000000	1	2	3
2	2017	310000000	2	3	5
2	2017	310000000	3	4	0
2	2018	300000000	0	1	2
2	2018	300000000	1	2	4
2	2018	300000000	2	3	0
2	2019	290000000	0	1	0
2	2019	290000000	1	2	0
2	2020	300000000	0	1	0

sev data.csv

year	co	age	incurred
2009	1	12	750000
2010	1	11	1000000
2014	1	7	960000
2015	1	6	900000
2018	1	3	1000000
2005	2	16	585606
2005	2	16	620319
2005	2	16	935383
2005	2	16	601001
2005	2	16	1000000
2006	2	15	1000000
2006	2	15	1000000
2006	2	15	650595
2006	2	15	1000000
2006	2	15	1000000
2006	2	15	660747
2006	2	15	1000000
2006	2	15	1000000
2006	2	15	1000000
2006	2	15	1000000
2006	2	15	1000000
2007	2	14	1000000
2007	2	14	677013
2007	2	14	682349
2007	2	14	551249
2008	2	13	1000000
2008	2	13	1000000
2008	2	13	833795
2008	2	13	1000000
2008	2	13	835682
2008	2	13	1000000
2008	2	13	563742
2009	2	12	1000000
2009	2	12	789589
2009	2	12	676473
2009	2	12	709377
2009	2	12	623965
2009	2	12	1000000
2009	2	12	1000000
2009	2	12	879431
2009	2	12	666018
2010	2	11	1000000
2010	2	11	785643
2010	2	11	818112

2010	2	11	937688
2011	2	10	573665
2011	2	10	626129
2011	2	10	730101
2012	2	9	728281
2012	2	9	1000000
2012	2	9	607633
2012	2	9	780319
2013	2	8	755611
2013	2	8	572497
2013	2	8	577085
2014	2	7	627388
2014	2	7	706466
2014	2	7	722980
2014	2	7	584997
2014	2	7	600185
2014	2	7	530892
2015	2	6	599592
2015	2	6	729452
2015	2	6	1000000
2015	2	6	679767
2015	2	6	702421
2015	2	6	1000000
2015	2	6	1000000
2016	2	5	1000000
2016	2	5	1000000
2018	2	3	600695

Below are the code files for the frequency and severity stan models, respectively, and called from the R code:

stan code freq model.stan

```

functions {
  real clayton_log_pdf(real u, real v, real theta) {
    return log((1+theta)*pow(u*v,-1-theta)*pow(pow(u,-theta)+pow(v,-theta)-1,-
((2*theta+1)/theta)));
  }
}
data {
  int<lower=0> N_obs;
  int<lower=0> N_obs_start;
  int<lower=0> N_obs_end;
  int<lower=0> N_companies;
  real beta;
  real premium[N_obs];

```

```

real co_lambda_prior[N_companies];
int<lower=0> year[N_obs];
real<lower=0> obs_start[N_obs];
real<lower=0> obs_end[N_obs];
int<lower=0> company[N_obs];
int<lower=0> incr_claims[N_obs];
}
parameters {
real<lower=0> company_freq[N_companies];
real<lower=0> weibull_c[N_companies];
real<lower=0> weibull_B[N_companies];
}
transformed parameters {
real pois_lambda[N_obs];
real normalized_weibull_c[N_companies];
real normalized_weibull_B[N_companies];
  for(i in 1 : N_obs) {
    pois_lambda[i] = company_freq[company[i]] * premium[i] * (exp(-
pow((obs_start[i]/weibull_B[company[i]]),weibull_c[company[i]]))
- exp(-
pow((obs_end[i]/weibull_B[company[i]]),weibull_c[company[i]])))*pow(1/1.1,2022-year[i]);
  }
  for ( i in 1 : N_companies) {
    normalized_weibull_c[i] = gamma_cdf(weibull_c[i], 13,9);
  }
  for ( i in 1: N_companies) {
    normalized_weibull_B[i] = gamma_cdf(weibull_B[i], 8,2);
  }
}
model {
for (i in 1:N_companies){
  weibull_c[i] ~ gamma(13,9);
}
for (i in 1:N_companies){
  weibull_B[i] ~ gamma(8,2);
}
for (i in 1:N_companies){
  target += clayton_log_pdf( normalized_weibull_c[i], normalized_weibull_B[i], 2.75 );
}
for (i in 1:N_companies){
  company_freq[i] ~ gamma(beta*co_lambda_prior[i],beta);
}
for (i in 1:N_obs){
  incr_claims[i] ~ poisson(pois_lambda[i]);
}
}
generated quantities {
}

```

stan code sev model.stan

```
data {
  int<lower=0> N_claims;
  int<lower=0> N_years;
  int<lower=0> N_ages;
  int<lower=0> N_companies;
  real threshold;
  real beta;
  real <upper = 1000000> incurred[N_claims];
  real age_factor[N_ages];
  real co_alpha_prior[N_companies];
  int<lower=0> year[N_claims];
  int<lower=0> age[N_claims];
  int<lower=0> co[N_claims];
  int<lower = 0, upper = 1> censored[N_claims];
}
parameters {
  real<lower=0> co_alpha[N_companies];
}
transformed parameters {
  real claim_alpha[N_claims];
  for( i in 1 : N_claims ) {
    claim_alpha[i] = co_alpha[co[i]] * age_factor[age[i]];
  }
}
model {
  for (i in 1:N_companies){
    co_alpha[i] ~ gamma(beta * co_alpha_prior[i], beta);
  }
  for (i in 1:N_claims){
    if (censored[i]) {
      target += pareto_lccdf( incurred[i] | threshold, claim_alpha[i]);
    } else {
      target += pareto_lpdf( incurred[i] | threshold, claim_alpha[i]);
    }
  }
}
generated quantities {
}
```

Next are the companion R code files for running the models:

r code freq model.R

```
library(dplyr)
library(distr)
library(fitdistrplus)
library(actuar)
library(rstan)
library(readr)
library(rstudioapi)
library(bayesplot)
library(ggplot2)
#setwd
setwd("C:/folder")
##import data
data <- read_csv("freq data.csv",
                 col_types = cols(company = col_integer(),
                                   year = col_integer(),
                                   premium = col_double(),
                                   obs_start = col_double(),
                                   obs_end = col_double(),
                                   incr_claims = col_integer()))

##make the data list
attach(data)
##co priors are for the prior ultimate freq per 10M premium for each company
co_lambda_prior <- c(1.5,2.5)
##universal beta param for ultimate claim count gamma distributions
beta<- 9
##transform data
##our frequencies are stated per $10M of premium so rescale the premium data to match
premium<-premium/10000000
data_list <- list(
  year = year,
  obs_start = obs_start,
  obs_end = obs_end,
  incr_claims = incr_claims,
  company = company,
  premium = premium,
  co_lambda_prior = co_lambda_prior,
  beta = beta,
  N_obs = length(incr_claims),
  N_obs_start = max(obs_start),
  N_obs_end = max(obs_end),
  N_companies = max(company))
##run stan
```

```

fit<- stan(file = 'stan code freq model.stan', data=data_list, iter=10000, chains = 4, cores = 4,
seed =123)
### plot output
print(fit, pars=c("company_freq", "weibull_c", "weibull_B"))
traceplot(fit, pars=c("company_freq", "weibull_c", "weibull_B"), inc_warmup = TRUE)

```

r code sev model.R

```

library(dplyr)
library(distr)
library(fitdistrplus)
library(actuar)
library(rstan)
library(readr)
#setwd
setwd("C:/folder/")
##import data
data <- read_csv("sev data.csv",
                 col_types = cols(age = col_integer(),
                                 incurred = col_double(), co = col_integer(), year = col_integer()))
##make the data list
attach(data)
##age factor is for relativity of early alphas to ultimate (multiplicative)
## must satisfy len(age_factor) = N_ages
age_factor <- c(.5, .75, .9, .95,1,1,1,1,1,1, 1, 1, 1,1, 1, 1)
##co priors are for the prior ultimate alpha for each company
co_alpha_prior <- c(0.95,1.05)
##universal beta param for all gamma distributions
beta<- 40
##implement 1M upper censoring on incurred
incurred[incurred>1000000]<-1000000
censored <- incurred==1000000
data_list <- list(
  incurred = incurred,
  year = year,
  age = age,
  co = co,
  N_claims = length(incurred),
  N_years = 16,
  N_ages = 16,
  N_companies = 2,
  threshold = 500000,
  age_factor = age_factor,
  co_alpha_prior = co_alpha_prior,
  beta = beta,
  censored = censored
)

```

```
##run stan
fit<- stan(file = 'stan code sev model.stan', data=data_list, iter=2000, chains = 1, cores = 4)
## plot output
print(fit, pars=c("co_alpha"))
traceplot(fit, pars=c("co_alpha"), inc_warmup = FALSE)
```