Focussed selection of the claim severity distribution

Yinzhi Wang^{a,*}, Ingrid Hobæk Haff^a

^a*Department of Mathematics, University of Oslo, Postboks 1053 Blindern, 0316 Oslo, Norway*

Abstract

Risk assessment is a core theme within non-life insurance and estimation of quantiles far out in the upper tail is therefore one of the main applications of the total loss distribution of a non-life insurance portfolio. The choice of claim severity distribution should therefore reflect this. Therefore, we have explored how the focussed information criterion, FIC, aimed at finding the best model for estimating a given parameter of interest, the focus parameter, works as a tool for selecting the claim size distribution. As a quantile cannot be used directly as a focus parameter, we have tried different proxy focus parameters. To see how the FIC performs in this setting, compared to the other commonly used model selection methods AIC and BIC, we have performed a simulation study. In particular, we wanted to investigate the effect of the heaviness of the tail of the claim size distribution and the amount of available data. The performance of the different model selection methods was then evaluated based on the quality of the resulting estimates of the quantiles. Our study shows the best of the focussed criteria is the FIC_ϵ , based on one single quantile from the claim severity distribution. Further, the performance of the FIC_ϵ is mostly either comparable to or considerably better than that of the BIC, which is the best performing of the state of the art approaches. In particular, the FIC_ϵ works well when the data are heavy-tailed, when the sample size is rather low and when the parameter of interest is the quantile far out in the tail of the total

*Corresponding author

Email addresses: yinzhiw@math.uio.no (Yinzhi Wang), ingrihaf@math.uio.no (Ingrid Hobæk Haff)

loss distribution.

Keywords: Loss distributions, Model selection, Risk measures

1. Introduction

The application of sound quantitative risk models is of major importance for the insurance industry, for instance for calculating premia, measuring tail risk, finding optimal reinsurance schemes. In non-life insurance, one must specify a stochastic model for the total amount the insurer paid due to claims from a portfolio of policies over a defined period of time. Typically, such models are split into separate models for the claim numbers $\mathcal{N}$ and the individual losses Z_i , called the claim frequency and the claim severity distribution, respectively. The collective risk model for total claim loss $\mathcal{X}$ is then given by

$$\mathcal{X} = \sum_{i=1}^{\mathcal{N}} Z_i.$$

The claim severities are commonly assumed identically distributed and independent of each other and of the claim number (Kaas et al., 2001; Klugman et al., 2012). Although the assumption of claim severities being identically distributed is unrealistic in practice, the heterogeneity between the claim severities for different policies will be averaged out when we consider the aggregate loss. Therefore, this assumption is adequate in this setting.

Choosing a suitable claim severity distribution is essential because it plays a fundamental role in measuring tail risk, namely, to a great extent it controls how heavy-tailed the distribution is. Some of the most important measures of tail risk are quantiles far out in the upper tail of the total loss distribution, as they play a role in risk management, among others in the determination of the reserve or solvency capital. These quantiles q_ϵ , are given by

$$\Pr(\mathcal{X} \geq q_\epsilon) = \epsilon,$$

where ϵ typically is 0.1 or 0.05, and tend to be much more influenced by the claim severity than by the claim frequency distribution (Klugman et al., 2012). Claim severities in non-life insurance are positive random variables, that tend to follow a skew distribution with a heavier right tail. Popular choices include two-parameter distributions like the Gamma, the log-normal, the log-Gamma,

the Weibull and the Pareto (Kleiber and Kotz, 2003; Klugman et al., 2012), three parameter distributions like the Burr (Kleiber and Kotz, 2003) and the extended Pareto (Bølviken, 2014).

The main question is then how to select a suitable distribution for a given data set. Possible strategies include graphical inspection, for instance of Q-Q plots, goodness-of-fit tests, like Smirnov-Kolmogorov and Anderson-Darling, or model selection criteria, like the Akaike and Bayesian information criteria (Gray and Pitts, 2013; Bakar et al., 2015; Reynkens et al., 2017; Brazauskas and Kleefeld, 2016). A disadvantage of such strategies is that they primarily test the fit in the middle of the distribution and these are measures of overall fit of the distributions to the data. Our main interest in this setting is however to estimate a quantile q_ϵ far out in the tail of the total loss distribution, and it is by no means certain that the best-fitting distribution overall is the one providing the most precise estimate of q_ϵ . Therefore, one could instead use the Focused Information Criterion (FIC) proposed by Claeskens and Hjort (2003), or the modified version of Jullum and Hjort (2017). The purpose of the FIC is to select the model that gives the most precise estimate of a certain focus parameter, in this setting the quantile q_ϵ , in terms of the mean squared error.

Limitation of data is common within certain business lines in non-life insurance, particularly for young companies. The question is whether the stochastic models and model selection methods that are optimal for large sample sizes are also optimal when data is limited. When data are abundant, a flexible model with many parameters may provide a better estimate of the quantile q_ϵ , while simpler models, with fewer parameters may be chosen when there are few data. The aim is to explore how the FIC performs for selecting the claim severity distribution compared to other commonly used selection criteria as the sample size varies, when q_ϵ is the focus and the claim frequency distribution is fixed. Since the focus in this paper is on the claim severity model, we therefore only consider the classic Poisson distribution with fixed intensity μ as claim frequency for all policies, although it is a simplification. In order to investigate the performance of the FIC relative to the other selection methods, we will perform an extensive

simulation study, where the methods are compared in terms of the quality of the quantile estimates from the model they selected.

The rest of the paper is organized as follows. The claim severity distributions that are considered are presented in Section 2. Then, the traditional model selection methods as well as the FIC are presented in Section 3. Further, the simulation study and its result are given in Section 4. In Section 5, the FIC is applied to a set of motor insurance claims. Finally, Section 6 provides a discussion and some concluding remarks.

2. Claim severity distributions

The claim severity distributions considered in the simulation study are presented below. Among those are the classic claim severity distributions, including the exponential, the Gamma, the log-normal, the Weibull, the log-Gamma and the Pareto. The log-Gamma distributions is given by $Z = \exp(Y) - 1$, where $Y \sim \text{Gamma}(\alpha, \beta)$, such that the distributions starts at 0. The Pareto family is a wide one, with several sub-families. Here we consider the sub-family that is most widely used as a model for claim sizes. It is the Pareto type II, or Lomax, distribution, sometimes called the “American Pareto”. This Pareto distribution has two parameters, $\alpha, \lambda > 0$, where α is a shape parameter and λ is a scale parameter. The pdf is given by

$$f(z) = \frac{\alpha/\lambda}{(1 + z/\lambda)^{\alpha+1}}, \quad z > 0.$$

Further, if $Y_1 \sim \text{Exp}(1)$ and $Y_2 \sim \text{Gamma}(\alpha, \lambda)$, then $Z = \frac{Y_1}{Y_2} \sim \text{Pareto}(\alpha, \lambda)$.

The extended Pareto distribution $\text{ext-Pareto}(\alpha, \beta, \theta)$, also called the Beta prime distribution, has three parameters, $\alpha, \beta, \theta > 0$, where α and θ are shape parameters and β is a scale parameter. The pdf is given by

$$f(z) = \frac{\Gamma(\alpha + \theta)}{\Gamma(\alpha)\Gamma(\theta)} \frac{1}{\beta} \frac{(z/\beta)^{\theta-1}}{(1 + z/\beta)^{\alpha+\theta}}, \quad z > 0.$$

This pdf is decreasing over the real line when $\theta \leq 1$ and has a single maximum for $\theta > 1$, which is illustrated in Figure 1.

Figure 1: The pdf of extended Pareto distribution for $\theta = 0.8$ (solid line) and $\theta = 1.5$ (dotted line).

When $\theta = 1$, the extended Pareto distribution takes the form of an ordinary Pareto distribution. When α becomes infinite and $\beta = \frac{\xi\alpha}{\theta}$ with $\xi = \mathbb{E}(Z) = \frac{\theta\beta}{\alpha-1}$ and θ being fixed, it becomes a $\text{Gamma}(\theta, \theta/\xi)$ distribution.

The Burr distribution used in the study is the Burr type XII distribution with three parameters $\alpha, \lambda, \tau > 0$, also called Singh-Maddala distribution. The pdf is given by

$$f(z) = \frac{\alpha\tau(z/\lambda)^\tau}{z\left(1 + (z/\lambda)^\tau\right)^{\alpha+1}}, \quad z > 0,$$

and is shown in Figure 2 for different sets of parameters. When $\tau = 1$, the distribution reduces to the Pareto distribution.

A more flexible parametric distribution could be the four-parameter distribution which is constructed based on a parameterized power transformation proposed by Bølviken and Haff (2018). It takes the form

$$Z = \beta X^\eta,$$

with

$$X = \frac{G_\theta}{G_\alpha}, \quad \theta, \alpha > 0,$$

where $G_\theta \sim \text{Gamma}(\theta, \theta)$ and $G_\alpha \sim \text{Gamma}(\alpha, \alpha)$ are independent, $\beta > 0$ is

Figure 2: Comparison of the Burr densities as the parameters vary.

a scale parameter and the additional parameter $\eta > 0$ adds flexibility to the model. The pdf of Z is given by

$$f_Z(z) = \frac{1}{\beta\eta} \left(\frac{\theta}{\alpha}\right)^\theta \frac{\Gamma(\theta + \alpha)}{\Gamma(\theta)\Gamma(\alpha)} \frac{(z/\beta)^{\theta/\eta-1}}{\left(1 + \frac{\theta}{\alpha}(z/\beta)^{1/\eta}\right)^{\theta+\alpha}}, \quad z > 0,$$

and shown in Figure 3 as the parameters vary.

3. Model selection methods

Statistical model selection plays an important role in almost any data analysis. The task of model selection is to choose a suitable statistical model, from a list of candidates. Tools used for selecting the claim severity distribution include Q-Q plots, formal hypothesis tests like goodness-of-fit tests and the information criteria, such as AIC and BIC (Claeskens, 2006; Claeskens and Hjort, 2008). Q-Q plots have the advantage of being rather easy to construct and interpret, but it is difficult to construct an automatic model selection procedure based on these plots. As for goodness-of-fit tests, they are not really model selection tools, that can be used to compare different models directly, but rather means of checking the fit to the data once the model has been chosen. The information

Figure 3: The pdf of four-parameter distribution for different sets of parameters.

criteria AIC and BIC are given by

$$\text{AIC} = 2|\boldsymbol{\theta}| - 2\ell(\boldsymbol{\theta})$$

and

$$\text{BIC} = \log(n)|\boldsymbol{\theta}| - 2\ell(\boldsymbol{\theta}),$$

where $\boldsymbol{\theta}$ are the model parameters, $|\boldsymbol{\theta}|$ the length of the parameter vector, n the sample size and $\ell(\boldsymbol{\theta})$ the log-likelihood function. These are indeed model selection methods, and are therefore included in our study. They have the advantage that they are quick and easy to compute. However, as mentioned earlier, they mainly assess the overall fit to given data, putting most of the emphasis on middle regions of the distributions, where there is a lot of data. In many situations, one would like to find the model which gives the best estimates of specific quantities that the model will be used for. The Focused Information Criterion (FIC), first proposed by Claeskens and Hjort (2003), has this precise aim.

In this paper we use the updated version of the FIC, which proposed by Jullum and Hjort (2017). The advantages of this version is that it is derived without the local neighbourhood asymptotics and that it extends the idea from

comparing and ranking sets of nested parametric models to general non-nested models with a non-parametric alternative.

Suppose we have i.i.d. observations $Z_1, \dots, Z_n$ from some unknown distribution G . The objective is to estimate a focus parameter $\mu = T(G)$, where T is a suitable functional which maps a distribution to a scalar. The focus parameter μ is estimated with the plug-in estimator $\hat{\mu}_{np} = T(\hat{G}_n)$, where $\hat{G}_n$ is the empirical estimate of the unknown distribution G , i.e. the empirical distribution of the data. It may be shown that

$$\sqrt{n}(\hat{\mu}_{np} - \mu) \xrightarrow{d} N(0, v_{np}),$$

with $v_{np} = E(IF(Z_i; G)^2)$, $IF(\cdot; G)$ being the influence function of $T(\cdot)$, given by

$$IF(z; G) = \lim_{t \rightarrow 0} \frac{T((1-t)G + t\delta_z) - T(G)}{t},$$

where δ_z is the distribution with unit point mass in z .

If one instead estimates μ using a parametric model with pdf $f(\cdot; \boldsymbol{\theta})$, the corresponding estimate $\hat{\mu}_{pm}$ will in general be biased due to the fact that the parametric model is not the true data generating process. When $\boldsymbol{\theta}$ is estimated by the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$, it may be shown that

$$\sqrt{n}(\hat{\mu}_{pm} - \mu_0) \xrightarrow{d} N(0, v_{pm}),$$

where μ_0 is the so-called least false estimate of μ based on the least false parameter $\boldsymbol{\theta}_0$ minimizing the Kullback-Leibler divergence from the true distribution to parametric one. Further, $v_{pm} = \mathbf{c}^T \mathbf{J}^{-1} \mathbf{K} \mathbf{J}^{-1} \mathbf{c}$, where $\mathbf{J} = -E(\mathbf{I}(Z_i; \boldsymbol{\theta}_0))$, $\mathbf{K} = E(\mathbf{u}(Z_i; \boldsymbol{\theta}_0)\mathbf{u}(Z_i; \boldsymbol{\theta}_0)^T)$ and $\mathbf{c} = \partial \mu_{pm}(\boldsymbol{\theta}_0) / \partial \boldsymbol{\theta}_0$, $\mathbf{u}(\cdot; \boldsymbol{\theta})$ and $\mathbf{I}(\cdot; \boldsymbol{\theta})$ being the first and second derivatives of $\log f(\cdot; \boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, respectively. This results in the following first order approximations of the mean squared errors of $\hat{\mu}_{np}$ and $\hat{\mu}_{pm}$:

$$\text{mse}_{np} = E((\hat{\mu}_{np} - \mu)^2) \approx \frac{v_{np}}{n} \text{ and } \text{mse}_{pm} = E((\hat{\mu}_{pm} - \mu_0)^2) \approx b^2 + \frac{v_{pm}}{n},$$

where $b = \mu_0 - \mu$. The FIC scores are then given by

$$\begin{aligned} \text{FIC}_{np} &= \frac{\hat{v}_{np}}{n}, \\ \text{FIC}_{pm} &= \widehat{bsq} + \frac{\hat{v}_{pm}}{n}, \end{aligned} \tag{1}$$

where $\hat{v}_{np}$, $\hat{v}_{pm}$ and $\widehat{bsq}$ are estimates of v_{np} , v_{pm} and b^2 , respectively. The two variance estimates $\hat{v}_{np}$ and $\hat{v}_{pm}$ are obtained by replacing theoretical means with empirical means based on data, G by $\hat{G}_n$ and θ_0 by $\hat{\theta}$. As for bsq , $\hat{b}^2 = (\hat{\mu}_{pm} - \hat{\mu}_{np})^2$ is a biased estimate of b^2 . A bias corrected estimate is obtained by $\widehat{bsq}_0 = \hat{b}^2 - n^{-1}(\hat{v}_{np} + \hat{v}_{pm} - 2\hat{c}^T \hat{J}^{-1} \hat{d})$, with $\hat{d} = n^{-1} \sum_{i=1}^n IF(Z_i; \hat{G}_n) \mathbf{u}(Z_i; \hat{\theta})$. Finally, one lets $\widehat{bsq} = \max(\widehat{bsq}_0, 0)$ to avoid negative estimates of the obviously nonnegative b^2 . Since the FIC scores are calculated by estimating the mean squared error of the different model estimates of the focus parameter, the focused model selection strategy is to select the model associated with the smallest FIC value. Note that the nonparametric alternative is only used as a benchmark in our case, and not as one of the candidate models.

As mentioned earlier, the desired focus parameter in this setting is the quantile q_ϵ in the total loss distribution, which depends on both the claim frequency and the claim size distributions. As there is no explicit expression for the pdf of this distribution, the FIC cannot be computed analytically when q_ϵ is the focus parameter. One would have to resort to Monte Carlo methods, which would result in a computationally heavy procedure. Further, the non-parametric estimate $\hat{\mu}_{np}$ would be computed based on observations of $\mathcal{X}$, of which there is only one per year, resulting in 20 observations in total at the very best. As q_ϵ is a quantile far out in the right tail of the distribution of $\mathcal{X}$, the uncertainty in $\hat{\mu}_{np}$ would be huge. For these reasons, we have chosen to use other focus parameters as proxies for q_ϵ . One possibility is to use a combination of the mean and standard deviation of $\mathcal{X}$. Since the number $\mathcal{N}$ of claims from all policies in the portfolio is assumed to be independent of the individual claim sizes Z_i , we have

$$\mathbb{E}(\mathcal{X}) = \mathbb{E}\left(\sum_{i=1}^{\mathcal{N}} Z_i\right) = \mathbb{E}(\mathcal{N}) \mathbb{E}(Z_i)$$

and

$$\text{sd}(\mathcal{X}) = \text{sd}\left(\sum_{i=1}^{\mathcal{N}} Z_i\right) = \sqrt{\text{E}(\mathcal{N}) \text{Var}(Z_i) + \text{Var}(\mathcal{N}) (\text{E}(Z_i))^2}.$$

As the aim is to select the claim severity distributions, and the claim frequency distribution thus is considered known with a known parameter, $\text{E}(\mathcal{N})$ and $\text{Var}(\mathcal{N})$ are known constants. In the study, we have assumed that $\mathcal{N} \sim \text{Poisson}(\lambda)$, where λ depends on the size of the portfolio, the claim intensity and the length of the considered period, here one year. This results in $\text{E}(\mathcal{X}) = \lambda \text{E}(Z_i)$ and $\text{sd}(\mathcal{X}) = \sqrt{\lambda \text{E}(Z_i^2)}$, but this can easily be modified to other claim frequency distributions, such as the negative binomial. Letting $\mu = \text{E}(\mathcal{X}) + k \text{sd}(\mathcal{X})$, the influence function, needed to compute $\hat{v}_{np}$ is then

$$IF(z; G) = \lambda(z - \text{E}(Z_i)) + k \sqrt{\frac{\lambda}{\text{E}(Z_i^2)}} (z^2 - \text{E}(Z_i^2)).$$

The estimates $\widehat{bsq}$ and $\hat{v}_{pm}$ for the different candidate severity distributions are obtained by letting $f(\cdot; \boldsymbol{\theta})$ be the pdf of the claim severity distribution in question. In the study, we have used three different versions of this focus parameter with $k = 1.5, 2$ and 3 .

Another option is to let the α -quantile q_α^Z of the claim severity distribution, for some $\alpha \in (0, 1)$, be the focus parameter. The influence function in this case is

$$IF(z; G) = \frac{\alpha - I(z \leq G^{-1}(\alpha))}{g(G^{-1}(\alpha))},$$

where G and g are the cdf and pdf of the true claim size distribution, respectively. To compute $\hat{v}_{np}$, one may use a kernel density estimator $\hat{g}_n$ for g . The estimates $\widehat{bsq}$ and $\hat{v}_{pm}$ as for $\mu = \text{E}(\mathcal{X}) + k \text{sd}(\mathcal{X})$. In the study, we have used q_ϵ^Z as a focus parameter.

Since the quantile q_ϵ in the total loss distribution is not just a function of the ϵ -quantile of the claim severity distribution, a better proxy focus parameter may be a weighted sum of quantiles in the right tail, i.e. $\sum_{i=1}^k w_i q_{\alpha_i}^Z$ for some weights w_i and $\alpha_i \in (0, 1)$, $i = 1, \dots, k$. The corresponding FIC scores are then

weighted, averaged FIC scores, called AFIC, given by

$$\begin{aligned} \text{AFIC}_{np} &= \sum_{i=1}^k w_i \text{FIC}_{np,i} \\ \text{AFIC}_{pm} &= \sum_{i=1}^k w_i \text{FIC}_{pm,i}, \end{aligned} \tag{2}$$

where $\text{FIC}_{np,i}$ and $\text{FIC}_{pm,i}$ are the FIC scores for the quantile $q_{\alpha_i}^Z$, computed as described above. In the study, we included AFIC scores with the 0.90, 0.91, $\dots$, 0.99-quantiles and $w_i = 1/10$, $i = 1, \dots, 9$.

4. Simulation study

When the claim size distribution is to be used for a given purpose, such as estimating the quantile q_ϵ , the method for selecting the distribution should reflect this. The goal of the FIC is precisely to select the distribution that provides the best estimate of the focus parameter. However, as explained in Section 3, the quantile q_ϵ cannot be used directly as a focus parameter. Instead, one must use a proxy. To investigate how the FIC works in practice with different proxies and varying sample size, we have performed a simulation study. The parameter settings for this study are given in Section 4.1 and the results are presented in Section 4.2.

4.1. Parameter settings

Table 1 presents the choice of parameter values for each nine of the distributions with corresponding standard deviations and 95% and 99% quantiles ($\epsilon = 0.05$ and 0.01 , respectively). These parameter values result in distributions with a mean of (approximately) 10 and with a range of tail weights from moderate to quite heavy.

As mentioned in Section 3, we will try FICs with $E(\mathcal{X}) + k \text{sd}(\mathcal{X})$ and q_ϵ^Z as focus parameters, as well as an average FIC over a range of quantiles of the claim size distribution, and compare their performance to the one of the AIC and BIC. Further, the number of simulations in each experiment is $N = 1000$.

The sample size is varied between $n = 5000, 500, 50$, representing a large, medium, small sample size, respectively. Further, the quantile q_ϵ in the total loss distribution is computed for $\epsilon = 0.05$ and 0.01 . The expected number of occurrences is set to $\lambda = 50, 500$. Finally, we use $m = 1000000$ simulations to compute Monte Carlo estimates of the quantiles.

The goal is to select the claim size distribution that produces the best estimate of the quantile q_ϵ , which is not necessarily the true distribution. Therefore, the performance of the model selection methods will not be measured by how they rank the different distributions, but rather in terms of the quality of the resulting quantile estimates. Each experiment is performed as follows. One of the 9 distributions is chosen to be the true claim size distribution with pdf $f_Z(\cdot; \theta)$, and a sample $z_1, \dots, z_n$ of size n is drawn from this distribution. Then, the parameters of each of the 9 distributions are estimated based on the sample. This results in the estimates $\hat{\theta}_1, \dots, \hat{\theta}_9$. The 7 model selection criteria are then computed for each of the 9 distributions, resulting in 7 different rankings of the distributions. For each model selection criterion, the distribution ranked as best is used to estimate q_ϵ by Monte Carlo, as described in Algorithm 1. This results in the estimates $\hat{q}_{\epsilon, ij}$, $i = 1, \dots, N$, $j = 1, \dots, 7$. The quality of these estimates is then evaluated by the finite sample bias $b_j = \frac{1}{N} \sum_{i=1}^N (\hat{q}_{\epsilon, ij} - q_\epsilon)$ and the root mean squared error $\text{RMSE}_j = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{q}_{\epsilon, ij} - q_\epsilon)^2}$.

4.2. Results

Tables 2 to 7 display the bias and RMSE, respectively, in the quantile estimates from the different model selection criteria when $\epsilon = 0.05$ and $\lambda = 50$ with varying sample size $n = 5000, 500, 50$. We see that among the state of the art methods, the BIC performs best, especially when the sample size decreases. The FIC with the overall best performance is somewhat surprisingly FIC_ϵ . As q_ϵ is a quantile in the total loss distribution, one would expect the best result to require several quantiles in the claim size distribution, and not just the upper ϵ -quantile. However, the AFIC, that is a function of several quantiles in the right tail of the claim severity distribution does not do that well. One reason may be

DISTR.	PAR. VAL.	STD. DEV.	QUANTILE			
			$\lambda = 50$		$\lambda = 500$	
			95%	99%	95%	99%
Exp(β)	0.1	10.00	672.8	753.3	5528.6	5758.1
Gamma(α, β)	(10, 1)	3.16	625.3	680.6	5389.4	5555.0
L. – normal(ξ, σ)	(1.5, 1.27)	16.06	784.0	991.8	5887.9	6355.6
Weibull(α, β)	(2, 11.28)	5.23	635.2	695.8	5417.4	5595.8
L. – Gamma(α, β)	(6, 3.04)	24.58	771.82	1007.0	5850.4	6470.2
Pareto(α, β)	(3, 20)	17.32	744.6	915.1	5758.7	6165.4
Ext.Pa(α, θ, β)	(3, 2, 6.7)	14.21	713.9	852.1	5681.5	6023.6
Burr(α, β, τ)	(30, 4, 25.7)	2.85	624.3	679.6	5388.3	5553.6
F. – par.($\alpha, \theta, \beta, \eta$)	(3, 1.3, 6, 1.1)	19.70	761.9	965.6	5835.4	6338.3

Table 1: Parameter values for the claim severity distributions.

that all quantiles are weighted equally. One might obtain better results with a stronger weight on the quantiles further out in the tail. Using $E(\mathcal{X}) + k \text{sd}(\mathcal{X})$ as a proxy for the quantile q_ϵ seems to work better, and the FIC scores based on this criterion for different values of k perform well in some cases. The problem with these criteria, is that different values of k give the best result for different distributions and sample sizes. Further, their performance is varying a lot. Relative to the BIC, the FIC_ϵ performs either comparatively well, or considerably better, except for a few cases. In particular, it tends to do well when the claim severity distribution is heavy-tailed and when the sample size becomes smaller.

Corresponding results for $\epsilon = 0.01$ are shown in Tables 8 to 11, though the simulations for $n = 500$ are not shown, as the pattern is similar to that for $\epsilon = 0.05$. Again, the BIC is the best among the state of the art model selection criteria, and FIC_ϵ the best among the focussed criteria. The relative performance of the FIC_ϵ compared to the BIC follows the same tendencies as for $\epsilon = 0.05$, but it is comparatively better. Hence, it seems like the advantage of using the focussed criterion becomes larger when the parameter of interest is

Algorithm 1

Input: $\lambda, \hat{\theta}, f_Z(\cdot; \theta), m$

```
1: for  $i = 1, \dots, m$  do
2:    $\mathcal{X}_i^* = 0$ 
3:   Draw  $\mathcal{N}^* \sim \text{Poisson}(\lambda)$ 
4:   for  $j = 1, \dots, \mathcal{N}^*$  do
5:     Draw  $Z^*$  from  $f_Z(z; \hat{\theta})$ 
6:      $\mathcal{X}_i^* = \mathcal{X}_i^* + Z^*$ 
7:   end for
8: end for
9: Sort as  $X_{(1)}^* \leq \dots \leq \mathcal{X}_{(m)}^*$ 
10: Return  $\hat{q}_\epsilon = X_{((1-\epsilon)m)}^*$ 
```

a quantile further out in the tail of the distribution of the total loss.

Tables 12 to 15 show the results from the simulations with a higher claim frequency $\lambda = 500$. Again the case $n = 500$ is not shown here, and neither is $\epsilon = 0.05$, since the corresponding results are very similar to the ones for $\lambda = 50$. As λ increases, the total loss becomes a sum of a larger number of independent, identically distributed random variables. Its distribution should therefore become more and more similar to the normal distribution. In light of that, one would expect the FIC_{mkd} , based on $E(\mathcal{X}) + k \text{sd}(\mathcal{X})$ to perform better for $\lambda = 500$ than for $\lambda = 50$. That does however not seem to be the case. Further, the relative performance of the different criteria follows the same patterns as for $\lambda = 50$.

	AIC	BIC	FIC _{M1.5D}	FIC _{M2D}	FIC _{M3D}	FIC _ε	AFIC
Exp.	-70.68	-0.97	-228.69	-228.69	-234.00	-5.08	-273.09
Ga.	-0.54	-0.55	0.04	-0.06	0.14	0.28	0.52
L.-n.	3.94	1.04	-8.92	-4.65	-2.79	-19.51	80.76
We.	-36.55	0.13	0.17	0.13	0.27	0.12	0.27
L.-G.	5.27	0.13	39.55	-22.24	-49.78	24.02	61.73
Pa.	-61.03	1.08	-2.74	-4.37	-3.91	0.09	-6.92
E. P.	-0.37	-7.29	49.57	16.33	-12.95	-1.94	21.31
Burr	-291.56	-4.81	0.21	0.26	0.16	0.09	0.31
F.-par.	-67.90	-52.25	-15.88	-17.20	-18.91	-0.42	-16.10

Table 2: Bias in the quantile estimates from the different model selection methods when $\epsilon = 0.05$, $n = 5000$ and $\lambda = 50$.

	AIC	BIC	FIC _{M1.5D}	FIC _{M2D}	FIC _{M3D}	FIC _ε	AFIC
Exp.	157.17	42.54	251.99	252.38	254.37	54.37	274.58
Ga.	20.15	20.17	3.69	3.58	3.66	4.01	3.79
L.-n.	45.23	42.37	59.62	56.55	53.08	78.30	187.47
We.	153.00	5.66	5.61	5.59	5.52	5.45	5.52
L.-G.	32.84	23.21	119.91	91.47	75.18	26.58	55.04
Pa.	216.26	23.38	32.78	26.51	27.73	28.51	32.83
E. P.	36.55	79.18	125.77	89.61	46.33	27.11	39.62
Burr	426.46	55.88	3.49	3.58	3.46	3.44	3.68
F.-par.	233.60	210.64	43.83	39.35	37.36	31.86	35.93

Table 3: RMSE in the quantile estimates from the different model selection methods when $\epsilon = 0.05$, $n = 5000$ and $\lambda = 50$.

	AIC	BIC	FIC _{M1.5D}	FIC _{M2D}	FIC _{M3D}	FIC _ε	AFIC
Exp.	-70.50	18.02	-219.16	-219.44	-220.99	-122.35	-286.41
Ga.	-30.63	-1.07	8.32	6.05	4.25	2.54	2.57
L.-n.	24.44	20.43	21.94	16.83	15.15	43.55	97.32
We.	-32.82	-1.78	19.60	17.69	9.50	0.69	1.58
L.-G.	119.54	123.96	29.39	18.07	8.86	21.14	61.16
Pa.	-59.31	-3.69	29.66	27.88	25.27	3.26	49.28
E. P.	-161.34	-93.80	12.94	11.97	8.05	-3.35	-74.58
Burr	-94.98	-3.25	5.50	-3.30	2.67	1.09	1.22
F.-par.	-120.29	-9.45	24.81	19.75	12.28	10.96	138.27

Table 4: Bias in the quantile estimates from the different model selection methods when $\epsilon = 0.05$, $n = 500$ and $\lambda = 50$.

	AIC	BIC	FIC _{M1.5D}	FIC _{M2D}	FIC _{M3D}	FIC _ε	AFIC
Exp.	198.42	71.70	261.36	261.80	263.13	207.26	293.29
Ga.	140.22	9.23	37.07	25.00	13.67	8.92	9.50
L.-n.	133.01	136.45	157.30	153.57	147.34	317.45	346.79
We.	145.42	37.74	68.74	64.66	44.65	14.19	14.25
L.-G.	270.10	188.26	178.64	170.83	172.35	208.41	255.39
Pa.	232.58	80.59	153.41	151.77	148.30	208.21	225.21
E. P.	338.40	269.40	133.68	131.33	125.34	8.05	241.02
Burr	243.89	29.90	30.32	14.45	8.02	7.94	7.98
F.-par.	316.04	96.55	156.45	154.69	146.09	220.51	236.52

Table 5: RMSE in the quantile estimates from the different model selection methods when $\epsilon = 0.05$, $n = 500$ and $\lambda = 50$.

	AIC	BIC	FIC _{M1.5D}	FIC _{M2D}	FIC _{M3D}	FIC _ε	AFIC
Exp.	-113.86	-38.70	-256.44	-313.60	-314.23	-163.16	-350.58
Ga.	-70.87	-58.05	31.43	-29.58	-30.09	-57.32	-58.19
L.-n.	86.73	30.99	-79.01	-261.08	-264.12	-32.34	-238.71
We.	-64.07	-62.37	26.16	-33.58	-34.87	-61.96	-62.79
L.-G.	-290.90	-251.18	-149.62	-401.17	-371.21	-172.03	-332.36
Pa.	-74.05	26.36	-22.89	-180.91	-168.71	0.21	182.90
E. P.	-124.98	33.70	-54.48	-186.18	-168.64	24.27	-148.17
Burr	-68.27	-55.89	22.91	-35.29	-33.71	-55.02	-55.48
F.-par.	-106.21	-15.31	-35.63	-242.11	-225.03	-48.71	-210.92

Table 6: Bias in the quantile estimates from the different model selection methods when $\epsilon = 0.05$, $n = 50$ and $\lambda = 50$.

	AIC	BIC	FIC _{M1.5D}	FIC _{M2D}	FIC _{M3D}	FIC _ε	AFIC
Exp.	231.76	150.32	297.40	347.95	346.91	262.96	365.84
Ga.	117.13	64.96	103.86	100.17	100.20	64.28	65.17
L.-n.	727.42	610.81	253.05	377.64	361.08	450.23	547.38
We.	85.06	77.35	108.34	110.82	108.11	81.71	77.82
L.-G.	583.15	478.83	226.93	437.82	419.45	337.35	420.09
Pa.	283.51	240.28	207.87	276.87	269.20	276.06	299.77
E. P.	316.45	218.32	165.00	245.96	235.56	234.63	259.27
Burr	112.25	61.49	90.28	90.56	93.51	61.15	61.19
F.-par.	372.68	311.75	234.60	336.60	324.02	347.39	453.31

Table 7: RMSE in the quantile estimates from the different model selection methods when $\epsilon = 0.05$, $n = 50$ and $\lambda = 50$.

	AIC	BIC	FIC _{M1.5D}	FIC _{M2D}	FIC _{M3D}	FIC _ε	AFIC
Exp.	-45.65	34.09	-227.37	-227.74	-232.96	5.91	227.91
Ga.	-2.67	-2.69	-1.98	-1.96	-2.01	-1.74	1.97
L.-n.	2.18	-4.34	-12.97	-2.70	4.68	19.57	20.59
We.	-8.98	1.40	1.29	1.24	1.26	1.31	1.39
L.-G.	14.68	-4.30	-57.14	-122.20	-155.22	17.72	90.26
Pa.	-15.37	9.96	-4.40	-6.04	-5.30	6.96	9.73
E. P.	-1.07	-7.40	23.97	-9.74	41.97	-1.71	-10.25
Burr	6.63	-4.37	0.87	-0.91	-0.91	0.83	0.87
F.-par.	-64.64	-58.32	-53.28	-53.92	-55.69	-3.71	-14.47

Table 8: Bias in the quantile estimates from the different model selection methods when $\epsilon = 0.01$, $n = 5000$ and $\lambda = 50$.

	AIC	BIC	FIC _{M1.5D}	FIC _{M2D}	FIC _{M3D}	FIC _ε	AFIC
Exp.	161.61	47.53	255.49	255.98	258.19	91.67	278.14
Ga.	21.92	20.32	3.75	3.72	3.76	3.69	3.85
L.-n.	62.75	52.32	125.02	118.88	115.41	86.64	63.32
We.	7.41	5.71	5.58	5.66	5.64	5.41	5.70
L.-G.	84.90	41.58	136.56	175.62	192.46	100.73	159.93
Pa.	65.40	39.99	48.92	47.71	46.90	42.17	54.97
E. P.	49.94	96.20	126.33	97.88	71.07	31.93	46.97
Burr	63.63	60.71	3.05	3.12	3.07	3.03	3.12
F.-par.	302.51	274.47	77.02	76.79	77.31	68.07	69.43

Table 9: RMSE in the quantile estimates from the different model selection methods when $\epsilon = 0.01$, $n = 5000$ and $\lambda = 50$.

	AIC	BIC	FIC _{M1.5D}	FIC _{M2D}	FIC _{M3D}	FIC _ε	AFIC
Exp.	-74.07	34.39	-235.65	-234.22	-292.02	-156.98	-256.96
Ga.	-16.25	-2.22	29.88	27.75	21.89	0.79	-1.67
L.-n.	1849.51	671.79	-211.68	-209.04	108.54	322.38	152.61
We.	-3.49	10.73	29.13	28.81	21.00	-2.09	269.25
L.-G.	163.28	55.40	-335.18	-335.53	42.31	28.47	87.81
Pa.	149.13	212.98	-103.47	-98.42	97.45	46.64	130.73
E. P.	32.73	232.75	-117.09	-111.88	45.57	42.12	109.12
Burr	-14.21	-0.88	23.89	21.16	-15.41	-0.64	-2.33
F.-par.	204.39	241.58	-163.95	-159.90	-110.21	96.06	170.40

Table 10: Bias in the quantile estimates from the different model selection methods when $\epsilon = 0.01$, $n = 50$ and $\lambda = 50$.

	AIC	BIC	FIC _{M1.5D}	FIC _{M2D}	FIC _{M3D}	FIC _ε	AFIC
Exp.	208.61	152.85	282.53	281.37	280.05	268.85	301.20
Ga.	102.87	31.82	107.60	103.81	103.50	31.81	32.82
L.-n.	2525.73	2439.54	331.36	329.75	321.90	624.25	1295.38
We.	60.87	77.82	114.45	114.51	110.46	50.24	323.91
L.-G.	2873.16	1748.88	382.34	380.16	367.47	723.98	1017.71
Pa.	620.75	534.80	240.15	240.98	239.30	388.84	613.63
E. P.	550.58	407.16	203.12	204.40	201.17	424.72	563.61
Burr	97.47	27.73	94.46	89.98	94.08	28.02	38.65
F.-par.	1162.50	1007.49	292.88	292.48	287.42	112.70	1258.48

Table 11: RMSE in the quantile estimates from the different model selection methods when $\epsilon = 0.01$, $n = 50$ and $\lambda = 50$.

	AIC	BIC	FIC _{M1.5D}	FIC _{M2D}	FIC _{M3D}	FIC _{ϵ}	AFIC
Exp.	-297.14	-95.15	-653.80	-654.04	-654.75	-32.23	-694.11
Ga.	-9.03	-8.66	23.66	16.30	6.27	-2.49	-3.01
L.-n.	31.23	3.83	171.21	171.91	155.17	115.14	150.08
We.	-4.33	0.36	0.82	0.75	0.87	1.01	1.07
L.-G.	150.29	38.26	-107.96	-69.32	-59.16	79.00	175.07
Pa.	-157.38	3.09	137.35	153.26	-14.79	-5.76	6.90
E. P.	-14.88	-76.47	40.49	36.81	46.55	-21.40	-26.22
Burr	-53.97	-43.50	0.96	0.87	0.81	0.32	0.93
F.-par.	-346.62	-391.70	33.86	-23.41	-155.82	9.35	-32.03

Table 12: Bias in the quantile estimates from the different model selection methods when $\epsilon = 0.01$, $n = 5000$ and $\lambda = 500$.

	AIC	BIC	FIC _{M1.5D}	FIC _{M2D}	FIC _{M3D}	FIC _{ϵ}	AFIC
Exp.	651.88	157.39	732.43	731.91	732.81	261.05	798.61
Ga.	177.78	177.72	36.98	33.42	25.91	25.28	26.43
L.-n.	356.28	322.14	357.97	347.14	351.96	455.48	467.34
We.	48.13	45.22	43.89	44.16	44.34	44.29	44.07
L.-G.	476.72	218.56	425.73	396.73	379.03	353.10	489.20
Pa.	395.45	196.76	342.77	344.33	224.17	191.42	244.77
E. P.	310.06	667.34	357.69	358.01	373.26	152.74	222.70
Burr	591.90	496.06	24.45	24.34	24.22	23.41	24.30
F.-par.	952.59	764.82	318.10	291.21	289.06	304.51	311.67

Table 13: RMSE in the quantile estimates from the different model selection methods when $\epsilon = 0.01$, $n = 5000$ and $\lambda = 500$.

	AIC	BIC	FIC _{M1.5D}	FIC _{M2D}	FIC _{M3D}	FIC _ε	AFIC
Exp.	-256.00	-98.29	-712.49	-712.08	-715.72	-402.16	-698.52
Ga.	-119.21	-5.89	89.37	88.46	85.07	3.26	4.76
L.-n.	5577.94	4757.56	-804.62	-808.98	-805.20	-1218.07	1711.67
We.	-39.20	-24.10	51.38	50.85	50.62	61.32	89.95
L.-G.	3051.94	1194.03	-1349.25	-1353.79	1349.96	89.95	612.50
Pa.	670.03	702.21	-497.10	-495.13	-502.32	244.67	468.60
E. P.	450.68	499.91	-544.46	-495.13	-506.32	363.33	384.39
Burr	-122.43	-13.09	63.65	62.86	60.80	-9.74	-19.60
F.-par.	1342.36	1342.32	-655.37	-653.21	-652.62	790.21	1075.99

Table 14: Bias in the quantile estimates from the different model selection methods when $\epsilon = 0.01$, $n = 50$ and $\lambda = 500$.

	AIC	BIC	FIC _{M1.5D}	FIC _{M2D}	FIC _{M3D}	FIC _ε	AFIC
Exp.	1080.13	904.56	1115.36	1116.16	1115.95	1142.78	1202.68
Ga.	836.90	258.67	377.62	375.75	368.81	257.87	258.30
L.-n.	1237.49	950.90	716.11	716.05	720.18	1031.62	1258.54
We.	501.72	407.38	486.15	483.97	480.80	409.19	871.77
L.-G.	4221.52	2729.74	1794.77	1794.31	1799.83	935.31	3023.61
Pa.	3664.67	3151.84	1389.07	1392.28	1384.50	1304.38	2073.68
E. P.	3679.09	2102.93	1141.17	1140.82	1133.71	1338.87	2863.31
Burr	805.66	229.70	346.08	344.26	341.55	326.06	330.54
F.-par.	9389.62	7494.26	1556.01	1559.69	1567.86	5228.72	9149.61

Table 15: RMSE in the quantile estimates from the different model selection methods when $\epsilon = 0.01$, $n = 50$ and $\lambda = 500$.

5. Real data example

In this section, we will apply the model selection methods considered in the simulation study to a set of motor insurance losses from a Norwegian insurance

company. These are discussed in Bølviken (2014), and consist of 6446 claims from the years 1997 to 2001. The claims, which are in 1000 Norwegian kroner (NOK), are given without the deductible. The average claim intensity in this data set is 5.7%. After fitting each of the nine distributions considered in the simulation study, we have computed the AIC, BIC, $\text{FIC}_{\text{m1.5d}}$, FIC_{m2d} , FIC_{m3d} , FIC_ϵ and AFIC for each of the distributions, assuming an expected number of $\lambda = 1000$ claims during the next year. The results are shown in Table 16. Note that the $\text{FIC}_{\text{m3d}} \cdot 10^{-5}$ cannot be computed for the fitted four-parameter distribution as its second moment is not defined.

We see that the classic model selection methods AIC and BIC rank the most complex and flexible distributions, i.e. the extended Pareto, the Burr and the four-parameter, the highest. This is not that surprising since there are abundant data. The focussed information criteria, on the other hand, prefer much simpler models; the exponential distribution is ranked first by $\text{FIC}_{0.05}$, $\text{FIC}_{0.01}$ and AFIC and second by $\text{FIC}_{\text{m1.5d}}$, FIC_{m2d} and FIC_{m3d} . The reason for this difference may be that these simple distributions capture the right tail of the distribution of the data rather well, but not the middle of the distribution. Further, the log-normal distribution is given a high rank by all criteria and may thus be a good compromise in this case.

6. Concluding remarks

Risk assessment is a core theme within non-life insurance. Estimation of quantiles far out in the upper tail is thus one of the main applications of the total loss distribution. The choice of claim severity distribution should reflect this. Therefore, we have explored how the focussed information criterion, FIC works as a tool for selecting the claim size distribution. As the quantile q_ϵ cannot be used directly as a focus parameter, we have tried different proxy focus parameters. To see how the FIC performs in this setting, compared to the other commonly used model selection methods AIC and BIC, we have performed a simulation study.

In particular, we wanted to investigate the effect of the heaviness of the tail of the claim size distribution and the amount of available data. Therefore, nine distributions with different tail properties were included in the study, and we used three different sample sizes. The performance of the different model selection methods was then evaluated based on the quality of the resulting estimates of q_ϵ for $\epsilon = 0.05$ and 0.01 , in terms of the bias and the RMSE.

Our study shows that among the state of the art approaches, the overall best performing model selection method for the claim severity distribution is the BIC, and the best of the focussed criteria is the simplest one, namely the FIC_ϵ , based on one single quantile from the claim severity distribution. Further, the performance of the FIC_ϵ is either comparable to or considerably better than that of the BIC, except for a few cases. In particular, the focussed criterion works well when the data are heavy-tailed, when the sample size is rather low and when the parameter of interest is a quantile far out in the tail of the total loss distribution. It is surprising that the AFIC, which takes several quantiles of the claim size distribution into account does not perform better. However, it might be improved by putting more weight on quantiles further out in the tail. Further, we would have expected the FIC_{mkd} to be better for larger claim frequencies, as the total loss distribution should approach the normal distribution. That is however not the case in our simulations.

In this study, we have restricted our attention to the choice of claim severity distributions, fixing the claim frequency distribution at the Poisson distribution. The FIC scores can however easily be modified to account for other claim frequency distributions, such as the negative binomial. Further, one could imagine doing a simultaneous selection of the claim number and size distributions using the FIC. As the upper tail of the total loss distribution typically is more influenced by the latter, it would be interesting to see whether quantiles provide information that enables to discriminate between different claim number distributions.

Acknowledgments

The authors would like to thank the referee and the editor for their useful comments and suggestions that improved the paper.

References

- Bakar, S.A., Hamzah, N., Maghsoudi, M., Nadarajah, S., 2015. Modeling loss data using composite models. *Insurance: Mathematics and Economics* 61, 146–154. doi:10.1016/j.insmatheco.2014.08.008.
- Bølviken, E., 2014. *Computation and Modelling in Insurance and Finance*. International Series on Actuarial Science, Cambridge University Press.
- Bølviken, E., Haff, I.H., 2018. One family, six distributions – A flexible model for insurance claim severity. *arXiv:1805.10854. Working Paper*.
- Brazauskas, V., Kleefeld, A., 2016. Modeling Severity and Measuring Tail Risk of Norwegian Fire Claims. *North American Actuarial Journal* 20(1), 1–16. doi:10.1080/10920277.2015.1062784.
- Claeskens, G., 2006. Information Criteria. *Encyclopedia of Actuarial Science* 2.
- Claeskens, G., Hjort, N.L., 2003. The Focused Information Criterion. *Journal of the American Statistical Association* 98(464), 900–916. doi:10.1198/016214503000000819.
- Claeskens, G., Hjort, N.L., 2008. *Model Selection and Model Averaging*. Cambridge Books, Cambridge University Press.
- Gray, R., Pitts, S., 2013. Risk Modelling in General Insurance: From Principles to Practice. *International Statistical Review* 81, 478–479.
- Jullum, M., Hjort, N., 2017. Parametric or nonparametric: The FIC approach. *Statistica Sinica* 27(3), 951–981. doi:10.5705/ss.202015.0364.

- Kaas, R., Goovaerts, M., Dhaene, J., Denuit, M., 2001. Modern Actuarial Risk Theory – Using R. Springer-Verlag Berlin Heidelberg.
- Kleiber, C., Kotz, S., 2003. Statistical Size Distributions in Economics and Actuarial Sciences. Wiley.
- Klugman, S., Panjer, H., Willmot, G., 2012. Loss Models: From Data to Decisions. Wiley Series in Probability and Statistics, Wiley.
- Reynkens, T., Verbelen, R., Beirlant, J., Antonio, K., 2017. Modelling censored losses using splicing: A global fit strategy with mixed Erlang and extreme value distributions. Insurance: Mathematics and Economics 77, 65–77. doi:10.1016/j.insmatheco.2017.08.005.

	AIC	BIC	$FIC_{M1.5D}$	FIC_{M2D}	FIC_{M3D}	$FIC_{0.05}$	$FIC_{0.01}$	AFIC
Exp.	5.38	5.38	1.47	1.53	1.66	1.16	2.74	1.30
Ga.	457.26	457.26	1.49	1.56	3.00	53.33	1901.33	339.92
L-n.	5.30	5.30	4.83	5.15	5.82	2.84	19.22	9.95
We.	5.37	5.38	0.92	0.96	1.09	15.93	1427.77	241.25
L-G.	7.48	7.48	480.06	581.41	813.20	370.41	8382.60	1292.96
Pa.	5.37	5.37	4.68	7.76	16.26	3.49	295.84	58.56
E. P.	5.27	5.27	606.62	635.09	693.99	14.04	104.22	44.00
Burr	5.26	5.26	35.08	53.42	101.62	15.79	99.46	21.93
F-par.	5.26	5.26	-	-	-	14.11	427.18	55.91

Table 16: Model selection measures for each of the nine considered distributions fitted to the motor insurance claims. The AIC and BIC scores are multiplied by 10^{-4} and the FIC_{mkl} scores by 10^{-5} .