Modeling Severity and Measuring Tail Risk of Norwegian Fire Claims

Vytaras Brazauskas[1]
University of Wisconsin-Milwaukee

Andreas Kleefeld[2]
Brandenburg University of Technology

To appear in *North American Actuarial Journal*
(*Submitted*: October 20, 2014 *Revised*: February 17, 2015 *Accepted*: May 3, 2015)

Abstract. The probabilistic behavior of the claim severity variable plays a fundamental role in calculation of deductibles, layers, loss elimination ratios, effects of inflation, and other quantities arising in insurance. Among several alternatives for modeling severity, the parametric approach continues to maintain the leading position, which is primarily due to its parsimony and flexibility. In this paper, several parametric families are employed to model severity of Norwegian fire claims for the years 1981 through 1992. The probability distributions we consider include: generalized Pareto, lognormal-Pareto (two versions), Weibull-Pareto (two versions), and folded-t. Except for the generalized Pareto distribution, the other five models are fairly new proposals that recently appeared in the actuarial literature. We use the maximum likelihood procedure to fit the models, and assess the quality of their fits using basic graphical tools (quantile-quantile plots), two goodness-of-fit statistics (Kolmogorov-Smirnov and Anderson-Darling), and two information criteria (AIC and BIC). In addition, we estimate the tail risk of 'ground up' Norwegian fire claims using the value-at-risk and tail-conditional median measures. We monitor the tail risk levels over time, for the period 1981 to 1992, and analyze predictive performances of the six probability models. In particular, we compute the next-year probability for a few upper tail events using the fitted models and compare them with the actual probabilities.

[1] Corresponding author: Vytaras Brazauskas, Ph.D., ASA, is a Professor in the Department of Mathematical Sciences, University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, WI 53201, USA. *e-mail*: `vytaras@uwm.edu`

[2] Andreas Kleefeld, Ph.D., is a Postdoctoral Fellow in the Institute for Applied Mathematics and Scientific Computing, Brandenburg University of Technology, Cottbus, Germany. *e-mail*: `kleefeld@tu-cottbus.de`

1 Introduction

The implementation of sound quantitative risk models is a vital concern for insurance industry, banks and other financial services companies. This process has accelerated over the last decade due to the revised regulatory frameworks such as *Solvency II* and *Basel II/III*. Among various methods used for measuring tail risk, those based on the Generalized Pareto Distribution (GPD) play a central role (see, e.g., McNeil, Frey, and Embrechts, 2005). Over the years, a number of authors have persuasively argued that the choice of GPD is natural and supported by rigorous theorems from extreme value theory, which is certainly true. In practice, however, implementation of this model is not as "smooth" as one would hope. Indeed, as discussed by Resnick (1997), the sensitivity of various model-fitting methods to the choice of threshold (that defines where the model tail starts) is a problem, and that the use of graphical tools such as the Hill plot is often "more guesswork than science." Some authors have tried to manage this problem by employing robust procedures to estimate the parameters of GPD (see, for example, Dupuis, 1998, Peng and Welsh, 2001, Juárez and Schucany, 2004, and Brazauskas and Kleefeld, 2009). In this paper, however, we take a different approach – we assume that the probability distributions under consideration are appropriate for all (extreme and non-extreme) data. And since Norwegian fire claims data were observed above a known and fairly large threshold, we fit *truncated* versions of the GPD and several alternative models.

As is well known, the probabilistic behavior of the claim severity variable plays fundamental role not only in measuring tail risk, but also in calculation of deductibles, layers, loss elimination ratios, effects of inflation, and other quantities arising in insurance. A number of reasons explaining why parametric models are preferred over other alternatives have been outlined by Klugman, Panjer, and Willmot (2012). These authors also provide a catalog of standard probability distributions that are used in actuarial work. And even though the catalog is indeed comprehensive, the search for, and development of, new distributions continues in the actuarial and statistical literatures. The chief objective of such attempts is to have parsimonious yet sufficiently flexible models that exhibit excellent goodness-of-fit properties for the entire range—not just the extreme right tail—of real insurance data. Also, if such distributions are successfully constructed or identified, the above-mentioned challenges encountered with the GPD would be eliminated.

Three major approaches have emerged in the actuarial literature for tackling these problems. The first approach is purely mathematical where one introduces a more complicated model and then checks if that solves the problem at hand. Prominent examples of this approach are skew-normal and skew-t models (see Eling, 2012) and log-phase-type distributions (see Ahn, Kim, and Ramaswami, 2012). Both papers used the well-known Danish fire claims data set (see McNeil, 1997) to illustrate that the proposed models provide reasonable fits. The second approach aims at fixing the issues associated with the GPD and pursues the composite models that are constructed using model splicing. In particular, researchers assume that smaller claims follow some standard distribution such as lognormal or Weibull, but larger claims are modeled with a type of Pareto distribution. As demonstrated in numerical examples of Scollnik (2007) and Scollnik and Sun (2012), the assumption of a Pareto-type model for the larger claims yields promising results. The third approach is guided by a different philosophy, the one of simplicity. The rationale of this approach is that the more complex the model, the higher the chance it will present analytical and/or computational challenges. Indeed, complex models are based on stronger assumptions and they have more parameters. This results in more uncertainty since usually the values of parameters are unknown and need to be estimated from the data. Currently, the most promising probability distributions in this research venue are the folded-symmetric models, which represent the positive half of normal, Cauchy, or, more generally, t distributions (for details, see Psarakis and Panaretoes, 1990, Brazauskas and Kleefeld, 2011, 2014, and Scollnik, 2014).

There is a vast literature on risk measures and their application to contract pricing, capital allocation, and risk management. For a quick introduction into these topics, the reader may be referred to Albrecht (2004), Tapiero (2004), and Young (2004). When measuring risk actuaries have to identify appropriate risk measures, collect reliable data, and also perform statistical estimation of the selected risk measures. In the actuarial literature, systematic studies of the statistical aspects in risk estimation were intitiated by Jones and Zitikis (2003), and then followed by Brazauskas and Kaiser (2004), Kaiser and Brazauskas (2006), and other authors. In this paper, we will adhere to the principles of sound statistical analysis emphasized by those authors. That is, when measuring tail risk via a parametrically-estimated risk measure, we will first perform extensive model validation and only then employ various tools for evaluation, monitoring and prediction of risk.

This paper could be characterized as a combination of a review article and a case study. Its objective is three-fold: (a) to review several recently proposed and most effective probability distributions for modeling heavy-tailed insurance losses; (b) to walk the reader through an extensive model validation process, with sufficient detail so the empirical findings could be easily reproduced; and (c) to identify the effects, if any, model selection has on the upper-tail risk measurement. To achieve these goals, we use the Norwegian fire claims (1981–1992) data which are widely-accessible and well-studied.

The rest of the paper is organized as follows. Section 2 presents key features of the data sets modeled in this paper. Section 3 introduces the probability distributions under consideration and provides their parameter estimates for each of the data sets. Quantile-quantile plots along with a number of model validation and selection criteria are presented in Section 4. Further, in Section 5, we measure the tail risk of Norwegian fire claims by utilizing value-at-risk, a popular risk metric, and tail-conditional median, an alternative to the well-known tail-conditional expectation. Final remarks are made and conclusions are drawn in Section 6.

2 Data

To illustrate how our selected probability distributions work on real data, we will use the well-studied Norwegian fire claims data which are available at the following website:

`http://lstat.kuleuven.be/Wiley` (in Chapter 1, file NORWEGIANFIRE.TXT).

The data represent the total damage done by fires in Norway for the years 1972 through 1992. Note that only damages in excess of a priority of 500,000 Norwegian krones are available. Also, it is not known whether the claims were inflation adjusted or not. Table 1 provides a summary of the data sets for the final 12 years, 1981–1992. This shorter period will suffice for our investigations and reduce possible effects of inflation, if any. We can see from the table that all typical features of insurance claims appear in these data sets. Specifically: most frequent claims are relatively small (from 40% of the observations in 1990 to 55% in 1982 fall between 500,000 and 1,000,000), but as severity grows, claim frequency declines (24%–39% between 1,000,000 and 2,000,000; 12%–18% between 2,000,000 and

5,000,000; etc.), and there is a small percentage of extremely large claims. (Notice how remote and spread out are the top three claims in 1988, with the largest one being almost half a *billion* Norwegian krones.) This exploratory analysis suggests that only highly-skewed and heavy-tailed models can capture most of the characteristics of the given data sets.

TABLE 1: Summary statistics for the *Norwegian Fire Claims* (1981–1992) data.

Claim Severity	Norwegian Fire Claims for Year											
(in 1,000,000's)	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992
Relative Frequencies (%)												
[0.5; 1.0)	54.5	54.7	53.8	48.3	49.6	50.7	41.9	41.2	41.9	40.4	43.3	45.5
[1.0; 2.0)	23.5	23.8	27.8	29.8	29.2	31.2	34.7	32.8	33.3	39.0	34.0	31.4
[2.0; 5.0)	13.5	14.3	12.0	15.4	14.0	11.9	17.1	16.9	18.0	16.2	17.9	16.4
[5.0; 10.0)	4.4	4.7	3.4	4.8	4.0	3.2	4.2	5.2	3.9	1.9	3.2	4.1
[10.0; 20.0)	2.3	2.1	2.2	1.1	1.6	1.7	1.2	1.9	1.9	1.4	1.3	1.6
20+	1.6	0.5	0.7	0.5	1.6	1.2	1.0	1.9	1.0	1.0	0.3	1.0
Top 3 Claims (in 1,000,000's)												
	43	19	22	22	60	87	35	84	45	26	17	45
	62	20	30	56	70	98	38	151	86	41	35	50
	78	23	51	106	135	188	45	465	145	79	50	102
Sample Size												
	429	428	407	557	607	647	767	827	718	628	624	615

3 Modeling Severity

The GPD, composite Pareto, and folded-t distributions will be used to fit the Norwegian fire claims data. To accomplish this task, we employ the truncated maximum likelihood procedures (denoted MLE). In this section, the most essential distributional properties of the models are introduced and their MLE parameter estimates for all the data sets are presented.

3.1 Generalized Pareto Distribution

The cumulative distribution function (cdf) of the GPD is given by

$$F_{\mathrm{GPD}(\sigma,\gamma)}(x) = \begin{cases} 1-(1-\gamma x/\sigma)^{1/\gamma}, & \gamma \neq 0 \\ 1-\exp(-x/\sigma), & \gamma = 0, \end{cases} \tag{3.1}$$

and the probability density function (pdf) by

$$f_{\mathrm{GPD}(\sigma,\gamma)}(x) = \begin{cases} \sigma^{-1}\left(1-\gamma x/\sigma\right)^{1/\gamma-1}, & \gamma \neq 0 \\ \sigma^{-1}\exp\left(-x/\sigma\right), & \gamma = 0, \end{cases} \tag{3.2}$$

where the pdf is positive for $x \geq 0$, when $\gamma \leq 0$, or for $0 \leq x \leq \sigma/\gamma$, when $\gamma > 0$. The parameters $\sigma > 0$ and $-\infty < \gamma < \infty$ control the scale and shape of the distribution, respectively. Note that when $\gamma = 0$ and $\gamma = 1$, the GPD reduces to the exponential distribution (with scale σ) and the uniform distribution on $[0, \sigma]$, respectively. If $\gamma < 0$, then the Pareto distributions are obtained. Also, the mean and variance of the GPD random variable, $X_{\mathrm{GPD}(\sigma,\gamma)}$, are given by

$$\mathbf{E}\big[X_{\mathrm{GPD}(\sigma,\gamma)}\big] = \frac{\sigma}{1+\gamma}, \quad \gamma > -1, \quad \text{and} \quad \mathbf{Var}\big[X_{\mathrm{GPD}(\sigma,\gamma)}\big] = \frac{\sigma^2}{(2\gamma+1)(\gamma+1)^2}, \quad \gamma > -1/2. \tag{3.3}$$

Further, besides functional simplicity of its cdf and pdf, another attractive feature of the GPD is that its quantile function (qf) has an explicit formula. This is especially useful for model diagnostics (e.g., quantile-quantile plots) and for portfolio risk evaluations based on value-at-risk measures. Specifically, for $0 < u < 1$, the qf is given by

$$F^{-1}_{\mathrm{GPD}(\sigma,\gamma)}(u) = \begin{cases} (\sigma/\gamma)\left(1-(1-u)^{\gamma}\right), & \gamma \neq 0 \\ -\sigma\log(1-u), & \gamma = 0. \end{cases} \tag{3.4}$$

Let $X_1, \ldots, X_n$ denote a sample of *independent and identically distributed* (*i.i.d.*) random variables from the GPD (σ, γ) model. Note that all the observed X_i's exceed a known threshold d. (For the Norwegian fire claims data, we have $d = 500,000$.) Then, a maximum likelihood estimator of (σ, γ) is found by numerically maximizing the left truncated (at $d = 500,000$) log-likelihood function:

$$\begin{aligned} \log \mathcal{L}_{\mathrm{GPD}}(\sigma,\gamma \,|\, X_1,\ldots,X_n) &= \sum_{i=1}^{n} \log\left(\frac{f_{\mathrm{GPD}(\sigma,\gamma)}(X_i)}{1-F_{\mathrm{GPD}(\sigma,\gamma)}(d)}\right) \\ &= -n\log\sigma + \frac{1-\gamma}{\gamma}\sum_{i=1}^{n}\log\left(1-\frac{\gamma}{\sigma}X_i\right) - n\log\left(1-F_{\mathrm{GPD}(\sigma,\gamma)}(d)\right), \end{aligned} \tag{3.5}$$

where expression (3.5) is derived by taking into account (3.2). For the data sets under consideration, the following MLE estimates resulted (see Table 2).

TABLE 2: Parameter MLEs of the GPD model for the *Norwegian Fire Claims* (1981–1992) data.

GPD	Norwegian Fire Claims for Year											
parameters	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992
σ ($\times 10^{-3}$)	70.5	160.5	127.3	344.4	170.8	174.8	489.0	382.5	488.2	555.0	526.0	396.6
γ	-0.83	-0.70	-0.71	-0.56	-0.75	-0.71	-0.51	-0.69	-0.56	-0.42	-0.42	-0.58

As is evident from Table 2, the distribution of Norwegian fire claims, when modeled using the GPD assumption, has a heavy right tail. Indeed, except for the years 1990 and 1991, the estimates of the shape parameter γ range from -0.83 in 1981 to -0.51 in 1987. According to (3.3), these models have infinite variance. The GPD models fitted for the years 1990 and 1991 do have finite variances, but their third and higher order moments are infinite.

3.2 Composite Pareto Models

Building on the work of Cooray and Ananda (2005) and Cooray (2009), Scollnik (2007) and later Scollnik and Sun (2012) proposed several composite models for modeling insurance losses. Specifically, they assumed that the claim severity variable has the following pdf:

$$f(x) = \begin{cases} w f_1(x)/F_1(\theta), & \text{if } \ 0 < x \leq \theta, \\ (1-w) f_2(x), & \text{if } \ x > \theta, \end{cases} \tag{3.6}$$

where w $(0 \leq w \leq 1)$ is a mixing weight, f_1 and F_1 are the pdf and cdf, respectively, of the "small" and "medium" claims distribution, and f_2 is the pdf of the "large" claims distribution. The cdf and qf are given by

$$F(x) = \begin{cases} w F_1(x)/F_1(\theta), & \text{if } \ 0 < x \leq \theta, \\ w + (1-w) F_2(x), & \text{if } \ x > \theta, \end{cases} \tag{3.7}$$

and

$$F^{-1}(u) = \begin{cases} F_1^{-1}\left(\dfrac{u F_1(\theta)}{w}\right), & \text{if } \ 0 < u \leq w, \\ F_2^{-1}\left(\dfrac{u-w}{1-w}\right), & \text{if } \ u > w, \end{cases} \tag{3.8}$$

respectively, where F_1^{-1} and F_2^{-1} denote the respective qf's. An appealing characteristic of this model is that the threshold level θ is treated as an unknown parameter, hence allowing the data to define which claims are small, medium, and which ones are large. Note also that the tail heaviness and the number of finite moments for f depends on the tail behavior of the second distribution f_2.

Further, by requiring that the pdf f be continuous and differentiable at θ, one can eliminate some of the parameters of f_1 and f_2 and thus simplify the composite model. That is, the number of free parameters in (3.6) can be reduced by imposing the following conditions:

$$w f_1(\theta)/F_1(\theta) = (1-w) f_2(\theta), \qquad w f_1'(\theta)/F_1(\theta) = (1-w) f_2'(\theta). \tag{3.9}$$

In this paper, we will consider only the most competitive composite models: *lognormal-Pareto*, with f_1 lognormal and f_2 either Pareto I (denoted LNPa2) or GPD (denoted LNPa3), and *Weibull-Pareto*, with f_1 Weibull and f_2 either Pareto I (denoted WePa2) or GPD (denoted WePa3).

3.2.1 Lognormal-Pareto Models

The pdf, cdf, and qf of the LNPa2 model are obtained by inserting the pdf's

$$f_1(x) = \frac{1}{\sigma x \sqrt{2\pi}} \exp\left\{ -\frac{1}{2}\left(\frac{\log x - \mu}{\sigma}\right)^2 \right\}, \qquad x > 0\,, \tag{3.10}$$

$$f_2(x) = \frac{\alpha \theta^\alpha}{x^{\alpha+1}}\,, \qquad x > \theta\,, \tag{3.11}$$

and the corresponding cdf's and qf's into (3.6), (3.7), and (3.8). Here $\theta > 0$ represents the threshold level at which splicing of the two distributions occurs, $\sigma > 0$ is a scale parameter for small and medium log-claims (i.e., claims below θ), and $\alpha > 0$ denotes the shape parameter which controls the heaviness of the tail. The remaining two parameters, $\mu > 0$ and $0 \le w \le 1$, are determined from the continuity and differentiability conditions (3.9) and given by

$$\mu = \log\theta - \alpha\sigma^2\,, \qquad w = \frac{\sqrt{2\pi}\,\alpha\sigma\,\Phi(\alpha\theta)\,\exp\{\alpha\sigma^2/2\}}{1 + \sqrt{2\pi}\,\alpha\sigma\,\Phi(\alpha\theta)\,\exp\{\alpha\sigma^2/2\}}\,,$$

where Φ denotes the cdf of the standard normal distribution. Note that the tail behavior of this model is driven by the Pareto distribution, which implies that the LNPa2 model has k finite moments when $\alpha > k$.

For the LNPa3 model, the pdf f_1 is given by (3.10) and f_2 is a truncated and reparametrized version of (3.2). More specifically, if in expression (3.2) one left truncates the pdf at θ and chooses

$\alpha = -1/\gamma$ and $\lambda = \sigma\alpha$, then the following density function will result:

$$f_2(x) = \frac{\alpha\,(\lambda+\theta)^{\alpha}}{(\lambda+x)^{\alpha+1}}\,, \qquad x > \theta\,, \tag{3.12}$$

where $\theta > 0$, $\sigma > 0$, $\alpha > 0$ have the same interpretation as in the case of LNPa2, and $\lambda > -\theta$ is a second location parameter. The remaining two parameters, $\mu > 0$ and $0 \leq w \leq 1$, are determined from the continuity and differentiability conditions (3.9) and given by

$$\mu = \log\theta - \left(\frac{\alpha\theta-\mu}{\lambda+\theta}\right)\sigma^2\,, \qquad w = \frac{\sqrt{2\pi}\,\alpha\theta\sigma\,\Phi\left(\frac{\log\theta-\mu}{\sigma}\right)\exp\left\{\left(\frac{\log\theta-\mu}{\sigma}\right)^2\Big/2\right\}}{\lambda+\theta+\sqrt{2\pi}\,\alpha\theta\sigma\,\Phi\left(\frac{\log\theta-\mu}{\sigma}\right)\exp\left\{\left(\frac{\log\theta-\mu}{\sigma}\right)^2\Big/2\right\}}\,.$$

Similar to the LNPa2 model, the tail behavior of this model is driven by the Pareto distribution, which implies that the LNPa3 model has k finite moments when $\alpha > k$.

3.2.2 Weibull-Pareto Models

The pdf of the WePa2 model is obtained by inserting the pdf's

$$f_1(x) = \left(\frac{\tau}{x}\right)\left(\frac{x}{\phi}\right)^{\tau}\exp\left\{-\left(\frac{x}{\phi}\right)^{\tau}\right\}\,, \qquad x > 0\,, \tag{3.13}$$

and f_2, given by (3.11), into (3.6). Then the corresponding cdf's and qf's are derived and inserted into (3.7) and (3.8). Here parameters $\theta > 0$ and $\alpha > 0$ have the same interpretation as before, i.e., as in Section 3.2.1, and $\tau > 0$ is the shape parameter for small and medium severity claims. The remaining two parameters, $\phi > 0$ and $0 \leq w \leq 1$, are again determined from the continuity and differentiability conditions (3.9) and given by

$$\phi = \theta\big(\alpha/\tau+1\big)^{-1/\tau}, \qquad w = \frac{\exp\{\alpha/\tau+1\}-1}{\exp\{\alpha/\tau+1\}+\tau/\alpha}\,.$$

Further, for the WePa3 model, we choose f_1 as in (3.13) and f_2 as in (3.12). The remaining steps are identical to those of the previously described composite models. Also, parameters $\theta > 0$, $\alpha > 0$, $\tau > 0$, $\lambda > -\theta$ have the same interpretations as before, and $\phi > 0$ and $0 \leq w \leq 1$ are given by

$$\phi = \theta\left(\frac{\alpha\theta-\lambda}{(\lambda+\theta)\,\tau}+1\right)^{-1/\tau}, \qquad w = \frac{\exp\{(\theta/\phi)^{\tau}\}-1}{(\tau/\alpha)\,(\lambda/\theta+1)\,(\theta/\phi)^{\tau}+\exp\{(\theta/\phi)^{\tau}\}-1}\,.$$

Finally, similar to the lognormal-Pareto models, the tail behavior of the Weibull-Pareto models is also driven by the Pareto distribution, which implies that these models have k finite moments when $\alpha > k$.

3.2.3 Numerical Results

Let $X_1, \ldots, X_n$ denote a sample of *i.i.d.* random variables from a composite Pareto model; all X_i's exceed $d = 500,000$. Then, a maximum likelihood estimator of θ and other unknown parameters of the model is found by numerically maximizing the left truncated (at $d = 500,000$) log-likelihood function:

$$\log \mathcal{L}\big(\theta, \text{other parameters} \,\big|\, X_1, \ldots, X_n\big) \;=\; \sum_{i=1}^{n} \log\left(\frac{f(X_i)}{1-F(d)}\right)$$

$$\begin{aligned} &= \sum_{i=1}^{n} \mathbf{1}\{X_i \leq \theta\} \log f_1(X_i) \;+\; \sum_{i=1}^{n} \mathbf{1}\{X_i > \theta\} \log f_2(X_i) \\ &\quad + \;\log\big(w/F_1(\theta)\big) \sum_{i=1}^{n} \mathbf{1}\{X_i \leq \theta\} \;+\; \log(1-w) \sum_{i=1}^{n} \mathbf{1}\{X_i > \theta\} \\ &\quad - \; n \log\big(1 - wF_1(d)/F_1(\theta)\big)\mathbf{1}\{d \leq \theta\} \;-\; n \log\big((1-w)(1-F_2(d))\big)\mathbf{1}\{d > \theta\}, \end{aligned} \tag{3.14}$$

where $\mathbf{1}\{\}$ denotes the indicator function and f and F are given by (3.6) and (3.7), respectively. Particular cases of expression (3.14) are obtained by selecting specific f_1, F_1 and f_2, F_2, as described in Sections 3.2.1 and 3.2.2. For the data sets under consideration, the following MLE estimates for the parameters of the composite Pareto models resulted (see Table 3).

Note that the three-parameter models, LNPa2 and WePa2, have two parameters in common (θ and α), which even have the same interpretations within each model. Likewise, the four-parameter models, LNPa3 and WePa3, have three parameters in common (θ, α, and λ), with identical interpretations in both models as well. This, of course, is due to the design of the proposed composite models. We notice from Table 3 that the shared parameters also take similar (but not identical) values at the corresponding models across all data sets. For example, in 1981, $\theta = 955,000$ for LNPa2 and $\theta = 934,000$ for WePa2; in 1984, $\alpha = 1.47$ for LNPa2 and $\alpha = 1.48$ for WePa2. Likewise, in 1991, $\theta = 2,677,000$ for LNPa3 and $\theta = 2,545,000$ for WePa3; $\alpha = 1.59$ for LNPa3 and $\alpha = 1.51$ for WePa3; $\lambda = -132,000$ for LNPa3 and $\lambda = -309,000$ for WePa3. It is important to understand that these shared parameters do not have to yield identical values for different models, as they are influenced by other features of the composite models. For example, parameter α governs the tail behavior in all four models, but it may take significantly different values in three- and four-parameter models. For instance, in 1985, it is equal to 1.21 for LNPa2 and 1.20 for WePa2, but 0.99 for LNPa3 and 0.98 for WePa3. On the other hand, in all four models $\alpha < 2$, implying that for all the years under consideration Norwegian fire claims are heavy tailed (i.e., their variances are infinite). Even more dramatically, according to LNPa3 and WePa3, in 1981 and 1985 the means were infinite as well (because $\alpha < 1$).

TABLE 3: Parameter MLEs of the composite Pareto models for the *Norwegian Fire Claims* (1981–1992) data.

Model parameters	Norwegian Fire Claims for Year											
	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992
Lognormal-Pareto I model, LNPa2 (θ, α, σ)												
θ $(\times 10^{-3})$	955	1018	1122	2428	1230	1412	1927	1839	2303	2057	2596	2297
α	1.19	1.25	1.29	1.47	1.21	1.31	1.43	1.21	1.40	1.63	1.67	1.40
σ	0.70	0.74	0.84	1.03	0.84	0.89	0.78	0.89	0.89	0.71	0.86	0.97
Weibull-Pareto I model, WePa2 (θ, α, τ)												
θ $(\times 10^{-3})$	934	994	1066	2297	1122	1324	1685	1685	2063	1778	2344	2023
α	1.19	1.26	1.29	1.48	1.20	1.31	1.42	1.21	1.39	1.63	1.68	1.40
τ	1.19	1.07	0.92	0.64	1.00	0.87	1.16	0.94	0.92	1.32	0.89	0.79
Lognormal-GPD model, LNPa3 $(\theta, \alpha, \sigma, \lambda)$												
θ $(\times 10^{-3})$	1085	991	1237	2700	2444	1789	1903	2035	3161	2543	2677	2588
α	0.98	1.33	1.19	1.33	0.99	1.02	1.44	1.14	1.21	1.27	1.59	1.32
σ	0.60	0.82	0.81	1.00	1.07	0.81	0.78	0.90	0.93	0.71	0.85	0.97
λ $(\times 10^{-3})$	-254	84	-124	-293	-550	-462	24	-145	-540	-635	-132	-198
Weibull-GPD model, WePa3 $(\theta, \alpha, \tau, \lambda)$												
θ $(\times 10^{-3})$	1041	966	1187	2621	2339	1668	1690	1785	2331	2153	2545	2327
α	0.97	1.32	1.18	1.30	0.98	1.00	1.42	1.14	1.29	1.27	1.51	1.31
τ	1.47	0.99	0.94	0.65	0.58	0.95	1.16	0.96	0.89	1.27	0.89	0.75
λ $(\times 10^{-3})$	-268	68	-137	-378	-561	-486	-8	-148	-254	-607	-309	-206

3.3 Folded-t Family

The folded-t family of distributions has been introduced by Psarakis and Panaretoes (1991). These models were designed for situations in which data measurements were recorded without their algebraic signs. Although such situations rarely (if ever) occur in insurance, it was noticed by Brazauskas and Kleefeld (2011) that they may be artificially constructed by appropriately transforming the observed claim amounts. In particular, if X denotes the claim sizes and d is the corresponding deductible, then the histogram of the log-folded claims, $\log(X/d)$, can be accurately approximated by the pdf of the folded-t distribution (see Figure 1).

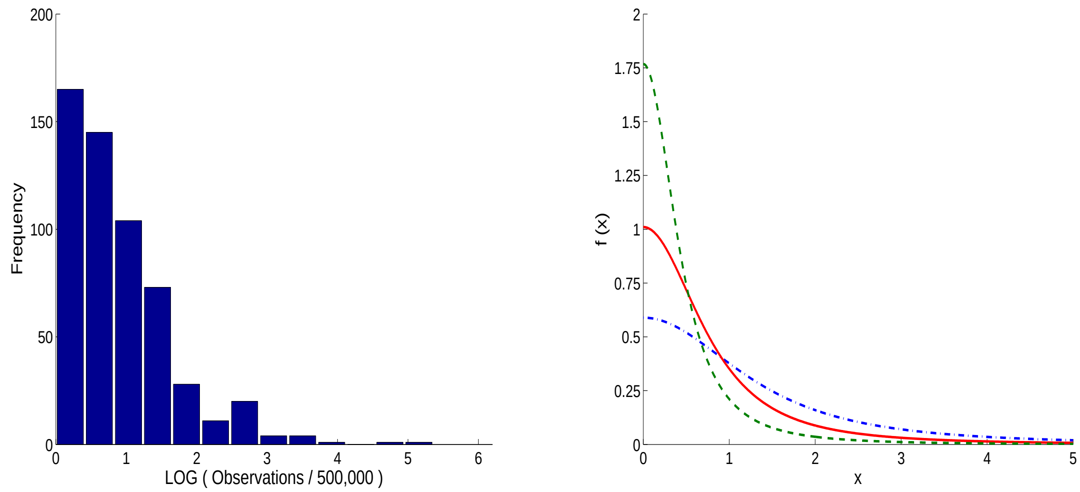

FIGURE 1: Left-hand panel: The histogram of the log-folded Norwegian fire claims for the year 1984. Right-hand panel: Shapes of the pdf of folded-t distributions for $\sigma = 0.4$ (dashed line), $\sigma = 0.7$ (solid line), $\sigma = 1.2$ (dash-dotted line) and $\nu = 2$.

Further, due to its close relation to the Student t distribution, the folded-t model has simple analytic expressions for its pdf, cdf, and qf:

$$\text{PDF:} \qquad f_{\text{FT}(\sigma,\nu)}(x) = (2/\sigma)\, f_{\text{T}(\nu)}(x/\sigma), \qquad x > 0, \tag{3.15}$$

$$\text{CDF:} \qquad F_{\text{FT}(\sigma,\nu)}(x) = 2\left[F_{\text{T}(\nu)}(x/\sigma) - 0.5\right], \qquad x > 0, \tag{3.16}$$

$$\text{QF:} \qquad F^{-1}_{\text{FT}(\sigma,\nu)}(u) = \sigma\, F^{-1}_{\text{T}(\nu)}((u+1)/2), \qquad 0 < u < 1, \tag{3.17}$$

where $f_{\text{T}(\nu)}$, $F_{\text{T}(\nu)}$, and $F^{-1}_{\text{T}(\nu)}$ denote the standard pdf, cdf, and qf, respectively, of the underlying Student t distribution with ν degrees of freedom. (Note that the standard pdf and cdf have location 0 and scale equal to 1.) Moreover, the choice $\nu = 1$ yields the folded Cauchy distribution, and the folded-t model converges to the folded normal as $\nu \to \infty$.

Let $X_1, \ldots, X_n$ denote a sample of *i.i.d.* random variables from the folded-$t\,(\sigma,\nu)$ model; all X_i's exceed $d = 500{,}000$. Then, a maximum likelihood estimator of (σ,ν) is found by numerically maximizing the left truncated (at $d = 500,000$) log-likelihood function:

$$\log \mathcal{L}_{\mathrm{FT}}(\sigma, \nu \,|\, X_1, \ldots, X_n) \;=\; \sum_{i=1}^{n} \log\left(\frac{f_{\mathrm{FT}(\sigma,\,\nu)}(X_i)}{1 - F_{\mathrm{FT}(\sigma,\,\nu)}(d)} \right)$$

$$= \; n \log\left(\frac{2\sqrt{\nu}\,\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\,\Gamma(\frac{1}{2})} \right) - \frac{\nu+1}{2} \sum_{i=1}^{n} \log\left(\sigma^2\,\nu + X_i^2\right) - n \log\left(1 - F_{\mathrm{FT}(\sigma,\,\nu)}(d)\right), \qquad (3.18)$$

where expression (3.18) is derived by taking into account (3.15) and the formula of $f_{\mathrm{T}(\nu)}$. For the data sets under consideration, the following MLE estimates resulted (see Table 4).

TABLE 4: Parameter MLEs of the folded-t model for the *Norwegian Fire Claims* (1981–1992) data.

Folded-t parameters	Norwegian Fire Claims for Year 1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992
σ ($\times 10^{-3}$)	266.2	341.1	324.8	565.2	400.0	411.4	722.0	630.9	728.0	803.5	744.4	626.7
ν	1.19	1.32	1.33	1.52	1.26	1.35	1.60	1.29	1.50	1.88	1.82	1.48

Similar to the previously studied models, the folded-t model also suggests that the Norwegian fire claims (1981–1992) data are heavy tailed. For $\nu \leq 2$, variance and higher order moments for the FT model are infinite, and as one can see from Table 4, the estimates of the tail parameter ν range from 1.19 in 1981 to 1.88 in 1990.

Of course, our conclusion about the heavy-tailed nature of Norwegian fire claims is an outcome of the assumed models. Although our *informal* prior knowledge hints at the appropriateness of the GPD, composite Pareto, and the folded-t models for the data sets at hand, that has to be augmented with a formal analysis. This leads us to the topic of the next section, i.e., model validation.

4 Model Validation

To evaluate the quality of fits, we first employ quantile-quantile plots, then two goodness-of-fit statistics (Kolmogorov-Smirnov, KS, and Anderson-Darling, AD), and finally two information-based criteria (Akaike information criterion, AIC, and Bayesian information criterion, BIC).

4.1 Quantile-Quantile Plots

To get a feel for how the fitted models "work" on the given data, we will start with basic visualization, namely, quantile-quantile plots. In Figure 2, we present plots of the fitted-versus-observed quantiles for the six models of Section 3. In order to avoid visual distortions due to large spacings between the most extreme observations, both axes in all the plots are measured on the logarithmic scale. That is, the points plotted in those graphs are the following pairs:

$$\left(\log\left(\widehat{F}^{-1}\big[u_{i_k} + \widehat{F}(d)(1-u_{i_k})\big]\right),\ \log\left(x_{(i_k)}\right)\right), \quad i_k = 1,\ldots,n_k,$$

where $\widehat{F}(d)$ is the estimated parametric cdf evaluated at $d = 500,000$, $\widehat{F}^{-1}$ is the estimated parametric qf, $x_{(1)} < \cdots < x_{(n_k)}$ denote the ordered claim severities, $u_{i_k} = (i_k - 0.5)/n_k$ is the quantile level, and n_k, for $k = 1,\ldots,12$, represent the sample sizes. Note that each data set has a relatively small number of observations that are exactly equal to the priority $d = 500,000$. For the purposes of parameter estimation, construction of quantile-quantile plots, and for computation of other model validation measures, such data clusters were de-grouped using the method described in Brazauskas and Serfling (2003). Also, the cdf and qf functions were evaluated using the MLE estimates from Tables 2–4.

As one can see from Figure 2, all the models do a reasonably good job for the 12 data sets. That is, most of the points in the plots do not deviate from the 45° line, but there are a few notable exceptions. For example, in years 1982 and 1987 (and to a lesser degree in year 1985), the 45° lines show a downward bending at the extreme right tail, which implies that all the models *overestimate* the observed extreme quantiles. This should be viewed as a good thing – the models chosen for the Norwegian fire claims data are conservative (i.e., they are capable of capturing heavier right tails than that of the observed losses). Also, since both axes in all the plots have the same range, one can immediately see for which years claims had very long tails (e.g., 1988) and for which shorter ones (e.g., 1982, 1983, 1987, 1990, 1991).

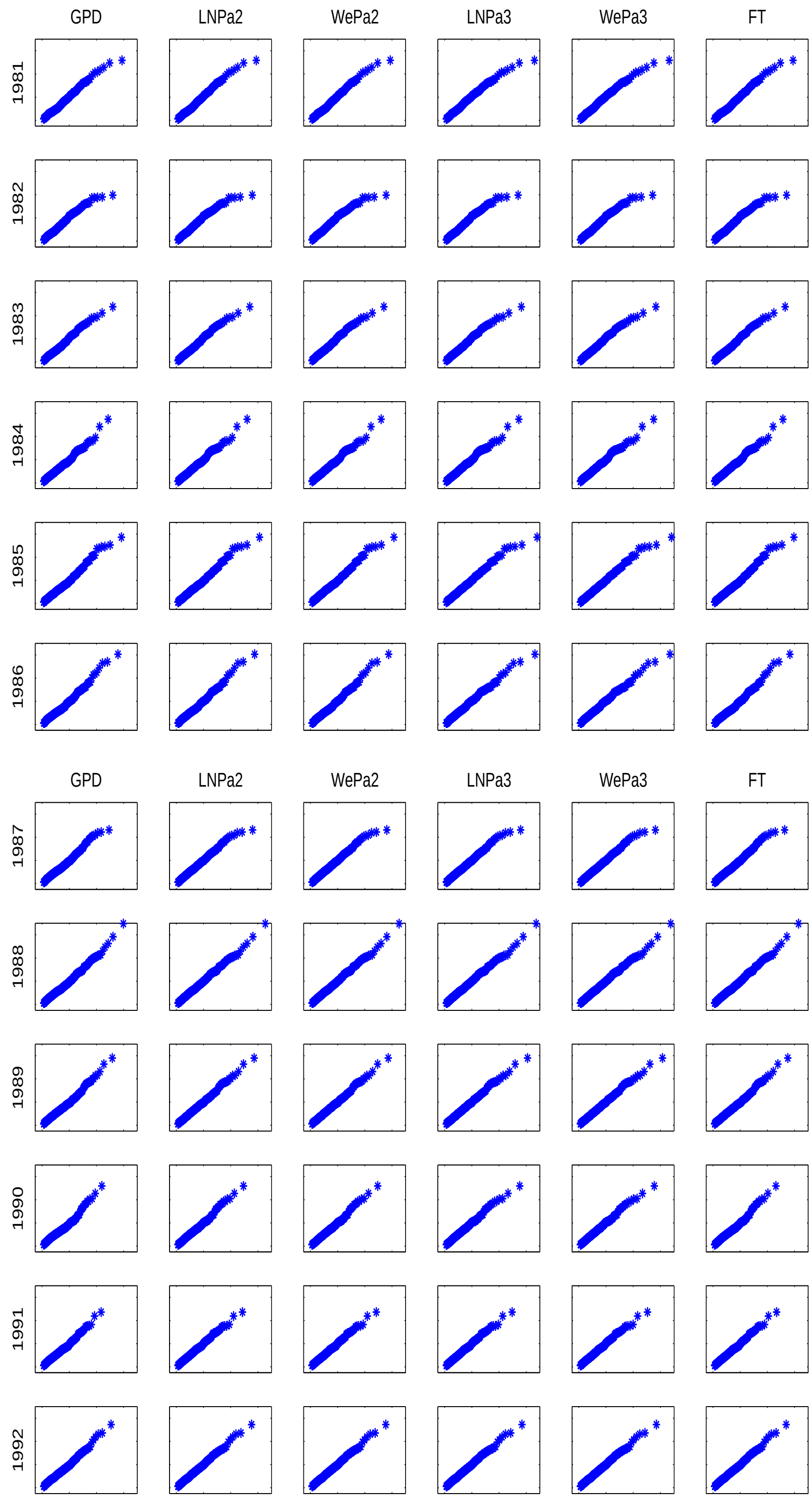

FIGURE 2: Fitted-versus-observed *log*-claims for the years 1981–1992. In all 72 plots, the horizontal and vertical axes range from 12.5 to 20 (with ticks at 13, 15, 17, 19).

4.2 Goodness-of-Fit Statistics

To formally assess the "closeness" of the fitted model to data, we will measure the distance (according to a selected measure) between the empirical distribution function $\widehat{F}_n(x) = n^{-1} \sum_{i=1}^n \mathbf{1}\{x_i \leq x\}$ and the parametrically estimated left-truncated (at $d = 500,000$) distribution function

$$\widehat{F}^*(x) = \frac{\widehat{F}(x) - \widehat{F}(d)}{1 - \widehat{F}(d)}, \qquad x \geq d.$$

There are multiple options available to accomplish this task, but in this paper we choose to work with two popular distances: (i) *maximum absolute* distance, and (ii) *cumulative weighted quadratic* distance, with more weight on the tails. The first measure leads to the well-known Kolmogorov-Smirnov statistic, which for the left-truncated data problem is defined as

$$D = \max_{x \geq d} \left| \widehat{F}_n(x) - \widehat{F}^*(x) \right|.$$

Note that for the computational purposes, the following formula is more convenient:

$$D_n = \max_{1 \leq j \leq n} \left\{ \left| \widehat{F}^*(x_{(j)}) - \frac{j-1}{n} \right|, \left| \widehat{F}^*(x_{(j)}) - \frac{j}{n} \right| \right\},$$

where $x_{(1)} < \cdots < x_{(n)}$ denote the ordered claim severities.

And the second measure leads to the Anderson-Darling statistic, which for the left-truncated data problem is defined as

$$A^2 = n \int_d^\infty \frac{\big(\widehat{F}_n(x) - \widehat{F}^*(x)\big)^2}{\widehat{F}^*(x)\big(1 - \widehat{F}^*(x)\big)} \, \mathrm{d}\widehat{F}^*(x).$$

For the computational purposes, the following formula is more convenient:

$$\begin{aligned} A_n^2 &= -n + n \sum_{j=1}^{n} \widehat{F}_n^2(x_{(j)}) \log \left(\frac{\widehat{F}^*(x_{(j+1)})}{\widehat{F}^*(x_{(j)})} \right) - n \sum_{j=0}^{n-1} \big(1 - \widehat{F}_n(x_{(j)})\big)^2 \log \left(\frac{1 - \widehat{F}^*(x_{(j+1)})}{1 - \widehat{F}^*(x_{(j)})} \right) \\ &= -n + n \sum_{j=1}^{n} \big(j/n\big)^2 \log \left(\frac{\widehat{F}^*(x_{(j+1)})}{\widehat{F}^*(x_{(j)})} \right) - n \sum_{j=0}^{n-1} \big(1 - j/n\big)^2 \log \left(\frac{1 - \widehat{F}^*(x_{(j+1)})}{1 - \widehat{F}^*(x_{(j)})} \right), \end{aligned}$$

where $d = x_{(0)} < x_{(1)} < \cdots < x_{(n)} < x_{(n+1)} = \infty$ denote the ordered claim severities.

Table 5 provides the goodness-of-fit measures of the six fitted models for the years 1981–1992. We first note that, for a fixed data set, comparison of the KS and AD values across the models confirms the well-known fact that models with more parameters adapt to data better, i.e., they have smaller values of the goodness-of-fit statistics. For instance, in 1988 the AD values for the 2-parameter models are 1.93 (for GPD) and 1.41 (for FT), but they range between 1.21 (for WePa3) and 1.28 (for LNPa2) for the 3- and 4-parameter models. On the other hand, such comparisons are not very useful as those statistics in general do not follow the same probability distribution when applied to different models. Therefore, we should examine the p-values (reported in Table 5), which we computed using parametric bootstrap with 1000 simulation runs. For a brief description of the parametric bootstrap procedure, see, for example, Klugman, Panjer, Willmot (2012, Section 20.4.5).

TABLE 5: Goodness-of-fit measures of the fitted models for the *Norwegian Fire Claims* (1981–1992) data.

Fitted	Norwegian Fire Claims for Year											
Model	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992
Kolmogorov-Smirnov statistic (p-value*)												
GPD	0.05	0.05	0.04	0.03	0.04	0.05	0.04	0.03	0.03	0.05	0.03	0.03
	(0.01)	(0.02)	(0.13)	(0.12)	(0.01)	(0.00)	(0.01)	(0.01)	(0.10)	(0.00)	(0.04)	(0.10)
LNPa2	0.04	0.04	0.03	0.02	0.02	0.04	0.02	0.02	0.02	0.02	0.02	0.02
	(0.03)	(0.07)	(0.44)	(0.43)	(0.27)	(0.01)	(0.23)	(0.29)	(0.76)	(0.25)	(0.49)	(0.47)
WePa2	0.04	0.04	0.03	0.03	0.02	0.04	0.02	0.02	0.02	0.02	0.02	0.02
	(0.03)	(0.08)	(0.52)	(0.26)	(0.29)	(0.01)	(0.27)	(0.31)	(0.67)	(0.35)	(0.53)	(0.33)
LNPa3	0.04	0.04	0.03	0.02	0.03	0.04	0.02	0.02	0.02	0.02	0.02	0.02
	(0.02)	(0.02)	(0.29)	(0.48)	(0.12)	(0.00)	(0.17)	(0.23)	(0.78)	(0.47)	(0.50)	(0.40)
WePa3	0.04	0.04	0.03	0.02	0.03	0.04	0.02	0.02	0.02	0.02	0.02	0.02
	(0.02)	(0.02)	(0.35)	(0.30)	(0.08)	(0.00)	(0.18)	(0.21)	(0.58)	(0.79)	(0.56)	(0.26)
FT	0.05	0.04	0.03	0.02	0.03	0.04	0.02	0.03	0.02	0.04	0.03	0.02
	(0.01)	(0.03)	(0.25)	(0.65)	(0.13)	(0.01)	(0.32)	(0.07)	(0.77)	(0.02)	(0.36)	(0.62)
Anderson-Darling statistic (p-value*)												
GPD	0.87	1.50	1.06	0.53	1.16	5.35	2.34	1.93	1.01	2.41	1.09	1.26
	(0.01)	(0.00)	(0.00)	(0.03)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
LNPa2	0.61	1.43	0.88	0.28	0.75	4.73	1.35	1.28	0.27	0.29	0.56	0.92
	(0.01)	(0.00)	(0.00)	(0.05)	(0.00)	(0.00)	(0.00)	(0.00)	(0.06)	(0.03)	(0.01)	(0.00)
WePa2	0.62	1.43	0.87	0.31	0.75	4.69	1.27	1.22	0.27	0.19	0.56	0.91
	(0.01)	(0.00)	(0.00)	(0.04)	(0.00)	(0.00)	(0.00)	(0.00)	(0.05)	(0.07)	(0.01)	(0.00)
LNPa3	0.36	1.44	0.84	0.27	0.68	4.57	1.35	1.26	0.25	0.22	0.56	0.92
	(0.01)	(0.00)	(0.00)	(0.04)	(0.00)	(0.00)	(0.00)	(0.00)	(0.05)	(0.04)	(0.01)	(0.00)
WePa3	0.35	1.44	0.83	0.29	0.68	4.50	1.27	1.21	0.26	0.10	0.55	0.91
	(0.01)	(0.00)	(0.00)	(0.04)	(0.00)	(0.00)	(0.00)	(0.00)	(0.04)	(0.11)	(0.01)	(0.00)
FT	0.76	1.46	0.96	0.29	0.88	4.96	1.53	1.41	0.35	0.83	0.59	0.96
	(0.01)	(0.00)	(0.01)	(0.10)	(0.00)	(0.00)	(0.00)	(0.00)	(0.07)	(0.01)	(0.02)	(0.00)

* The p-values are reported in parentheses; they are computed using parametric bootstrap with 1000 simulation runs.

Several conclusions emerge from Table 5. First, the p-values are higher for the KS statistic than the corresponding p-values for the AD statistic. (Note that a higher p-value means a closer fit.) This is not surprising because usually the AD test is more powerful than the KS test. Second, while according to the KS criterion most models are acceptable for most of the data sets, one would reach the opposite conclusion if the AD criterion were used. This is also not surprising and could be explained by the fact that the data sets we analyzed are fairly large (see Table 1), in conjunction with the AD test being more powerful. Third, performance of the GPD model, which is a 2-parameter model, is not competitive when compared to the other five models. Fourth, it might be surprising but the most flexible 4-parameter models, LNPa3 and WePa3, are not dominating the competition – the top performers are the 3-parameter models, LNPa2 and WePa2. Their advantage, however, is quite minimal. Indeed, according to the KS test, LNPa2 is ranked *first* or *second* six times, WePa2 seven times, FT four times, LNPa3 three times, and WePa3 two times. Plus, in 1986, there is a three-way tie for the first spot between LNPa2, WePa2, and FT. If one uses the AD criterion, then for half of the data sets all six models are tied (and strongly rejected). However, when there are more or less meaningful differences among the models (1984, 1989, 1990), LNPa2 is ranked second twice, WePa2 is second once, WePa3 is first once, and FT is ranked first twice.

4.3 Information Criteria

One way for selecting the best model among those under consideration is to evaluate their likelihood functions and choose the model with the highest value of the likelihood. This approach, however, is flawed because models with more parameters tend to have a higher likelihood value, all other things being equal. Therefore, in such situations information-based decision rules such as the AIC and BIC, which extract a penalty for introducing additional parameters, come in handy.

The Akaike Information Criterion, AIC, penalizes the log-likelihood function with the number of parameters and is defined as follows:

$$\mathrm{AIC} = 2\,\mathrm{NLL} + 2\dim(\boldsymbol{\theta}),$$

where NLL stands for 'negative log-likelihood' and $\dim(\boldsymbol{\theta})$ denotes the dimension of $\boldsymbol{\theta}$, i.e., the number of components of the vector parameter $\boldsymbol{\theta}$. Models with smaller values of AIC are preferred.

The Bayesian Information Criterion, BIC, penalizes the log-likelihood function with the number of parameters times the logarithm of the sample size n:

$$\mathrm{BIC} = 2\,\mathrm{NLL} + \dim(\boldsymbol{\theta})\log(n),$$

where NLL and $\dim(\boldsymbol{\theta})$ have the same meaning as in the AIC definition, and models with smaller values of BIC are preferred.

TABLE 6: Information measures of the fitted models for the *Norwegian Fire Claims* (1981–1992) data.

Fitted	Norwegian Fire Claims for Year											
Model	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992
Negative Log-Likelihood, NLL												
GPD	3,439	3,393	3,214	4,457	4,891	5,160	6,232	6,849	5,886	5,066	5,008	4,985
LNPa2	3,437	3,393	3,213	4,456	4,889	5,157	6,228	6,845	5,882	5,057	5,005	4,983
WePa2	3,437	3,393	3,213	4,456	4,889	5,157	6,227	6,844	5,882	5,056	5,005	4,983
LNPa3	3,436	3,393	3,212	4,456	4,888	5,155	6,228	6,845	5,882	5,056	5,005	4,983
WePa3	3,436	3,393	3,212	4,456	4,888	5,155	6,227	6,844	5,882	5,055	5,005	4,983
FT	3,438	3,394	3,213	4,456	4,890	5,159	6,229	6,846	5,883	5,060	5,006	4,983
Akaike Information Criterion, AIC												
GPD	6,881	6,791	6,431	8,919	9,786	10,325	12,469	13,701	11,776	10,137	10,020	9,973
LNPa2	6,880	6,793	6,431	8,918	9,784	10,320	12,462	13,696	11,770	10,119	10,017	9,972
WePa2	6,880	6,792	6,431	8,918	9,784	10,320	12,460	13,695	11,770	10,118	10,016	9,971
LNPa3	6,880	6,794	6,433	8,920	9,785	10,318	12,464	13,698	11,772	10,120	10,019	9,973
WePa3	6,879	6,794	6,433	8,920	9,785	10,317	12,462	13,696	11,772	10,118	10,018	9,973
FT	6,881	6,791	6,430	8,916	9,784	10,321	12,462	13,696	11,770	10,124	10,016	9,970
Bayesian Information Criterion, BIC												
GPD	6,889	6,799	6,439	8,927	9,795	10,334	12,478	13,711	11,785	10,146	10,029	9,982
LNPa2	6,892	6,805	6,443	8,931	9,798	10,333	12,476	13,710	11,784	10,133	10,030	9,985
WePa2	6,892	6,804	6,443	8,931	9,798	10,333	12,474	13,709	11,784	10,132	10,030	9,985
LNPa3	6,896	6,810	6,449	8,937	9,802	10,336	12,482	13,716	11,790	10,138	10,036	9,991
WePa3	6,895	6,810	6,449	8,937	9,802	10,335	12,481	13,715	11,790	10,136	10,036	9,991
FT	6,889	6,799	6,438	8,925	9,792	10,330	12,471	13,706	11,779	10,133	10,024	9,979

Using the log-likelihood functions and parameter estimates from Section 3, we evaluated the information-based measures of the fitted models for all the data sets under consideration. The results are summarized in Table 6, where several patterns can be observed. First, under the NLL criterion, the 4-parameter models perform better than the other models for most of the data sets.

But when the AIC criterion is used the 2- and 3-parameter models start to appear more competitive. Finally, when the BIC criterion is employed, the 2-parameter models take over the leading positions, with the FT model exhibiting extremely strong performance: 9 times absolute first, 2 times tied for first with GPD, and 1 time tied for second with LNPa2.

Clearly, the BIC criterion favors more parsimonious models than the AIC does. To make a choice between the two criteria is not a trivial exercise, but the following quote, taken from Claeskens (2004), may shed some light about the BIC's properties: "If the true data generating model belongs to the finite parameter family of models under investigation, the Bayesian information criterion consistently selects the correct model ... If this assumption does not hold, models selected by the Bayesian information criterion will tend to underfit, that is, will use too few parameters."

5 Measuring Tail Risk

In this section we focus on measuring the tail risk of 'ground up' Norwegian fire claims for the years 1981 through 1992. In Section 5.1, point estimates of the value-at-risk and tail-conditional median measures are provided. In Section 5.2, the probability of a few upper tail events is calculated and compared to the actual next-year probabilities.

5.1 Risk over Time

Point estimates for the value-at-risk at the 90% confidence level,

$$\mathrm{VaR}_{0.90}[X] = F_X^{-1}(0.90),$$

are presented in Table 7. In addition, we also provide estimates of the 90% tail-conditional median:

$$\mathrm{TCM}_{0.90}[X] = \mathrm{median}\big[X \,\big|\, X > \mathrm{VaR}_{0.90}[X]\big] = F_X^{-1}(0.95).$$

The estimates of these measures are obtained using the six fitted models of Section 3. That is, they are computed using the following formulas of $F_X^{-1}(u)$ and the MLE values: for GPD, equation (3.4) and Table 2; for composite models, equation (3.8) and Table 3; for FT, equation (3.17) and Table 4.

The TCM risk measure helps to address the traditional 'what-if' question: If the VaR event happens, what is the expected loss? The answer is that in case such an unlikely event occurs, half of the time the loss is expected to be between the VaR and TCM values and the remaining 50% of the time to exceed the TCM value. Notice that for the Norwegian fire claims data, the TCM is a better suited risk measure than the well-established tail-conditional expectation, $\mathbf{E}\big[X \,\big|\, X > \mathrm{VaR}_\alpha[X]\big]$, as for some of the fitted models the latter is sometimes infinite; hence, non-informative.

Further, in our examples the chosen time horizon, 1 year, is in line with a typical (re)insurance company's business cycle and consistent with Solvency Capital Requirement (see European Insurance and Occupational Pensions Authority, 2014, pp. 6–7). However, the confidence level we use (90%) is more appropriate for the model-backtesting purposes, not capital reserving. Finally, we note in passing that the underlying assumptions in Solvency Capital Requirement specify a confidence level at 99.5%. Such a high level is interpreted as a 1-in-200 year event, or for some risks it may be related to historical credit-default probabilities (used by the leading agencies in issuing companies' credit ratings).

TABLE 7: Value-at-Risk, $\mathrm{VaR}_{0.90}[X]$, and Tail-Conditional Median, $\mathrm{TCM}_{0.90}[X]$ (in parentheses), for the *Norwegian Fire Claims* (1981–1992) data. The risk estimates, measured in millions of Norwegian krones, are based on the six fitted models of Section 3, using a 1-year time horizon.

Fitted	Norwegian Fire Claims for Year											
Model	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992
GPD	0.49	0.92	0.74	1.62	1.06	1.01	2.15	2.16	2.28	2.17	2.05	1.91
	(0.94)	(1.63)	(1.33)	(2.69)	(1.94)	(1.81)	(3.47)	(3.83)	(3.77)	(3.36)	(3.17)	(3.18)
LNPa2	2.42	2.16	1.87	2.12	2.33	2.09	2.93	3.20	3.00	2.83	2.57	2.52
	(4.33)	(3.76)	(3.20)	(3.39)	(4.14)	(3.56)	(4.75)	(5.68)	(4.93)	(4.33)	(3.89)	(4.13)
WePa2	1.94	1.66	1.43	1.43	1.88	1.60	2.55	2.59	2.44	2.54	2.10	1.94
	(3.47)	(2.89)	(2.45)	(2.32)	(3.34)	(2.72)	(4.14)	(4.59)	(4.02)	(3.90)	(3.17)	(3.20)
LNPa3	2.65	1.98	1.92	2.15	1.87	2.17	2.93	3.16	2.89	2.75	2.57	2.50
	(5.11)	(3.39)	(3.33)	(3.43)	(3.23)	(3.83)	(4.75)	(5.67)	(4.71)	(4.28)	(3.89)	(4.09)
WePa3	2.20	1.57	1.45	1.47	1.12	1.68	2.54	2.59	2.25	2.43	2.09	1.88
	(4.21)	(2.70)	(2.49)	(2.36)	(1.92)	(2.87)	(4.14)	(4.64)	(3.57)	(3.76)	(3.14)	(3.06)
FT	1.29	1.45	1.37	2.07	1.79	1.70	2.52	2.77	2.69	2.45	2.33	2.35
	(2.33)	(2.50)	(2.33)	(3.34)	(3.15)	(2.90)	(3.98)	(4.81)	(4.36)	(3.68)	(3.53)	(3.84)

Several patterns can be noticed by examining the VaR and TCM values in Table 7. First, for a fixed model, the riskiness of Norwegian fire claims does not exhibit any obvious trend, neither upward nor downward. Second, for a fixed year, the models yield wide ranging estimates of risk. Surely, there are years when the estimates are fairly close to each other (e.g., in 1991, $\mathrm{VaR}_{0.90}[X]$ ranges from 2.05 for GPD to 2.57 for LNPa2 and LNPa3), but most of the time they are quite different, in some cases substantially (e.g., in 1981, $\mathrm{VaR}_{0.90}[X]$ ranges from 0.49 for GPD to 2.65 for LNPa3). As expected, these ranges are even wider for $\mathrm{TCM}_{0.90}[X]$ which is a more extreme quantile than $\mathrm{VaR}_{0.90}[X]$. Third, the GPD-based risk estimates are usually the lowest and those based on LNPa2 and LNPa3 are the highest, with the other models producing risk estimates in the middle between the two extremes. Fourth, according to the GPD model, in 1981 there was observed less than 10% of claims (because $\mathrm{VaR}_{0.90}[X] = 0.49$ is below the priority point of 0.50 millions).

5.2 Risk Prediction

To see how successful the six probability models are at predicting next-year risk, we evaluate the following upper tail probabilities:

$$\mathbf{P}\Big\{X > \widehat{\mathrm{VaR}}_\alpha[X] \,\Big|\, X > 500,000\Big\} = \frac{1-\alpha}{\mathbf{P}\big\{X > 500,000\big\}}, \qquad \widehat{\mathrm{VaR}}_\alpha[X] > 500,000,$$

for $\alpha = 0.90,\ 0.95$. (If $\widehat{\mathrm{VaR}}_\alpha[X] \leq 500,000$, the conditional probability is 1.) The next-year predictions of the tail probability $\mathbf{P}\big\{X > 500,000\big\}$ are obtained using the six fitted models of Section 3. That is, they are computed using the following formulas of $F_X(x)$ and the MLE values: for GPD, equation (3.1) and Table 2; for composite models, equation (3.7) and Table 3; for FT, equation (3.16) and Table 4. These computations are based on year t data and then compared with the actual probability for year $t+1$ ($t = 1981, \ldots, 1991$). Note that the actual probability, $\mathbf{P}\big\{X > \widehat{\mathrm{VaR}}_\alpha[X] \,\big|\, X > 500,000\big\}$, can be evaluated directly from the data (no model fitting is needed) using the empirical approach:

$$\widehat{\mathbf{P}}\,\Big\{X > \widehat{\mathrm{VaR}}_\alpha[X] \,\Big|\, X > 500,000\Big\} = \frac{1}{n_{t+1}} \sum_{i=1}^{n_{t+1}} \mathbf{1}\big\{X_i > \widehat{\mathrm{VaR}}_\alpha[X]\big\},$$

where n_{t+1} denotes the sample size for year $t+1$ ($t = 1981, \ldots, 1991$) and the value-at-risk estimates for year t, $\widehat{\mathrm{VaR}}_\alpha[X]$, are available in Table 7. However, since $\widehat{\mathrm{VaR}}_\alpha[X]$ values are model dependent, each distribution has its own 'actual' probability.

The results of the computations described above are summarized in Table 8. Comparison of the predicted and actual probabilities reveals several patterns. First, for a fixed year, the models yield wide ranging predictions of the tail probabilities. The predictions are more spread out for the level $\alpha = 0.90$ than for $\alpha = 0.95$. Also, those differences are more pronounced in the earlier years than in the later ones. Second, the tail risk predictions based on the GPD model are typically quite remote from those of the other five models, which yield (more or less) similar predictions. Third, for each model the next-year predictions are fairly accurate for all the years. For example, if we take the relative prediction error, RPE $= \left(\frac{\text{predicted}}{\text{actual}} - 1\right) 100\%$, as a measure of accuracy, then the medians of the *absolute* values of RPE are 7% (GPD), 13% (LNPa2), 12% (WePa2), 15% (LNPa3), 8% (WePa3), and 9% (FT) for $\alpha = 0.90$. Likewise, they are 14% (GPD), 11% (LNPa2), 8% (WePa2), 11% (LNPa3), 11% (WePa3), and 10% (FT) for $\alpha = 0.95$. To get a sense of how volatile RPE's are, we will look at the first and third quartile (i.e., the middle 50%) of the absolute values of RPE. For $\alpha = 0.90$, they are: 3% and 16% (GPD), 4% and 16% (LNPa2), 6% and 13% (WePa2), 8% and 16% (LNPa3), 4% and 14% (WePa3), 8% and 14% (FT). And for $\alpha = 0.95$: 6% and 25% (GPD), 4% and 27% (LNPa2), 3% and 22% (WePa2), 3% and 27% (LNPa3), 4% and 23% (WePa3), 7% and 19% (FT). In summary, all six models are reasonably successful at predicting next-year's tail probabilities.

TABLE 8: Predicted and actual (in parentheses) upper tail probabilities, $\mathbf{P}\big\{X > \widehat{\mathrm{VaR}}_\alpha[X] \,\big|\, X > 500{,}000\big\}$, for the *Norwegian Fire Claims* (1982–1992) data.

Confidence Level	Fitted Model	Norwegian Fire Claims for Year										
		1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992
$\alpha = 0.90$	GPD	1.00	0.52	0.65	0.29	0.47	0.48	0.23	0.25	0.22	0.21	0.22
		(1.00)	(0.51)	(0.72)	(0.27)	(0.46)	(0.57)	(0.23)	(0.22)	(0.16)	(0.18)	(0.22)
	LNPa2	0.17	0.18	0.21	0.19	0.18	0.19	0.13	0.15	0.14	0.12	0.14
		(0.17)	(0.17)	(0.23)	(0.19)	(0.15)	(0.22)	(0.16)	(0.12)	(0.11)	(0.11)	(0.16)
	WePa2	0.22	0.24	0.29	0.34	0.23	0.27	0.16	0.19	0.19	0.15	0.20
		(0.21)	(0.22)	(0.33)	(0.33)	(0.20)	(0.31)	(0.19)	(0.17)	(0.15)	(0.13)	(0.21)
	LNPa3	0.15	0.20	0.19	0.19	0.23	0.17	0.13	0.15	0.15	0.12	0.14
		(0.14)	(0.19)	(0.23)	(0.19)	(0.20)	(0.20)	(0.16)	(0.12)	(0.11)	(0.11)	(0.16)
	WePa3	0.18	0.27	0.28	0.32	0.44	0.23	0.16	0.19	0.21	0.15	0.20
		(0.17)	(0.25)	(0.33)	(0.32)	(0.42)	(0.28)	(0.19)	(0.17)	(0.16)	(0.15)	(0.21)
	FT	0.36	0.30	0.31	0.20	0.25	0.25	0.17	0.18	0.17	0.17	0.17
		(0.33)	(0.28)	(0.35)	(0.20)	(0.21)	(0.28)	(0.19)	(0.16)	(0.12)	(0.15)	(0.18)
$\alpha = 0.95$	GPD	0.51	0.26	0.33	0.14	0.23	0.24	0.11	0.13	0.11	0.11	0.11
		(0.49)	(0.23)	(0.37)	(0.14)	(0.18)	(0.26)	(0.13)	(0.10)	(0.07)	(0.09)	(0.11)
	LNPa2	0.08	0.09	0.10	0.10	0.09	0.09	0.07	0.07	0.07	0.06	0.07
		(0.09)	(0.09)	(0.10)	(0.11)	(0.08)	(0.10)	(0.09)	(0.05)	(0.04)	(0.06)	(0.09)
	WePa2	0.11	0.12	0.14	0.17	0.12	0.13	0.08	0.10	0.09	0.07	0.10
		(0.11)	(0.12)	(0.14)	(0.17)	(0.10)	(0.15)	(0.11)	(0.07)	(0.06)	(0.08)	(0.11)
	LNPa3	0.07	0.10	0.10	0.09	0.11	0.08	0.07	0.07	0.07	0.06	0.07
		(0.07)	(0.10)	(0.09)	(0.11)	(0.11)	(0.09)	(0.09)	(0.05)	(0.04)	(0.06)	(0.09)
	WePa3	0.09	0.13	0.14	0.16	0.22	0.12	0.08	0.09	0.11	0.07	0.10
		(0.09)	(0.13)	(0.14)	(0.16)	(0.19)	(0.14)	(0.11)	(0.07)	(0.08)	(0.08)	(0.11)
	FT	0.18	0.15	0.16	0.10	0.13	0.13	0.09	0.09	0.09	0.08	0.09
		(0.17)	(0.14)	(0.17)	(0.11)	(0.11)	(0.14)	(0.11)	(0.07)	(0.05)	(0.08)	(0.10)

6 Concluding Remarks

In this paper, six parametric families have been employed to model severity and measure tail risk of Norwegian fire claims for the years 1981 through 1992. Among the probability distributions we have used, two are 2-parameter (generalized Pareto, GPD, and folded-t, FT), two are 3-parameter (composite lognormal, LNPa2, and composite Weibull, WePa2; both with Pareto I upper tail), and two are 4-parameter models (composite lognormal, LNPa3, and composite Weibull, WePa3; both with GPD upper tail). Overall, all six models have done a good job. However, as a formal model-validation analysis has demonstrated, the most flexible 4-parameter models, LNPa3 and WePa3, do not consistently yield a *statistically* closer fit when compared to that of the simpler models. Similar conclusions have also emerged from the analysis of risk predictions. That is, minor advantages in terms of model fits have not produced more accurate predictions by the corresponding models.

Further, it is certainly tempting to conclude that simpler distributions, such as GPD and FT, are preferred for the task of measuring tail risk of the Norwegian fire claims (1981–1992) data. However, one has to be careful here. A primary reason for caution is that the six fitted models, even after undergoing extensive statistical validation (e.g., quantile-quantile plots, goodness-of-fit tests, information criteria), have lead to substantially different risk evaluations, which in turn would produce wide ranging estimates of reserves. Hence, a more appropriate conclusion should be that formal statistical analysis is a necessary but not sufficient condition for measuring and pricing tail risk. Model uncertainty is difficult to eliminate, but using multiple models for sensitivity checks is reassuring.

Finally, the risk prediction exercise considered in this paper can be extended into a number of directions. For instance, one could use all available data (i.e., for all previous years) to make a next-year risk prediction. This approach has various choices too. For example: (a) data for all previous years is combined into one data set that is used to build a model (this effectively makes the *i.i.d.* assumption); (b) data from year-to-year is not *i.i.d.* but stationary; (c) data from year-to-year is dependent with numerous options to model dependency. An alternative way to make next-year risk predictions would be to bypass the model-fitting step and go directly into combining all available annual estimates (e.g., make a regression-type prediction using tail probability estimates from earlier years). Yet another way to extend and refine the models would be to construct confidence intervals and track how often they cover future VaR or tail probability values. Such extensions are beyond the scope of the present paper, but certainly interesting venues for future research.

Acknowledgments

The authors are very appreciative of useful comments and valuable insights provided by two anonymous referees that helped to improve the paper. Also, the first author gratefully acknowledges the support provided by a grant from The Actuarial Foundation.

References

[1] Ahn, S., Kim, J.H.T., and Ramaswami, V. (2012). A new class of models for heavy tailed distributions in finance and insurance risk. *Insurance: Mathematics and Economics*, **51**(1), 43–52.

[2] Albrecht, P. (2004). Risk measures. In *Encyclopedia of Actuarial Science* (B. Sundt and J. Teugels, eds.), volume **3**, 1493–1501; Wiley, London.

[3] Brazauskas, V. and Kaiser, T. (2004). Discussion of "Empirical estimation of risk measures and related quantities" by Jones and Zitikis. *North American Actuarial Journal*, **8**(3), 114–117.

[4] Brazauskas, V. and Kleefeld, A. (2009). Robust and efficient fitting of the generalized Pareto distribution with actuarial applications in view. *Insurance: Mathematics and Economics*, **45**(3), 424–435.

[5] Brazauskas, V. and Kleefeld, A. (2011). Folded and log-folded-t distributions as models for insurance loss data. *Scandinavian Actuarial Journal*, **2011**(1), 59–79.

[6] Brazauskas, V. and Kleefeld, A. (2014). Authors' reply to "Letter to the Editor: Regarding folded models and the paper by Brazauskas and Kleefeld (2011)" by Scollnik. *Scandinavian Actuarial Journal*, **2014**(8), 753–757.

[7] Brazauskas, V. and Serfling, R. (2003). Favorable estimators for fitting Pareto models: A study using goodness-of-fit measures with actual data. *ASTIN Bulletin*, **33**(2), 365–381.

[8] Claeskens, G. (2004). Information criteria. In *Encyclopedia of Actuarial Science* (B. Sundt and J. Teugels, eds.), volume **2**, 879–882; Wiley, London.

[9] Cooray, K. (2009). The Weibull-Pareto composite family with applications to the analysis of unimodal failure rate data. *Communications in Statistics: Theory and Methods*, **38**, 1901–1915.

[10] Cooray, K. and Ananda, M.A. (2005). Modeling actuarial data with a composite lognormal-Pareto model. *Scandinavian Actuarial Journal*, **2005**(5), 321–334.

[11] Dupuis, D. J. (1998). Exceedances over high thresholds: a guide to threshold selection. *Extremes*, **1**(3), 251–261.

[12] Eling, M. (2012). Fitting insurance claims to skewed distributions: Are the skew-normal and skew-student good models? *Insurance: Mathematics and Economics*, **51**, 239–248.

[13] European Insurance and Occupational Pensions Authority (2014). *The underlying assumptions in the standard formula for the Solvency Capital Requirement calculation*. EIOPA-14-322, Frankfurt, Germany.

[14] Jones, B.L. and Zitikis, R. (2003). Empirical estimation of risk measures and related quantities. *North American Actuarial Journal*, **7**(4), 44–54.

[15] Juárez, S.F. and Schucany, W.R. (2004). Robust and efficient estimation for the generalized Pareto distribution. *Extremes*, **7**(3), 237–251.

[16] Kaiser, T. and Brazauskas, V. (2006). Interval estimation of actuarial risk measures. *North American Actuarial Journal*, **10**(4), 249–268.

[17] Klugman, S.A., Panjer, H.H., and Willmot, G.E. (2012). *Loss Models: From Data to Decisions*, 4th edition. Wiley, New York.

[18] McNeil, A.J. (1997). Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin*, **27**, 117–137.

[19] McNeil A.J., Frey R., and Embrechts P. (2005). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press.

[20] Peng, L. and Welsh, A.H. (2001). Robust estimation of the generalized Pareto distribution. *Extremes*, **4**(1), 53–65.

[21] Psarakis, S. and Panaretoes, J. (1990). The folded t distribution. *Communications in Statistics: Theory and Methods*, **19**, 2717–2734.

[22] Resnick, S.I. (1997). Discussion of the Danish data on large fire insurance losses. *ASTIN Bulletin*, **27**(1), 139–151.

[23] Scollnik, D.P.M. (2007). On composite lognormal-Pareto models. *Scandinavian Actuarial Journal*, **2007**(1), 20–33.

[24] Scollnik, D.P.M. (2014). Letter to the Editor: Regarding folded models and the paper by Brazauskas and Kleefeld (2011). *Scandinavian Actuarial Journal*, **2014**(3), 278–281.

[25] Scollnik, D.P.M. and Sun, C. (2012). Modeling with Weibull-Pareto models. *North American Actuarial Journal*, **16**(2), 260–272.

[26] Tapiero, C.S. (2004). Risk management: An interdisciplinary framework. In *Encyclopedia of Actuarial Science* (B. Sundt and J. Teugels, eds.), volume **3**, 1483–1493; Wiley, London.

[27] Young, V.R. (2004). Premium principles. In *Encyclopedia of Actuarial Science* (B. Sundt and J. Teugels, eds.), volume **3**, 1322–1331; Wiley, London.