Authors' Reply to "Letter to the Editor regarding folded models and the paper by Brazauskas and Kleefeld (2011)"

Vytaras Brazauskas and Andreas Kleefeld

We thank Dr. Scollnik for reading our paper carefully and for pointing out an important issue in the numerical example of the paper. Yes, we agree that our comparison of the newly proposed model with its closest competitors was "quick" and a bit unfair to the other distributions. Indeed, using the log-transformed data instead of original data for comparing the fits of various distributions gives a home-court advantage to the folded-t_7 (FT_7) model. As one can see from Tables 1 and 2 in Scollnik (2012), the logarithmic transformation changes the values of statistical performance measures for the truncated generalized Pareto distribution (GPD), as it should, and makes the GPD a much more competitive model for the data under consideration. Moreover, we see that the fit of the truncated lognormal model is borderline and that of the truncated composite lognormal-Pareto (LNPa) is excellent.

Using the `fminsearch` function in MATLAB for finding maximum likelihood estimators, we were able to replicate (within a small margin of rounding error) all numbers in Table 2 of the discussion paper. The direct fit of the GPD to the Norwegian fire claims now clearly passes the χ^2 test and the values of its (appropriately transformed) negative log-likelihood, NLL, and the Akaike information criterion (AIC) are substantially smaller. However, while the GPD looks more competitive now, it is still *uniformly* outperformed by the FT_7 model, according to the NLL, AIC and the χ^2 criteria. Consequently, since the truncated lognormal model yields inferior fit when compared to that of the GPD, it is also uniformly outperformed by the FT_7 model.

Further, since the LNPa model has three parameters (all other distributions under consideration have at most two parameters), it was not viewed in our paper as one of the "closest competitors". Nonetheless, it certainly fits the Norwegian data very well and thus merits further investigation. To this end, we first note that, under reasonable circumstances, one would expect the model with more parameters to fit the given data set better and to have a smaller NLL than a more parsimonious model. Therefore, in such situations information-based decision rules, such as the AIC and Schwarz Bayesian criterion (SBC), come in handy. According to the AIC measure, the penalty to the LNPa model for having an additional parameter is relatively small and thus the LNPa outperforms the FT_7 model ($AIC_{LNPa} = 1688.521 < 1690.834 = AIC_{FT_7}$). On the other hand, according to the SBC measure, the conclusion is opposite: the FT_7 outperforms the LNPa model ($SBC_{FT_7} = 1700.270 < 1702.674 = SBC_{LNPa}$). Note that in all these comparisons the FT_7 was treated as a two-parameter model, although its degrees of freedom were fixed ($\nu = 7$). When both parameters are estimated using the maximum likelihood approach, the NLL, AIC and SBC for the folded-t model are practically unchanged: $\text{NLL}_{\text{FT}} = 843.413$, $\text{AIC}_{\text{FT}} = 1690.827$, $\text{SBC}_{\text{FT}} = 1700.262$. For extensive discussion on the appropriateness of the AIC and SBC measures in actuarial science, see Brockett (1991) and a discussion to that paper by Bradley Carlin.

Furthermore, apart from the formal statistical measures, in practice we also want to get a feel for how the model "works" on the given data. For such purposes, various graphical tools and simple prediction of claims (typically, most extreme claims) are useful. Therefore, in Figure 1 below we present six plots of the fitted-versus-observed claim sizes for the GPD, LNPa, and the log-folded-t (LFT, with both parameters treated as unknown and estimated from the data using the maximum likelihood method) models. In order to avoid visual distortions due to large spacings between the most extreme observations, plots in the left-hand column are restricted to *approximately* the lowest 90% of data (measured on the original scale, i.e., in millions of Norwegian krones, NOK) and those in the right-hand column focus on the top 10% of data (measured on the logarithmic scale). In addition, the percentage values marked on the data show where the 50th, 75th, 90th, 95th, and 99th percentiles occur. The points plotted in those graphs are the following pairs:

$$\left(\widehat{F}^{-1}(u_i),\ x_{(i)}\right), \quad i = 1, 2, \ldots, 827,$$

where $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(827)}$ denote the ordered claim sizes and $\widehat{F}^{-1}(u_i)$ is the parametric quantile function evaluated at $u_i = (i - 0.5)/827$. The cluster of observations at 500,000 (i.e., $x_{(1)} = \cdots = x_{(14)} = 500,000$) was de-grouped using the method described in Brazauskas and Serfling (2003). For evaluation of the quantile functions we used the following maximum likelihood estimates of the parameters: $\widehat{\sigma} = 382.490$, $\widehat{\gamma} = 0.691$ (GPD), $\widehat{\theta} = 1838.982$, $\widehat{\alpha} = 1.205$, $\widehat{\sigma} = 0.893$ (LNPa), and $\widehat{\sigma} = 1.154$, $\widehat{\nu} = 6.866$ (LFT). As one can see from Figure 1, all three distributions do a good job for the lowest 99% of the data (i.e., those points do not deviate from the 45° line). The differences in their performance emerge at the top 1% of the claim sizes—the segment that is most uncertain, most costly, and thus most interesting in practice. For the latter segment of claim sizes, we observe that while the GPD and LNPa distributions underestimate the risk, the LFT model stays right on target.

Just to make sure the looks do not deceive us, in Table 1 we list the eight largest claims, observed and predicted, which corresponds to the top 1% of the data. (In the table, the above-described formulas and parameter values were not changed.) We see once again that the LFT model yields more accurate predictions for the extreme claims than its competitors.

It is of course premature to draw far-reaching conclusions about the LFT distribution because all these favorable remarks are based on only one data set. However, the design of the model, combined with its mathematical tractability and computational attractiveness, makes us optimistic about the model's future. We believe the LFT can add value to the actuarial practice and thus it deserves a place in the actuary's toolbox along with the GPD, LNPa, and other well-established distributions.

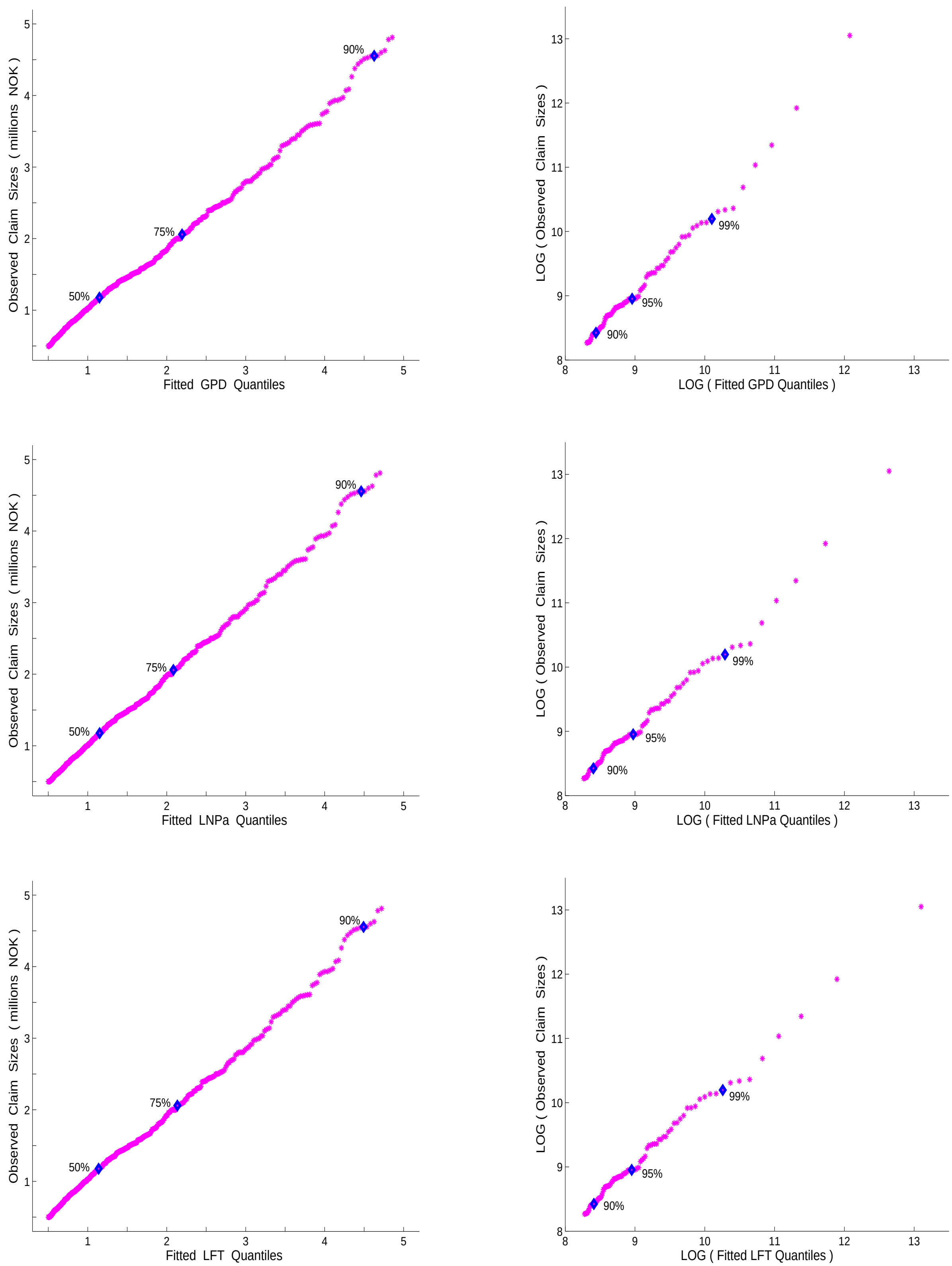

FIGURE 1: Fitted-versus-observed claim sizes for the GPD, LNPa, and LFT models. Left-hand column: lowest 90% of claim sizes. Right-hand column: top 10% of claim sizes.

Table 1: Top 1% of claim sizes (in 1000s), observed and predicted.

Observed Claim Sizes			Parametrically Estimated Claim Sizes		
			GPD$(\widehat{\sigma}, \widehat{\gamma})$	LNPa$(\widehat{\theta}, \widehat{\alpha}, \widehat{\sigma})$	LFT$(\widehat{\sigma}, \widehat{\nu})$
$x_{(820)}$	=	30,000	26,609.4	32,644.2	31,794.2
$x_{(821)}$	=	30,849	29,432.6	36,760.5	36,074.1
$x_{(822)}$	=	31,628	33,101.7	42,226.9	41,890.0
$x_{(823)}$	=	43,752	38,107.5	49,878.5	50,283.1
$x_{(824)}$	=	61,937	45,439.6	61,445.5	63,522.6
$x_{(825)}$	=	84,464	57,478.4	81,237.7	87,680.3
$x_{(826)}$	=	150,597	82,043.5	124,126.5	146,273.2
$x_{(827)}$	=	465,365	175,910.2	308,898.6	488,864.5

Finally, we shall note that there is *no error* in an equation on page 64 of Brazauskas and Kleefeld (2011), as stated by Dr. Scollnik. According to the following mathematical derivations, both log-likelihood equations are equivalent though they look markedly different. One approach leads to:

$$
\begin{aligned}
\sum_{i=1}^{n} \frac{\partial}{\partial \sigma}\Big[\log f(Y_i|\sigma)\Big] &= \sum_{i=1}^{n} \frac{1}{\sigma}\left(\frac{\nu(Y_i^2-\sigma^2)}{\nu\sigma^2+Y_i^2}\right) \qquad (1)\\
&= \frac{1}{\sigma}\sum_{i=1}^{n}\left(\frac{\nu Y_i^2+Y_i^2-\nu\sigma^2-Y_i^2}{\nu\sigma^2+Y_i^2}\right)\\
&= \frac{1}{\sigma}\sum_{i=1}^{n}\left(\frac{Y_i^2(\nu+1)}{\nu\sigma^2+Y_i^2}-1\right)\\
&= \frac{1}{\sigma}\left(\sum_{i=1}^{n}\frac{Y_i^2(\nu+1)}{\nu\sigma^2+Y_i^2}-n\right) \qquad (2)\\
&\overset{!}{=} 0
\end{aligned}
$$

Starting with (1), we can establish:

$$
\begin{aligned}
\sum_{i=1}^{n} \frac{1}{\sigma}\left(\frac{\nu(Y_i^2-\sigma^2)}{\nu\sigma^2+Y_i^2}\right) &= -\frac{\nu}{\sigma}\sum_{i=1}^{n}\left(\frac{-Y_i^2+\sigma^2}{\nu\sigma^2+Y_i^2}\right)\\
&= -\frac{\nu}{\sigma}\sum_{i=1}^{n}\left(\frac{-Y_i^2+\sigma^2+\sigma^2\nu-\sigma^2\nu}{\nu\sigma^2+Y_i^2}\right)\\
&= -\frac{\nu}{\sigma}\sum_{i=1}^{n}\left(\frac{\sigma^2+\sigma^2\nu}{\nu\sigma^2+Y_i^2}-\frac{Y_i^2+\sigma^2\nu}{\nu\sigma^2+Y_i^2}\right)\\
&= -\frac{\nu}{\sigma}\sum_{i=1}^{n}\left(\frac{\sigma^2(1+\nu)}{\nu\sigma^2+Y_i^2}-1\right)\\
&= -\frac{\nu}{\sigma}\left(\sum_{i=1}^{n}\frac{\sigma^2(1+\nu)}{\nu\sigma^2+Y_i^2}-n\right) \qquad (3)\\
&\overset{!}{=} 0
\end{aligned}
$$

Thus, the roots of (2) and (3) are the same. Therefore, either equation can be used to find $\widehat{\sigma}$.

References

[1] 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.

[2] 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.

[3] Brockett, P. (1991). Information theoretic approach to actuarial science: A unification and extension of relevant theory and applications (with discussion). *Transactions of the Society of Actuaries*, **XLIII**, 73–135.

[4] Scollnik, D. (2012). Letter to the Editor regarding folded models and the paper by Brazauskas and Kleefeld (2011). *Scandinavian Actuarial Journal*, **2012**(?), ???–???.