TEMPERED PARETO-TYPE MODELLING USING WEIBULL DISTRIBUTIONS

BY

HANSJÖRG ALBRECHER, JOSÉ CARLOS ARAUJO-ACUNA AND
JAN BEIRLANT

ABSTRACT

In various applications of heavy-tail modelling, the assumed Pareto behaviour is tempered ultimately in the range of the largest data. In insurance applications, claim payments are influenced by claim management and claims may, for instance, be subject to a higher level of inspection at highest damage levels leading to weaker tails than apparent from modal claims. Generalizing earlier results of Meerschaert *et al.* (2012) and Raschke (2020), in this paper we consider tempering of a Pareto-type distribution with a general Weibull distribution in a peaks-over-threshold approach. This requires to modulate the tempering parameters as a function of the chosen threshold. Modelling such a tempering effect is important in order to avoid overestimation of risk measures such as the value-at-risk at high quantiles. We use a pseudo maximum likelihood approach to estimate the model parameters and consider the estimation of extreme quantiles. We derive basic asymptotic results for the estimators, give illustrations with simulation experiments and apply the developed techniques to fire and liability insurance data, providing insight into the relevance of the tempering component in heavy-tail modelling.

KEYWORDS

Weibull tempering, heavy tails, tail estimation, peaks-over-threshold.

JEL codes: 62G32

1. INTRODUCTION

Probability distributions with power-law tails are extensively used in various fields of applications including insurance, finance, information technology, mining of precious stones and language studies (see, e.g., Nair *et al.* (2020)

for a recent overview). In the extreme value methodology, such applications are appropriately modelled using the concept of Pareto-type models such that a variable X of interest satisfies

$$\mathbb{P}(X > x) = x^{-\alpha}\ell(x), \tag{1.1}$$

with $\alpha > 0$ and some slowly varying function ℓ satisfying

$$\frac{\ell(tx)}{\ell(t)} \to 1 \text{ as } t \to \infty \text{ for every } x > 0. \tag{1.2}$$

In addition to the (pure) Pareto distribution, further examples from this model are the Burr, Fréchet, t and log-gamma distribution (see Beirlant *et al.* (1996), Chapter 2 for an overview). Often the power-law behaviour does not extend indefinitely due to some truncation or tapering effects. In Beirlant *et al.* (2016), estimation of truncated tails was developed in a peaks-over-threshold (POT) approach for Pareto-type tails, and other max-domains of attraction were dealt with in Beirlant *et al.* (2017). Inspired by applications in geophysics and finance, Meerschaert *et al.* (2012) discussed parameter estimation under exponential tempering of a simple Pareto law with survival function

$$\mathbb{P}(X > x) = cx^{-\alpha}e^{-\beta x}, \tag{1.3}$$

where $\alpha, \beta > 0$ and $c > 0$ is a scale parameter. In the context of insurance data, Raschke (2020) recently discussed the use of the more general Weibull tempering of a simple power law with survival function

$$\mathbb{P}(X > x) = cx^{-\alpha}e^{-(\beta x)^{\tau}}, \tag{1.4}$$

with $c, \alpha, \beta, \tau > 0$.

However, typically the power-law behaviour only sets in from some threshold t on, rather than from the lowest measurements as assumed when using the simple Pareto model. The Pareto-type model (1.1) allows for flexible modelling of this behaviour. In this paper, we therefore want to study *Weibull tempered Pareto-type distributions* with survival function

$$\mathbb{P}(X > x) = x^{-\alpha}\ell(x)e^{-(\beta x)^{\tau}}, \tag{1.5}$$

with ℓ a slowly varying function, $\alpha = 1/\gamma > 0$ controlling the power-law tail with extreme value index γ, and β, τ governing the Weibull tempering.

We illustrate the need for such Weibull tempering of a Pareto-type tail with the Norwegian fire insurance data set discussed in Beirlant *et al.* (1996), which contains the year of occurrence of the claim and the claim value (in thousand Krones) from 1972 until 1992, see also Brazauskas and Kleefeld (2016, Section 2) for a detailed description of the data. In Figure 1, these data are plotted by year of occurrence, next to a log–log plot (Pareto QQ-plot)

$$\left(-\log\left(1 - \frac{j}{n+1}\right), \log X_{j,n}\right), j = 1, \ldots, n,$$
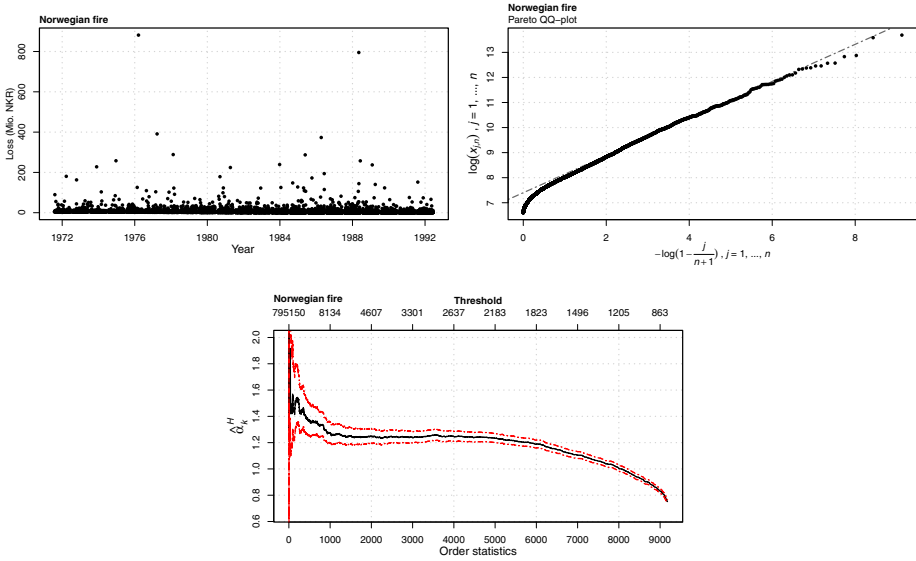
FIGURE 1: Norwegian fire claim data: claim sizes as a function of occurrence time (top left), log–log plot (top right) and $\hat{\alpha}_k^H$ estimates with 95% confidence interval (bottom).
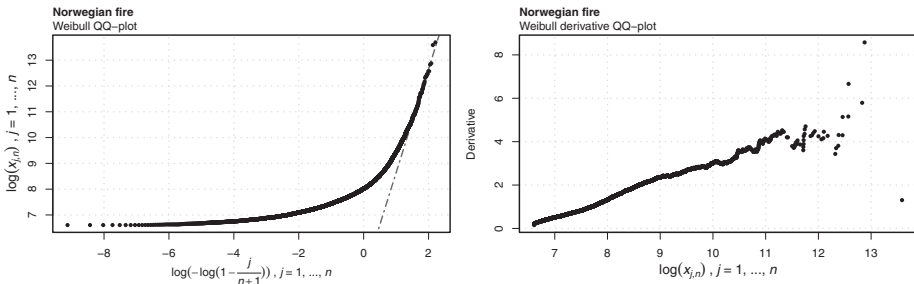

FIGURE 2: Norwegian fire claim data: Weibull QQ-plot (left) and Weibull derivative plot (right).

where $X_{1,n} \leq X_{2,n} \leq \ldots \leq X_{n,n}$ denote the ordered data from a sample of size n. Strict Pareto behaviour corresponds to an overall linear log–log plot, but linearity only arises approximately at the top 5000 observations. The bending at the largest observations in the upper right corner in the log–log plot was also noted. This tapering near the highest observations often occurs with insurance claim data and typically is due to a stricter claim management policy for the larger claims. This tapering is also visible when plotting the pseudo maximum likelihood estimator $\hat{\alpha}_k^H = 1/H_{k,n}$ of α under (1.1) (cf. bottom plot in Figure 1), where $H_{k,n}$ denotes the Hill estimator (1975)

$$H_{k,n} = \frac{1}{k} \sum_{j=1}^{k} \log \frac{X_{n-j+1,n}}{X_{n-k,n}}. \tag{1.6}$$
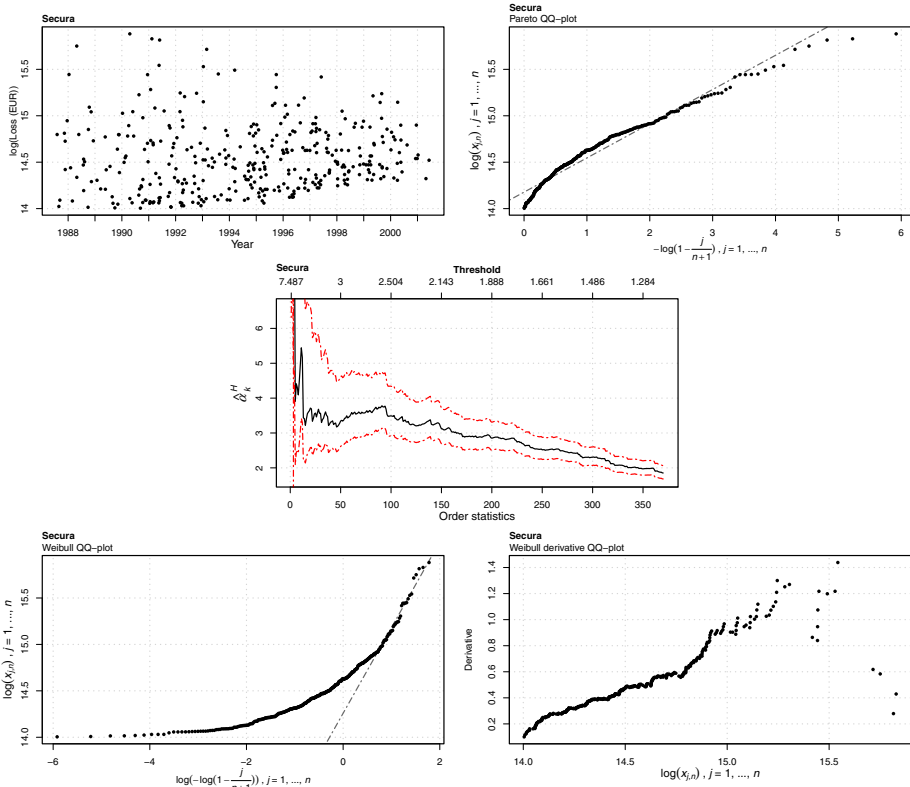
FIGURE 3: Secura Belgian Re claim data: log(Claim sizes) as a function of the year of occurrence (top left), log–log plot (top right), $\hat{\alpha}_k^H$ estimates with 95% confidence interval (middle), Weibull QQ-plot (bottom left) and Weibull derivative plot (bottom right).

The latter can be considered as an estimator of the slope in the log–log plot when restricting to the top $k + 1$ observations. In that sense, the statistics $H_{k,n}$ can be considered as derivatives of the Pareto QQ-plot at the top k observations. Here, the values $\hat{\alpha}_k$ exhibit a stable area for $1000 \leq k \leq 5000$ which expresses power-law behaviour beyond $X_{n-100,n}$, and make a sharp increase at the smallest k values due to tapering. Following the QQ- and derivative plot methodology from Chapter 4 in Albrecher *et al.* (2017), one can construct a Weibull QQ-plot ($\log(-\log(1 - \frac{j}{n+1}))$, $\log X_{j,n}$), $j = 1, \ldots, n$, and its derivative plot in order to verify the Weibull nature of the tempering as proposed in (1.5). A Weibull tail is observed when a linear behaviour is apparent in that QQ-plot at some top portion of the data, which can then be confirmed by a constant derivative plot in that region. For the present case, Figure 2 shows that the derivative plot becomes constant on average when

$\log X > 11$, corresponding to a linear Weibull pattern in the QQ-plot at the top observations with vertical coordinate larger than 11.

As a second example, a tapering effect is also observed in the Secura Belgian Re data set from Beirlant *et al.* (1996). We refer the reader to Beirlant *et al.* (2006, Sections 1.3.3 and 6.2) for further details about the data set. The Pareto QQ-plot in Figure 3 shows a linear pattern from $\log X > 15$, but bending is visible near the top 10 observations, leading to higher values of $\hat{\alpha}_k^H$ at $k \leq 10$. The Weibull derivative plot shows an ultimately decreasing behaviour at the largest 10 observations. This then could lead to truncated Pareto modelling rather than Weibull tempering of a Pareto-type tail, as discussed in detail in Beirlant *et al.* (2016).

In this paper, we complement the graphical and exploratory analysis of Weibull tempering of Pareto-type tails as illustrated above with a mathematical analysis of model (1.5). This can be considered as an alternative to the truncated Pareto-type distributions X discussed in Beirlant *et al.* (2016) which were defined by $X =_d Y | Y < T$ for some high value of T and Y satisfying Pareto-type behaviour (1.1). Truncation also leads to tapering and appears, for instance, in modelling of earthquake energy levels on the basis of the Gutenberg-Richter law. From the viewpoint of truncation, model (1.5) corresponds to $X = \min(Y, W)$ with Y and W independent, Y being Pareto-type distributed and W Weibull distributed with $\mathbb{P}(W > x) = e^{-(\beta x)^\tau}$. Such a model is intended to describe situations where a gradual transit from a power-law decay to an exponentially fast decay is observed as one goes further into the tail. In view of the general nature of the Pareto-type models (1.1), this approach will not be able to capture the characteristics over the whole range of the distribution but focuses rather on the largest observations above some threshold $X_{n-k,n}$. However, if appropriate such tempered tail fits could be spliced with different methods to describe the data below the chosen $X_{n-k,n}$, as it was done before to obtain composed models with a Pareto or generalized Pareto tail fit, see, for instance, Reynkens *et al.* (2017) for mixed Erlang compositions with Pareto tails, Brazauskas and Kleefeld (2016) for log-normal and Weibull models spliced with Pareto tail fits, and Raschke (2020) for Pareto–Pareto or cascade Pareto modelling. Albrecher *et al.* (2020) considered a parsimonious and versatile family of distributions for the modelling of heavy-tailed risks using the class of matrix Mittag-Leffler distributions.

In Section 2, we position the tempered Pareto-Weibull model in a POT approach allowing $\beta \to 0$ as the threshold $t \to \infty$ and study pseudo maximum likelihood estimation providing basic asymptotic theory. We also discuss estimation of extreme return levels and return periods. Proofs of mathematical results are deferred to the Appendix. In Section 3, we provide simulation results, and in Section 4 we complete the analysis of the Norwegian fire and the Belgian liability insurance data sets based on the obtained results. Section 5 concludes.

2. TEMPERED PARETO-TYPE MODELLING AND ESTIMATION

Let $X = \min(Y, W)$ with Y and W independent, where Y is Pareto-type distributed following (1.1) and

$$\mathbb{P}(W > x) = e^{-(\beta x)^{\tau}} \text{ for } x > 0.$$

The survival function of X is then given by

$$\mathbb{P}(X > x) := \overline{F}(x) = x^{-\alpha} \ell(x) e^{-(\beta x)^{\tau}}.$$

For the POT distribution $\frac{X}{t} | X > t$ for some threshold $t > 0$, we obtain for $x > 1$

$$\overline{F}_t(x) := \mathbb{P}\left(\frac{X}{t} > x | X > t\right)$$

$$= \frac{\mathbb{P}(X > tx)}{\mathbb{P}(X > t)}$$

$$= \frac{(tx)^{-\alpha}}{t^{-\alpha}} \frac{\ell(xt)}{\ell(t)} \frac{e^{-(\beta xt)^{\tau}}}{e^{-(\beta t)^{\tau}}}$$

$$= x^{-\alpha} \frac{\ell(xt)}{\ell(t)} e^{-(\beta t)^{\tau}(x^{\tau}-1)}.$$

By definition $\ell(xt)/\ell(t) \approx 1$ for large enough thresholds t. We then assume that at some large values of t, the parameter β is inversely proportional to t, so that a simple Pareto-Weibull model (1.4) provides an appropriate fit to the POTs X/t ($X > t$), at least better than the simple Pareto fit with distribution function $1 - x^{-\alpha}$ as used in classical extreme value methodology for Pareto-type tails. In order to formalize the above, one takes the limit for $t \to \infty$ which necessarily requires $\beta = \beta_t \downarrow 0$ as $t \uparrow \infty$. The model considered in this paper is then formally given by ($\mathcal{M}$). The POT distribution $\overline{F}_t$ satisfies

$$\overline{F}_t(x) \to \overline{F}_{\alpha,\beta_{\infty},\tau}(x) := x^{-\alpha} e^{-\beta_{\infty}^{\tau}(x^{\tau}-1)}, \text{ as } t \to \infty \text{ for every } x > 1,$$

where

(a) **(rough tempering)** $\beta = \beta_t$ satisfies $\beta_t t \to \beta_{\infty} > 0$, corresponding to the situation where the deviation from the Pareto behaviour due to Weibull tempering will be visible in the data from t on and the approximation of the POT distribution using the limit distribution $\overline{F}_{\alpha,\beta_{\infty},\tau}$ appears more appropriate than using $\overline{F}_{\alpha,0,\tau} = x^{-\alpha}$, the simple Pareto distribution;

(b) **(light tempering)** $\beta = \beta_t$ satisfies $\beta_t t \to 0$, corresponding to

$$\overline{F}_t(x) \to x^{-\alpha}, \ x > 1,$$

in which case the tempering is hardly or not visible in the data above t. It will then be practically impossible to discriminate light tempering from no tempering.
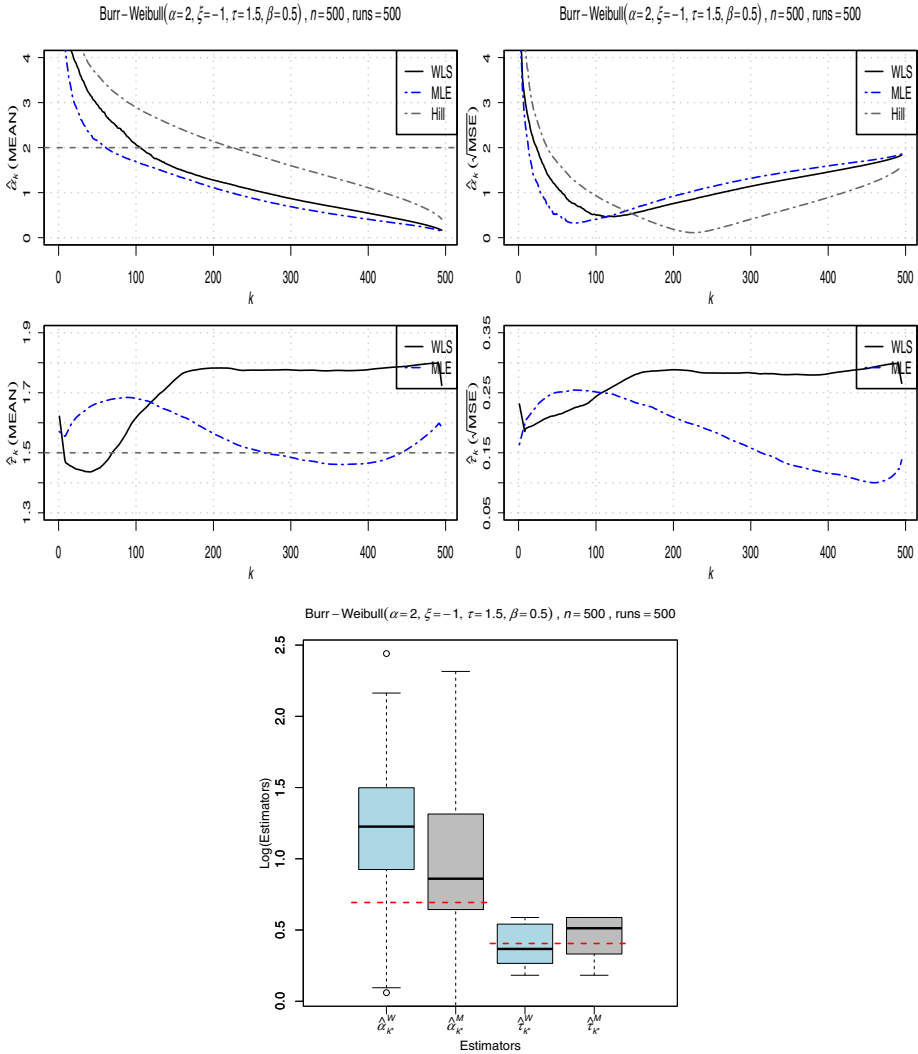
FIGURE 4: Burr-Weibull$(2.0, -1.0, 1.5, 0.5)$. *Top:* Mean (left) and RMSE (right) of $\widehat{\alpha}_k^W, \widehat{\alpha}_k^M$ and $H_{k,n}$ as a function of k ; *Middle:* Mean (left) and RMSE (right) of $\widehat{\tau}_k^W$ and $\widehat{\tau}_k^M$ as a function of k; *Bottom:* Boxplots of $\widehat{\alpha}_k^W, \widehat{\alpha}_k^M, \widehat{\tau}_k^W$ and $\widehat{\tau}_k^M$ (log-scale). Horizontal dashed lines indicate the real parameters.

Given a particular threshold t, the quasi-likelihood procedure consists of fitting the limit distribution in $(\mathcal{M})$ to the POT data

$$\frac{X_j}{t} \text{ when } X_j > t, j = 1, \ldots, n.$$

We also use the notation $\lambda = \beta_\infty^\tau$, so that the limit distribution in $(\mathcal{M})$ is given by

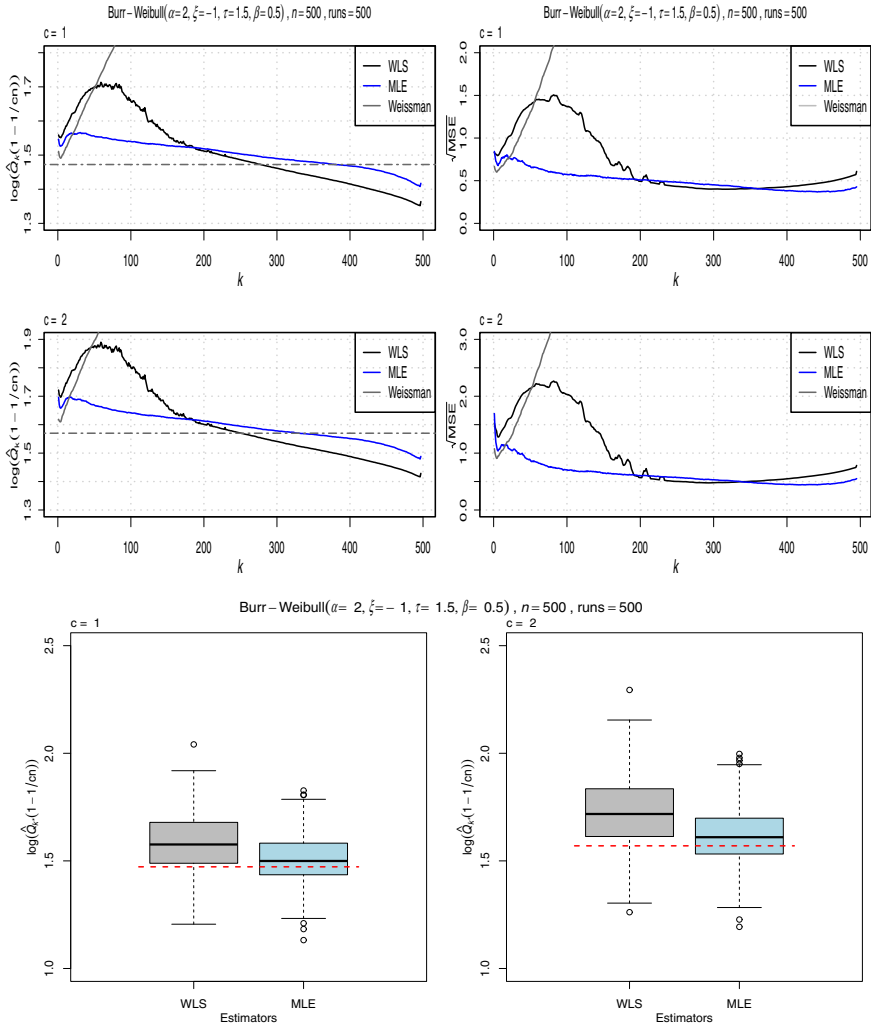$$\overline{F}_{\alpha,\lambda,\tau}(x) = x^{-\alpha} e^{-\lambda(x^\tau - 1)}, \ x > 1.$$

FIGURE 5: Burr-Weibull(2.0, −1.0, 1.5, 0.5): quantile estimates $\widehat{Q}_{p,k}^W$, $\widehat{Q}_{p,k}^M$ with $p = \frac{1}{cn}$ with $c = 1$ (top) and $c = 2$ (middle). Means (left) and RMSE (right) as a function of k. Bottom line: boxplots of $\widehat{Q}_{p,\hat{k}}^W$, $\widehat{Q}_{p,\hat{k}}^M$ with $c = 1$ (left) and $c = 2$ (right). Horizontal dashed lines indicate the real parameters.

The log-likelihood is then given by

$$
\log L(\alpha, \lambda, \tau) = -(1+\alpha) \sum_{j=1}^{n} \log\left(\frac{X_j}{t}\right) 1_{(X_j > t)} - \lambda \sum_{j=1}^{n} \left(\left(\frac{X_j}{t}\right)^\tau - 1\right) 1_{(X_j > t)}
$$

$$
+ \sum_{j=1}^{n} \log\left(\alpha + \lambda\tau\left(\frac{X_j}{t}\right)^\tau\right) 1_{(X_j > t)}.
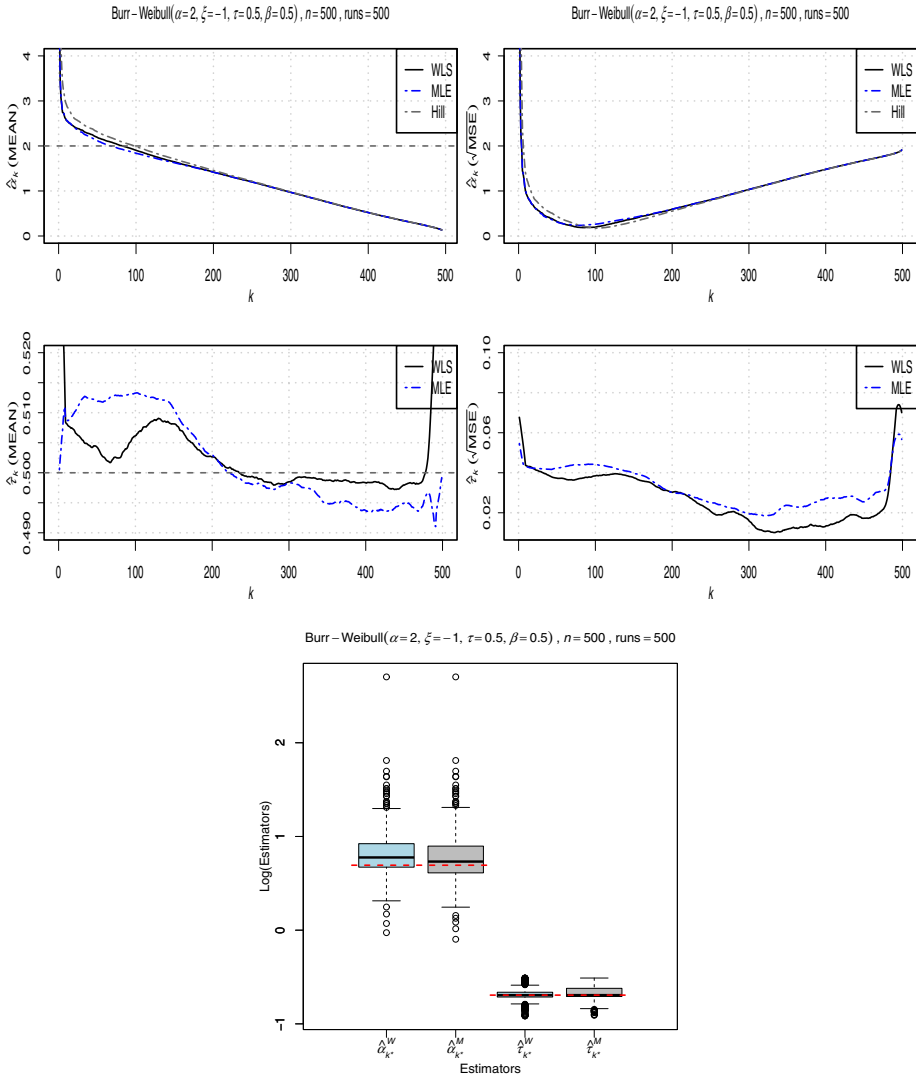$$

(2.1)

FIGURE 6: Fréchet-Weibull(2.0, 2.0, 0.2). *Top:* Mean (left) and RMSE (right) of $\widehat{\alpha}_k^W$, $\widehat{\alpha}_k^M$ and $H_{k,n}$ as a function of k; *Middle:* Mean (left) and RMSE (right) of $\widehat{\tau}_k^W$ and $\widehat{\tau}_k^M$ as a function of k; *Bottom:* Boxplots of $\widehat{\alpha}_{\hat{k}}^W$, $\widehat{\alpha}_{\hat{k}}^M$, $\widehat{\tau}_{\hat{k}}^W$ and $\widehat{\tau}_{\hat{k}}^M$ (log-scale). Horizontal dashed lines indicate the real parameters.

In extreme value methodology, the choice of a threshold t is an important matter. A common practice is to select the $(k + 1)$-largest observation $x_{n−k,n}$ for some $k \in \{4, \dots, n − 1\}$ as the threshold t, and to plot the resulting estimates as a function of the inverse rank k. Many authors then suggest to find k in a stable portion of these plots, if available. Data-driven choices of k are sometimes available minimizing the asymptotic mean-squared error based on asymptotic results that describe the bias and variance for intermediate k sequences. While
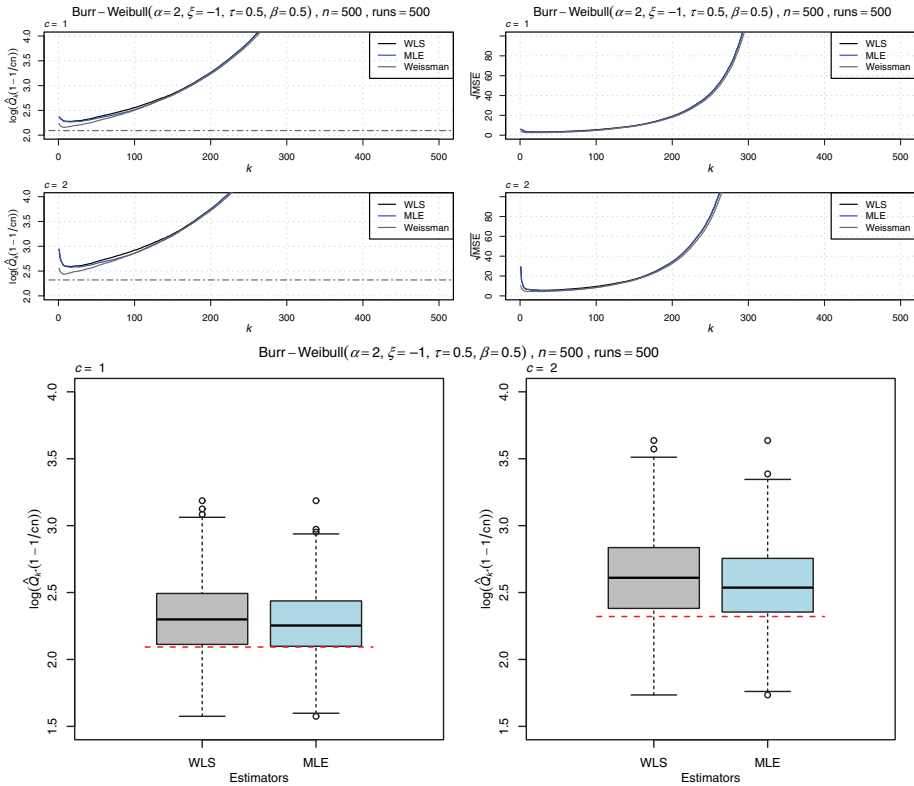
FIGURE 7: Frechet-Weibull(2.0, 2.0, 0.2): quantile estimates $\widehat{Q}_{p,k}^W$, $\widehat{Q}_{p,k}^M$ with $p = \frac{1}{cn}$ with $c = 1$ (top) and $c = 2$ (middle). Means (left) and RMSE (right) as a function of k. Bottom line: boxplots of $\widehat{Q}_{p,\hat{k}}^W$, $\widehat{Q}_{p,\hat{k}}^M$ with $c = 1$ (left) and $c = 2$ (right). Horizontal dashed lines indicate the real parameters.

an asymptotic result is presented below in Theorem 2.1, we here present an approach focusing on the goodness-of-fit of the tempering model to the POT data above the different thresholds $x_{n-k,n}$, using a QQ-plot approach. Then, for a given value of τ, one finds the least-squares line that minimizes

$$\left(-\log\left(1 - \hat{F}_k\left(V_{j,k}\right)\right), \alpha \log V_{j,k} + \tau \beta_\infty^\tau h_\tau\left(V_{j,k}\right) \right), \quad j = 1, \ldots, k, \qquad (2.2)$$

with $h_\tau(x) = (x^\tau - 1)/\tau$, the POT data $V_{j,k} = X_{n-j+1,n}/X_{n-k,n}$, $j = 1, \ldots, k$, and $\hat{F}_k$ denoting the empirical distribution function based on those POTs.

Therefore, since $\hat{F}_k\left(V_{j,k}\right) = \frac{j}{k+1}$, one is led to minimize

$$WLS(V_{j,k}; \alpha_k, \delta_k, \tau_k) := \sum_{j=1}^k w_{j,k}\left(\frac{1}{\alpha}\log\frac{k+1}{k-j+1} - \log V_{j,k} - \delta h_\tau\left(V_{j,k}\right)\right)^2,$$
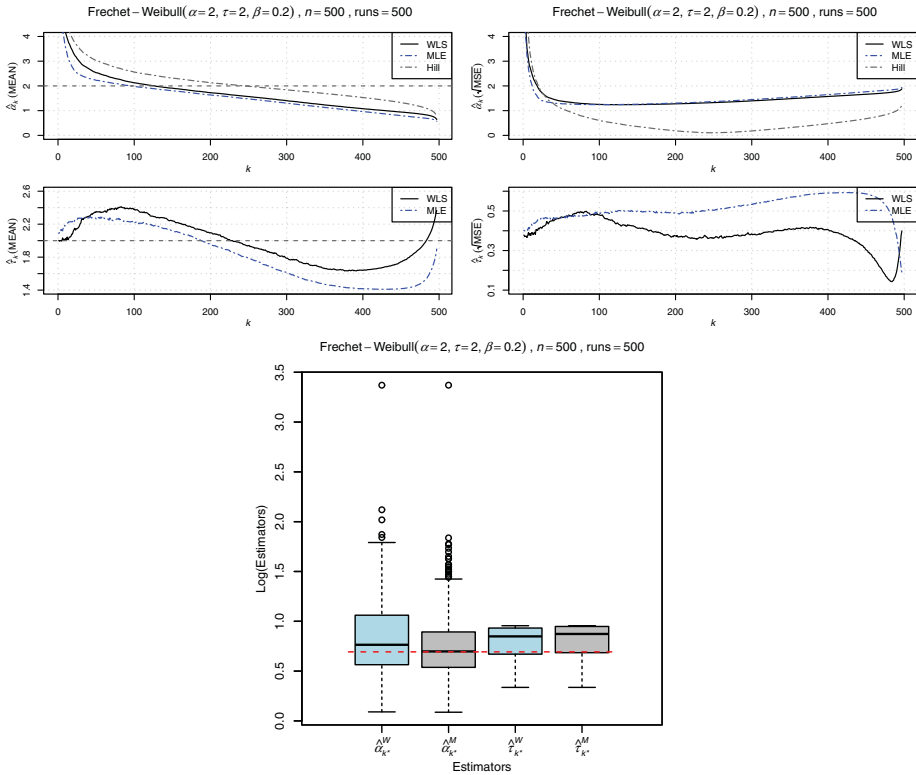
$$(2.3)$$

FIGURE 8: Pareto-Weibull(1.0, 2.0, 0.2).*Top:* Mean (left) and RMSE (right) of $\hat{\alpha}_k^W$, $\hat{\alpha}_k^M$ and $H_{k,n}$ as a function of k ; *Middle:* Mean (left) and RMSE (right) of $\hat{\tau}_k^W$ and $\hat{\tau}_k^M$ as a function of k; *Bottom:* Boxplots of $\hat{\alpha}_{\hat{k}}^W$, $\hat{\alpha}_{\hat{k}}^M$, $\hat{\tau}_{\hat{k}}^W$ and $\hat{\tau}_{\hat{k}}^M$ (log-scale). Horizontal dashed lines indicate the real parameters.

with respect to α and $\delta = \tau \beta_\infty^\tau$, where $\{w_{j,k}, j = 1, \ldots, k\}$ are appropriate weights. In particular, if $w_{j,k} = 1/\log\left(\frac{k+1}{k-j+1}\right)$ when $\delta \downarrow 0$, that is without tempering, we recover the classical Hill estimator $H_{k,n}$.

Optimization using (2.3) also leads to an adaptive selection method for choosing k which gives appropriate estimates for $(\alpha, \tau, \beta_\infty)$, choosing the k for which the weighted least-squares (WLS) value is minimal:

$$\hat{k} = \underset{k}{\arg\min} \, SS_k \tag{2.4}$$

with

$$SS_k = \sum_{j=1}^{k} \frac{1}{\log\left(\frac{k+1}{k-j+1}\right)} \left(\frac{1}{\hat{\alpha}_k^W} \log\left(\frac{k+1}{k-j+1}\right) - \log V_{j,k} - \hat{\delta}_k^W h_{\hat{\tau}_k^W}(V_{j,k})\right)^2 .$$
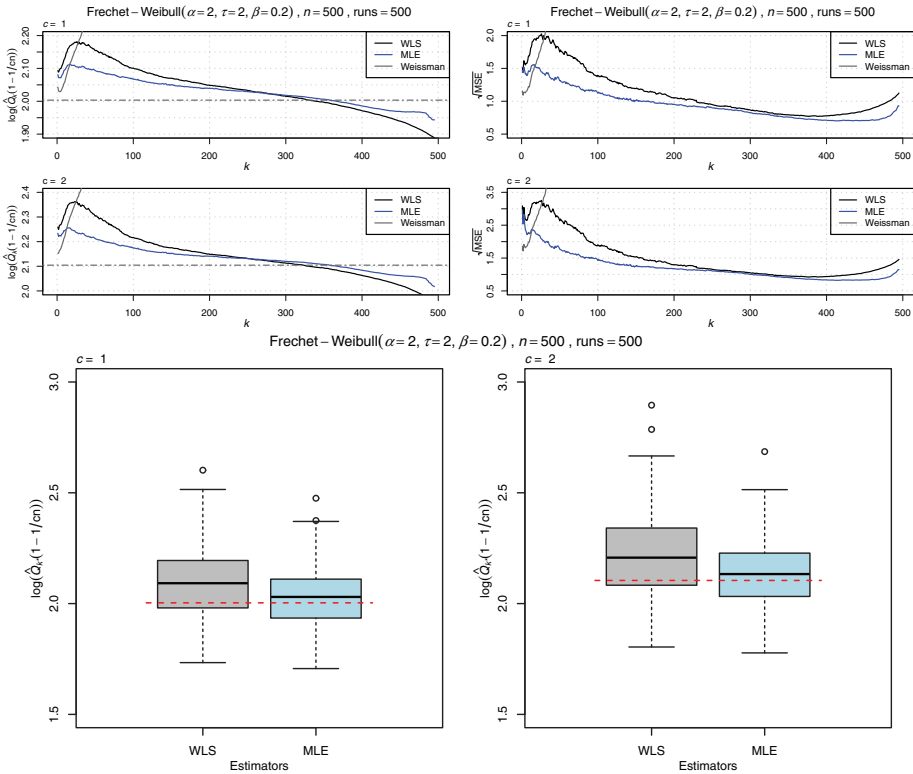
$$\tag{2.5}$$

FIGURE 9: Pareto-Weibull(1.0, 2.0, 0.2): quantile estimates $\widehat{Q}_{p,k}^W$, $\widehat{Q}_{p,k}^M$ with $p = \frac{1}{cn}$ with $c = 1$ (top) and $c = 2$ (middle). Means (left) and RMSE (right) as a function of k. Bottom line: boxplots of $\widehat{Q}_{p,\hat{k}}^W$, $\widehat{Q}_{p,\hat{k}}^M$ with $c = 1$ (left) and $c = 2$ (right). Horizontal dashed lines indicate the real parameters.

Since for $\tau \to 0$ the parameters α and τ become non-identifiable, numerical issues will arise during the statistical estimation procedure when directly optimizing the likelihood, or when minimizing (2.3). However, fixing a value of τ during the calibration procedure reduces numerical instabilities. The optimization procedure Algorithm 1 which is used in the simulations and cases leads to WLS estimates $(\widehat{\alpha}_k^W, \widehat{\beta}_{\infty,k}^W, \widehat{\tau}_k^W)$ and maximum likelihood estimates $(\widehat{\alpha}_k^M, \widehat{\beta}_{\infty,k}^M, \widehat{\tau}_k^M)$, starting from a grid of m initial τ values $\tilde{\tau}_1 < \tilde{\tau}_2 < \cdots < \tilde{\tau}_m$, $m \in \mathbb{N}$.

In order to estimate return periods of the type $1/\mathbb{P}(X > z)$ for some large outcome level z, we use the approximation

$$\frac{\mathbb{P}(X > tx)}{\mathbb{P}(X > t)} \approx x^{-\alpha} e^{-\lambda \tau h_\tau(x)}$$
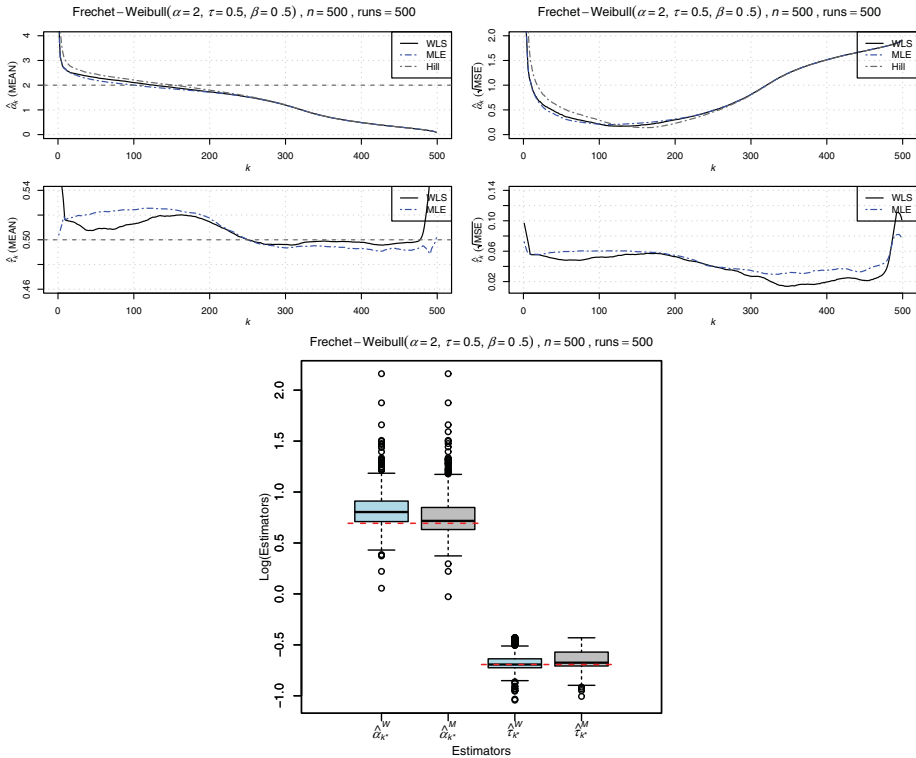
FIGURE 10: log-normal-Weibull(0.0, 100, 1.5, 0.5). *Top left:* $\hat{\alpha}_k^W$, $\hat{\alpha}_k^M$ and $H_{k,n}$ mean estimates as a function of k. *Middle left:* $\hat{\tau}_k^W$, $\hat{\tau}_k^M$ mean estimates as a function of k. *Right:* quantile estimates $\widehat{Q}_{p,k}^W$, $\widehat{Q}_{p,k}^M$ with $p = \frac{1}{cn}$ with $c = 0.2$ (top) and $c = 0.4$ (middle). *Bottom:* boxplots of $\hat{\alpha}_{\hat{k}}^W$, $\hat{\alpha}_{\hat{k}}^M$, $\hat{\tau}_{\hat{k}}^W$ and $\hat{\tau}_{\hat{k}}^M$ (log-scale). Horizontal dashed lines indicate the real parameters.

with t large, so that setting $tx = z$ and $t = x_{n-k,n}$ for some k, we obtain the estimators for $\mathbb{P}(X > z)$

$$\hat{P}_{z,k}^W = \frac{k+1}{n+1}\left(\frac{z}{x_{n-k,n}}\right)^{-\hat{\alpha}_k^W}\exp\left(-\hat{\lambda}_k\hat{\tau}_k^W h_{\hat{\tau}_k^W}(z/x_{n-k,n})\right) \qquad (2.6)$$

and similarly $\hat{P}_{z,k}^M$, where $\mathbb{P}(X > t) = \mathbb{P}(X > x_{n-k,n})$ is estimated using the empirical proportion $(k+1)/(n+1)$.

The value z solving the equation

$$\frac{k+1}{n+1}\left(\frac{z}{x_{n-k,n}}\right)^{-\hat{\alpha}_k^W}\exp\left(-\hat{\lambda}_k^W\hat{\tau}_k^W h_{\hat{\tau}_k^W}(z/x_{n-k,n})\right) = p, \qquad (2.7)$$

for a given value $p \leq \frac{1}{n}$, then yields an estimator $\widehat{Q}_{p,k}^W$ for the extreme quantile or return level $Q(1-p)$, and hence for the value-at-risk (VaR$_p$) at extreme quantile levels $1-p$. Similarly, one obtains the estimator $\widehat{Q}_{p,k}^M$.
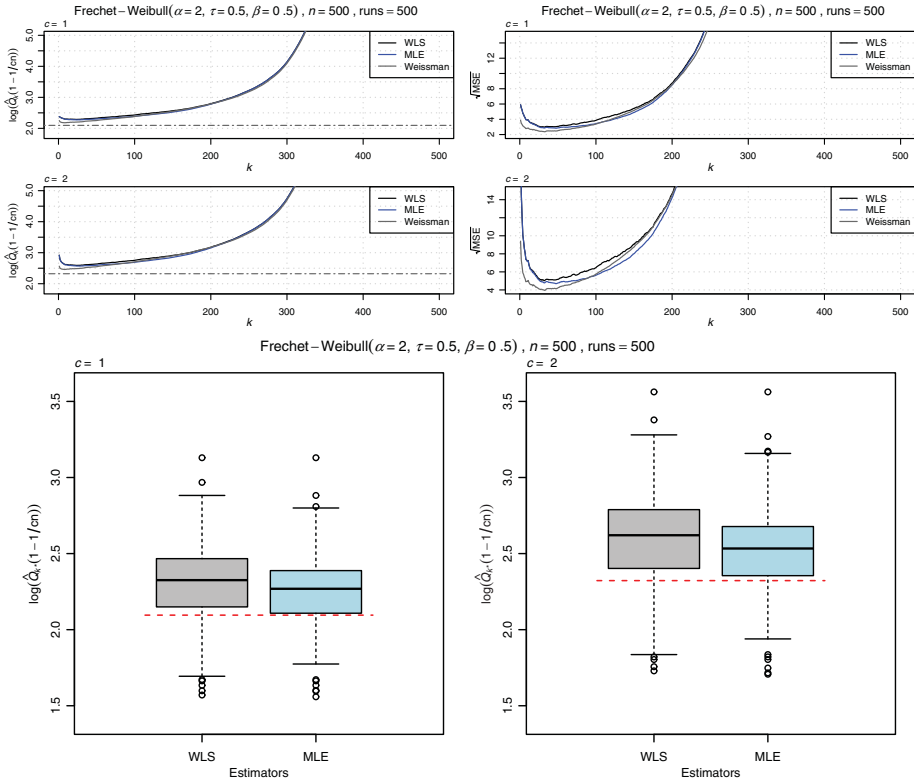
FIGURE 11: Log-normal-Weibull(0.0, 100, 1.5, 0.5): Boxplots of $\widehat{Q}_{p,\hat{k}}^{W}$, $\widehat{Q}_{p,\hat{k}}^{M}$ with $c = 0.2$ (left) and $c = 0.4$ (right). Horizontal dashed lines indicate the real values.

We end this section stating the asymptotic distribution of the maximum likelihood estimators $\hat{\alpha}_t$, $\hat{\lambda}_t$, $\hat{\tau}_t$. The likelihood equations in (α, λ, τ) are given by

$$\sum_{j=1}^{n} \left\{ \alpha + \lambda \tau \left(\frac{X_j}{t} \right)^{\tau} \right\}^{-1} 1_{(X_j > t)} = \sum_{j=1}^{n} \log \left(\frac{X_j}{t} \right) 1_{(X_j > t)},$$

$$\sum_{j=1}^{n} \frac{\left(\frac{X_j}{t} \right)^{\tau}}{\alpha + \lambda \tau \left(\frac{X_j}{t} \right)^{\tau}} 1_{(X_j > t)} = \sum_{j=1}^{n} h_{\tau} \left(\frac{X_j}{t} \right) 1_{(X_j > t)},$$

$$\sum_{j=1}^{n} \frac{\left(\frac{X_j}{t} \right)^{\tau} \log \left(\frac{X_j}{t} \right)}{\alpha + \lambda \tau \left(\frac{X_j}{t} \right)^{\tau}} 1_{(X_j > t)} = \sum_{j=1}^{n} \left(\frac{X_j}{t} \right)^{\tau} \log \left(\frac{X_j}{t} \right) 1_{(X_j > t)}.$$

Algorithm 1 Estimation of $(\widehat{\alpha}_k^W, \widehat{\beta}_{\infty,k}^W, \widehat{\tau}_k^W)$ and $(\widehat{\alpha}_k^M, \widehat{\beta}_{\infty,k}^M, \widehat{\tau}_k^M)$

1: **set** $\tilde{\tau}_1 < \tilde{\tau}_2 < \cdots < \tilde{\tau}_m, m \in \mathbb{N}$
2: **for** $k = 1, 2,$ **to** $n - 1$ **do**
3:      **for** $i = 1, 2,$ **to** m **do**
4:          **Optimization step.** Set

$$\left(\widehat{\alpha}_{k,\tilde{\tau}_i}, \widehat{\delta}_{k,\tilde{\tau}_i}\right) := \underset{(\alpha>0,\delta>0)}{\arg\min}\ WLS(V_{j,k};\alpha, \delta, \tilde{\tau}_i)$$

5:          $\widehat{WLS}_{k,\tilde{\tau}_i} \leftarrow WLS(V_{j,k};\widehat{\alpha}_{k,\tilde{\tau}_i}, \widehat{\delta}_{k,\tilde{\tau}_i}, \tilde{\tau}_i)$
6:          $\widehat{\lambda}_{k,\tilde{\tau}_i} \leftarrow \widehat{\delta}_{k,\tilde{\tau}_i}/\tilde{\tau}_i$
7:          $\widehat{\log L}_{k,\tilde{\tau}_i} \leftarrow \log L(V_{j,k};\widehat{\alpha}_{k,\tilde{\tau}_i}, \widehat{\lambda}_{k,\tilde{\tau}_i}, \tilde{\tau}_i)$
8:      Set

$$\left(\widehat{\alpha}_k^W, \widehat{\delta}_k^W, \widehat{\tau}_k^W\right) := \underset{(\widehat{\alpha}_{k,\tilde{\tau}_i}, \widehat{\delta}_{k,\tilde{\tau}_i}, \tilde{\tau}_i)}{\arg\min}\left\{\widehat{WLS}_{k,\tilde{\tau}_i}; i = 1, \ldots, m\right\}$$

9:          $\widehat{\beta}_{\infty,k}^W \leftarrow \left(\widehat{\delta}_k^W/\tilde{\tau}_k^W\right)^{1/\tilde{\tau}_k^W}$
10:     Set

$$\left(\widehat{\alpha}_k^M, \widehat{\lambda}_k^M, \widehat{\tau}_k^M\right) := \underset{(\widehat{\alpha}_{k,\tilde{\tau}_i}, \widehat{\lambda}_{k,\tilde{\tau}_i}, \tilde{\tau}_i)}{\arg\max}\left\{\widehat{\log L}_{k,\tilde{\tau}_i}; i = 1, \ldots, m\right\}$$

11:         $\widehat{\beta}_{\infty,k}^M \leftarrow \left(\widehat{\lambda}_k^M\right)^{1/\tilde{\tau}_k^M}$
12: **return** $(\widehat{\alpha}_k^W, \widehat{\beta}_{\infty,k}^W, \widehat{\tau}_k^W)$ and $(\widehat{\alpha}_k^M, \widehat{\beta}_{\infty,k}^M, \widehat{\tau}_k^M)$, for $k = 1, 2, \ldots, n - 1$.

We further assume classical second-order slow variation

$$\frac{\ell(ty)}{\ell(t)} = 1 + Dt^\rho h_\rho(y),\ \text{with}\ D \in \mathbb{R}, \rho < 0, \tag{2.8}$$

and set $\hat{\boldsymbol{\theta}}_t = (\hat{\alpha}_t, \hat{\lambda}_t, \hat{\tau}_t)^t$ and $\boldsymbol{\theta} = (\alpha, \lambda, \tau)^t$.

Theorem 2.1. *Under* $\overline{F}(x) = x^{-\alpha}\ell(x)e^{-\beta x^\tau}$ *satisfying* $(\mathcal{M})$ *with* $\beta_\infty > 0$ *and* ℓ *satisfying* (2.8), *we have* $n, t \to \infty$ *such that* $n\overline{F}(t) \to \infty$ *and* $\sqrt{n\overline{F}(t)}t^\rho \to \nu > 0$ *that*

$$\sqrt{n\overline{F}(t)}\left(\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}\right) \to_d \mathcal{N}_3\left(D\nu\mathbf{I}^{-1}\mathbf{b}, \mathbf{I}^{-1}\right)$$

with $\mathbf{I} \in \mathbb{R}^{3\times3}$ *symmetric and* $\mathbf{b} \in \mathbb{R}^{3\times1}$ *and*

$$I_{1,1} = \int_1^\infty \frac{u^{-\alpha-1}e^{-\lambda\tau h_\tau(u)}}{\alpha + \lambda\tau u^\tau}du,$$

$$I_{2,2} = \tau^2 \int_1^\infty \frac{u^{2\tau-\alpha-1}e^{-\lambda\tau h_\tau(u)}}{\alpha + \lambda\tau u^\tau}du,$$

$$I_{3,3} = \lambda \int_1^\infty \left\{ \log^2(u) - \frac{2\log u}{\alpha + \lambda\tau u^\tau} + \frac{\lambda u^\tau(1 + 2\tau\log u) - \alpha\tau(\log u)^2}{(\alpha + \lambda\tau u^\tau)^2} \right\}$$
$$u^{\tau-\alpha-1} e^{-\lambda\tau h_\tau(u)} (\alpha + \lambda\tau u^\tau) du,$$

$$I_{1,2} = \tau \int_1^\infty \frac{u^{\tau-\alpha-1} e^{-\lambda\tau h_\tau(u)}}{\alpha + \lambda\tau u^\tau} du,$$

$$I_{1,3} = \lambda \int_1^\infty (1 + \tau\log u) \frac{u^{\tau-\alpha-1} e^{-\lambda\tau h_\tau(u)}}{\alpha + \lambda\tau u^\tau} du,$$

$$I_{2,3} = \int_1^\infty \left\{ \log u - \frac{\alpha(1 + \tau\log u)}{(\alpha + \lambda\tau u^\tau)^2} \right\} u^{\tau-\alpha-1} e^{-\lambda\tau h_\tau(u)} (\alpha + \lambda\tau u^\tau) du,$$

$$b_1 = \int_1^\infty \left(\frac{1}{\alpha + \lambda\tau u^\tau} - \log u \right) u^{-\alpha-1} e^{-\lambda\tau h_\tau(u)} [h_\rho(u)(\alpha + \lambda\tau u^\tau) - u^\rho] du,$$

$$b_2 = \int_1^\infty \left(\frac{\tau u^\tau}{\alpha + \lambda\tau u^\tau} - \tau h_\tau(u) \right) u^{-\alpha-1} e^{-\lambda\tau h_\tau(u)} [h_\rho(u)(\alpha + \lambda\tau u^\tau) - u^\rho] du,$$

$$b_3 = \lambda \int_1^\infty \left(\frac{1 + \tau\log u}{\alpha + \lambda\tau u^\tau} - \log u \right) u^{\tau-\alpha-1} e^{-\lambda\tau h_\tau(u)} [h_\rho(u)(\alpha + \lambda\tau u^\tau) - u^\rho] du.$$

The derivation of this result is postponed to the Appendix.

3. SIMULATION RESULTS

The finite sample behaviour of the estimators $(\hat{\alpha}_k^W, \hat{\tau}_k^W)$ and $(\hat{\alpha}_k^M, \hat{\tau}_k^M)$ and the resulting tail probabilities $\hat{P}_{z,k}^W$, $\hat{P}_{z,k}^M$ and extreme quantiles $\hat{Q}_{p,k}^W$, $\hat{Q}_{p,k}^M$ resulting from Algorithm 1, (2.6) and (2.7), respectively, have been studied through an extensive Monte Carlo simulation procedure. For each setting, 500 runs with sample size $n = 500$ were performed. The mean and root mean squared error (RMSE) of the estimators are presented for the following models:

(a) Burr-Weibull$(\alpha, \xi, \tau, \beta)$ model with Burr distribution given by

$$F_Y(y) = 1 - \left(1 + y^{-\xi\alpha}\right)^{1/\xi}, \ y > 0, \ \alpha > 0, \ \xi < 0.$$

Here, (2.8) is satisfied with $\rho = \xi\alpha$. We used $(\alpha, \xi, \tau, \beta) = (2, -1, 1.50, 0.50)$ and $(2, -1, 0.50, 0.20)$.

(b) Fréchet-Weibull(α, τ, β) model with the Fréchet distribution function

$$F_Y(y) = \exp\left(-y^{-\alpha}\right), \ y > 0, \ \alpha > 0.$$

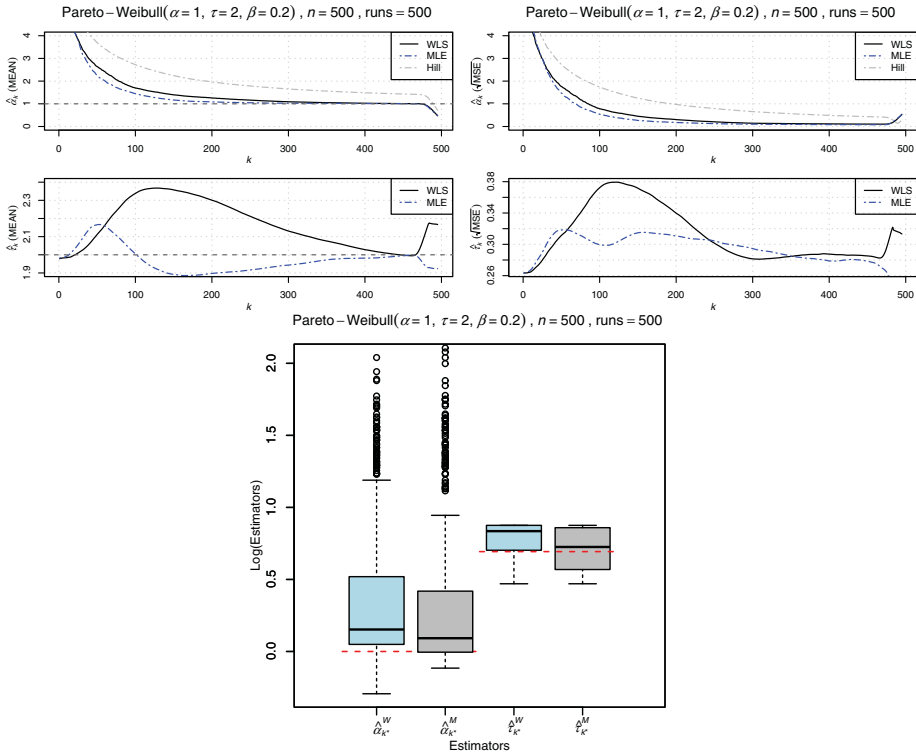Here, (2.8) is satisfied with $\rho = -\alpha$. We used $(\alpha, \tau, \beta) = (2, 2, 0.50)$ and $(2, 0.50, 0.20)$.

FIGURE 12: Burr-Weibull(2.0, −1.0, 0.5, 0.5). *Top:* Mean (left) and RMSE (right) of $\widehat{\alpha}_k^W$, $\widehat{\alpha}_k^M$ and $H_{k,n}$ as a function of k; *Middle:* Mean (left) and RMSE (right) of $\widehat{\tau}_k^W$ and $\widehat{\tau}_k^M$ as a function of k; *Bottom:* Boxplots of $\widehat{\alpha}_{\hat{k}}^W$, $\widehat{\alpha}_{\hat{k}}^M$, $\widehat{\tau}_{\hat{k}}^W$ and $\widehat{\tau}_{\hat{k}}^M$ (log-scale). Horizontal dashed lines indicate the real parameters.

(c) Pareto-Weibull(α, τ, β) model using the Pareto distribution

$$1 - F_Y(y) = y^{-\alpha}, \ y > 1, \ \alpha > 0.$$

Here $\ell(x) = 1$. We used $(\alpha, \tau, \beta) = (1, 2, 0.20)$.

(d) In order to study the behaviour of the estimators under Weibull tempering of a heavy-tailed distribution outside the Pareto-type family, we simulated from a tempered log-normal distribution with parameters $\mu = 0$ and $\sigma = 10$.

In the plots concerning the estimation of α, we also plot the results for the Hill estimator $H_{k,n}$, while in case of the tail quantile estimates $\hat{Q}_{p,k}^W$ and $\hat{Q}_{p,k}^M$ we also provide the results for the Weissman (1978) estimator $\hat{Q}_{p,k}^H = X_{n-k,n} \left(\frac{k}{np} \right)^{1/\hat{\alpha}_k^H}$. Finally, we also present the boxplots of the estimates when using the adaptive choice $\hat{k}$ given in (2.4) for k. The characteristics for the tail probability estimators $\hat{P}_{z,k}^W$, $\hat{P}_{z,k}^M$ are quite comparable to those of the extreme quantiles and are omitted here.
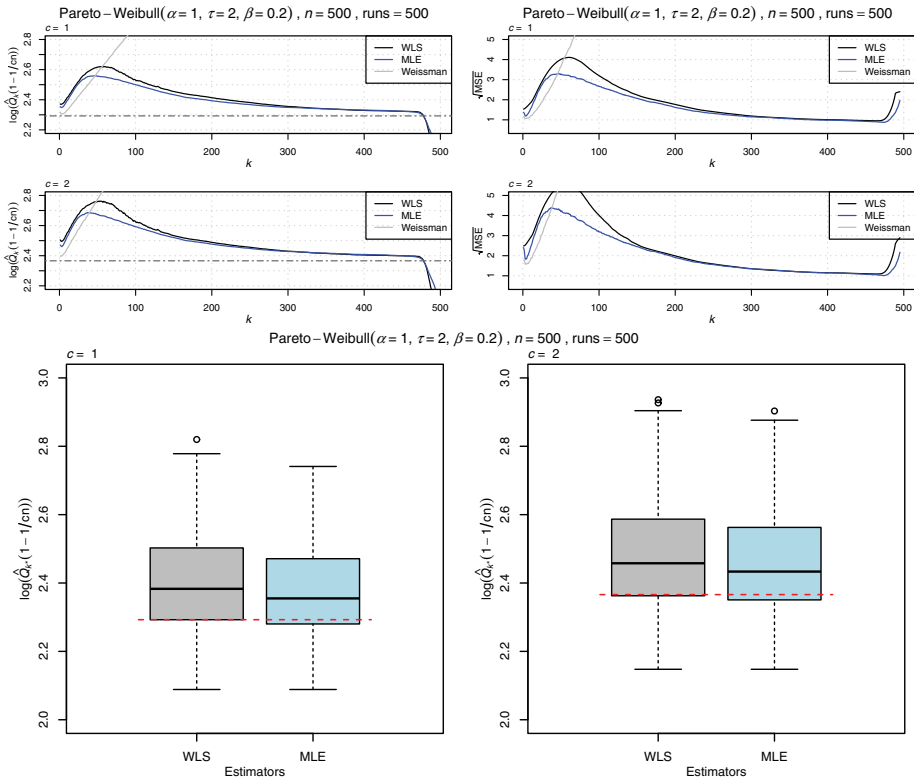
FIGURE 13: Burr-Weibull(2.0, −1.0, 0.5, 0.5): quantile estimates $\widehat{Q}_{p,k}^{W}$, $\widehat{Q}_{p,k}^{M}$ with $p = \frac{1}{cn}$ with $c = 1$ (top) and $c = 2$ (middle). Means (left) and RMSE (right) as a function of k. Bottom line: boxplots of $\widehat{Q}_{p,\hat{k}}^{W}$, $\widehat{Q}_{p,\hat{k}}^{M}$ with $c = 1$ (left) and $c = 2$ (right). Horizontal dashed lines indicate the real parameters.

Clearly, the results for the MLE results $\hat{\alpha}^{M}$, $\hat{\tau}^{M}$ and $\hat{Q}_{p}^{M}$ improve upon the weighted least squares-based results. The results with the adaptive choice $\hat{k}$ of k are promising, and again best for the MLE results. In case $\tau > 1$ (see Figures 4, 5, 6, 7, 8, 9, 10 and 11) when the tempering is quite strong, the results for the proposed methods are clearly improving upon the classical estimators $H_{k,n}$ and $\hat{Q}_{p,k}^{H}$. Note that in these cases the VaR estimates based on the MLE parameters taken at the adaptive value $\hat{k}$ show a rather small bias, even in case of the log-normal model which is situated outside our Pareto-type model assumption.

In case $\tau < 1$ (see Figures 12, 13, 14 and 15), hence under weaker tempering, the bias and RMSE results are comparable with the classical estimators. The VaR estimates at $\hat{k}$ tend to overestimate the correct value. As will become clear from the case studies in the next section, the Pareto and tempered Pareto fits can lead to quite different extreme tail fits *per sample*.

We conclude that the use of classical estimators ignoring the tempering effect leads to serious overestimation of the risk measures, while the proposed
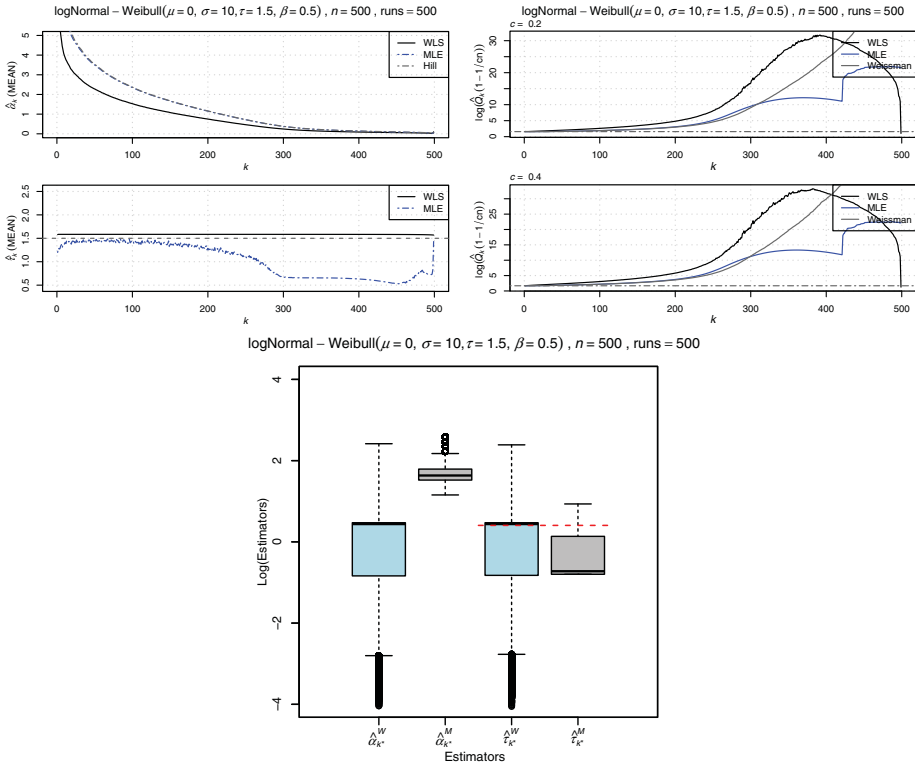
FIGURE 14: Fréchet-Weibull(2.0, 0.50, 0.5). *Top:* Mean (left) and RMSE (right) of $\widehat{\alpha}_k^W$, $\widehat{\alpha}_k^M$ and $H_{k,n}$ as a function of k ; *Middle:* Mean (left) and RMSE (right) of $\widehat{\tau}_k^W$ and $\widehat{\tau}_k^M$ as a function of k; *Bottom:* Boxplots of $\widehat{\alpha}_{\hat{k}}^W$, $\widehat{\alpha}_{\hat{k}}^M$, $\widehat{\tau}_{\hat{k}}^W$ and $\widehat{\tau}_{\hat{k}}^M$ (log-scale). Horizontal dashed lines indicate the real parameters.
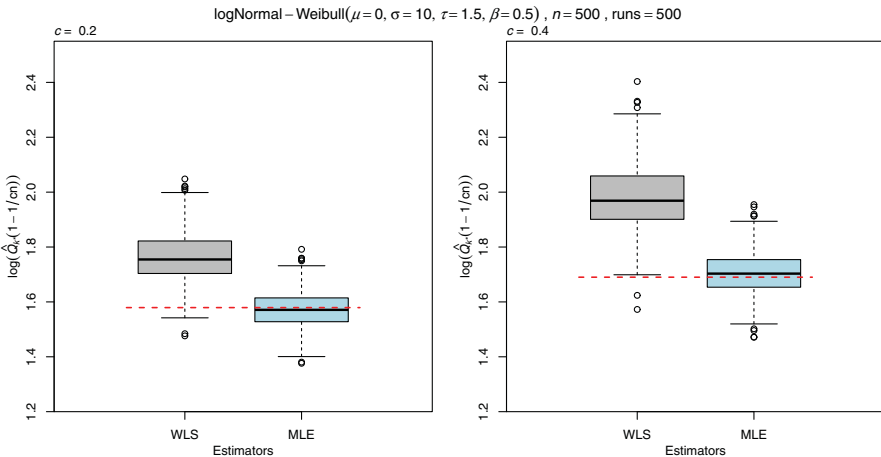

FIGURE 15: Fréchet-Weibull(2.0, 0.5, 0.5): quantile estimates $\widehat{Q}_{p,k}^W$, $\widehat{Q}_{p,k}^M$ with $p = \frac{1}{cn}$ with $c = 1$ (top) and $c = 2$ (middle). Means (left) and RMSE (right) as a function of k. Bottom line: boxplots of $\widehat{Q}_{p,\hat{k}}^W$, $\widehat{Q}_{p,\hat{k}}^M$ with $c = 1$ (left) and $c = 2$ (right). Horizontal dashed lines indicate the real parameters.
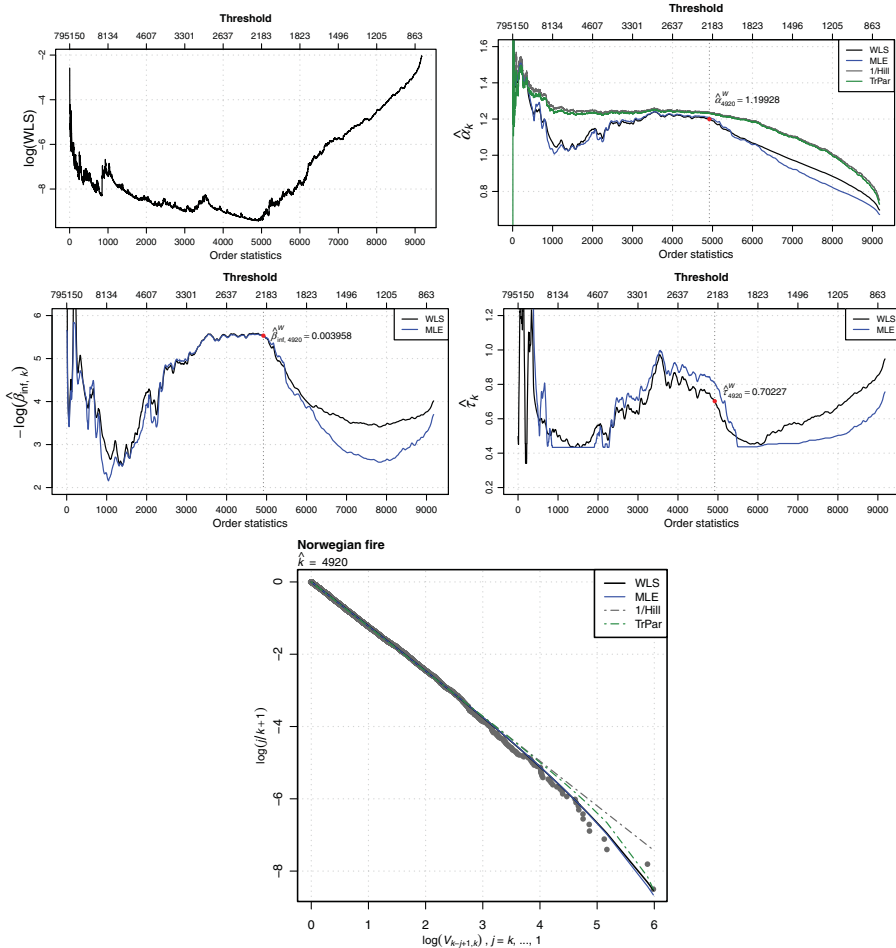
FIGURE 16: *Norwegian fire insurance data:* Top left: SS_k from (2.5); Top right: $\hat{\alpha}_k^W$, $\hat{\alpha}_k^M$, $H_{k,n}$ and $\hat{\alpha}_k^T$; Middle left: $-\log \hat{\beta}_{\infty,k}^W$, $-\log \hat{\beta}_{\infty,k}^M$; Middle right: $\hat{\tau}_k^W$, $\hat{\tau}_k^M$; Bottom: log-log plot with fit obtained from (2.2) with $k = \hat{k} = 4915$ using MLE and WLS estimates, next to Pareto and truncated Pareto fit.

method provides reasonable VaR estimates especially for larger values of $\tau > 1$. In case of smaller tempering with a heavier Weibull tail, improvements can be made concerning the adaptive choice of k. Another possibility is to search for bias reduced estimators as available in the non-tempering literature (see, for instance, Chapters 3 and 4 in Beirlant *et al.* (2006)).

4. INSURANCE CASES

We now apply the presented methods to the Norwegian and the Secura Re Belgian data sets introduced in Section 1. In addition, we contrast the tail index
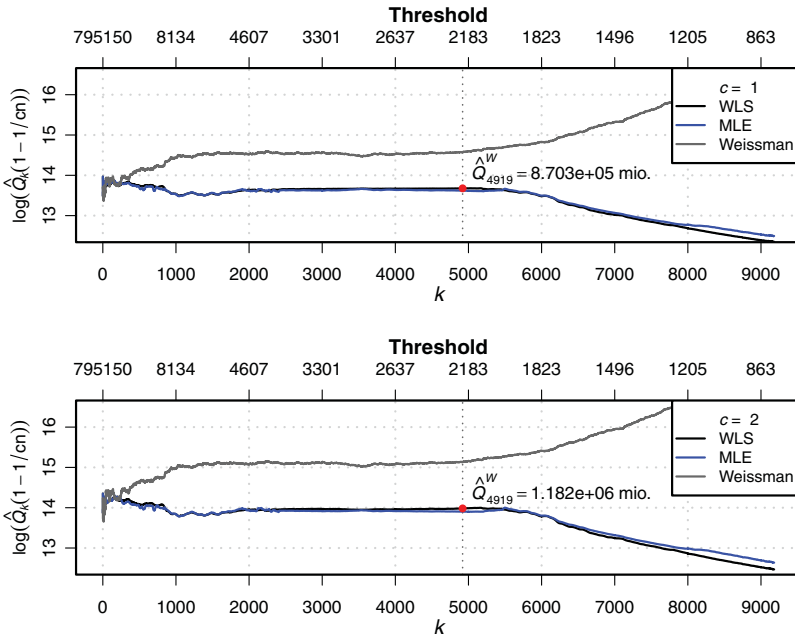
FIGURE 17: *Norwegian fire insurance data*: $\hat{Q}_{p,k}^W$, $\hat{Q}_{p,k}^M$ and $\hat{Q}_{p,k}^H$ quantile estimates with $p = 1/n$ (top) and $p = 1/(2n)$ (bottom).

estimates $\widehat{\alpha}_{k,n}^W$ and $\widehat{\alpha}_{k,n}^M$ with the values obtained for the truncated Pareto-type model proposed in Beirlant *et al.* (2016), where $\widehat{\alpha}_{k,n}^T$ is obtained as the solution to

$$H_{k,n} = \frac{1}{\alpha_{k,n}^T} + \frac{R_{k,n}^{\alpha_{k,n}^T} \log\left(R_{k,n}\right)}{1 - R_{k,n}^{\alpha_{k,n}^T}},$$

with $R_{k,n} = X_{n-k,n}/X_{n,n}$. The latter estimator was first proposed in Aban *et al.* (2006) as the conditional MLE based on the $k + 1$ ($0 \le k < n$) largest order statistics representing only the portion of the tail where the truncated Pareto approximation holds, see also Albrecher *et al.* (2017, Section 4.2.3).

We then also measure the goodness-of-fit using QQ-plot (2.2) and the analogous expression for the truncated model.

For the Norwegian fire insurance data set, we find $\hat{k} = 4920$ from the plot of SS_k from (2.5) in Figure 16, where also the different parameter estimates as a function of k can be found. The log–log plot based on (2.2) at $k = 4920$ shows a good tail fit for the tempered Pareto model, in contrast to the simple Pareto fit which will overfit tail probabilities and quantiles. This can be seen from Figure 17 where for larger k, the classical Weissman estimates $\hat{Q}_{1/(cn),k}^H$ ($c = 1, 2$) lead to much larger estimates than those based on the proposed tempering modelling. Only when k is really small, that is, when restricting to the data situated in the
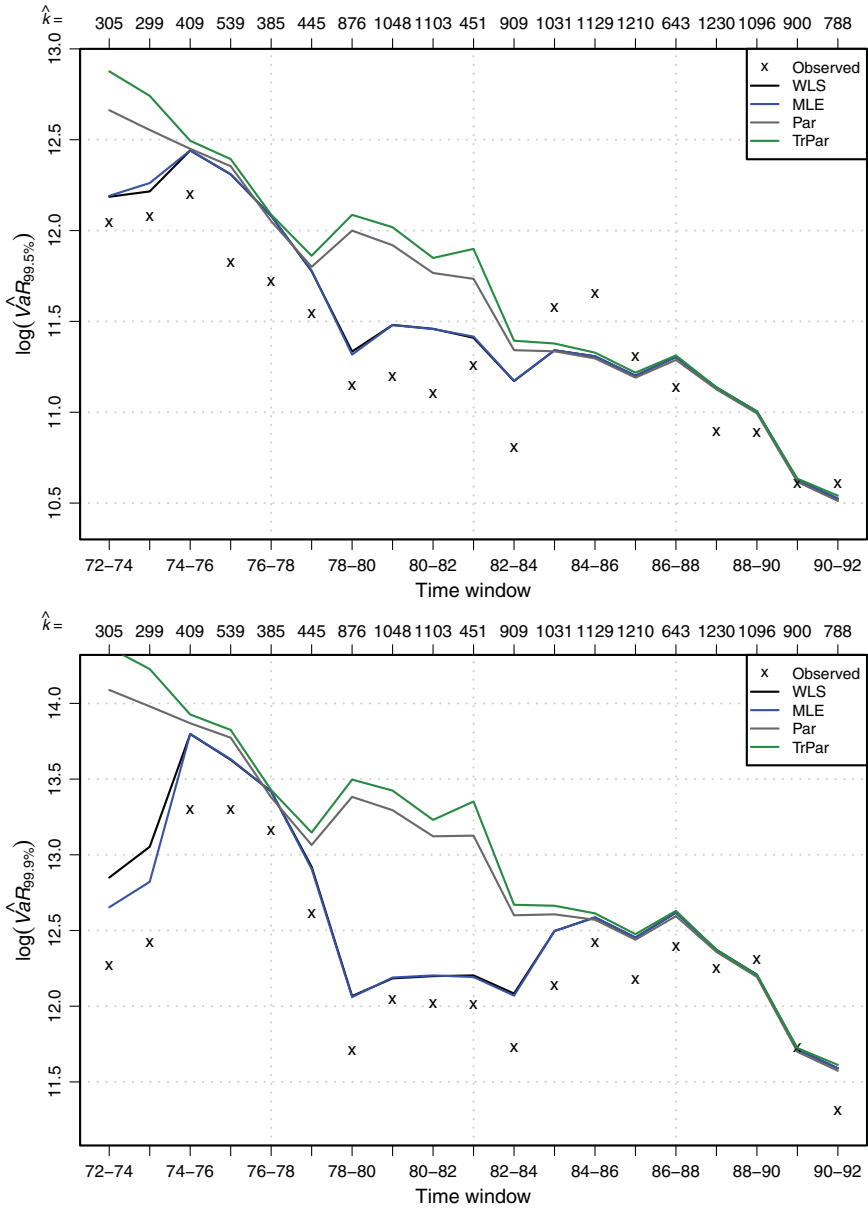
FIGURE 18: *Norwegian fire insurance data*: log-VaR(99.5%) *(top)* and log-VaR(99.5%) *(bottom)* at $\hat{k}$ for tempered model (black and dark grey lines), Pareto (dashed grey), truncated Pareto (dotted grey) and observed values (x). For each time window, $\hat{k}$ is displayed at the top margin.
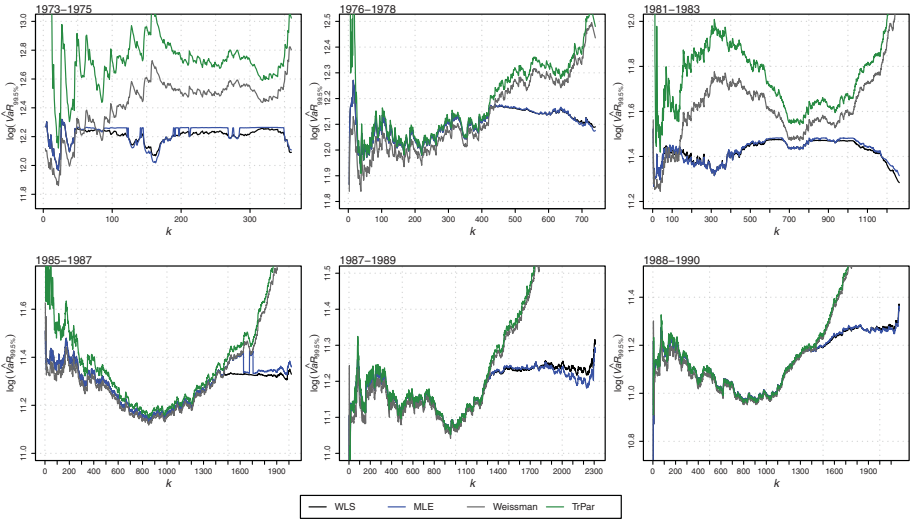
FIGURE 19: *Norwegian fire insurance data*: log VaR(99.5%) for tempered model (black and dark grey lines), Pareto (dashed grey) and truncated Pareto (dotted grey) for selected time windows.
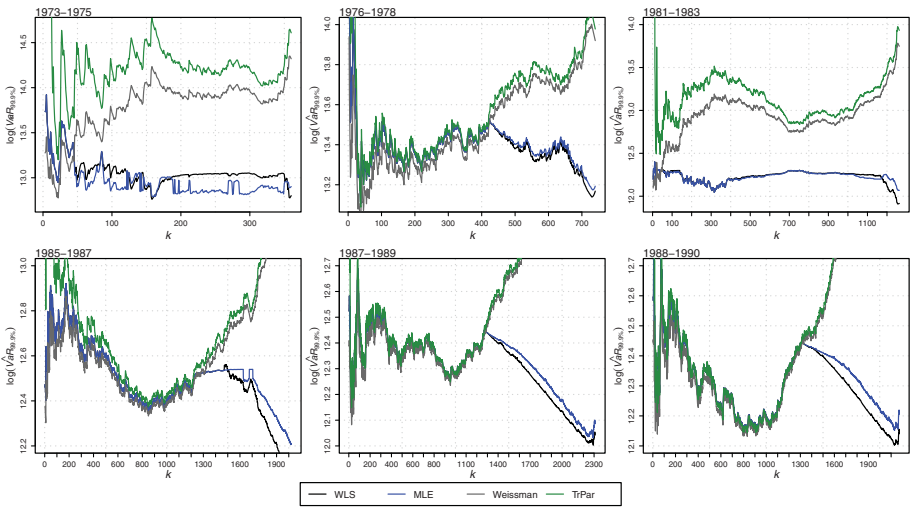

FIGURE 20: *Norwegian fire insurance data*: log VaR(99.9%) for tempered model (black and dark grey lines), Pareto (dashed grey) and truncated Pareto (dotted grey) for selected time windows.

bottom curved area of the log–log plot, the classical linear Pareto fit is able to provide a reasonable representation of the most extreme data. Finally, note from the log–log plot in Figure 16 that the truncated Pareto fit follows the linear Pareto fit except for the two final extreme points after which a sharp deviation is observed up to an estimated finite truncation point T estimated at
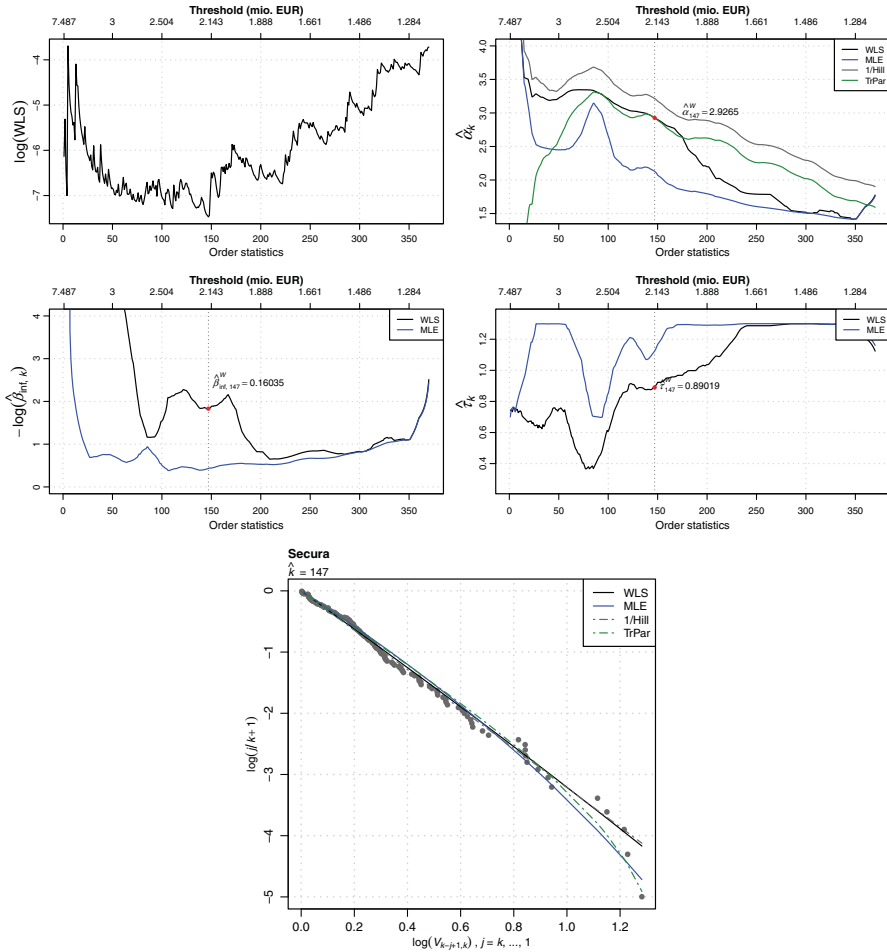
FIGURE 21: *Secura data set:* Top left: SS_k from (2.5); Top right: $\hat{\alpha}_k^W$, $\hat{\alpha}_k^M$, $H_{k,n}$ and $\hat{\alpha}_k^T$; Middle left: $-\log\hat{\beta}_{\infty,k}^W$, $-\log\hat{\beta}_{\infty,k}^M$; Middle right: $\hat{\tau}_k^W$, $\hat{\tau}_k^M$; Bottom: log-log plot with fit obtained from (2.2) with $k = \hat{k} = 147$ using MLE and WLS estimates, next to Pareto and truncated Pareto fit.

$\hat{T}_{\hat{k}} = 1,211,106$, when using the estimation method proposed in Beirlant *et al.* (2016, Section 3, Equation (19)).

In order to illustrate the possibility of extending the proposed method in a time-dependent regression context, we fitted the approach to 3-year sliding time windows. The size of the windows was selected to have at least 300 observations at each point in time. Figure 18 shows the estimated VaR at 99.5% (top) and 99.9% (bottom) using the tempered Pareto approach with $\hat{k}$ selected using the proposed adaptive procedure, next to simple Pareto and truncated Pareto modelling. We also compare with the observed quantiles obtained using the standard R function, which estimates the quantiles as weighted averages of consecutive order statistics. The VaR values based on
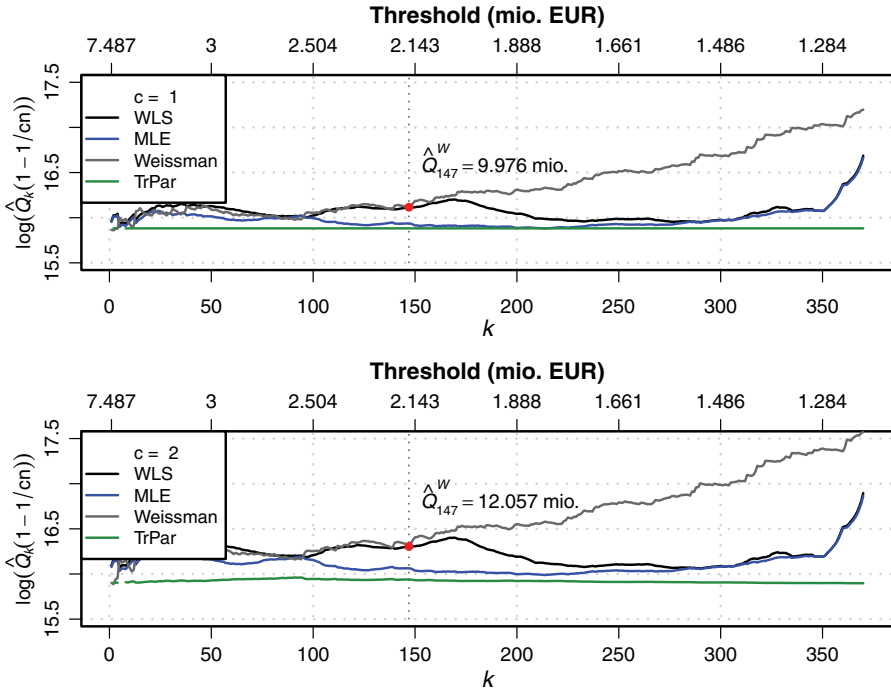
FIGURE 22: *Secura data set*: $\hat{Q}_{p,k}^W$, $\hat{Q}_{p,k}^M$ and $\hat{Q}_{p,k}^H$ quantile estimates with $p = 1/n$ (top) and $p = 1/(2n)$ (bottom).

the tempered Pareto model are situated between the observed and the Pareto and truncated Pareto fits, from which one can conclude that the tempered tail behaviour observed for the complete data set in the bottom frame in Figure 16 is also present conditional on a time window, leading to overestimation when using classical methods that ignore the proposed tempering. It is also worth noticing that the VaR at 99.5% values exhibits an overall decreasing trend with some stable behaviour between 1979 and 1987. Figures 19 and 20 show the respective VaR estimates for all values of k for some selected time windows.

In Figure 21, the respective results are given for the Secura Re Belgium data set. Here, the best tempered Pareto fit is found at $\hat{k} = 147$, with the corresponding log–log plot given in the bottom figure. Here, the tempered Pareto WLS fit closely follows the linear Pareto fit, while the MLE fit shows too much bending near the largest data. Both the Pareto and WLS tempered Pareto fit do miss the deviation at the top two data, which however is taken into account in the truncated Pareto analysis with $\hat{T}_{\hat{k}} = 8,967,620 = e^{16.009}$. While this deviation can be considered as statistically non-significant, it makes sense to consider the truncated Pareto fit here since Belgian car insurance contracts do show explicit upper limits. Another motivation for a truncated model is that the extreme quantile estimates $\hat{Q}_{1/(cn),\hat{k}}^M$ hardly change from $c = 1$ to $c = 2$, namely around the value e^{16}, see Figure 22.

5. Conclusion

In this paper, we addressed the fitting of Pareto-type distributions with a tempering component of Weibull type at large values. We extend earlier results for exponential tempering on strict Pareto tails, provide a POT approach, develop estimation procedures and provide asymptotic properties of the proposed estimators. Finally, we present a simulation study and also apply the developed methods to actual insurance data, discussing challenges in the implementation and how to overcome them. The estimation of VaR values at extreme quantile levels shows improvements compared to more classical extreme value estimation methods that ignore the considered tempering effect. These improvements are more pronounced with growing tempering effect.

Further research concerning the generalization to a regression context and the use of tempered Pareto-Weibull models in composed or splicing models will be taken up in the future.

Acknowledgements

References

ABAN, I.B., MEERSCHAERT, M.M. and PANORSKA, A.K. (2006) Parameter estimation for the truncated Pareto distribution. *Journal of the American Statistical Association*, **101**(473), 270–277.

ALBRECHER, H., BEIRLANT, J. and TEUGELS, J. (2017) *Reinsurance: Actuarial and Statistical Aspects*. Wiley Series in Probability and Statistics. Chichester: John Wiley & Sons.

ALBRECHER, H., BLADT, M. and BLADT, M. (2020) Matrix Mittag-Leffler distributions and modeling heavy-tailed risks. *Extremes*, **23**(3), 425–450.

BEIRLANT, J., FRAGA ALVES, M.I. and GOMES, I. (2016) Tail fitting for truncated and non-truncated Pareto-type distributions. *Extremes*, **19**(3), 429–462.

BEIRLANT, J., FRAGA ALVES, M.I. and REYNKENS, T. (2017) Fitting tails affected by truncation. *Electronic Journal of Statistics*, (1), 2026–2065.

BEIRLANT, J., GOEGEBEUR, Y., SEGERS, J. and TEUGELS, J. (2006) *Statistics of Extremes: Theory and Applications*. Chichester: John Wiley & Sons.

BEIRLANT, J., TEUGELS, J. and VYNCKIER, P. (1996) *Practical Analysis of Extreme Values*. Leuven: Leuven University Press.

BRAZAUSKAS, V. and KLEEFELD, A. (2016) Modeling severity and measuring tail risk of Norwegian fire claims. *North American Actuarial Journal*, **20**(1), 1–16.

HILL, B.M. (1975) A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, **3**(5), 1163–1174.

MEERSCHAERT, M.M., ROY, P. and SHAO, Q. (2012) Parameter estimation for exponentially tempered power law distributions. *Communications in Statistics. Theory and Methods*, **41**(10), 1839–1856.

NAIR, J., WIERMAN, A. and ZWART, B. (2020) *The Fundamentals of Heavy Tails: Properties, Emergence, and Estimation*. Preprint, California Institute of Technology.

RASCHKE, M. (2020) Alternative modelling and inference methods for claim size distributions. *Annals of Actuarial Science*, **14**(1), 1–19.

REYNKENS, T., VERBELEN, R., BEIRLANT, J. and 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.

WEISSMAN, I. (1978) Estimation of parameters and large quantiles based on the *k* largest observations. *Journal of the American Statistical Association*, **73**(364), 812–815.

HANSJÖRG ALBRECHER
Department of Actuarial Science
Faculty of Business and Economics
University of Lausanne
Lausanne, Switzerland
Swiss Finance Institute
Lausanne, Switzerland
E-Mail: hansjoerg.albrecher@unil.ch

JOSÉ CARLOS ARAUJO-ACUNA (Corresponding author)
Department of Actuarial Science
Faculty of Business and Economics
University of Lausanne
Lausanne, Switzerland
E-Mail: josecarlos.araujoacuna@unil.ch

JAN BEIRLANT
Department of Mathematics
LStat and LRisk
KU Leuven
Leuven, Belgium
Department of Mathematical Statistics and Actuarial Science
University of the Free State
Bloemfontein, South Africa
E-Mail: jan.beirlant@kuleuven.be

APPENDIX A. PROOF OF THEOREM 2.1

Using Taylor expansions of the likelihood equations in $\hat{\boldsymbol{\theta}}_t$ around the correct value $\boldsymbol{\theta}$ leads to the following system of three equations, with $\tilde{\boldsymbol{\theta}} = (\tilde{\alpha}, \tilde{\lambda}, \tilde{\tau})$ situated in between $\hat{\boldsymbol{\theta}}_t$ and $\boldsymbol{\theta}$:

$$\sqrt{n\overline{F}(t)}(\widehat{\alpha}_t - \alpha)\frac{1}{n\overline{F}(t)}\sum_{j=1}^{n}\frac{1}{\left(\tilde{\alpha} + \tilde{\lambda}\tilde{\tau}\left(\frac{X_j}{t}\right)^{\tilde{\tau}}\right)^2}1_{(X_j > t)}$$

$$+ \sqrt{n\overline{F}(t)}(\hat{\lambda}_t - \lambda)\frac{1}{n\overline{F}(t)}\sum_{j=1}^{n}\frac{\tilde{\tau}\left(\frac{X_j}{t}\right)^{\tilde{\tau}}}{\left(\tilde{\alpha} + \tilde{\lambda}\tilde{\tau}\left(\frac{X_j}{t}\right)^{\tilde{\tau}}\right)^2}1_{(X_j > t)}$$

$$+ \sqrt{n\overline{F}(t)}(\hat{\tau}_t - \tau) \frac{1}{n\overline{F}(t)} \sum_{j=1}^{n} \tilde{\lambda} \frac{\left(\frac{X_j}{t}\right)^{\tilde{\tau}} \left(1 + \tilde{\tau} \log \frac{X_j}{t}\right)}{\left(\tilde{\alpha} + \tilde{\lambda}\tilde{\tau}\left(\frac{X_j}{t}\right)^{\tilde{\tau}}\right)^2} 1_{(X_j > t)}$$

$$= \sqrt{n\overline{F}(t)} \left(\frac{1}{n\overline{F}(t)} \sum_{j=1}^{n} \left\{ \frac{1}{\alpha + \lambda\tau\left(\frac{X_j}{t}\right)^{\tau}} - \log \frac{X_j}{t} \right\} 1_{(X_j > t)} \right) \tag{A1}$$

$$\sqrt{n\overline{F}(t)} \left(\widehat{\alpha}_t - \alpha\right) \frac{1}{n\overline{F}(t)} \sum_{j=1}^{n} \frac{\tilde{\tau}\left(\frac{X_j}{t}\right)^{\tilde{\tau}}}{\left(\tilde{\alpha} + \tilde{\lambda}\tilde{\tau}\left(\frac{X_j}{t}\right)^{\tilde{\tau}}\right)^2} 1_{(X_j > t)}$$

$$+ \sqrt{n\overline{F}(t)} \left(\hat{\lambda}_t - \lambda\right) \frac{1}{n\overline{F}(t)} \sum_{j=1}^{n} \frac{\tilde{\tau}^2\left(\frac{X_j}{t}\right)^{2\tilde{\tau}}}{\left(\tilde{\alpha} + \tilde{\lambda}\tilde{\tau}\left(\frac{X_j}{t}\right)^{\tilde{\tau}}\right)^2} 1_{(X_j > t)}$$

$$+ \sqrt{n\overline{F}(t)} \left(\hat{\tau}_t - \tau\right) \frac{1}{n\overline{F}(t)} \sum_{j=1}^{n} \left(\frac{\tilde{\alpha}\left(\frac{X_j}{t}\right)^{\tilde{\tau}} \left(1 + \tilde{\tau}\log\frac{X_j}{t}\right)}{\left(\tilde{\alpha} + \tilde{\lambda}\tilde{\tau}\left(\frac{X_j}{t}\right)^{\tilde{\tau}}\right)^2} - \left(\frac{X_j}{t}\right)^{\tilde{\tau}} \log\frac{X_j}{t} \right) 1_{(X_j > t)}$$

$$= \sqrt{n\overline{F}(t)} \left(\frac{1}{n\overline{F}(t)} \sum_{j=1}^{n} \left\{ \frac{\tau\left(\frac{X_j}{t}\right)^{\tau}}{\alpha + \lambda\tau\left(\frac{X_j}{t}\right)^{\tau}} - \left(\frac{X_j}{t}\right)^{\tau} + 1 \right\} 1_{(X_j > t)} \right) \tag{A2}$$

$$\sqrt{n\overline{F}(t)} \left(\widehat{\alpha}_t - \alpha\right) \frac{1}{n\overline{F}(t)} \sum_{j=1}^{n} \frac{\tilde{\lambda}\left(\frac{X_j}{t}\right)^{\tilde{\tau}} \left(1 + \tilde{\tau}\log\frac{X_j}{t}\right)}{\left(\tilde{\alpha} + \tilde{\lambda}\tilde{\tau}\left(\frac{X_j}{t}\right)^{\tilde{\tau}}\right)^2} 1_{(X_j > t)}$$

$$+ \sqrt{n\overline{F}(t)}(\hat{\lambda}_t - \lambda) \frac{1}{n\overline{F}(t)} \sum_{j=1}^{n} \left(\frac{\tilde{\alpha}\left(\frac{X_j}{t}\right)^{\tilde{\tau}} \left(1 + \tilde{\tau}\log\frac{X_j}{t}\right)}{\left(\tilde{\alpha} + \tilde{\lambda}\tilde{\tau}\left(\frac{X_j}{t}\right)^{\tilde{\tau}}\right)^2} - \left(\frac{X_j}{t}\right)^{\tilde{\tau}} \log\frac{X_j}{t} \right) 1_{(X_j > t)}$$

$$+ \sqrt{n\overline{F}(t)} \left(\hat{\tau}_t - \tau\right) \frac{\tilde{\lambda}}{n\overline{F}(t)} \sum_{j=1}^{n} \left(\frac{\tilde{\lambda}\left(\frac{X_j}{t}\right)^{\tilde{\tau}} \left(1 + 2\tilde{\tau}\log\frac{X_j}{t}\right) + \tilde{\alpha}\tilde{\tau}\left(\log\frac{X_j}{t}\right)^2}{\left(\tilde{\alpha} + \tilde{\lambda}\tilde{\tau}\left(\frac{X_j}{t}\right)^{\tilde{\tau}}\right)^2} \right.$$

$$\left. - \frac{2\log\frac{X_j}{t}}{\tilde{\alpha} + \tilde{\lambda}\tilde{\tau}\left(\frac{X_j}{t}\right)^{\tilde{\tau}}} + (\log\frac{X_j}{t})^2 \right) \left(\frac{X_j}{t}\right)^{\tilde{\tau}} 1_{(X_j > t)}$$

$$= \sqrt{n\overline{F}(t)} \left(\frac{\lambda}{n\overline{F}(t)} \sum_{j=1}^{n} \left\{ \frac{\left(\frac{X_j}{t}\right)^{\tau} \left(1 + \tau\log\frac{X_j}{t}\right)}{\alpha + \lambda\tau\left(\frac{X_j}{t}\right)^{\tau}} - \left(\frac{X_j}{t}\right)^{\tau} \log\frac{X_j}{t} \right\} 1_{(X_j > t)} \right) \tag{A3}$$

The coefficients of $\sqrt{n\overline{F}(t)}(\hat{\alpha}_t - \alpha)$, $\sqrt{n\overline{F}(t)}(\hat{\lambda}_t - \lambda)$ and $\sqrt{n\overline{F}(t)}(\hat{\tau}_t - \tau)$ on the left-hand sides of (A1), (A2) and (A3) now converge in probability to the corresponding elements of $\mathbf{I}$. For instance for

$$\mathbf{I}_{1,1,n,t}(\alpha, \lambda, \tau) := \frac{1}{n\overline{F}(t)} \sum_{j=1}^{n} \frac{1}{\left(\alpha + \lambda\tau\left(\frac{X_j}{t}\right)^\tau\right)^2} 1_{(X_j > t)}$$

we have

$$
\begin{aligned}
\mathbb{E}(\mathbf{I}_{1,1,n,t}(\alpha, \lambda, \tau)) &= -\int_t^\infty \frac{1}{\left(\alpha + \lambda\tau\left(\frac{x}{t}\right)^{\tilde{\tau}}\right)^2} d\frac{\overline{F}(x)}{\overline{F}(t)} \\
&= -\int_1^\infty \frac{1}{(\alpha + \lambda\tau u^\tau)^2} d\overline{F}_t(u) \\
&\to -\int_1^\infty \frac{1}{(\alpha + \lambda\tau u^\tau)^2} d\overline{F}_{\alpha,\lambda,\tau}(u) = I_{1,1},
\end{aligned}
$$

as $t \to \infty$ using the consistency of maximum likelihood estimators and assumption $(\mathcal{M})$. The convergence of $\mathbf{I}_{1,1,n,t}(\tilde{\alpha}, \tilde{\lambda}, \tilde{\tau})$ to $I_{1,1}$ then follows from

$$\text{Var}\left(\mathbf{I}_{1,1,n,t}(\alpha, \lambda, \tau)\right) = O\left(\left(n\overline{F}(t)\right)^{-1}\right) \text{ and } \mathbf{I}_{1,1,n,t}\left(\tilde{\alpha}, \tilde{\lambda}, \tilde{\tau}\right) - \mathbf{I}_{1,1,n,t}(\alpha, \lambda, \tau) = o_p(1)$$

as $n, t \to \infty$ using the consistency of the maximum likelihood estimators.

Next the asymptotic normal distribution of the right-hand sides of (A1)–(A3)

$$
\begin{aligned}
\sqrt{n\overline{F}(t)}\left(\frac{1}{n\overline{F}(t)} \sum_{j=1}^{n} \left\{\frac{1}{\alpha + \lambda\tau\left(\frac{X_j}{t}\right)^\tau} - \log\frac{X_j}{t}\right\} 1_{(X_j > t)},\right. \\
\frac{1}{n\overline{F}(t)} \sum_{j=1}^{n} \left\{\frac{\tau\left(\frac{X_j}{t}\right)^\tau}{\alpha + \lambda\tau\left(\frac{X_j}{t}\right)^\tau} - \left(\frac{X_j}{t}\right)^\tau + 1\right\} 1_{(X_j > t)}, \\
\left.\frac{\lambda}{n\overline{F}(t)} \sum_{j=1}^{n} \left\{\frac{\left(\frac{X_j}{t}\right)^\tau \left(1 + \tau\log\frac{X_j}{t}\right)}{\alpha + \lambda\tau\left(\frac{X_j}{t}\right)^\tau} - \left(\frac{X_j}{t}\right)^\tau \log\frac{X_j}{t}\right\} 1_{(X_j > t)}\right)
\end{aligned}
\tag{A4}
$$

is derived.

Concerning the first component

$$
\begin{aligned}
\frac{1}{n\overline{F}(t)}\mathbb{E}\left(\sum_{j=1}^{n} \left\{\frac{1}{\alpha + \lambda\tau\left(\frac{X_j}{t}\right)^\tau} - \log\frac{X_j}{t}\right\} 1_{(X_j > t)}\right) &= -\frac{1}{\overline{F}(t)}\int_t^\infty \left\{\frac{1}{\alpha + \beta_\infty\frac{x}{t}} - \log\left(\frac{x}{t}\right)\right\} d\overline{F}(x) \\
&= -\int_1^\infty \left\{\frac{1}{\alpha + \lambda\tau u^\tau} - \log u\right\} d\overline{F}_t(u),
\end{aligned}
$$

with $\overline{F}_t(u) = \mathbb{P}(X/t > u | X > t) = u^{-\alpha}(1 + Dt^\rho h_\rho(u))e^{-\lambda(u^\tau - 1)}$ using the second-order slow variation condition (2.8), so that

$$-\frac{d\overline{F}_t(u)}{du} = u^{-\alpha-1}e^{-\lambda(u^\tau-1)}(\alpha + \lambda\tau u^\tau) + Dt^\rho u^{-\alpha-1}e^{-\lambda(u^\tau-1)}\{h_\rho(u)[\alpha + \lambda\tau u^\tau] - u^\rho\}.$$

Using partial integration, one easily checks that

$$\int_1^\infty \left\{ \frac{1}{\alpha + \lambda\tau u^\tau} - \log u \right\} u^{-\alpha-1}e^{-\lambda(u^\tau-1)}(\alpha + \lambda\tau u^\tau)\,du = 0,$$

so that the expected value of the first component is given by $Dt^\rho\, b_1$, leading to the asymptotic bias expression of $\hat{\alpha}_t$ as given in Theorem 2.1, and similar calculations lead to the bias of $\hat{\lambda}_t$ and $\hat{\tau}_t$.

So it remains to derive the asymptotic variances and covariances of the vector in (A4). The variance of the first component is derived from

$$\frac{1}{n\overline{F}(t)} \sum_{j=1}^n \mathbb{E}\left\{ \frac{1}{\alpha + \lambda\tau\left(\frac{X_j}{t}\right)^\tau} - \log\frac{X_j}{t} \right\}^2 1_{(X_j>t)}$$

$$= \frac{1}{\overline{F}(t)} \mathbb{E}\left(\left\{ \frac{1}{\alpha + \lambda\tau\left(\frac{X}{t}\right)^\tau} - \log\frac{X}{t} \right\}^2 1_{(X>t)} \right)$$

$$= -\int_1^\infty \left(\frac{1}{(\alpha + \lambda\tau u^\tau)^2} - \frac{2\log u}{\alpha + \lambda\tau u^\tau} + (\log u)^2 \right) d\overline{F}_t(u)$$

$$\to -\int_1^\infty \left(\frac{1}{(\alpha + \lambda\tau u^\tau)^2} - \frac{2\log u}{\alpha + \lambda\tau u^\tau} + (\log u)^2 \right) du^{-\alpha}e^{-\lambda(u^\tau-1)},$$

as $n, t \to \infty$. Using partial integration, one finds that $\int_1^\infty \left(\frac{2\log u}{\alpha + \lambda\tau u^\tau} - (\log u)^2\right) du^{-\alpha}e^{-\lambda(u^\tau-1)} = 0$, so that the asymptotic variance of the first component in (A4) equals $\mathbf{I}_{1,1}$. In the same way, one finds that the asymptotic variance covariance matrix of (A4) equals $\mathbf{I}$.
Hence

$$(\mathbf{I} + o_p(1))\sqrt{n\overline{F}(t)}(\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}) = \mathcal{N}_3\left((Dv)\mathbf{b}, \mathbf{I}\right) + o_p(1), \tag{A5}$$

from which the result follows.