Institute of Insurance Economics

University of St.Gallen

EXTREME VALUE MIXTURE MODELS: ARE THEY ABLE TO COPE WITH EXTREME INSURANCE LOSSES

DANIELA LAAS

WORKING PAPERS ON RISK MANAGEMENT AND INSURANCE NO. 174

EDITED BY HATO SCHMEISER

CHAIR FOR RISK MANAGEMENT AND INSURANCE

APRIL 2016

Extreme Value Mixture Models: Are they Able to Cope with Extreme Insurance Losses?

Daniela Laas

This version: April 4th 2016

Abstract

This paper provides a critical analysis of extreme value mixture models for the approximation of heavy-tailed insurance loss distributions. In a Bayesian framework, we consider various extreme value mixture models with different degrees of complexity. Based on an empirically calibrated distribution of windstorm losses, we perform a comprehensive simulation study and examine the accuracy of the tail approximation, extrapolation to out-of-sample quantiles, variation of results across different data sets, and influence of outliers. Moreover, the models are applied to samples of U.S. hurricane losses and fire insurance losses reported by Aon Re Belgium. Our analysis shows a substantial variation of the results across different samples and a substantial impact of extreme observations. In comparison to the peaks-over-threshold approach, the extreme value mixture model approach leads to a good tail fit for the distribution of windstorm losses. For the samples of hurricane and fire losses, in contrast, the estimation errors are substantially higher than under the peaks-over-threshold method. The tail fit is not robust with regard to the adequacy of the bulk model distribution and the choice of an inappropriate bulk component can lead to very low thresholds and biased tail estimates.

Keywords: Extreme Value Mixture Models, Peaks-Over-Threshold-Approach, Bayesian Analysis, Extreme Insurance Losses, Heavy Tails, Simulation Study

1 Introduction

The risk of extreme losses due to natural disasters or other extreme events (e.g., fire, medical malpractice, business interruption) results in heavy-tailed loss distributions in various insurance lines. An exact estimation of the probability of large losses is important for economic capital calculations, reinsurance pricing, and pricing of insurance derivatives (e.g., cat bonds), but very difficult because only very few extreme observations are available. Since the publication of McNeil (1997), the peaks-over-threshold (POT) approach of classical extreme value theory has become a popular method for the approximation of the tails of unknown distributions, especially in the case of heavy tails. Under the POT approach, the distribution of excesses over a high threshold is approximated by a generalized pareto distribution (GPD). While theoretically founded, the approach is often difficult to implement because of the need to choose an adequate threshold. In many cases, it is impossible to identify a unique threshold and the tail fit strongly depends on the chosen threshold (see, e.g., Scarrott and MacDonald, 2012).

In the meantime, numerous threshold selection methods have been proposed (for an overview, see Scarrott and MacDonald, 2012). One major class is formed by the so-called extreme value mixture models (EVMMs). These model the whole distribution by combining a certain distribution for the non-extreme range (the *bulk*, see, e.g., Scarrott and MacDonald, 2012) and a GPD for the tail. The threshold is an additional parameter that separates the central part and the tail and is estimated from the data together with the remaining parameters. Thus, the threshold is not fixed ex ante and not the result of the subjective choice of the statistician (see, e.g., Scarrott and MacDonald, 2012).

The development of the EVMM approach has been largely driven by the work of Behrens et al. (2004). They proposed one of the first EVMMs, a combination of a gamma distribution and the GPD. Subsequently, various alternatives with different or more flexible bulk models were introduced (see, e.g., Tancredi et al., 2006, MacDonald et al., 2011, Do Nascimento et al., 2012, Fúquene Patiño, 2015). Moreover, alternative forms of EVMMs were developed, such as mixtures with a GPD for both the upper and the lower tail (see, e.g., Zhao et al., 2010, and MacDonald et al., 2013), mixtures with a pareto tail (instead of a GPD tail, see, e.g., Cooray and Ananda, 2005, Scollnik, 2007, Scollnik and Sun, 2012), or the mixture of hybrid paretos by Carreau and Bengio (2009). Very comprehensive and structured overviews of the whole range of EVMMs are given by Scarrott and MacDonald (2012) and Scarrott (2016).

The accuracy of the EVMM approximation is shown by a series of simulation studies and applications to empirical data sets. The simulation studies are based on samples generated by the mixture models themselves (see, e.g., Behrens et al., 2004) or from some standard parametric distributions (e.g., gamma, inverse gamma, chi-squared in MacDonald et al., 2013), which are not necessarily representative for insurance loss distributions. In addition, empirical applications to insurance losses are scarce and the majority of extreme value mixture models are only applied to the well-known data set of Danish fire insurance losses (see, e.g., Scollnik, 2007, Bee, 2012, Lee et al., 2012) and a sample of reinsured automobile losses (see Wong and Li, 2010, and Laas, 2016). There are numerous analyses for environmental data sets (see, e.g., Tancredi et al., 2006, Do Nascimento et al., 2012, Oumow et al., 2012, Fúquene Patiño, 2015), but the distributions of several environmental variables are not as heavy-tailed as typical insurance loss distributions for claims from fire and natural catastrophes (see, e.g., Reiss and Thomas, 2007, p. 355, Schmutz and Doerr, 1998, p. 12, and the estimated tail indices in the aforementioned papers). Moreover, various papers deal with financial returns (see, e.g., Cabras and Castellanos, 2009, Zhao et al., 2010,

Zhao et al., 2011, So and Chan, 2014), but the possibility of negative returns involves the choice of bulk models with unlimited support (in contrast to a bounded support for insurance loss distributions). The application to financial returns further requires an extended EVMM approach that accounts for the high temporal dependence via a GARCH model or an alternative method (see, e.g., Zhao et al., 2010, Zhao et al., 2011, So and Chan, 2014).

Comparisons between different EVMMs are also scarce. The most comprehensive comparison is given in the paper of Hu (2013). He also examines the effect of the model flexibility in terms of the introduction of a continuity constraint and the definition of the tail fraction. However, his analysis does not include mixtures with lognormal bulk models and the approach of Cabras and Castellanos (2011). In addition, due to several shortcomings of the maximum likelihood (ML) estimation of Hu (2013) (see Scarrott, 2016), further analyses are necessary. The ML method has the disadvantage that it provides only one single estimate for the threshold (see, e.g., Scarrott and MacDonald, 2012). Thus, the plausibility of various thresholds, which is typical for EVMMs and the POT approach, is neglected (see, e.g., Scarrott and MacDonald, 2012, and Scarrott, 2016). The Bayesian approach, in contrast, leads to a full distribution of possible thresholds (see, e.g., Scarrott and MacDonald, 2012). In addition, in many cases, the optimisation of the likelihood is difficult and may only lead to a local maximum (see, e.g., Scarrott, 2016). This problem may especially arise in the study of Hu (2013), as he only considers the 90% quantile as initial value for the threshold in the optimisation (see also Scarrott, 2016).

A potential problem of the EVMM approach is the mutual dependence of the fit of the bulk and tail models (see Scarrott, 2016). If the bulk model is inadequate, the estimate for the threshold might be rather low and the tail approximation unsatisfactory. In order to check whether this is a relevant issue for insurance loss distributions, further analyses are necessary. As the estimation of extreme value mixture models requires substantial efforts, the EVMM approach is only worthwhile if it leads to a considerably higher accuracy than alternative, less complex methods. We therefore apply the EVMM method to two empirical data sets of U.S. hurricane losses and fire insurance claims (reported by Aon Re Belgium) and compare the resulting fit to the fit obtained via other methods (the POT approach and standard parametric approximations). To the knowledge of the author, the EVMM approach has not been applied to hurricane losses so far. Moreover, the distribution of Aon Re Belgium fire insurance losses seems to involve a considerably higher risk of very extreme claims than the Danish fire insurance loss distribution. We further perform a comprehensive simulation study based on a realistic distribution of European windstorm losses that was generated by Risk Management Solutions, Inc., by means of a catastrophe model.[1] This way, we combine the benefits of an empirical analysis and a simulation study and get further insights into the effects of the sample size, extrapolation, effects of outliers, and variability of the results.

Within a Bayesian framework, we examine and compare various EVMMs with different complexities of the bulk distribution. In particular, the following bulk models are taken into account: a gamma distribution as proposed by Behrens et al. (2004), a lognormal distribution (see, e.g., Scollnik, 2007, and Cabras and Castellanos, 2011), a weibull distribution (see, e.g., Cabras and Castellanos, 2011, and Scollnik and Sun, 2012), the boundary corrected kernel density proposed in MacDonald et al. (2013), and the semiparametric density estimator in Cabras and Castellanos (2011). The models are considered both under a continuity constraint for the mixture density and without the requirement of continuity. In addition, different assumptions with regard to the tail fraction are analysed.

[1] I am very grateful to RMS and Laurent Marescot for providing the data. For the disclaimer, see Appendix B.

The main result of our analysis is that the accuracy in the tail approximation under the EVMM approach varies substantially between different applications. Compared to the POT method with a "naive" selection of the threshold, the EVMM approach permits a good tail estimation and extrapolation for the windstorm losses, but fails for the hurricane and fire losses (at least for the models considered here). Although a more flexible model parameterisation can slightly improve the fit, it cannot compensate the inadequacy of the bulk model. Moreover, in the case of a relatively small sample, we observe a great impact of the composition of the data set at hand and the existence of outliers on the estimation results.

Our paper contributes to the aforementioned literature on EVMMs and the literature on the approximation of heavy-tailed insurance loss distributions (see, e.g., Cummins et al., 1990, Bolancé et al., 2003, Balasooriya and Low, 2008, Guillen et al., 2011, Ahn et al., 2012, Eling, 2012). It provides new insights into the mechanics of the EVMM approach (e.g., its robustness with regard to outliers) that may be helpful in further developments of extreme value mixture models. In addition, our results and implications can support insurers and other potential users (e.g., regulators for the design of regulatory capital standards) in their decision whether and in which form they should implement the EVMM approach.

This paper is structured as follows: Sections 2 and 3 give a short overview of EVMMs and their Bayesian estimation. The main part of the paper, Section 4, presents the simulation study and empirical applications. Section 5 provides several implications and Section 6 contains our conclusions.

2 Theoretical Framework

2.1 The Peaks-Over-Threshold Approach

Before introducing the EVMM method, it is helpful to give a short description of the POT approach of classical extreme value theory. For this, let X be a random variable with distribution function F and upper end point $x_F := \sup\{x : F(x) < 1\} \leq \infty$ (for this section, refer to McNeil et al., 2005). In the center of the POT models is the generalized pareto distribution. Its distribution function is given by:

$$G_{\sigma,\xi}(y) = \begin{cases} 1 - (1 + \xi \cdot \frac{y}{\sigma})^{-1/\xi}, & \xi \neq 0 \\ 1 - \exp(-\frac{y}{\sigma}), & \xi = 0 \end{cases} \tag{1}$$

with $y \geq 0$ for $\xi \geq 0$ and $y \in [0, -\sigma/\xi]$ for $\xi < 0$. The parameters $\sigma > 0$ and $\xi \in \mathbb{R}$ determine the scale and shape of the GPD, respectively. It has a short tail for $\xi < 0$, an exponentially decaying tail for $\xi = 0$, and a heavy tail in the case of $\xi > 0$. Moreover, the rth moment only exists if $r < 1/\xi$.

The POT approach is based on the work of Balkema and de Haan (1974) and Pickands (1975). They show that for a large class $\mathcal{M}$ of distributions, the conditional distribution $F_u(y) := \mathbb{P}(X - u \leq y | X > u)$ of excesses of X over a threshold u converges to a GPD as the threshold is raised towards x_F. The class $\mathcal{M}$ is called Maximum Domain of Attraction of the generalized extreme value (GEV) distribution, and it essentially comprises all common continuous distributions.[2]

[2]More formally, they show the following (see, e.g., Embrechts et al., 2008, for details): Let $\mathcal{M}_\xi$ be the maximum domain of attraction of the generalized extreme value distribution with shape parameter $\xi \in \mathbb{R}$. Then $F \in \mathcal{M}_\xi$ if and only if $\lim_{u \uparrow x_F} \sup_{0<y<x_F-u} |F_u(y) - G_{\xi,\sigma(u)}(y)| = 0$ for some positive function $\sigma(u)$.

Justified by this theoretical result, under the POT method, the excess distribution of X over a high threshold u is approximated by a GPD. Thus, it is assumed that there exists a value $u < x_F$ with:

$$\mathbb{P}(X - u \leq y | X > u) = G_{\xi,\sigma_u}(y),$$

approximately for all *excesses* $0 < y < x_F - u$ and appropriate parameters $\sigma_u > 0$ and $\xi \in \mathbb{R}$, or, equivalently:

$$\mathbb{P}(X \leq x | X > u) = G_{\sigma_u,\xi}(x - u) =: G(x|u, \sigma_u, \xi),$$

approximately for all *exceedances* $x > u$. Setting $\phi_u := \mathbb{P}(X > u)$, this implies for the unconditional distribution (see, e.g., MacDonald et al., 2013):

$$\mathbb{P}(X \leq x) = (1 - \phi_u) + \phi_u \cdot G(x|u, \sigma_u, \xi). \tag{2}$$

It can be easily shown that if F_u is a GPD with shape ξ and scale σ_u, then for all higher thresholds $v > u$, the distribution of excesses over v is also a GPD with the same shape parameter ξ and scale $\sigma_v = \sigma_u + \xi \cdot (v - u)$. This property is called *threshold stability* of the GPD and shows the threshold dependence of the scale parameter.

The POT approach has the advantage that no specific distribution has to be assumed. Instead, it suffices that the unknown distribution belongs to $\mathcal{M}$. A major drawback is the need to choose an appropriate threshold u. There is frequently high uncertainty concerning the location of the threshold, and more thresholds are plausible (see, e.g., Scarrott and MacDonald, 2012). Furthermore, as shown in several applications, the estimated parameters and extreme quantiles may depend considerably on the selected threshold (see, e.g., Scarrott and MacDonald, 2012). Traditionally, graphical tools such as the mean-excess-plot or threshold-stability-plot have been used. However, decisions based on these plots are rather subjective (see, e.g., Scarrott and MacDonald, 2012). In order to resolve the threshold selection problem, a variety of methods have been proposed (for a comprehensive overview see Scarrott and MacDonald, 2012). These range from simple rules of thumb to complex resampling methods and mixture models. The latter will be described in the following section.

2.2 Extreme Value Mixture Models

2.2.1 General Form of EVMMs

Extreme value mixture models provide a way to automate the threshold selection. Although their specific forms differ, most EVMMs extend the GPD model of the POT approach to the whole range of the distribution by combining it with certain model for the non-extreme part (see, e.g., Scarrott and MacDonald, 2012, and Scarrott, 2016). The two components are connected at the threshold u, which is an additional parameter that has to be determined in the estimation process (see, e.g., Scarrott and MacDonald, 2012, and Scarrott, 2016). Thus, the distribution function can be written as (see, e.g., Scarrott, 2016):

$$F_{MM}(x|\boldsymbol{\theta}, u, \sigma_u, \xi) = \begin{cases} (1 - \phi_u) \cdot \frac{H(x|\boldsymbol{\theta})}{H(u|\boldsymbol{\theta})}, & x \leq u \\ (1 - \phi_u) + \phi_u \cdot G(x|u, \sigma_u, \xi), & x > u. \end{cases} \tag{3}$$

Here, $H(\cdot|\boldsymbol{\theta})$ denotes the distribution function for the bulk range of the data and $\boldsymbol{\theta} = (\theta_1, ..., \theta_K)$ is the corresponding parameter vector. The bulk model may be parametric, semiparametric or nonparametric

(see Scarrott and MacDonald, 2012). The parameter ϕ_u determines the tail fraction. Frequently, it is specified by the bulk model by setting $\phi_u = 1 - H(u|\boldsymbol{\theta})$ (see, e.g., Behrens et al., 2004, and Do Nascimento et al., 2012). However, some authors also treat it as an extra parameter that has to be estimated from the data (see, e.g., MacDonald et al., 2011). In the case of maximum likelihood estimation, ϕ_u then corresponds to the proportion of observations above the threshold u (see, e.g., Scarrott, 2016). Following Scarrott (2016), we refer to the two approaches as *bulk model based tail fraction* (BTF) and *parameterised tail fraction* (PTF) approach, respectively.

The densities of the whole mixture model (3), the bulk component, and the GPD are denoted with f_{MM}, h, and g, respectively. Figure 1 gives an example for the density and distribution function of a mixture model with gamma bulk component. In general, the density f_{MM} is not continuous at the threshold. As continuity seems physically reasonable (see, e.g., Scarrott and MacDonald, 2012, and Scarrott, 2016), several EVMMs include a continuity constraint (see, e.g., Scollnik, 2007, and, Carreau and Bengio, 2009). A continuous mixture density can be obtained via a restriction of the parameters of the bulk or tail model (see Scarrott, 2016). If the continuity constraint is imposed on σ_u, this parameter is given by (see Scarrott, 2016):

$$\sigma_u^c = c(\boldsymbol{\theta}, u, \phi_u) = \frac{H(u|\boldsymbol{\theta})}{1-\phi_u} \cdot \frac{\phi_u}{h(u|\boldsymbol{\theta})}. \tag{4}$$

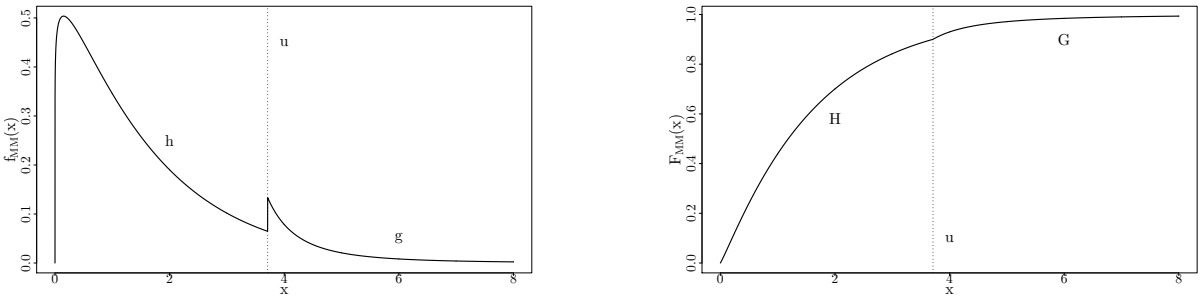


Figure 1: EVMM Density and Cumulative Distribution Function

This figure shows the density (left plot) and cumulative distribution function (right plot) for an EVMM. The bulk model is a gamma distribution with shape 1.1 and scale 1.5, the threshold u is located at the 90% quantile of the bulk distribution (i.e., $u = 3.7$), the GPD parameters are $\sigma_u = 0.75$ and $\xi = 0.5$, and the BTF approach is used.

The likelihood function for an independent sample $\boldsymbol{x} = (x_1, ..., x_n)$ from F_{MM} can be easily calculated by deriving the density from (3). For $\xi \neq 0$ (the case $\xi = 0$ is analogous), it is the product (see, e.g., Scarrott, 2016):

$$\begin{aligned} L_{MM}(\boldsymbol{\theta}, u, \sigma_u, \xi|\boldsymbol{x}) &= \prod_{i:x_i \leq u} (1-\phi_u) \cdot \frac{h(x_i|\boldsymbol{\theta})}{H(u|\boldsymbol{\theta})} \cdot \prod_{i:x_i > u} \phi_u \cdot g(x_i|u, \sigma_u, \xi) \\ &= \left(\frac{1-\phi_u}{H(u|\boldsymbol{\theta})}\right)^{n_b} \cdot (\phi_u)^{n_u} \cdot \prod_{i:x_i \leq u} h(x_i|\boldsymbol{\theta}) \cdot \prod_{i:x_i > u} \frac{1}{\sigma_u} \left(1 + \frac{\xi(x_i - u)}{\sigma_u}\right)^{-\frac{1+\xi}{\xi}}. \end{aligned} \tag{5}$$

In this formula, n_u denotes the number of threshold exceedances $x_i > u$, and $n_b = n - n_u$. If a continuity constraint is introduced via σ_u, the likelihood function is given by (see, e.g., Scarrott and Hu, 2015):

$$L_{MMC}(\boldsymbol{\theta}, u, \xi|\boldsymbol{x}) = L_{MM}(\boldsymbol{\theta}, u, \sigma_u^c, \xi|\boldsymbol{x}) \qquad \text{with} \qquad \sigma_u^c = c(\boldsymbol{\theta}, u, \phi_u). \tag{6}$$

2.2.2 Details about the Mixture Models Considered in the Paper

In the meantime, a multitude of EVMMs with various bulk distributions and of different forms have been proposed (see Scarrott, 2016). In this paper, we focus on mixtures defined by (3) and bulk models with support $\mathbb{R}^+$. In detail, the following models are considered:[3]

- Gammagpd(con) mixtures (models 1a to 1d): EVMMs with distribution function (3) and gamma bulk model H (as proposed by Behrens et al., 2004). The ending *con* marks the model with continuity constraint (4).
- Lognormgpd(con) mixtures (models 2a to 2d): The bulk component H in (3) is a lognormal distribution (see, e.g., Scollnik, 2007, Cabras and Castellanos, 2011, Bee, 2012).
- Weibullgpd(con) mixtures (models 3a to 3d): Mixture models of the form (3) with a weibull distribution H (see, e.g., Cabras and Castellanos, 2011, and Scollnik and Sun, 2012).
- Semipargpd mixture (model 4): The semiparametric mixture of Cabras and Castellanos (2011). This mixture slightly differs from the remaining models. For a given value of u, they set $\phi_u = n_u/n$, approximate the truncated density $h(x)\mathbf{1}_{\{x \leq u\}}/H(u)$ semiparametrically, and replace the likelihood function (5) by the profile likelihod (for details, see Cabras and Castellanos, 2011). The truncated bulk density is approximated by means of the method proposed by Lindsey (1974a,b) (see also Efron and Tibshirani, 1996).
- Kernelgpd(con) mixtures (models 5a to 5d): Extreme value mixture distributions as defined by (3) with a boundary corrected kernel density estimator for the bulk component H. The parameter $\boldsymbol{\theta}$ is the bandwidth λ. This model is proposed by MacDonald et al. (2013). In order to reduce the computational burden, we use a boundary correction method that does not require renormalization (the reflection method of Boneva et al., 1971).

The lowercase letter a in the model number refers to the model version with PTF approach (i.e., ϕ_u is an additional parameter) and no continuity constraint. The letter b denotes the non-continuous mixture with BTF approach (i.e., $\phi_u = 1 - H(u|\boldsymbol{\theta})$). The models with continuity constraint are referred to by c and d in the case of the PTF and BTF approach, respectively.

3 Bayesian Inference for EVMMs

3.1 Prior Distributions

Under the Bayesian framework, a whole distribution of the unknown parameters (the so-called posterior distribution) is derived based on the information provided by the data (in form of the likelihood) and prior knowledge about the parameters or related quantities (for an introduction into the Bayesian methodology, see, e.g., Gelman et al., 2013). The prior information is expressed in form of so-called prior distributions. We choose the same priors as in Laas (2016).[4]

[3] For the mixtures with gamma, lognormal, and weibull bulk components, we use the model names from the R-package evmix (see, Scarrott and Hu, 2015). The kernelgpd(con) mixture in this paper corresponds to the bckdengpd(con) model in the evmix package.

[4] In our paper, we assume that no expert information is available. Thus, the prior parameters are calibrated so as to obtain very low-informative prior distributions with high variances (see Section 4.1.2). In this case, the choice of the specific type of the prior distribution (e.g., lognormal or gamma) should not influence the posterior distribution (see Gelman et al., 2013, p. 54). We therefore rely on the prior distributions commonly used in the literature.

Priors for the Bulk Parameters The prior $\pi_B(\boldsymbol{\theta}|\boldsymbol{\eta})$ for the parameters $\boldsymbol{\theta}$ of the bulk model depends on the considered EVMM ($\boldsymbol{\eta}$ denotes the prior parameters). For the lognormgpd(con) mixture, we choose the product of a normal distribution for the mean and an inverse gamma distribution for the variance. This prior is used in several papers with Bayesian estimation of the parameters of a normal or lognormal distribution (see, e.g. Richardson and Green, 1997, and de Alba, 2006). With regard to the gammagpd(con) mixture, we follow Wiper and Rios (2001) and Do Nascimento et al. (2012) and assume a gamma prior for the shape parameter and an inverse gamma distribution for the mean (i.e., the ratio of the shape and rate). In order to estimate the posterior for the weibullgpd(con) EVMM, a gamma prior is specified for the weibull shape parameter (see, e.g., Tsionas, 2002, and Marín et al., 2005). Moreover, a normal prior is defined for the logarithm of the scale. We also try gamma priors for both weibull parameters and the vague prior derived by Sinha (1986), but this leads to convergence problems of the Markov chains for various samples. For the Bayesian estimation of the kernelgpd(con) mixture, an inverse gamma prior for λ^2 is used by MacDonald et al. (2013), as proposed by Brewer (1998) and Brewer (2000). However, this prior may lead to problems if $\lambda^2 < 0.5$ is very likely (see MacDonald, 2011). We therefore refer to Zhang et al. (2006) and assume a Cauchy distribution for the bandwidth λ. In order to take the positivity of λ into account, we truncate the Cauchy distribution at zero.

Prior for the Threshold The prior π_u for u is frequently assumed to be a normal distribution that is truncated from below at some low value, e.g., the minimum of the data (see, e.g., Behrens et al., 2004, MacDonald et al., 2011, Do Nascimento et al., 2012). To ensure the existence of sufficient data for the estimation both below and above the threshold, Cabras and Castellanos (2011) choose a uniform prior over the range $\left[x_{(K+1)}; x_{(n-2)}\right]$. Combining both approaches, we use a truncated normal $N(m_u, s_u^2)$ distribution with lower boundary $x_{(K+1)}$ and upper boundary $x_{(n-2)}$.[5]

Prior for the GPD Parameters For the EVMM without continuity constraint, a joint prior for σ_u and ξ has to be defined. In the meantime, several priors have been proposed (some examples can be found in Stephenson and Ribatet, 2006, and Northrop and Attalides, 2014). If expert knowledge is available, a common approach is the method of Coles and Tawn (1996). They derive a prior for σ_u and ξ from expert information about the extreme quantiles of the underlying distribution. In the non-informative case (assumed in our paper), Jeffrey's prior as developed by Castellanos and Cabras (2007) is a widely-used alternative (see, e.g., Do Nascimento et al., 2012, and Fúquene Patiño, 2015). It is given by:

$$\pi_T(\sigma_u, \xi) \propto \sigma_u^{-1} \cdot (1+\xi)^{-1} \cdot (1+2\xi)^{-1/2}, \qquad \xi > -0.5,\ \sigma_u > 0 \tag{7}$$

and always leads to a proper posterior density (see Castellanos and Cabras, 2007). Due to our focus on heavy-tailed distributions, the restriction $\xi > -0.5$ should be appropriate (see Villa, 2015).

If a continuity constraint is imposed, σ_u is determined by the bulk model and u. For the shape parameter ξ, we choose a normal distribution with positive mean m_ξ and variance s_ξ^2. This prior is denoted by π_{TC}.

3.2 Posterior Distribution

The posterior distribution is proportional to the product of the likelihood and the joint prior of all parameters (see, e. g., Gelman et al., 2013). Typically, the parameters of the different model components are assumed to be independent (see, e.g., Behrens et al., 2004, Cabras and Castellanos, 2011,

[5] $x_{(k)}$ denotes the kth order statistic of the sample.

Do Nascimento et al., 2012). Thus, the posterior π_{MM} of the unconstrained EVMM fulfills (see, e.g., Cabras and Castellanos, 2011):

$$\pi_{MM}(\boldsymbol{\theta}, u, \sigma_u, \xi|\boldsymbol{\eta}, m_u, s_u^2, \boldsymbol{x}) \propto L_{MM}(\boldsymbol{\theta}, u, \sigma_u, \xi|\boldsymbol{x}) \cdot \pi_B(\boldsymbol{\theta}|\boldsymbol{\eta}) \cdot \pi_u(u|m_u, s_u^2) \cdot \pi_T(\sigma_u, \xi). \tag{8}$$

In the case of the mixture with continuity constraint, we obtain:

$$\pi_{MMC}(\boldsymbol{\theta}, u, \xi|\boldsymbol{\eta}, m_u, s_u^2, m_\xi, s_\xi^2, \boldsymbol{x}) \propto L_{MMC}(\boldsymbol{\theta}, u, \xi|\boldsymbol{x}) \cdot \pi_B(\boldsymbol{\theta}|\boldsymbol{\eta}) \cdot \pi_u(u|m_u, s_u^2) \cdot \pi_{TC}(\xi|m_\xi, s_\xi^2). \tag{9}$$

For the (pseudo)posterior of the semipargpd mixture, we refer to Cabras and Castellanos (2011).

The assumption of independent priors for u and (σ_u, ξ) negates the threshold dependence of the scale parameter (see, e.g., Scarrott and MacDonald, 2012). The parameter dependence can be resolved by means of the Poisson point process approach (see, e.g., Tancredi et al., 2006, MacDonald et al., 2011, and Oumow et al., 2012) or a GPD reparameterisation (see, e.g., Scarrott and MacDonald, 2012, and Laas, 2016). According to the study by Laas (2016), the reparameterisation is not necessary for the applications considered in this paper (samples that are not from an exact EVMM density or of limited size), especially if parallel tempering is applied (see below). We therefore continue with the original model, as also done by Behrens et al. (2004) and Do Nascimento et al. (2012).

As the posteriors cannot be calculated analytically, they are approximated via Markov chain Monte Carlo (MCMC) simulation (for a detailed description of this technique, see, e.g., Gamerman and Lopes, 2006). The Markov chains are generated by means of the algorithm proposed by Metropolis et al. (1953) and Hastings (1970) (MH algorithm) and parallel tempering (PT, see, e.g., Gamerman and Lopes, 2006, and Craiu and Rosenthal, 2014). For the model without continuity constraint, our MH algorithm is similar to that of Behrens et al. (2004) and Do Nascimento et al. (2012) and described in detail in Appendix B of the paper by Laas (2016). In the case of the parameterised tail fraction approach, the chains for ϕ_u are generated indirectly from the chains for u by setting ϕ_u at the proportion of threshold exceedances, as also done by MacDonald et al. (2011) and Oumow et al. (2012). According to Laas (2016), this simplification has only negligible effects on the results and improves the convergence behaviour of the Markov chains. The MH steps for the continuous mixture are analogous to those for the model without continuity requirement. The method of PT (see Laas, 2016, Appendix B) extends the MH algorithm to better account for multimodal posterior distributions and can substantially improve the convergence and mixing behaviour of the Markov chains for the EVMM parameters (see Laas, 2016).

The Markov chains $(Y^{(t)})_t$ describing the posteriors of quantities $Y = g(\boldsymbol{\Psi})$, which are functions of the parameters $\boldsymbol{\Psi}$ (here $\boldsymbol{\Psi} = (\boldsymbol{\theta}, u, \sigma_u, \xi)$), can be directly calculated from the parameter chains $(\boldsymbol{\Psi}^{(t)})_t$.

The posterior distribution is the basis for Bayesian inference. Common point estimates for parameters and other quantities of interest (e.g., quantiles) are the mean or median of the posterior distribution of the respective quantity (see, e.g., Gelman et al., 2013). The uncertainty can be expressed in form of the range that comprises $100(1-\alpha)\%$ of the posterior probability, the so-called $100(1-\alpha)\%$ posterior interval (see, e.g., Gelman et al., 2013).

In the case of a unimodal posterior distribution that is based on uninformative priors, the Bayesian posterior mean is similar to the ML estimate (see, e.g., Condgon, 2006). The Bayesian approach is especially useful if the likelihood function has several maxima and various thresholds are plausible. In this situation, the ML estimation is not only difficult, it also neglects several conceivable thresholds (see also Scarrott and MacDonald, 2012). Under the Bayesian method, in contrast, all thresholds are taken into account according to their posterior probability. Furthermore, even though the MCMC simulation also sometimes leads to convergence problems, in our analyses it fails far less frequently than the optimisation of the loglikelihood.

4 EVMM Approximation of Insurance Loss Distributions

4.1 Simulation Study Based on the RMS Windstorm Loss Distribution

4.1.1 Data Description

Our simulation study is based on a distribution of aggregate windstorm losses in Europe that we obtained from Risk Management Solutions (RMS), Inc., one of the world's leading companies in the modeling of catastrophic risks. The distribution was generated by means of a catastrophe model (cat model), using an analysis run on an anonymized subset of territories and industry exposure in Europe. Cat models combine information on the considered peril (e.g., data on wind speeds and frequency of events for the peril *European Windstorm*) and properties at risk (e.g., location, building type) in order to derive the vulnerability of the risk portfolio and subsequently the respective loss distribution (see Grossi et al., 2005).

The loss distribution is given in form of the aggregate exceedance probability (EP) curve (see, e.g., Grossi et al., 2005). This curve gives the exceedance probabilities $\overline{F}(x_i)$ of 263 losses x_i (given in million EUR). Thus, the values of the cumulative distribution function are given by $F(x_i) = 1 - \overline{F}(x_i)$. In order to obtain a continuous distribution, we interpolate linearly between two adjacent data points. In addition, we perform linear interpolations between $(0; 0)$ and the first given data point $(6.55; 0.000184)$ and between the last given data point $(91,658.28; 0.999982)$ and $(200,000; 1)$. For the following analysis, we assume that the generated distribution function corresponds to the true one.

Figure 2 shows the generated distribution function and the histogram of a sample of size 100,000. The mean and standard deviation of the sample are 3,300 and 3,383, respectively. The distribution is skewed to the right and leptocurtic (the skewness and kurtosis of the sample are 5 and 89, respectively). Moreover, according to the Anderson-Darling and Cramer-von Mises tests for various samples of size 10,000, the whole distribution differs from a GPD. In contrast, for some sufficiently high threshold u, the null hypothesis of a GPD excess distribution cannot be rejected. The quantiles for $\alpha = 0.975, 0.99, 0.999$, and 0.9999 are 11,970, 16,328, 30,500, and 52,970, respectively.

For our analysis, we draw $N = 50$ samples of size n from the distribution of windstorm losses. Two sample sizes are considered: $n = 1,000$ and $n = 10,000$. Due to substantially higher computational efforts, the semipargpd mixture is only estimated for the samples of size $n = 1,000$. For the same reason, the models with kernel bulk density are only applied to the first 250 observations of the generated data sets.

Finally, we note that if a company indeed knows that the losses are distributed according to the RMS distribution, it obviously does not need an approximation via the EVMM approach. Our study therefore

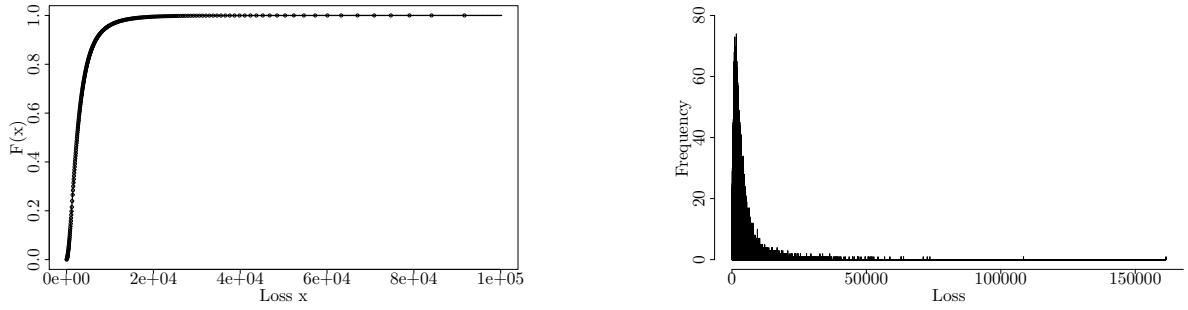


Figure 2: Distribution of European Windstorm Losses

This figure illustrates the distribution of European windstorm losses. In the subfigure on the left, the circles correspond to the data points derived from the aggregate EP curve provided by RMS. The line shows the assumed continuous distribution function. The plot on the right shows the histogram of a sample of 100,000 observations from the generated distribution.

assumes a case in which an insurer only has a limited sample of observations and approximates the tail via the EVMM approach (e.g., for the estimation of out-of-sample quantiles). However, in order to evaluate the accuracy of the tail approximation by means of a comparison of the estimated tail and the true underlying tail, the true underlying distribution has to be known.[6]

4.1.2 Calibration and MCMC Simulations

The Bayesian estimation requires the calibration of the prior distributions and the variances of the proposal distributions used under the MH and PT algorithms. In our examination, we assume that no expert knowledge is available, and therefore choose prior parameters that lead to flat diffuse priors with high variances (see also MacDonald et al., 2013, among others). For the priors of the bulk models, the parameters given in Table A.1 in Laas (2016) are used.[7] The mean m_u and standard deviation s_u of the prior for u are set at the 90% quantile of the distribution of windstorm losses and 10,000, respectively (see, Behrens et al., 2004, MacDonald et al., 2013, Laas, 2016). In the case of the models with continuity constraint, we further specify $m_\xi = 1$ and $s_\xi = 10$ for the prior of ξ.

For the calibration of the proposal variances, we use the optimal asymptotic acceptance rate of 0.23 derived by Gelman et al. (1996) and Roberts et al. (1997) as a reference point (as was done by MacDonald et al., 2011), but also allow lower or higher values if appropriate. According to Roberts and Rosenthal (2001) and Gelman et al. (1996), deviations from the optimal value do not lead to substantial loss in efficiency of the sampler or can even increase its efficiency in the case of multimodal posteriors.

Following Laas (2016), the PT algorithm is implemented for four temperatures $\tau_1 = 1$, $\tau_2 = 1.5$, $\tau_3 = 2$, and $\tau_4 = 2.5$. In addition, swaps between two chains are proposed after each $s = 10$ steps.

In order to assess the convergence of the Markov chains, several chains with overdispersed initial vectors are generated and compared both visually (as recommended by Gelman, 1996) and by means of the Gelman-Rubin statistic (see Gelman and Rubin, 1992, and Brooks and Gelman, 1998). In our final

[6] This is not necessary in the case of a goodness-of-fit test, but a statistical test has the disadvantage that in the case of a low power, a non-rejection of the null hypothesis does not necessarily imply a good fit.

[7] Table A.1 in Laas (2016) specifies the following priors: Gamma(1,0.01) and inverse gamma (1.5,5) distributions for the shape and mean of the gamma density, normal(1,1000^2) and inverse gamma (2.5,5) distributions for the parameters of the lognormal distribution, and gamma(1,0.01) and normal(0,2^2) distributions for the weibull shape and log-scale.

simulations, for each model and data set, we generate two Markov chains with both the MH algorithm and PT and subsequently merge them to one MH chain and one PT chain.[8] For the subsequent inference, we then choose the chains from the algorithm that has led to better convergence and mixing properties. While the MH sampler leads to convergence or mixing problems for various data sets, the method of PT works quite well in the case of $n = 1,000$. We only remove a few samples for which a good approximation of the posterior is not entirely certain for all mixtures. For the sample size $n = 10,000$, PT provides well-behaving Markov chains for the mixtures with weibull and gamma bulk components (with a few problems only for the gammagpd(con) models under the PTF approach). If a lognormal bulk distribution is chosen, in contrast, the Markov chains for the bulk model parameters exhibit substantial mixing problems for various data sets.[9]

For the comparison of the fit of the EVMM approach with the approximation under alternative methods, we also consider the POT method (models 6a to 6c) with three "naive" thresholds (the 85% data quantile, the 90% quantile proposed by DuMouchel, 1983, and the 95% data quantile). Moreover, we fit three simple parametric distributions (models 7 to 9) that are commonly applied to insurance losses with heavy tails (the lognormal, pareto, and burr distribution, see, e.g., Cummins et al., 1999, and Milidonis and Grace, 2008). The MH steps for σ_u and ξ under the POT approach correspond to those for the EVMM parameters. The Markov chains for the posteriors of the parametric distributions are generated in JAGS using very flat priors.

The analyses in this paper focus on the tail fit (measured by means of the deviations of the quantile estimates from the true quantiles), not the overall goodness-of-fit. A few remarks about the overall fit can be found in Sections 5.1 and 6. We do not apply goodness-of-fit tests as, besides the difficulties to implement them in the Bayesian framework, they might lead to erratic conclusions with regard to the tail fit. First, an inappropriate model for the main part of the distribution might lead to a model rejection even if the GPD provides a good tail fit. Second, a non-rejection does not necessarily imply a good approximation.

4.1.3 Results

The estimation results are illustrated in Figures 3 and 4 for $n = 10,000$ and $n = 1,000$, respectively. The boxplots in Subfigure (a) summarize the distributions of errors in the estimation of the 97.5% quantile ($Q_{0.975}$) across all data sets under the different approaches. For each sample and model, the quantile is estimated via the median of the respective posterior distribution (see, e.g., Cabras and Castellanos, 2011) and the error is calculated as the absolute value of the relative deviation from the true quantile. Analogously, Subfigures (b) to (d) show the deviations for the 99%, 99.9%, and 99.99% quantiles ($Q_{0.99}$, $Q_{0.999}$, and $Q_{0.9999}$). The two subfigures at the bottom contain the boxplots of the estimates (i.e., the posterior medians) of the shape ξ and threshold u.[10]

[8] The chain lengths differ, but all chains are very long in order to ensure a good approximation of the posterior. For example, for the samples of size $n = 1,000$ and the gammagpd(con) mixtures, each of the two MH chains consists of 75,000 iterations (excluding the burn-in period). The two PT chains have a length of 50,000 iterations.

[9] To maintain a large number of samples of size $n = 10,000$, the data sets with problems for one of the lognormgpd(con) mixtures are not removed for the estimation of the remaining models. Thus, the sample bases slightly differ.

[10] Please note: For a given sample, the quantity of interest (e.g. the shape ξ or quantile) is estimated via the posterior median, i.e., the median of the posterior distribution. The boxplot summarizes the results for the N samples by showing the minimum, median, maximum, etc. of the set of N posterior medians. Following Cabras and Castellanos (2011), we use the posterior median (not the mean), as some Markov chains include a few very extreme outliers.

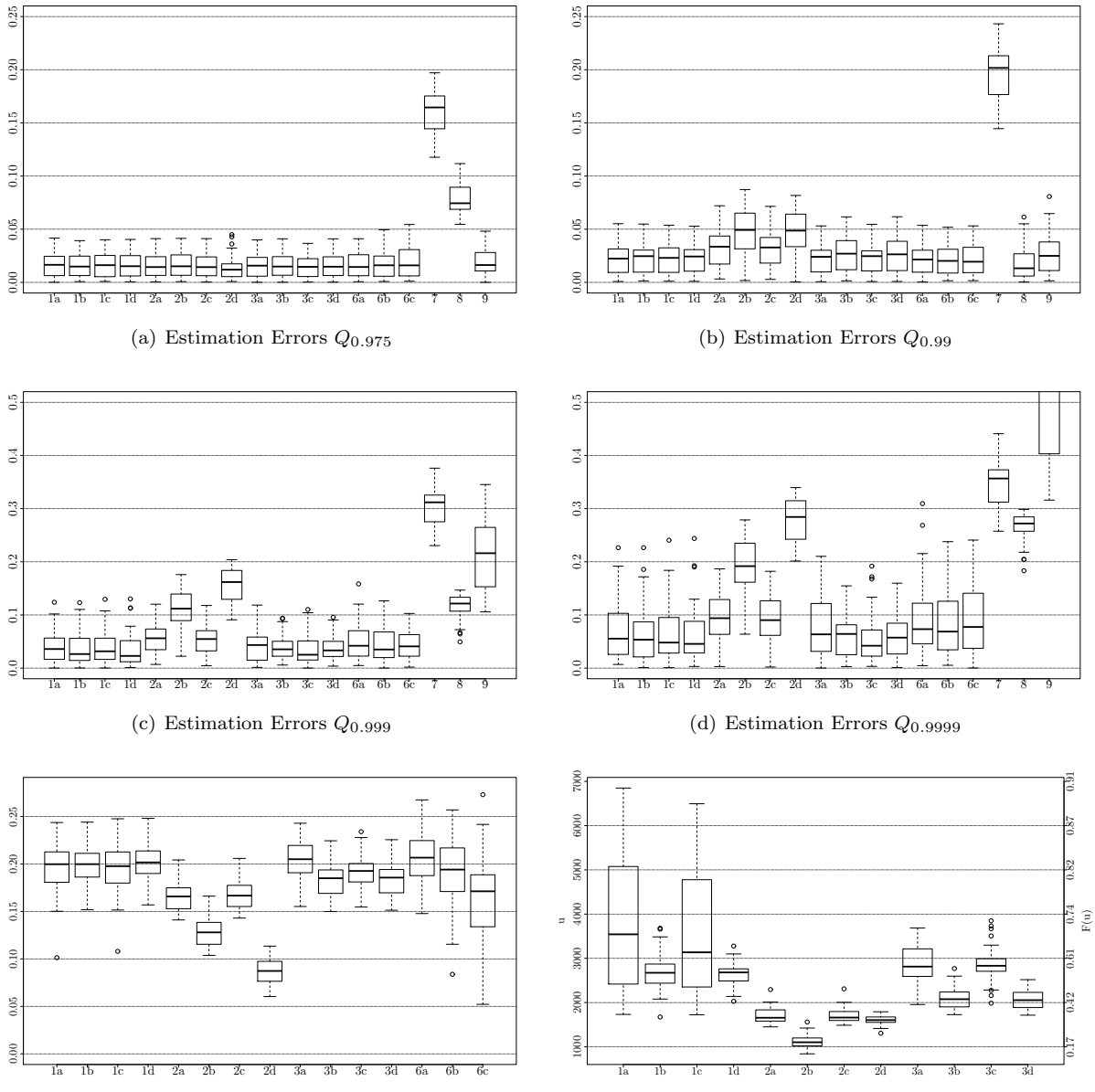
$n = 10,000$ For a large sample size, the results show a similar and good approximation of the 97.5% quantile by all EVMMs (1a) to (3d) (see Figure 3a). For the majority of data sets, deviations of 2% or less are possible and all errors range between 0 and 5%. At higher quantile levels, the lognormgpd(con) mixtures (2a to 2d) cannot keep up with the models with a gamma or weibull bulk component, especially under the BTF approach (see Subfigures b to d). The gammagpd(con) and weibullgpd(con) mixtures lead to median errors of around 3% at $Q_{0.99}$, 4% at $Q_{0.999}$, and 5% at $Q_{0.9999}$. The maximum errors increase from around 6% at the 99% quantile level to around 12% at the 99.9% level and around 20% at the 99.99% level. For the models with weibull and gamma bulk distributions, the tail fraction approach and requirement of continuity hardly influence the tail approximation.

With regard to the alternative approaches, the EVMMs with gamma and weibull distributions (models 1a to 1d and 3a to 3d) achieve an equally good approximation as the POT approach (models 6a to 6c) up to the 99.9% quantile and a slightly better estimation of the 99.99% quantile. In addition, in the extreme tail, the upper error quartiles and maximum errors are lower under the weibullgpd(con) mixtures (3b) to (3d) than under the POT approach. The simple parametric distributions (models 7 to 9) are also clearly outperformed at very high quantile levels.

According to the boxplots in Subfigure (e), the estimates of the shape ξ under the gammagpd(con) and weibullgpd(con) EVMMs range between 0.15 and 0.25 for all data sets. The medians above all samples lie between 0.18 and 0.20, similar to the median estimates under the POT approach. However, the estimates under the POT models (6b) and (6c) vary substantially more between the data sets. For the lognormgpd(con) mixtures, considerably lower GPD shape parameters are obtained, in particular under the BTF approach. This explains the worse tail approximation under these models and is the result of rather low thresholds (see Subfigure f). The thresholds under the remaining models are not located in the tail either (see the values $F(u)$ of the RMS distribution function at the estimated thresholds on the right axis in Subfigure f), but they are higher than the thresholds under the lognormgpd(con) mixtures.

$n = 1,000$ Figure 4 shows the results for $n = 1,000$. With the exception of model (2d), all mixtures achieve the same level of accuracy for the 97.5% and 99% quantiles. In the extreme tail (i.e., at $Q_{0.999}$ and $Q_{0.9999}$), some differences exist and the mixtures (2a) to (2c) (lognormal bulk) and (3b) to (3d) (weibull bulk) permit the closest approximations. The median estimation errors are higher than in the case of $n = 10,000$, but comparable to or lower than the deviations under the POT approach. The 99.9% quantile, which cannot be reliably estimated by the empirical quantile in a sample of $n = 1,000$ observations, is approximated under the "best" EVMMs (2a, 2c, and 3d) with a deviation of approximately 10% or less for the majority of data sets. Out-of-sample extrapolations to the 99.99% quantile involve median errors below 20% under the mixtures with the best tail fit. This value is substantially undercut by the pareto distribution (model 8). However, the pareto approximation fails at $Q_{0.99}$ and below.

The boxplots further show a substantial variation in the quantile estimates under all EVMMs. At the 99.9% quantile, the maximum deviations are 50% or higher and at the 99.99% quantile, errors of more than 100% are possible. This problem also arises under the POT approach. The pareto distribution leads to a lower variation in the extreme tail, but not at lower quantile levels. The occurrence of very high deviations implies that even if the median errors may be acceptable, the present sample might lead to a highly flawed tail approximation. Moreover, an analysis of the errors with sign (i.e., not in absolute values) shows that both substantial underestimations and high overestimations are possible.



(a) Estimation Errors $Q_{0.975}$

(b) Estimation Errors $Q_{0.99}$

(c) Estimation Errors $Q_{0.999}$

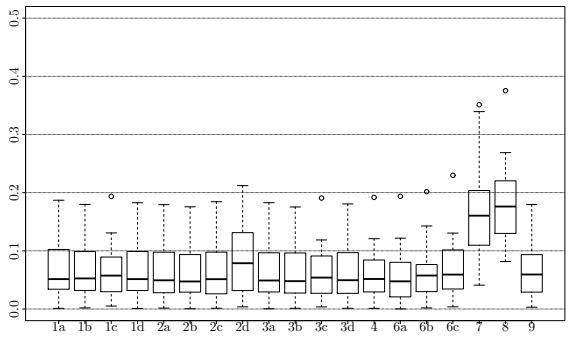
(d) Estimation Errors $Q_{0.9999}$

(e) Estimates of ξ

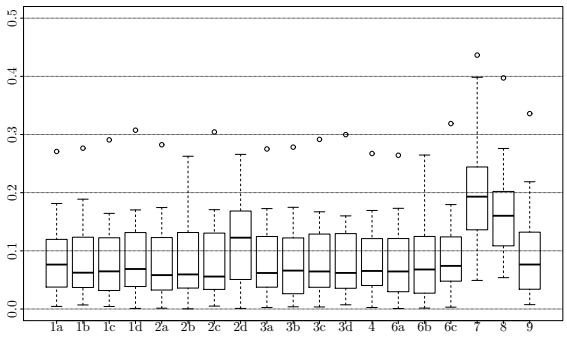
(f) Estimates of u

Figure 3: Tail Approximation of the Distribution of European Windstorm Losses, $n = 10,000$

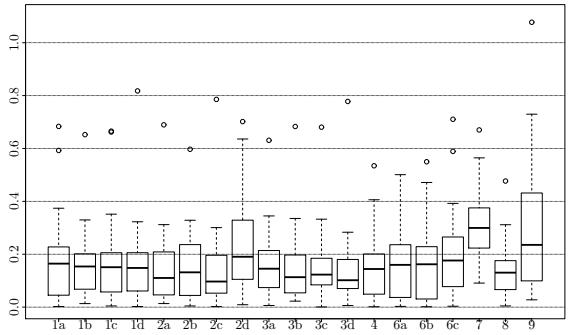
This figure shows the results of the tail approximation of the distribution of European windstorm losses in the simulation study with sample sizes $n = 10,000$. The boxplots in Subfigure (a) summarize the distributions of the (absolute values of the) relative errors in the estimation of the 97.5% quantile across all data sets under the different approaches. Models 1a to 1d / 2a to 2d / 3a to 3d are the EVMMs with gamma / lognormal / weibull bulk components (see page 7). 6a to 6c refers to the POT approach and models 7, 8, and 9 are the lognormal, pareto, and burr distribution, respectively. A value of 0.05 at the y-axis corresponds to an error of 5%. Similarly, Subfigures (b), (c), and (d) show the deviations in the estimation of the 99%, 99.9%, and 99.99% quantiles, respectively. In Subfigure (d), the y-axis is limited to the range $(0; 0.5)$ and the boxplot for model (9) is shown only partially. The boxplots in Subfigures (e) and (f) sum up the distributions of estimates of ξ and u for the different models. The parameters are estimated via the medians of the respective posterior distributions.



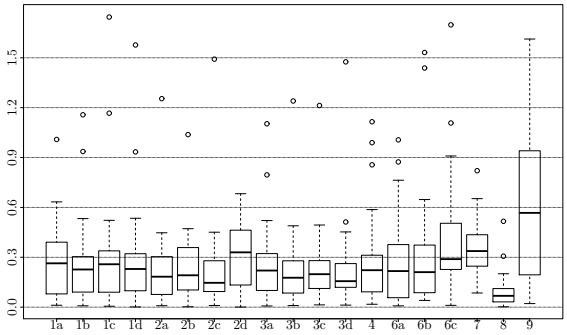
(a) Estimation Errors $Q_{0.975}$



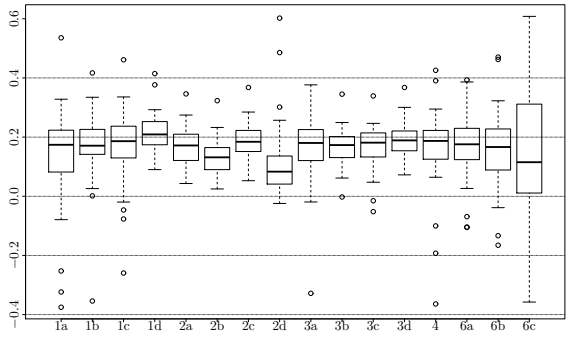
(b) Estimation Errors $Q_{0.99}$



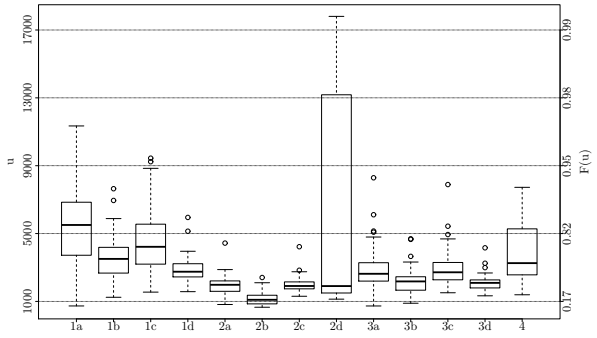
(c) Estimation Errors $Q_{0.999}$



(d) Estimation Errors $Q_{0.9999}$



(e) Estimates of ξ



(f) Estimates of u

Figure 4: Tail Approximation of the Distribution of European Windstorm Losses, $n = 1,000$

This figure shows the results of the tail approximation of the distribution of European windstorm losses in the simulation study with sample sizes $n = 1,000$. The boxplots in Subfigure (a) summarize the distributions of the (absolute values of the) relative errors in the estimation of the 97.5% quantile across all data sets under the different approaches. Models 1a to 1d / 2a to 2d / 3a to 3d / 4 are the EVMMs with gamma / lognormal / weibull / semiparametric bulk components (see page 7). 6a to 6c refers to the POT approach and models 7, 8, and 9 are the lognormal, pareto, and burr distribution, respectively. A value of 0.05 at the y-axis corresponds to an error of 5%. Similarly, Subfigures (b), (c), and (d) show the deviations in the estimation of the 99%, 99.9%, and 99.99% quantiles, respectively. To improve the illustration, only errors up to 1.75 are shown in Subfigure (d). The boxplots in Subfigures (e) and (f) sum up the distributions of estimates of ξ and u for the different models. The parameters are estimated via the medians of the respective posterior distributions.

The variation in the quantile estimates results from the substantial variability of the parameter estimates. As indicated by the boxplots in Subfigure (e), the medians of the N estimates of ξ are around 0.2, but there are also some samples that lead to posterior medians close to or lower than zero. Furthermore, the location of the threshold varies considerably. For all models, the difference between the values $F(u_{max})$ and $F(u_{min})$ of the RMS distribution function at the highest and lowest estimated thresholds is at least 0.4, i.e., the estimates are located in different areas of the underlying distribution. Under model (2d), the threshold estimates even range between the 20% and 99% quantile of the underlying distribution.

$\boldsymbol{n = 250}$ Due to the smaller sample size, the output for the kernelgpd(con) mixtures is not shown in Figure 4. Under all four model versions (5a) to (5d), the 97.5% and 99% quantiles can be estimated with median errors of 9% and 14%, respectively. Extrapolations to the 99.9% and 99.99% quantiles lead to median deviations of approximately 23% and 41%, respectively, if a continuous density is required, and around 30% and 48%, respectively, for the models without continuity constraint. These values exceed the median estimation errors under the remaining mixtures shown in Figure 4, but this is the result of the reduced sample size. We also apply the mixtures (1a), (2a), and (3a) with gamma, lognormal, and weibull bulk models to the data sets with $n = 250$ and obtain similar results as for the kernelgpd(con) mixtures. The required level of accuracy depends on the application, but approximation errors of 40% or more are presumably not satisfactory in most cases. However, a "naive" application of the POT approach yields to even higher median deviations (36% for $Q_{0.999}$ and 54% for $Q_{0.9999}$ if the threshold is set at the 90% data quantile).

4.1.4 Influence of Outliers

The maximum losses in the two sets of samples are 87,601 ($n = 10,000$) and 79,328 ($n = 1,000$). In order to examine the influence of extreme outliers on the estimation results, we add an extreme loss between 100,000 and 200,000 to each of the first $N_1 = 15$ samples and re-estimate all models.[11]

$\boldsymbol{n = 10,000}$ Some core results for the case $n = 10,000$ are illustrated in Figure 5. The remaining results (e.g., for the 97.5% and 99.99% quantiles) are available upon request from the author. The boxplots in Subfigures (a) and (b) summarize the N_1 relative changes in the estimates of the 99% and 99.9% quantiles under the different models. Subfigures (c) and (d) show the median relative approximation errors of $Q_{0.99}$ and $Q_{0.999}$ above all N_1 samples before (circles) and after (squares) the inclusion of the extreme losses. In order to improve the illustration, we only show the medians here, but an analysis of the boxplots largely confirms the results indicated by the medians. The two subfigures at the bottom provide the boxplots for the relative changes in the estimates of the parameters ξ and u.[12]

The existence of outliers in the samples leads to an increase in the estimates of ξ under all models (see Subfigure e). Under the EVMM approach, the increases mainly range between 5% and 10%, except for model (2d). For (2d) and the POT method, higher increases can be observed, especially in the case of a very high threshold (as the estimates are based on only few observations). The remaining tail parameters u (see Subfigure f), σ_u, and ϕ_u (not shown here) are less affected and mainly change between 0 and 5%.

[11] For data set i with maximum loss M_i, the extreme loss L_i^* is determined by: $L_i^* = \min\{\max\{3 \cdot M_i; 100,000\}; 200,000\}$.

[12] Due to mixing problems of the Markov chains, the results for the lognormgpd(con) mixtures (2a) to (2c) are again based on a smaller subset of samples and have to be treated with caution.

As a result of the higher shape parameters, the estimates for the tail quantiles increase. For the majority of EVMMs and data sets, the relative changes rise from some value between 0.5% and 1% for $Q_{0.975}$ (not shown here), to approximately 1.5% for $Q_{0.99}$ (see Subfigure a), to some value in the range from 3% to 5% for $Q_{0.999}$ (see Subfigure b), to some increase between 5% and 8% for $Q_{0.9999}$ (not shown here). Only the 97.5% and 99% quantile estimates from model (2d) are subject to higher changes, and there is a higher variation in the results for model (1a) (see Subfigure b). Depending on the threshold, the POT approach leads to similar or slightly lower increases than the EVMM method in the lower tail area (i.e., at $Q_{0.975}$ and $Q_{0.99}$), but to higher changes in the extreme tail. The quantile estimates from the parametric models (7) to (9) change only marginally.

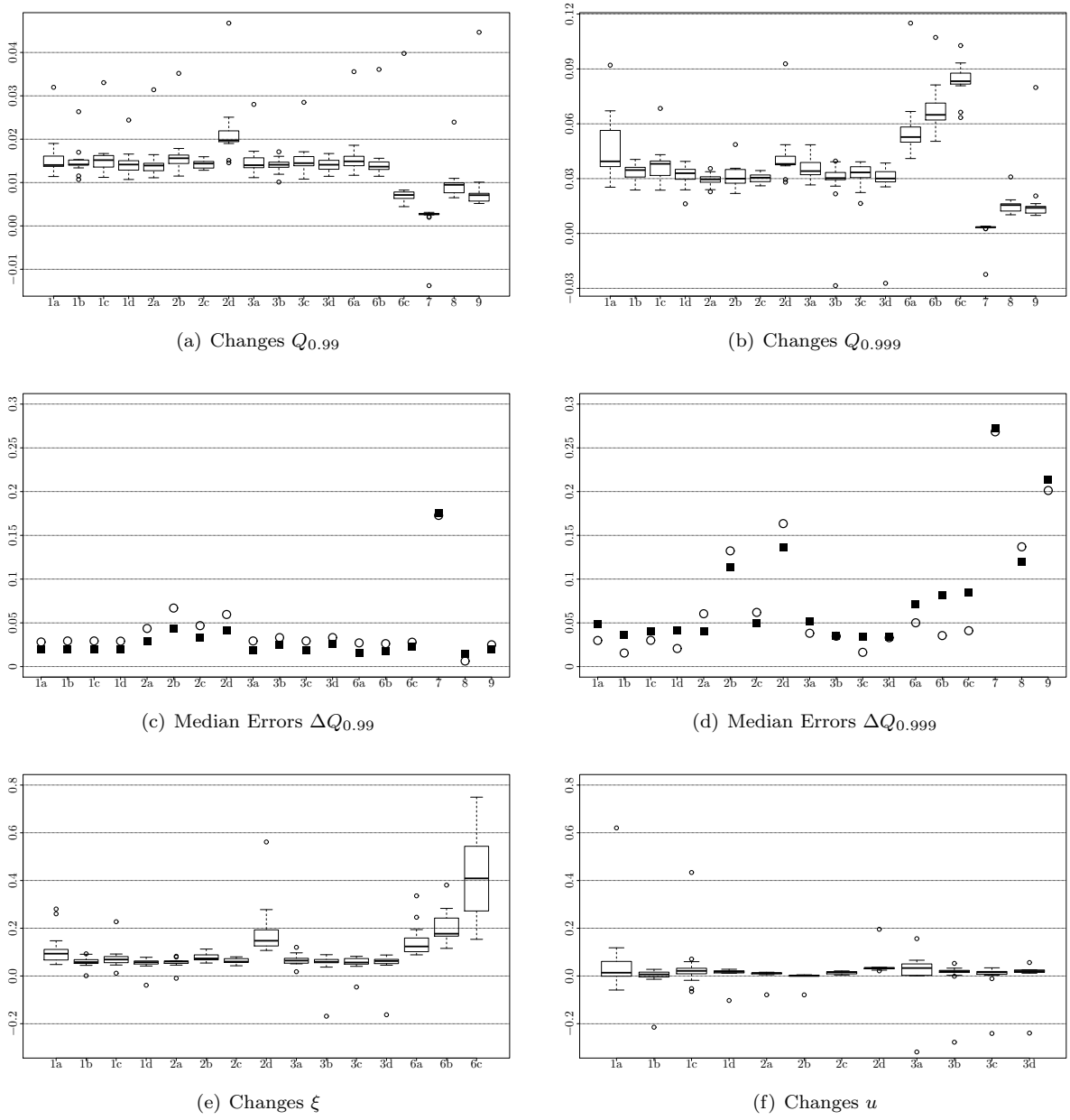
The effects on the tail fit depend on the tail area. In the region of the 97.5% and 99% quantiles, there is a tendency towards underestimation for the samples without outliers. Consequently, the increase in the quantile estimates due to the extreme losses slightly improves the tail approximation (see Subfigure c).[13] At higher quantile levels, the approximation errors only decline under the lognormgpd(con) mixtures (for the same reason). The deviations under models (3b) and (3d) remain unchanged. For the remaining EVMMs, higher errors result, as the quantiles are overestimated to a larger extent. However, the POT approximation also deteriorates and the relative errors (around 8% at $Q_{0.999}$) exceed the deviations (5% or lower) under all mixtures except of (2b) and (2d). In addition, the EVMM approach remains substantially better than the three parametric models (7) to (9).

$\boldsymbol{n = 1,000}$ Figure 6 shows the results for $n = 1,000$. Due to substantially higher changes and approximation errors under the lognormgpdcon mixture (2d), the results for this model are excluded from the plots and following interpretations. As expected, in the case of a rather small sample, single outliers have a greater impact on the tail approximation. The median increases of the estimates of ξ vary between the mixture models and range from approximately 35% for models (2c), (3b), and (3d) to almost 60% for the gammagpd mixture (1a) (see Subfigure e). Moreover, the interquartile ranges indicate a substantial variation of the results for different samples. As the POT estimates of ξ are based on a lower number of observations, they are more sensitive to outliers and the increases substantially exceed those under the EVMMs. The median changes of the threshold estimates (see Subfigure f) are between 2% (for model 1b) and 12% (for model 1d), but for some models (e.g., models 1a, 3a, 4), the changes vary substantially.

Under the EVMM approach, the changes in the parameters translate into higher estimates of the 97.5% and 99% quantiles by around 5% and 10%, respectively (for $Q_{0.99}$, see Subfigure a). The median increases for $Q_{0.999}$ range from around 20% for various mixtures with a lognormal or weibull bulk component to 38% for model (1a) (see Subfigure b). In the extreme tail, slightly higher changes result under the PTF approach and the PTF approach also leads to a higher variation between different samples (compare, e.g., the results for models 1a and 1c with those of 1b and 1d in Subfigure b). Furthermore, the analysis confirms the above results of a higher sensitivity of the extreme tail quantiles ($Q_{0.999}$ or higher) under the POT approach and relatively small changes under the parametric models.

In the case of $n = 1,000$, the application of the EVMM approach to the original set of samples (i.e., without outliers) does not only lead to an overestimation of very high tail quantiles, but for various samples also of $Q_{0.975}$ and $Q_{0.99}$. The existence of extremely high losses therefore induces higher approximation

[13] Please note that we only consider the deviations in absolute values, i.e., under- and overestimations are evaluated equally.



(a) Changes $Q_{0.99}$

(b) Changes $Q_{0.999}$

(c) Median Errors $\Delta Q_{0.99}$

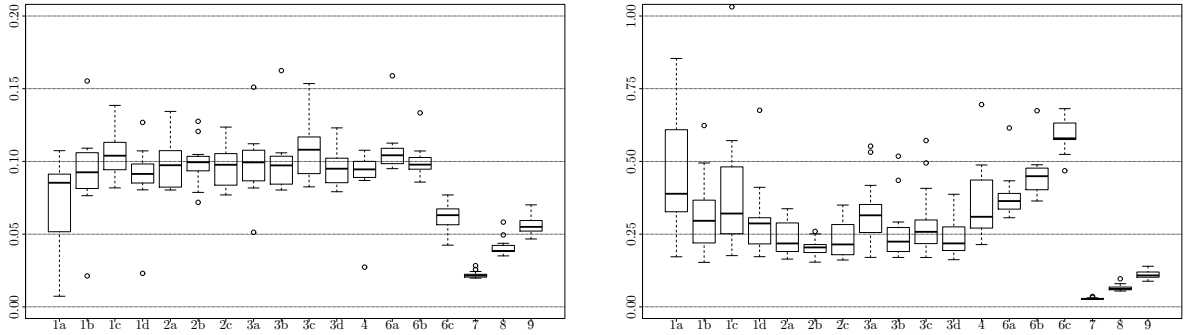
(d) Median Errors $\Delta Q_{0.999}$

(e) Changes ξ

(f) Changes u

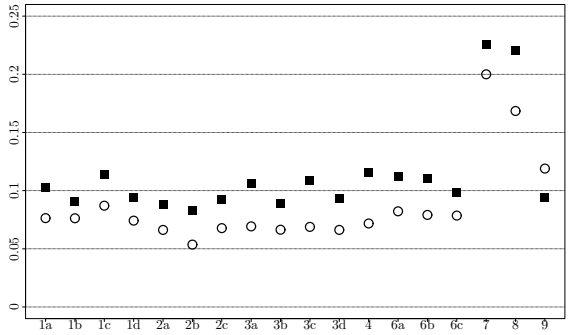
Figure 5: Effects of Outliers on the Results for the Distribution of Windstorm Losses, $n = 10,000$

This figure shows the influence of outliers on the approximation of the distribution of European windstorm losses for $n = 10,000$. The results are based on two sets of $N_1 = 15$ samples. The samples in the first set do not include outliers. The samples in the second set comprise the same losses plus an additional extreme loss. For each model, the boxplots in Subfigures (a) and (b) summarize the N_1 relative changes in the estimates of the 99% and 99.9% quantiles due to the outliers. Subfigures (c) and (d) show the median relative errors in the estimation of the 99% and 99.9% quantiles for the N_1 samples without outliers (circles) and the N_1 samples with outliers (squares). The relative errors are considered in absolute values (i.e., without signs). The last two subfigures illustrate the relative N_1 changes in the estimates of ξ and u as a result of the extreme loss. In both subfigures, one outlier exceeds the upper limit of the y-axis of 0.8. In all subfigures, a value of y_0 at the y-axis (e.g., $y_0 = 0.1$) corresponds to a relative change / error of $(y_0 \cdot 100)\%$ (e.g., 10%). For the model numbers, see pages 7 and 12.

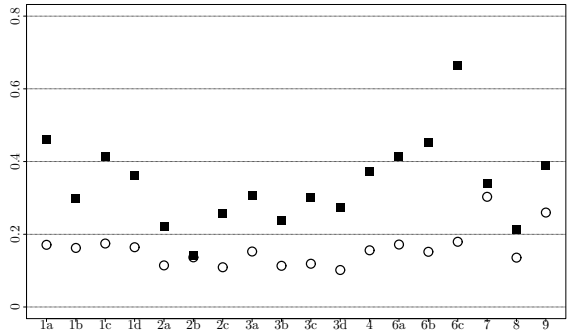


(a) Changes $Q_{0.99}$

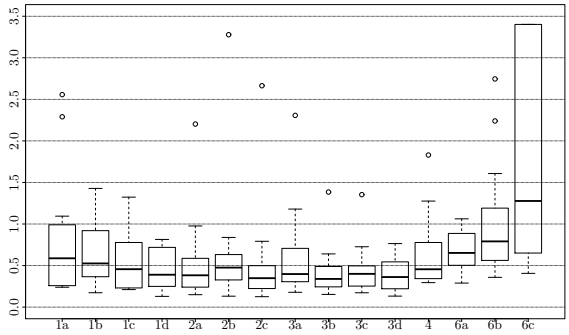
(b) Changes $Q_{0.999}$



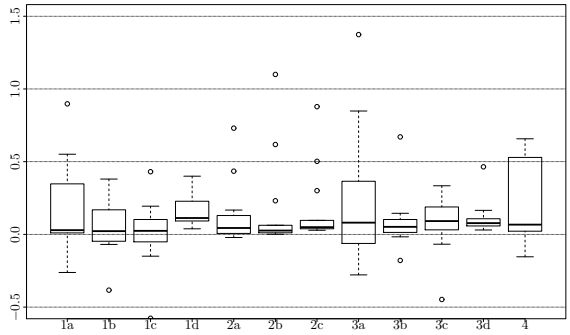
(c) Median Errors $\Delta Q_{0.99}$



(d) Median Errors $\Delta Q_{0.999}$



(e) Changes ξ



(f) Changes u

Figure 6: Effects of Outliers on the Results for the Distribution of Windstorm Losses, $n = 1,000$

This figure shows the influence of outliers on the approximation of the distribution of European windstorm losses for $n = 1,000$. The results are based on two sets of $N_1 = 15$ samples. The samples in the first set do not include outliers. The samples in the second set comprise the same losses plus an additional extreme loss. For each model, the boxplots in Subfigures (a) and (b) summarize the N_1 relative changes in the estimates of the 99% and 99.9% quantiles due to the outliers. Subfigures (c) and (d) show the median relative errors in the estimation of the 99% and 99.9% quantiles for the N_1 samples without outliers (circles) and the N_1 samples with outliers (squares). The relative errors are considered in absolute values (i.e., without signs). The last two subfigures illustrate the relative N_1 changes in the estimates of ξ and u as a result of the extreme loss. In Subfigure (e), several extreme increases above 3.5 are not shown. Moreover, one change of the threshold u exceeds the upper limit of the y-axis (1.5) in Subfigure (f). In all subfigures, a value of y_0 at the y-axis (e.g., $y_0 = 0.1$) corresponds to a relative change / error of $(y_0 \cdot 100)\%$ (e.g., 10%). For the model numbers, see pages 7 and 12.

errors at all quantile levels $\alpha = 0.975, 0.99, 0.999, 0.9999$ (see Subfigures c and d). Due to smaller changes in the estimates, the BTF approach outperforms the PTF approach. The best fit is achieved under the lognormgpd mixture (2b), followed by the remaining mixtures with lognormal bulk components and the weibullgpd(con) mixtures (3b) and (3d). These models also provide a better approximation than the POT approach and a similar level of accuracy as the "best" parametric alternative (model 8).

4.2 Empirical Applications

4.2.1 Data Description

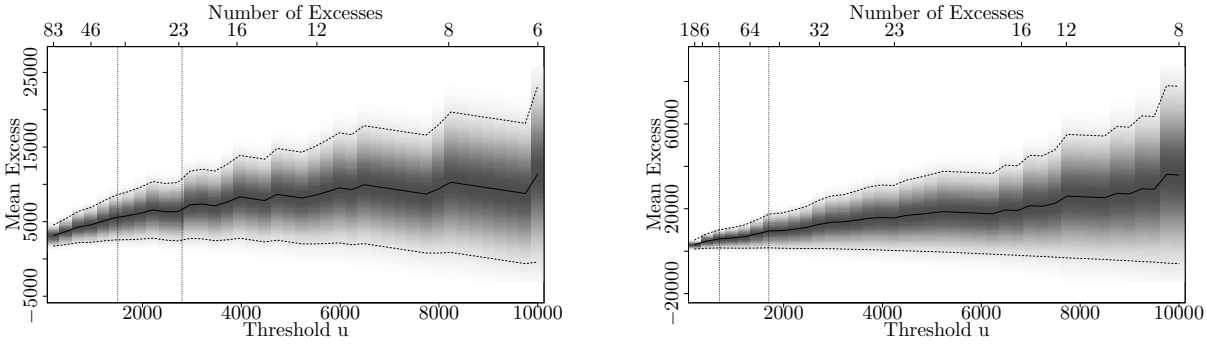
In order to get further insights, the EVMM approach is applied to two empirical data sets of insurance losses. The first data set, the macro validation data set for U.S. hurricane models of Collins and Lowe (2001), comprises insured losses from hurricanes in the United States between 1900 and 1999, adapted to year 2000 exposure and price levels. This data set has been used in some previous analyses (see Jagger et al., 2008), but to the knowledge of the author, not in the context of EVMMs.[14] The whole sample consists of 163 observations and is available in Collins and Lowe (2001). For the purpose of our analysis, we divide the reported losses (given thousands USD) by 1,000. Thus, the losses in our data set represent the adjusted insured losses in millions of USD. The descriptive statistics are summarized in Table 1. The losses range between 0.5 and 49,729, the average is 1,731, and the median is 217 (million USD, respectively). The distribution is skewed to the right and the occurrence of a small number of very large losses indicates the existence of heavy tails. In line with this, the mean excess plot is (slightly) increasing (see Figure 7a).

	U.S. Hurricane Losses	Aon Re Belgium Fire Insurance Losses
Sample Size	163	1,823
Mean	1,731.21	363.47
Median	217.22	9.15
Std. Dev.	4,978.09	4,868.25
Skewness	6.60	33.90
Kurtosis	57.59	1,289.32
Minimum	0.50	0.19
90% Quantile	3,912.10	205.80
99% Quantile	24,486.69	6,437.07
Maximum	49,728.84	190,541.50

Table 1: Descriptive Statistics for the Empirical Data

Our second sample comprises 1,823 fire insurance losses reported by Aon Re Belgium. The data set includes information on the building type, claim size, and sum insured.[15] We focus on the claim size and again divide the losses by 1,000. In addition, due to a substantial number of identical claims, a small amount of noise is added to each loss. As shown in Table 1, the average claim size is 363, with a standard deviation of 4,868. Moreover, the skewness and kurtosis are 34 and 1,289, respectively. The sample includes very extreme losses: the maximum loss and 90% quantile are 20,831 and 22 times the median loss, respectively. These characteristics differ substantially from the characteristics of the Danish fire insurance data set, which has typically been analysed in the context of EVMMs.[16]

[14] Jagger et al. (2008) apply the POT approach to the logarithm of the losses and develop a model for the prediction of the annual aggregate losses based on climate data, i.e., their results are not directly comparable to the results in our paper.
[15] The data set is available at http://lstat.kuleuven.be/Wiley/Data/aon.txt (accessed February 2, 2016).
[16] For the sample of 2,492 Danish fire insurance losses, the skewness and kurtosis are 20 and 550, respectively, and the maximum loss and 90% quantile are 161 and 3 times the median loss, respectively.



(a) U.S. Hurricane Losses

(b) Aon Re Belgium Fire Losses

Figure 7: Sample Mean Excess Plots

This figure presents the mean excess plots for the U.S. hurricane losses and Aon Re Belgium fire losses. The dark line shows the sample mean excesses with respect to u, and the dotted lines indicate the corresponding 95% confidence intervals. In addition, the vertical dotted lines indicate the thresholds used under the POT approach (1,500 and 2,800 for the U.S. hurricane losses and 700 and 1,700 for the Aon Re Belgium fire losses). The thresholds u are given at the lower x-axis, the number of exceedances of u at the upper x-axis.

4.2.2 Results

The posterior distributions are derived under the assumption of the same flat priors and with the same MCMC algorithms as in the previous section. The POT approach is implemented for three thresholds. Based on the respective mean-excess-plots (see Figure 7), we select $u = 1,500$ and $u = 2,800$ for the U.S. hurricane data and $u = 700$ and $u = 1,700$ for the Aon Re Belgium fire losses. In addition, "naive" thresholds of the 90% data quantiles (3,912 and 206, respectively) are used.

For the Aon Re Belgium data set, the Markov chains generated by PT converge and mix well except for the gammagpdcon mixtures (1c) and (1d). In the case of the U.S. hurricane data, we assume that convergence has been achieved for all models except the weibullgpdcon mixture with BTF approach (model 3d). However, in view of the Markov chains generated by PT for high temperatures τ (see the description of the PT algorithm in Laas, 2016), we note that there might be certain problems for the hurricane losses and some mixtures with a weibull or gamma bulk component.

U.S. Hurricane Losses Table 2 contains the parameter and quantile estimates (i.e., the posterior medians) for the U.S. hurricane data. As the true underlying distribution is unknown, the estimated quantiles are compared to the sample quantiles (see, e.g., Do Nascimento et al., 2012), and ΔQ_α indicates the relative deviation of the estimated α-quantile from the empirical quantile. In addition, Table A1 in Appendix A shows the uncertainty in the estimates in form of the 95% posterior intervals.

Among the class of EVMMs, the lognormgpd mixture with PTF approach (model 2a) provides the best tail approximation. The estimated 90% and 95% quantiles almost match the empirical ones and the calculated 97.5% and 99% quantiles deviate by 19% and 13%, respectively. In comparison to the remaining EVMMs, rather small approximation errors can also be obtained by means of the lognormgpd(con) models (2b) and (2c) as well as the semipargpd mixture (4). The kernelgpd mixtures (5a) and (5b) perform worse than the "best" models, but better than the EVMMs with weibull or gamma bulk components. Especially under the BTF approach, the gammagpd(con) mixtures (1b) and (1d) fail completely.

		ξ	u	$Q_{0.95}$	$\Delta Q_{0.95}$	$Q_{0.975}$	$\Delta Q_{0.975}$	$Q_{0.99}$	$\Delta Q_{0.99}$
	Empirical			$8.20 \cdot 10^3$		$1.15 \cdot 10^4$		$2.45 \cdot 10^4$	
(1a)	Gammagpd, PTF	0.93	102	$7.81 \cdot 10^3$	−0.05	$1.56 \cdot 10^4$	0.35	$3.75 \cdot 10^4$	0.53
(1b)	Gammagpd, BTF	1.30	21	$9.47 \cdot 10^3$	0.16	$2.37 \cdot 10^4$	1.06	$7.94 \cdot 10^4$	2.24
(1c)	Gammagpdcon, PTF	0.94	109	$7.98 \cdot 10^3$	−0.03	$1.60 \cdot 10^4$	0.39	$3.89 \cdot 10^4$	0.59
(1d)	Gammagpdcon, BTF	1.90	10	$1.80 \cdot 10^4$	1.20	$6.77 \cdot 10^4$	4.87	$3.88 \cdot 10^5$	14.84
(2a)	Lognormgpd, PTF	0.66	$9.51 \cdot 10^3$	$8.06 \cdot 10^3$	−0.02	$1.37 \cdot 10^4$	0.19	$2.77 \cdot 10^4$	0.13
(2b)	Lognormgpd, BTF	0.53	$1.08 \cdot 10^4$	$8.54 \cdot 10^3$	0.04	$1.54 \cdot 10^4$	0.34	$2.83 \cdot 10^4$	0.15
(2c)	Lognormgpdcon, PTF	0.81	$3.17 \cdot 10^3$	$8.16 \cdot 10^3$	0.00	$1.43 \cdot 10^4$	0.24	$3.04 \cdot 10^4$	0.24
(2d)	Lognormgpdcon, BTF	0.65	$9.01 \cdot 10^3$	$8.34 \cdot 10^3$	0.02	$1.59 \cdot 10^4$	0.38	$3.29 \cdot 10^4$	0.35
(3a)	Weibullgpd, PTF	0.98	95	$8.01 \cdot 10^3$	−0.02	$1.64 \cdot 10^4$	0.43	$4.14 \cdot 10^4$	0.69
(3b)	Weibullgpd, BTF	1.30	21	$9.39 \cdot 10^3$	0.15	$2.33 \cdot 10^4$	1.03	$7.71 \cdot 10^4$	2.15
(3c)	Weibullgpdcon, PTF	0.94	111	$7.96 \cdot 10^3$	−0.03	$1.59 \cdot 10^4$	0.38	$3.85 \cdot 10^4$	0.57
(4)	Semipargpd	0.83	391	$7.70 \cdot 10^3$	−0.06	$1.44 \cdot 10^4$	0.25	$3.16 \cdot 10^4$	0.29
(5a)	Kernelgpd, PTF	0.90	90	$7.61 \cdot 10^3$	−0.07	$1.48 \cdot 10^4$	0.29	$3.48 \cdot 10^4$	0.42
(5b)	Kernelgpd, BTF	0.90	90	$7.63 \cdot 10^3$	−0.07	$1.49 \cdot 10^4$	0.29	$3.50 \cdot 10^4$	0.43
(5c)	Kernelgpdcon, PTF	0.96	95	$7.90 \cdot 10^3$	−0.04	$1.59 \cdot 10^4$	0.38	$3.90 \cdot 10^4$	0.59
(5d)	Kernelgpdcon, BTF	0.95	95	$7.84 \cdot 10^3$	−0.04	$1.58 \cdot 10^4$	0.37	$3.86 \cdot 10^4$	0.58
(6a)	POT, $u = 1500$	0.52		$8.24 \cdot 10^3$	0.00	$1.38 \cdot 10^4$	0.20	$2.48 \cdot 10^4$	0.01
(6b)	POT, $u = 2800$	0.50		$8.31 \cdot 10^3$	0.01	$1.40 \cdot 10^4$	0.22	$2.52 \cdot 10^4$	0.03
(6c)	POT, $u = 3912$	0.32		$8.76 \cdot 10^3$	0.07	$1.52 \cdot 10^4$	0.32	$2.60 \cdot 10^4$	0.06
(7)	Lognormal			$8.47 \cdot 10^3$	0.03	$1.69 \cdot 10^4$	0.47	$3.78 \cdot 10^4$	0.54
(8)	Pareto			$1.37 \cdot 10^4$	0.67	$4.44 \cdot 10^4$	2.85	$2.09 \cdot 10^5$	7.55
(9)	Burr			$8.41 \cdot 10^3$	0.03	$1.80 \cdot 10^4$	0.57	$4.69 \cdot 10^4$	0.92

Table 2: Estimation Results for the U.S. Hurricane Data Set

This table presents the results for the sample of U.S. hurricane losses. The first and second data columns show the medians of the posterior distributions for the GPD shape parameter ξ and threshold u. The remaining columns contain the estimates for the α-quantiles Q_α (i.e., the posterior medians) and their relative deviations ΔQ_α from the empirical quantiles. The results for model (3d) are not shown due to convergence problems of the Markov chains.

Our results of the quantile estimation further show that except for the EVMM with weibull bulk model and PTF approach (models 3a and 3c), the introduction of a continuity constraint deteriorates the approximation of $Q_{0.975}$ and $Q_{0.99}$ (if the tail fraction approach is held fixed). Moreover, for each mixture model, the tail approximation is better under the parameterised tail fraction approach (PTF) than under the bulk model based tail fraction approach (BTF). This confirms the results of Hu (2013) that the more flexible models without the requirement of continuity at the threshold and with the additional parameter ϕ_u tend to provide a better fit, especially if the bulk model is misspecified.

The varying tail fit for different choices of the bulk model can be explained by means of the results of the parameter estimation. In the case of the lognormgpd(con) mixtures, the posterior distributions of the thresholds have a large part of their probability mass at high values for u. For example, under model (2a), 73% of the sampled thresholds exceed the 80% data quantile and the median threshold (9,506) corresponds to a tail proportion of 4.9%. The thresholds in the semipargpd mixture (the second best model) tend to be lower than in the models with lognormal body, but higher than the remaining mixtures. Due to the incompatibility of the gamma and weibull distributions with large losses, the thresholds are rather small, especially under the BTF approach. This leads to high overestimations of the shape parameter ξ and consequently of the extreme quantiles.

A comparison of all models shows that the simple parametric models (7) to (9) cannot keep up with the EVMM method. In contrast, for all three thresholds, the POT approach (models 6a to 6c) leads to a comparable approximation of the 97.5% quantile and a substantially better estimation of the 99% quantile. For $u = 1,500$, the POT approach even permits an almost perfect match of the 99% quantile.

According to the 95% posterior intervals, the quantile estimates are subject to great uncertainty (see Table A1 in Appendix A). For example, for the lognormgpd mixture (2a), the length of the 95% posterior interval for the 95% quantile (97.5% /99% quantile) amounts to 72% (136% / 286%) of the respective empirical quantile. This results from the small sample size, heavy tails, and flat priors. However, the posterior intervals for $Q_{0.95}$ and $Q_{0.975}$ obtained under the simple parametric models (which have a lower number of parameters) and the POT approach (where the uncertainty regarding the threshold is negated) are not substantially shorter either. For $Q_{0.99}$, the POT approach leads to shorter posterior intervals.

		ξ	u	$Q_{0.975}$	$\Delta Q_{0.975}$	$Q_{0.99}$	$\Delta Q_{0.99}$	$Q_{0.995}$	$\Delta Q_{0.995}$
	Empirical			$1.96 \cdot 10^3$		$6.44 \cdot 10^3$		$9.55 \cdot 10^3$	
(1a)	Gammagpd, PTF	1.65	3.74	$1.86 \cdot 10^3$	−0.05	$8.42 \cdot 10^3$	0.31	$2.64 \cdot 10^4$	1.76
(1b)	Gammagpd, BTF	1.69	3.02	$1.96 \cdot 10^3$	0.00	$9.24 \cdot 10^3$	0.43	$2.98 \cdot 10^4$	2.12
(2a)	Lognormgpd, PTF	1.65	3.62	$1.87 \cdot 10^3$	−0.05	$8.53 \cdot 10^3$	0.33	$2.69 \cdot 10^4$	1.81
(2b)	Lognormgpd, BTF	1.65	3.66	$1.85 \cdot 10^3$	−0.06	$8.37 \cdot 10^3$	0.30	$2.62 \cdot 10^4$	1.74
(2c)	Lognormgpdcon, PTF	1.64	6.42	$1.86 \cdot 10^3$	−0.05	$8.38 \cdot 10^3$	0.30	$2.62 \cdot 10^4$	1.74
(2d)	Lognormgpdcon, BTF	1.76	2.36	$2.22 \cdot 10^3$	0.13	$1.12 \cdot 10^4$	0.74	$3.79 \cdot 10^4$	2.97
(3a)	Weibullgpd, PTF	1.65	3.82	$1.87 \cdot 10^3$	−0.05	$8.53 \cdot 10^3$	0.32	$2.68 \cdot 10^4$	1.81
(3b)	Weibullgpd, BTF	1.70	2.99	$2.01 \cdot 10^3$	0.02	$9.54 \cdot 10^3$	0.48	$3.10 \cdot 10^4$	2.25
(3c)	Weibullgpdcon, PTF	1.66	3.99	$1.89 \cdot 10^3$	−0.04	$8.64 \cdot 10^3$	0.34	$2.73 \cdot 10^4$	1.86
(3d)	Weibullgpdcon, BTF	1.74	1.65	$2.12 \cdot 10^3$	0.08	$1.04 \cdot 10^4$	0.62	$3.47 \cdot 10^4$	2.63
(4)	Semipargpd	1.67	2.99	$1.93 \cdot 10^3$	−0.02	$8.93 \cdot 10^3$	0.39	$2.84 \cdot 10^4$	1.98
(6a)	POT, $u = 206$	1.33		$1.78 \cdot 10^3$	−0.09	$5.51 \cdot 10^3$	−0.14	$1.26 \cdot 10^4$	0.32
(6b)	POT, $u = 700$	1.16		$2.10 \cdot 10^3$	0.07	$5.62 \cdot 10^3$	−0.13	$1.05 \cdot 10^4$	0.10
(6c)	POT, $u = 1700$	0.77		$1.84 \cdot 10^3$	−0.06	$5.62 \cdot 10^3$	−0.13	$1.11 \cdot 10^4$	0.16
(7)	Lognormal			$6.50 \cdot 10^2$	−0.67	$1.34 \cdot 10^3$	−0.79	$2.19 \cdot 10^3$	−0.77
(8)	Pareto			$1.37 \cdot 10^3$	−0.30	$5.28 \cdot 10^3$	−0.18	$1.46 \cdot 10^4$	0.53
(9)	Burr			$3.35 \cdot 10^3$	0.71	$2.09 \cdot 10^4$	2.25	$8.36 \cdot 10^4$	7.75

Table 3: Estimation Results for the Aon Re Data Set

This table presents the results for the sample of Aon Re Belgium fire insurance losses. The first and second data columns show the medians of the posterior distributions for the shape parameter ξ and threshold u. The remaining columns contain the estimates for the α-quantiles Q_α (i.e., the posterior medians) and their relative deviations ΔQ_α from the empirical quantiles. The results for models (1c) and (1d) are not shown due to convergence problems of the Markov chains.

Aon Re Belgium Fire Losses The results for the sample of fire insurance claims are given in Table 3. The 95% posterior intervals are provided in Appendix A in Table A2. Due to the higher number of observations, we now consider the 97.5%, 99%, and 99.5% quantiles. For this data set, rather low thresholds are estimated under all EVMMs and the posterior medians for the shape parameter ξ substantially exceed the estimates under the POT approach. The 97.5% quantile can be approximated with errors of around 5%, but the estimates are subjected to great uncertainty. Under all mixture models, the length of the 95% posterior interval is 70% or more of the value of the empirical quantile. At higher quantile levels, the models with parameterised tail fraction ϕ_u and without continuity constraint tend to permit

a slightly better tail approximation than the less flexible mixtures (in line with the results for the U.S. hurricane data). However, all EVMMs lead to high overestimations of at least 30% at $Q_{0.99}$ and 174% at $Q_{0.995}$. In addition, the 95% posterior intervals for the 99.5% quantile do not even cover the empirical 99.5% quantile loss. The POT method provides a substantially higher accuracy at the 99% and 99.5% quantile levels, especially for the thresholds that are selected based on the mean-excess-plot. The pareto distribution also leads to lower estimation errors, but the deviation at $Q_{0.995}$ is also high (53%).

5 Implications

5.1 Model Choice

The simulation study and empirical applications show that the tail fit of the considered EVMMs varies substantially between different data sets. General advice for the selection of certain bulk model is therefore difficult. The mixtures with lognormal bulk components provide the best tail approximation for the U.S. hurricane losses, by far (among the class of EVMMs), but the mixing problems and high estimation errors (compared to the remaining models) in the simulation study for $n = 10,000$ indicate that this model is inadequate for the distribution of windstorm losses. By contrast, the weibullgpd(con) mixtures permit a rather good estimation of the tails of the distribution of windstorm losses, but yield unsatisfactory results for the empirical data sets. The gammagpd(con) mixtures frequently lead to similar results as the models with weibull bulk distribution and may be an alternative in the case of convergence problems of the Markov chains for the weibullgpd(con) models. Furthermore, the use of the more flexible (but therefore also more time-consuming) semiparametric approach of Cabras and Castellanos (2011) or kernelgpd(con) mixture of MacDonald et al. (2013) does not necessarily lead to a better tail approximation. For example, both models cannot keep up with the lognormgpd(con) mixtures in the approximation of the extreme U.S. hurricane losses.

With regard to the tail fraction approach and continuity constraint, our analyses confirm the findings of Hu (2013) and presumption of Scarrott (2016) that the use of a more flexible approach can improve the approximation in the case of an inadequate bulk model (see, e.g., the results for the mixtures with weibull or gamma bulk components for the empirical data sets and the results for the lognormgpd(con) mixtures in the simulation study). However, the effects of an inappropriate bulk distribution can only be partly offset and the approximation frequently remains rather bad. If the bulk model provides a good fit and no continuity constraint is imposed (sometimes also in the case with continuity requirement), the choice of the tail fraction approach only marginally influences the tail fit (see, e.g., the results for models 1a to 1d and 3a to 3d in the simulation study). The BTF approach may even slightly improve the approximation in the extreme tail, as also indicated by Hu (2013) and Scarrott (2016). Moreover, according to the results in Section 4.1.4, the estimates of extreme tail quantiles may be less sensitive to outliers under the BTF approach. For various data sets and an appropriate bulk distribution, the requirement of a continuous density does not have substantial effects either (see also Hu, 2013). However, examples exist where the continuity requirement induces substantially higher deviations in the tail (see, e.g., the results for the mixtures with lognormal bulk models (2a) to (2d) for the U.S. hurricane data). If the focus lies on the tail estimation, we therefore think that leaving out the continuity requirement lowers the risk of high estimation errors. In addition, as shown by Do Nascimento et al. (2012), the problem of a discontinuous density can largely be resolved by means of a calculation of the posterior predictive density.

So far, we have only examined the tail fit. A comparison of the overall goodness-of-fit based on the medians of the posterior distributions of the AIC and BIC values (as proposed by Brooks, 2002) does not lead to general guidelines either.[17] For the three loss distributions considered in this paper, the different parametric bulk models mainly lead to similar AIC and BIC values, and no bulk distribution is always "best". Furthermore, the posterior medians of the loglikelihood values are similar under the semiparametric approaches and parametric mixtures. In some cases, the goodness-of-fit statistics support the use of a more flexible model parmeterization (i.e., the PTF approach and no continuity constraint), as the improvements in the loglikelihoods (more precisely, the posterior medians of the loglikelihoods) compensate the "costs" of the additional parameters. In various other cases, in contrast, the likelihood values differ only marginally and the models with less parameters are promoted, especially under the BIC.

The results of our analyses imply that, in general, several EVMMs have to be applied to the present data set. Indicators for an inadequate bulk model are mixing problems of the Markov chains, substantial differences in the results between the BTF and PTF approach, and, in some circumstances, very low threshold estimates (see the next section). The final decision for one model has to be based on careful analyses, such as a comparison of the estimated and empirical quantiles via the qq-plot, the relative errors, or statistical measures that summarize the errors at several quantile levels (e.g., the average scaled absolute error ASAE, see, e.g. Wong and Li, 2010).

5.2 Threshold Estimation

The modeling of the tail via the GPD is based on the asymptotic argument that the excess distribution of various distributions converges to a GPD if the threshold is raised towards the upper end of the range (see, e.g., McNeil et al., 2005). In order to ensure unbiased estimates of the GPD parameters and a good tail approximation, rather high thresholds are therefore typically chosen under the POT approach. The application of EVMMs to empirical insurance losses shows that for some data sets and bulk models, the EVMM approach also leads to thresholds around the 90% quantile or above (i.e., the posterior distribution for u is mainly located in the tail area). For example, for the Secura Belgian Re data set considered in Laas (2016), the estimated tail fraction is 10% or lower for the gammagpd and weibullgpd mixtures. In the analyses in this paper, high thresholds result for the lognormgpd(con) mixtures and the U.S. hurricane data set and for very few data sets in the simulation study with sample size $n = 1,000$. Thus, the parameters of the GPD distribution are estimated only based on the observations in the tail. In other words, the GPD is fitted only to that range of the distribution where the GPD approximation is theoretically justified.

For various other data sets, however, we observe rather low estimates for the threshold. Examples from this paper include the Aon Re data, the U.S. hurricane data in combination with mixtures other than the lognormgpd(con) models, and the majority of samples from the distribution of European windstorm losses. Bee (2012) and Scollnik and Sun (2012) also obtain rather low thresholds in the application of weibullgpdcon and lognormgpdcon mixtures to the Danish fire insurance losses.

There are two possible reasons for a location of the posterior for u in the body of the distribution. First, the GPD may already provide a good fit in the non-extreme part of the distribution. In this case, the EVMM approach may lead to a good approximation of the tail. The quantile estimates may also be

[17] The results for the U.S. hurricane data and fire insurance losses are shown in Table A3 in Appendix A.

more robust to outliers than under the POT approach, which is typically based on higher thresholds and therefore a lower number of observations. In view of the estimation results, we assume that this is the case, for example, for the RMS distribution of European windstorm losses.

The second reason is an inadequate bulk model choice. Under the mixture model approach, the whole distribution is approximated and the estimation leads to that threshold u and parameters $\boldsymbol{\theta}$, σ_u, and ξ (or, more exactly, to their posterior distributions) that provide the best model fit to the whole sample, not to the tail. If the bulk model is inadequate and the GPD permits a better fit to the body, rather low estimates for the threshold may result. In consequence, the estimates of the GPD parameters σ_u and ξ may be biased and inappropriate for the tail approximation. This explains, for instance, the clear superiority of the POT approach in the approximation of the tails of the Aon Re Belgium fire losses.

This problem can be illustrated by means of the profile likelihood function. The latter is given by $PL_{MM}(u|\boldsymbol{x}) = \max_{\boldsymbol{\theta},\sigma_u,\xi} L_{MM}(\boldsymbol{\theta}, u, \sigma_u, \xi|\boldsymbol{x})$. Figure 8 shows the profile log likelihood for the Aon Re data and various EVMMs. The profile likelihood functions all have their maximum at low values for u. As the maximum likelihood estimate corresponds to that threshold $\hat{u}$ that maximizes the profile likelihood (see, e.g., Murphy and van der Vaart, 2000), and the Bayesian posterior mean is close to $\hat{u}$ for flat uninformative priors (see, e.g., Condgon, 2006), the posterior distribution for u is centered around a low threshold.

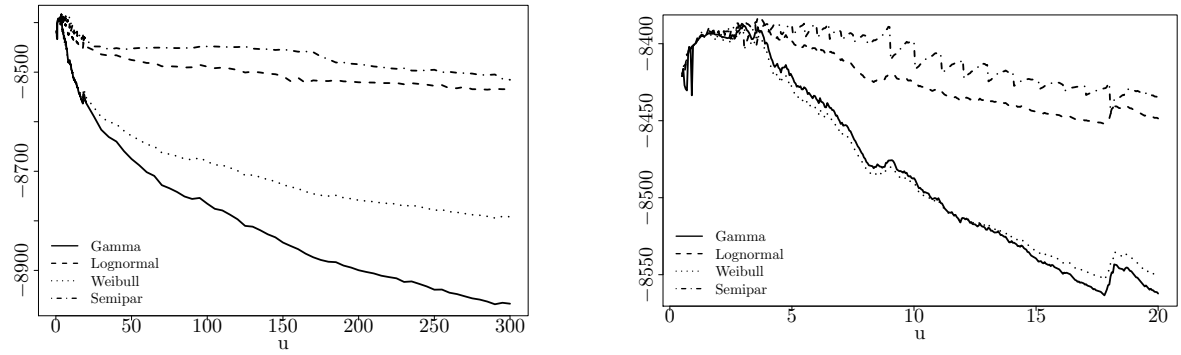


Figure 8: Profile Likelihood Functions for the Aon Re Belgium Fire Losses

This figure shows the profile likelihood functions for the Aon Re Belgium fire insurance losses for the ranges $u \in (0.5; 300)$ (left plot) and $u \in (0.5; 20)$ (right plot). Besides the semipargpd mixture (model 4), only the mixtures using the BTF approach without continuity constraint are taken into account. The solid line shows the profile likelihood for the gammagpd mixture (1b), the dashed line for model (2b) with lognormal bulk and the dotted line for the weibullgpd mixture (3b). The profile likelihood for model (4) is given by the dotdashed line.

Lee et al. (2012) restrict the range of possible thresholds to values above the 75% quantile. In the Bayesian context, this can be achieved by means of a truncation of the prior for u. However, the introduction of a lower boundary for u can only solve the problem some of the time. For instance, for the semipargpd model and Aon Re data, a restriction of u to values above the 80% data quantile leads to a posterior median of u of 106 (the 86% quantile) and an error of 10% in the estimation of $Q_{0.99}$. For the models with parametric bulk distribution, in contrast, the truncation basically corresponds to an ex ante fixing of the threshold, as the resulting posterior distributions are highly concentrated at the truncation point. This can also be expected from the plots of the profile likelihood functions of the parametric mixtures, which are steadily declining after the global maximum (see Figure 8).

6 Conclusions

In this paper, we provide a critical analysis of the accuracy of the EVMM approach in the approximation of the tails of heavy-tailed insurance loss distributions. Based on an empirically calibrated distribution of windstorm losses in Europe, we perform a comprehensive simulation study and evaluate the tail fit for different choices of the bulk model, tail fraction approaches, and continuity requirements. The results show a good approximation under the mixture models with a gamma or weibull bulk density for the large majority of data sets of size $n = 10,000$. The estimation errors tend to be higher for small sample sizes, but the approximation is comparable to or slightly better than under the POT method. In addition, under both the EVMM and POT approach, the results differ substantially between different samples of relatively small size, and outliers may have a great impact.

We further apply the EVMM method to two data sets of empirical insurance losses. For the sample of U.S. hurricane losses, the lognormgpd(con) mixtures lead to a much closer approximation of the largest observations than the remaining EVMMs, but cannot keep up with the POT approach. The tails of the distribution of fire insurance losses cannot be adequately modeled by any of the considered EVMMs. The POT approach, in contrast, permits a rather good approximation, and a simple pareto approximation leads to lower estimation errors, as well. These results are only based on the empirical quantiles. The quantiles of the true underlying distribution may differ, but the very low threshold estimates, exaggerated GPD shape parameters, and posterior distributions that do not even cover the empirical quantiles indicate that the EVMM approach is inadequate for this loss distribution.

The relatively good fit for the simulated data in contrast to the substantial estimation errors in the empirical applications is not the result of the different types of analyses. There are several empirical data sets which can be well approximated under the EVMM approach. For example, according to Wong and Li (2010) and Laas (2016), the weibullgpd mixture leads to a good approximation of the tails of a sample of reinsured automobile losses. Moreover, in the simulation study, we also compare the estimated quantiles with the empirical quantiles of the respective data set and obtain even slightly lower estimation errors (the results are not included in the paper). Our results therefore only reflect the considerable variation of the EVMM tail fit between different applications. A comparison of the estimates of ξ for the distributions of windstorm losses and reinsured automobile losses in Laas (2016) (around 0.2) with the estimates for the U.S. hurricane and Aon Re Belgium fire insurance losses indicates that the tail heaviness may be an influence factor on the adequacy of the EVMM approach. However, this requires further analyses. A fundamental problem of the EVMM method is the occurrence of rather low threshold estimates in the case of an inadequate bulk model in combination with a slow convergence of the excess distribution to the GPD (i.e., the GPD approximation is only valid in the tail). For the bulk distributions considered in this paper, the EVMM approach is therefore only sometimes an appropriate tool to circumvent the threshold selection problem of the POT method.

In comparison to the simple parametric distributions, the EVMM approach permits a comparable or better tail approximation for most applications considered in this paper (with very few exceptions for the pareto distribution at the highest quantile levels). Moreover, an analysis of the overall goodness-of-fit shows similar or better AIC values under the extreme value mixture distributions. If the goal is the approximation of the whole distribution by means of one single model with a relatively good tail

approximation up to certain quantile level, the EVMM method therefore seems to be a good (but also costly) alternative to the use of a simple parametric distribution.

In view of the substantial computational efforts, an insurer has to consider carefully whether it is worth implementing the EVMM approach. An advantage is the applicability of the EVMM approach to various risks an insurance company is exposed to. Thus the implementation costs can be split and the expenses are not completely in vain if the approach turns out to be inappropriate for the data set at hand. As mentioned in the introduction, the EVMM approach can also be applied to financial risks (in an extendend version). In addition, the method can be used for the approximation of operational loss distributions (see Bee, 2012, and Ergashev et al., 2013). Due to the very heavy tails of many distributions of operational losses (see, e.g., Moscadelli, 2004, and Chapelle et al., 2008), we assume that the EVMM tail approximation also varies substantially between different data sets. This requires further examination.

Our paper focuses on three parametric EVMMs, the semiparametric approach of Cabras and Castellanos (2011), as well as the kernelgpd mixture of MacDonald et al. (2013). Future work could extend the analyses to other semi- and nonparametric alternatives, such as the models of Tancredi et al. (2006), Do Nascimento et al. (2012), and Fúquene Patiño (2015). They approximate the bulk distribution (or some part of the bulk distribution) by means of mixtures of uniform distributions, a finite number of gamma densities, and a Dirichlet process mixture of gamma densities, respectively. This increases the flexibility of the bulk model compared to the fully parametric EVMMs, but also complicates the estimation and increases the computational burden.

Appendix A

		ξ	u	$Q_{0.95}$	$Q_{0.975}$	$Q_{0.99}$
	Empirical			$8.20 \cdot 10^3$	$1.15 \cdot 10^4$	$2.45 \cdot 10^4$
(1a)	Gammagpd, PTF	$(0.51; 1.51)$	$(20; 189)$	$(4.3 \cdot 10^3; 1.5 \cdot 10^4)$	$(6.7 \cdot 10^3; 3.9 \cdot 10^4)$	$(1.0 \cdot 10^4; 1.5 \cdot 10^5)$
(1b)	Gammagpd, BTF	$(0.74; 1.84)$	$(19; 81)$	$(4.1 \cdot 10^3; 2.1 \cdot 10^4)$	$(7.1 \cdot 10^3; 7.0 \cdot 10^4)$	$(1.1 \cdot 10^4; 3.5 \cdot 10^5)$
(1c)	Gammagpdcon, PTF	$(0.53; 1.44)$	$(38; 173)$	$(4.2 \cdot 10^3; 1.5 \cdot 10^4)$	$(6.9 \cdot 10^3; 3.8 \cdot 10^4)$	$(1.1 \cdot 10^4; 1.3 \cdot 10^5)$
(1d)	Gammagpdcon, BTF	$(1.41; 2.45)$	$(4; 23)$	$(5.0 \cdot 10^3; 5.1 \cdot 10^4)$	$(1.0 \cdot 10^4; 2.6 \cdot 10^5)$	$(2.6 \cdot 10^4; 2.3 \cdot 10^6)$
(2a)	Lognormgpd, PTF	$(-0.50; 3.23)$	$(70; 1.6 \cdot 10^4)$	$(5.1 \cdot 10^3; 1.1 \cdot 10^4)$	$(8.2 \cdot 10^3; 2.4 \cdot 10^4)$	$(1.1 \cdot 10^4; 8.1 \cdot 10^4)$
(2b)	Lognormgpd, BTF	$(-0.50; 4.80)$	$(2.6 \cdot 10^3; 1.6 \cdot 10^4)$	$(4.7 \cdot 10^3; 1.4 \cdot 10^4)$	$(8.1 \cdot 10^3; 2.9 \cdot 10^4)$	$(1.0 \cdot 10^4; 1.0 \cdot 10^5)$
(2c)	Lognormgpdcon, PTF	$(-0.23; 3.52)$	$(95; 1.5 \cdot 10^4)$	$(4.5 \cdot 10^3; 1.4 \cdot 10^4)$	$(6.7 \cdot 10^3; 3.4 \cdot 10^4)$	$(9.5 \cdot 10^2; 1.6 \cdot 10^5)$
(2d)	Lognormgpdcon, BTF	$(-0.40; 4.04)$	$(1.8 \cdot 10^3; 1.6 \cdot 10^4)$	$(4.5 \cdot 10^3; 1.4 \cdot 10^4)$	$(7.0 \cdot 10^3; 3.6 \cdot 10^4)$	$(1.2 \cdot 10^3; 4.9 \cdot 10^5)$
(3a)	Weibullgpd, PTF	$(0.52; 1.62)$	$(19; 250)$	$(4.2 \cdot 10^3; 1.6 \cdot 10^4)$	$(6.6 \cdot 10^3; 4.7 \cdot 10^4)$	$(1.1 \cdot 10^4; 1.9 \cdot 10^5)$
(3b)	Weibullgpd, BTF	$(0.70; 1.81)$	$(19; 82)$	$(4.1 \cdot 10^3; 2.1 \cdot 10^4)$	$(7.1 \cdot 10^3; 7.0 \cdot 10^4)$	$(1.2 \cdot 10^4; 3.4 \cdot 10^5)$
(3c)	Weibullgpdcon, PTF	$(0.52; 1.44)$	$(33; 201)$	$(4.2 \cdot 10^3; 1.5 \cdot 10^4)$	$(6.8 \cdot 10^3; 3.8 \cdot 10^4)$	$(1.1 \cdot 10^4; 1.3 \cdot 10^5)$
(4)	Semipargpd	$(0.32; 1.39)$	$(65; 1.1 \cdot 10^3)$	$(4.4 \cdot 10^3; 1.3 \cdot 10^4)$	$(7.1 \cdot 10^3; 3.3 \cdot 10^4)$	$(1.1 \cdot 10^4; 1.1 \cdot 10^5)$
(5a)	Kernelgpd, PTF	$(0.52; 1.36)$	$(65; 274)$	$(4.3 \cdot 10^3; 1.4 \cdot 10^4)$	$(6.7 \cdot 10^3; 3.3 \cdot 10^4)$	$(1.1 \cdot 10^4; 1.1 \cdot 10^5)$
(5b)	Kernelgpd, BTF	$(0.50; 1.35)$	$(65; 291)$	$(4.4 \cdot 10^3; 1.4 \cdot 10^4)$	$(6.7 \cdot 10^3; 3.3 \cdot 10^4)$	$(1.1 \cdot 10^4; 1.1 \cdot 10^5)$
(5c)	Kernelgpdcon, PTF	$(0.56; 1.42)$	$(79; 301)$	$(4.4 \cdot 10^3; 1.5 \cdot 10^4)$	$(7.0 \cdot 10^3; 3.7 \cdot 10^4)$	$(1.1 \cdot 10^4; 1.2 \cdot 10^5)$
(5d)	Kernelgpdcon, BTF	$(0.55; 1.40)$	$(79; 276)$	$(4.3 \cdot 10^3; 1.4 \cdot 10^4)$	$(6.9 \cdot 10^3; 3.6 \cdot 10^4)$	$(1.1 \cdot 10^4; 1.2 \cdot 10^5)$
(6a)	POT, $u = 1500$	$(0.07; 1.14)$		$(5.2 \cdot 10^3; 1.2 \cdot 10^4)$	$(7.9 \cdot 10^3; 2.5 \cdot 10^4)$	$(1.2 \cdot 10^4; 6.4 \cdot 10^4)$
(6b)	POT, $u = 2800$	$(-0.04; 1.30)$		$(5.3 \cdot 10^3; 1.2 \cdot 10^4)$	$(7.9 \cdot 10^3; 2.4 \cdot 10^4)$	$(1.2 \cdot 10^4; 6.8 \cdot 10^4)$
(6c)	POT, $u = 3912$	$(-0.20; 1.16)$		$(5.8 \cdot 10^3; 1.3 \cdot 10^4)$	$(8.8 \cdot 10^3; 2.5 \cdot 10^4)$	$(1.3 \cdot 10^4; 5.9 \cdot 10^4)$
(7)	Lognormal			$(4.7 \cdot 10^3; 1.4 \cdot 10^4)$	$(8.6 \cdot 10^3; 2.9 \cdot 10^4)$	$(1.8 \cdot 10^4; 6.9 \cdot 10^4)$
(8)	Pareto			$(4.4 \cdot 10^3; 3.4 \cdot 10^4)$	$(9.0 \cdot 10^3; 1.4 \cdot 10^5)$	$(2.2 \cdot 10^4; 9.8 \cdot 10^5)$
(9)	Burr			$(4.1 \cdot 10^3; 1.7 \cdot 10^4)$	$(7.1 \cdot 10^3; 4.8 \cdot 10^4)$	$(1.2 \cdot 10^4; 1.9 \cdot 10^5)$

Table A1: Posterior Intervals for the U.S. Hurricane Data Set

This table shows the 95% posterior intervals of the estimates for the sample of U.S hurricane losses. The table contains the posterior intervals for the shape parameter ξ, the threshold u, as well as the quantiles Q_α for $\alpha = 0.95/0.975/0.99$.

		ξ	u	$Q_{0.975}$	$Q_{0.99}$	$Q_{0.995}$
	Empirical			$1.96 \cdot 10^3$	$6.44 \cdot 10^3$	$9.55 \cdot 10^3$
(1a)	Gammagpd, PTF	(1.50; 1.80)	(2.9; 5.1)	$(1.3 \cdot 10^3; 2.6 \cdot 10^3)$	$(4.9 \cdot 10^3; 1.3 \cdot 10^4)$	$(1.3 \cdot 10^4; 4.6 \cdot 10^4)$
(1b)	Gammagpd, BTF	(1.54; 1.84)	(2.8; 3.8)	$(1.3 \cdot 10^3; 2.8 \cdot 10^3)$	$(5.3 \cdot 10^3; 1.5 \cdot 10^4)$	$(1.5 \cdot 10^4; 5.2 \cdot 10^4)$
(2a)	Lognormgpd, PTF	(1.51; 1.81)	(2.7; 4.1)	$(1.3 \cdot 10^3; 2.7 \cdot 10^3)$	$(4.9 \cdot 10^3; 1.4 \cdot 10^4)$	$(1.3 \cdot 10^4; 4.7 \cdot 10^4)$
(2b)	Lognormgpd, BTF	(1.50; 1.79)	(3.0; 3.8)	$(1.3 \cdot 10^3; 2.6 \cdot 10^3)$	$(4.9 \cdot 10^3; 1.3 \cdot 10^4)$	$(1.4 \cdot 10^4; 4.6 \cdot 10^4)$
(2c)	Lognormgpdcon, PTF	(1.47; 1.84)	(1.5; 7.9)	$(1.2 \cdot 10^3; 2.8 \cdot 10^3)$	$(4.5 \cdot 10^3; 1.4 \cdot 10^4)$	$(1.2 \cdot 10^4; 5.0 \cdot 10^4)$
(2d)	Lognormgpdcon, BTF	(1.62; 1.91)	(1.7; 2.9)	$(1.5 \cdot 10^3; 3.2 \cdot 10^3)$	$(6.3 \cdot 10^3; 1.8 \cdot 10^4)$	$(1.9 \cdot 10^4; 6.7 \cdot 10^4)$
(3a)	Weibullgpd, PTF	(1.50; 1.81)	(3.0; 4.5)	$(1.2 \cdot 10^3; 2.6 \cdot 10^3)$	$(4.8 \cdot 10^3; 1.4 \cdot 10^4)$	$(1.3 \cdot 10^4; 4.7 \cdot 10^4)$
(3b)	Weibullgpd, BTF	(1.55; 1.84)	(1.4; 3.2)	$(1.3 \cdot 10^3; 2.8 \cdot 10^3)$	$(5.4 \cdot 10^3; 1.5 \cdot 10^4)$	$(1.6 \cdot 10^4; 5.4 \cdot 10^4)$
(3c)	Weibullgpdcon, PTF	(1.51; 1.81)	(3.2; 4.6)	$(1.3 \cdot 10^3; 2.7 \cdot 10^3)$	$(4.9 \cdot 10^3; 1.4 \cdot 10^4)$	$(1.3 \cdot 10^4; 4.8 \cdot 10^4)$
(3d)	Weibullgpdcon, BTF	(1.61; 1.88)	(1.3; 2.0)	$(1.4 \cdot 10^3; 3.0 \cdot 10^3)$	$(6.0 \cdot 10^3; 1.7 \cdot 10^4)$	$(1.8 \cdot 10^4; 6.0 \cdot 10^4)$
(4)	Semipargpd	(1.51; 1.84)	(2.7; 6.5)	$(1.3 \cdot 10^3; 2.8 \cdot 10^3)$	$(5.0 \cdot 10^3; 1.5 \cdot 10^4)$	$(1.3 \cdot 10^4; 5.1 \cdot 10^4)$
(6a)	POT, $u = 206$	(1.05; 1.63)		$(1.4 \cdot 10^3; 2.3 \cdot 10^3)$	$(3.4 \cdot 10^3; 8.5 \cdot 10^3)$	$(6.5 \cdot 10^3; 2.3 \cdot 10^4)$
(6b)	POT, $u = 700$	(0.83; 1.51)		$(1.7 \cdot 10^3; 2.6 \cdot 10^3)$	$(4.0 \cdot 10^3; 7.7 \cdot 10^3)$	$(6.8 \cdot 10^3; 1.7 \cdot 10^4)$
(6c)	POT, $u = 1700$	(0.46; 1.13)		$(1.8 \cdot 10^3; 1.9 \cdot 10^3)$	$(4.0 \cdot 10^3; 7.6 \cdot 10^3)$	$(6.9 \cdot 10^3; 1.7 \cdot 10^4)$
(7)	Lognormal			$(5.5 \cdot 10^2; 7.6 \cdot 10^2)$	$(1.1 \cdot 10^3; 1.6 \cdot 10^3)$	$(1.8 \cdot 10^3; 2.6 \cdot 10^3)$
(8)	Pareto			$(1.0 \cdot 10^3; 1.8 \cdot 10^3)$	$(3.4 \cdot 10^3; 7.6 \cdot 10^3)$	$(8.7 \cdot 10^3; 2.3 \cdot 10^4)$
(9)	Burr			$(2.2 \cdot 10^3; 4.9 \cdot 10^3)$	$(1.2 \cdot 10^4; 3.4 \cdot 10^4)$	$(4.1 \cdot 10^4; 1.4 \cdot 10^5)$

Table A2: Posterior Intervals for the Aon Re Data Set

This table presents the 95% posterior intervals of the estimates for the sample of Aon Re Belgium fire insurance losses. The columns show the posterior intervals for the shape parameter ξ, the threshold u, as well as the quantiles Q_α for $\alpha = 0.975/0.99/0.99.5$.

		US Hurricane Losses			Aon Re Belgium Fire Losses		
		$-L$	AIC	BIC	$-L$	AIC	BIC
(1a)	Gammagpd, PTF	1,241	2,494	2,512	8,387	16,786	16,819
(1b)	Gammagpd, BTF	1,241	2,491	2,507	8,391	16,791	16,819
(1c)	Gammagpdcon, PTF	1,241	2,492	2,507	no conv.	no conv.	no conv.
(1d)	Gammagpdcon, BTF	1,251	2,511	2,523	no conv.	no conv.	no conv.
(2a)	Lognormgpd, PTF	1,244	2,500	2,519	8,386	16,785	16,818
(2b)	Lognormgpd, BTF	1,245	2,497	2,510	8,386	16,782	16,809
(2c)	Lognormgpdcon, PTF	1,244	2,498	2,513	8,390	16,790	16,817
(2d)	Lognormgpdcon, BTF	1,245	2,498	2,510	8,394	16,795	16,817
(3a)	Weibullgpd, PTF	1,241	2,494	2,512	8,388	16,788	16,821
(3b)	Weibullgpd, BTF	1,240	2,490	2,506	8,394	16,797	16,825
(3c)	Weibullgpdcon, PTF	1,241	2,493	2,508	8,389	16,788	16,815
(3d)	Weibullgpdcon, BTF	no conv.	no conv.	no conv.	8,396	16,801	16,823
(4)	Semipargpd	1,240	not def.	not def.	8,388	not def.	not def.
(5a)	Kernelgpd, PTF	1,239	not def.	not def.	not calc.	not def.	not def.
(5b)	Kernelgpd, BTF	1,238	not def.	not def.	not calc.	not def.	not def.
(5c)	Kernelgpdcon, PTF	1,238	not def.	not def.	not calc.	not def.	not def.
(5d)	Kernelgpdcon, BTF	1,238	not def.	not def.	not calc.	not def.	not def.
(7)	Lognormal	1,245	2,493	2,500	8,580	17,165	17,176
(8)	Pareto	1,251	2,507	2,513	8,503	17,010	17,021
(9)	Burr	1,250	2,506	2,515	8,406	16,817	16,834

Table A3: Goodness-of-Fit Statistics U.S. Hurricane and Aon Re Belgium Fire Losses

This table shows the goodness-of-fit results for the samples of U.S. hurricane losses and fire insurance losses. The columns show the posterior medians of the negative loglikelihood $-L$, the AIC, and BIC. The fit of model (3d) to the hurricane data and of models (1c) and (1d) to the fire losses is not shown due to convergence problems of the Markov chains. The AIC and BIC are not defined for models (4) and (5a) to (5d) and the kernelgpd EVMMs are not applied to the Aon Re Belgium fire losses.

Appendix B

For the data provided by Risk Management Solutions, Inc., please consider the following disclaimer:

This Information is compiled using proprietary computer risk assessment technology of Risk Management Solutions, Inc. ("RMS"). The technology and data used in providing this Information is based on the scientific data, mathematical and empirical models, and encoded experience of scientists and specialists (including without limitation: earthquake engineers, wind engineers, structural engineers, geologists, seismologists, meteorologists, geotechnical specialists and mathematicians). As with any model of physical systems, particularly those with low frequencies of occurrence and potentially high severity outcomes, the actual losses from catastrophic events may differ from the results of simulation analyses. Furthermore, the accuracy of predictions depends largely on the accuracy and quality of the data used. The Information is provided under license to the Institut für Versicherungswirtschaft I.VW-HSG and is RMS proprietary information and may not be reproduced by any third party without the prior written consent RMS. Furthermore, this Information may not be used under any circumstances in the development or calibration of any product or service offering that competes with RMS. The recipient of this Information is further advised that RMS is not engaged in the insurance, reinsurance, or related industries, and that the Information provided is not intended to constitute professional advice. RMS SPECIFICALLY DISCLAIMS ANY AND ALL RESPONSIBILITIES, OBLIGATIONS AND LIABILITY WITH RESPECT TO ANY DECISIONS OR ADVICE MADE OR GIVEN AS A RESULT OF THE INFORMATION OR USE THEREOF, INCLUDING ALL WARRANTIES, WHETHER EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO, WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. IN NO EVENT SHALL RMS (OR ITS PARENT, SUBSIDIARY, OR OTHER AFFILIATED COMPANIES) BE LIABLE FOR DIRECT, INDIRECT, SPECIAL, INCIDENTAL, OR CONSEQUENTIAL DAMAGES WITH RESPECT TO ANY DECISIONS OR ADVICE MADE OR GIVEN AS A RESULT OF THE CONTENTS OF THIS INFORMATION OR USE THEREOF.

References

Ahn, S., J. H. T. Kim, and V. Ramaswami, 2012. A New Class of Models for Heavy Tailed Distributions in Finance and Insurance Risk. *Insurance: Mathematics and Economics*, 51(1):43–52.

Balasooriya, U. and C.-K. Low, 2008. Modeling Insurance Claims with Extreme Observations: Transformed Kernel Density and Generalized Lambda Distribution. *North American Actuarial Journal*, 12(2):129–142.

Balkema, A. A. and L. de Haan, 1974. Residual Life Time at Great Age. *The Annals of Probability*, 2(5):792–804.

Bee, M., 2012. Statistical Analysis of the Lognormal-Pareto Distribution using Probability Weighted Moments and Maximum Likelihood, University of Trento, Department of Economics, Working Paper No. 08/2012.

Behrens, C. N., H. F. Lopes, and D. Gamerman, 2004. Bayesian Analysis of Extreme Events with Threshold Estimation. *Statistical Modelling*, 4(3):227–244.

Bolancé, C., M. Guillen, and J. P. Nielsen, 2003. Kernel Density Estimation of Actuarial Loss Functions. *Insurance: Mathematics and Economics*, 32:19–36.

Boneva, L. I., D. G. Kendall, and I. Stefanov, 1971. Spline Transformations: Three New Diagnostic Aids for the Statistical Data Analyst (With Discussion). *Journal of the Royal Statistical Society. Series B1*, 33:1–70.

Brewer, M. J., 1998. A Modelling Approach for Bandwidth Selection in Kernel Density Estimation. In R. Payne and P. Green, editors, *Proceedings of COMPSTAT*, pages 203–208. Physica Verlag, Heidelberg, Germany.

Brewer, M. J., 2000. A Bayesian Model for Local Smoothing in Kernel Density Estimation. *Statistics and Computing*, 10:299–309.

Brooks, S. P., 2002. Discussion on the Paper by Spiegelhalter, Best , Carlin and van der Linde. *Journal of the Royal Statistical Society. Series B (Methodological)*, 64(4):616–618.

Brooks, S. P. and A. Gelman, 1998. General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics*, 7(4):434–455.

Cabras, S. and M. E. Castellanos, 2009. An Objective Bayesian Approach for Threshold Estimation in the Peaks Over the Threshold Model, Working Paper.

Cabras, S. and M. E. Castellanos, 2011. A Bayesian Approach for Estimating Extreme Quantiles under a Semiparametric Mixture Model. *ASTIN Bulletin*, 41(1):87–106.

Carreau, J. and Y. Bengio, 2009. A Hybrid Pareto Model for Asymmetric Fat-tailed Data: The Univariate Case. *Extremes*, 12(1):53–76.

Castellanos, E. M. and S. Cabras, 2007. A Default Bayesian Procedure for the Generalized Pareto Distribution. *Journal of Statistical Planning and Inference*, 137:473–483.

Chapelle, A., Y. Crama, G. Hübner, and J.-P. Peters, 2008. Practical Methods for Measuring and Managing Operational Risk in the Financial Sector: A Clinical Study. *Journal of Banking & Finance*, 32(6):1049–1061.

Coles, S. G. and J. A. Tawn, 1996. A Bayesian Analysis of Extreme Analysis Rainfall Data. *Journal of the Royal Statistical Society: Series C*, 45(4):463–478.

Collins, D. J. and S. P. Lowe, 2001. A Macro Validation Dataset for U.S. Hurricane Models. (Available at: http://www.casact.org/pubs/forum/01wforum/01wf217.pdf, accessed September 25$^{\text{th}}$ 2015).

Condgon, P., 2006. *Bayesian Statistical Modelling*. John Wiley & Sons, Chichester, United Kingdom, 2nd edition.

Cooray, K. and M. M. A. Ananda, 2005. Modeling Actuarial Data with a Composite Lognormal-Pareto Model. *Scandinavian Actuarial Journal*, (5):321–334.

Craiu, R. V. and J. S. Rosenthal, 2014. Bayesian Computation via Markov Chain Monte Carlo. *Annual Review of Statistics and Its Application*, 1:179–201.

Cummins, D. J., C. M. Lewis, and R. D. Phillips, 1999. Pricing Excess-of-Loss Reinsurance Contracts against Catastrophic Loss. In K. A. Froot, editor, *The Financing of Catastrophe Risk*, pages 93–148. University of Chicago Press, Chicago, Illinois.

Cummins, J., G. Dionne, J. B. McDonald, and M. M. Pritchett, 1990. Applications of the GB2 Family of Distributions in Modeling Insurance Loss Processes. *Insurance: Mathematics and Economics*, 9(4):257–272.

de Alba, E., 2006. Claims Reserving when there are Negative Values in the Development Triangle. *North American Actuarial Journal*, 10(3):45–59.

Do Nascimento, F. F., D. Gamerman, and H. F. Lopes, 2012. A Semiparametric Bayesian Approach to Extreme Value Estimation. *Statistics and Computing*, 22(2):661–675.

DuMouchel, W. H., 1983. Estimating the Stable Index α in order to Measure Tail Thickness: A Critique. *The Annals of Statistics*, 11(4):1019–1031.

Efron, B. and R. Tibshirani, 1996. Using Specially Designed Exponential Families for Density Estimation. *Annals of Statistics*, 24(6):2431–2461.

Eling, M., 2012. Fitting Insurance Returns to Skewed Distributions: Are the Skew-normal and Skew-student Good Models? *Insurance: Mathematics and Economics*, 51(2):239–248.

Embrechts, P., C. Klüppelberg, and T. Mikosch, 2008. *Modelling Extremal Events for Insurance and Finance*. Springer Verlag, Berlin, Heidelberg, Germany, 4th edition.

Ergashev, B., S. Mittnik, and E. Sekeris, 2013. A Bayesian Approach to Extreme Value Estimation in Operational Risk Modeling. *Journal of Operational Risk*, 4:55–81.

Fúquene Patiño, J. A., 2015. A Semi-parametric Bayesian Extreme Value Model Using a Dirichlet Process Mixture of Gamma Densities. *Journal of Applied Statistics*, 42(2):267–280.

Gamerman, D. and H. F. Lopes, 2006. *Markov Chain Monte Carlo in Practice, Stochastic Simulation for Bayesian Inference*. Chapman & Hall/CRC, Boca Raton, Florida, 2nd edition.

Gelman, A., 1996. Inference and Monitoring Convergence. In W. Gilks, S. Richardson, and D. J. Spiegelhalter, editors, *Markov Chain Monte Carlo in Practice*, pages 131–143. Chapman & Hall / CRC, Boca Raton, Florida.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, 2013. *Bayesian Data Analysis (with Corrections for Third Printing, 7 Apr 2015)*. CRC Press, Boca Raton, Florida, 3rd edition.

Gelman, A., G. O. Roberts, and W. R. Gilks, 1996. Efficient Metropolis Jumping Rules. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors, *Bayesian Statistics 5*, pages 599–607. Oxford University Press, Oxford, United Kingdom.

Gelman, A. and D. B. Rubin, 1992. Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, 7(4):457–472.

Grossi, P., H. Kunreuther, and D. Windeler, 2005. An Introduction to Catastrophe Models and Insurance. In P. Grossi and H. Kunreuther, editors, *Catastrophe Modeling: A New Approach to Managing Risk*, pages 23–42. Springer, Boston, Massachusetts.

Guillen, M., F. Prieto, and J. M. Sarabia, 2011. Modelling Losses and Locating the Tail with the Pareto Positive Stable Distribution. *Insurance: Mathematics and Economics*, 49(3):454–461.

Hastings, W. K., 1970. Monte Carlo Sampling Methods using Markov Chains and their Applications. *Biometrika*, 57(1):97–109.

Hu, Y., 2013. Extreme Value Mixture Modelling with Simulation Study and Applications in Finance and Insurance, University of Canterbury, Master Thesis.

Jagger, T. H., J. B. Elsner, and M. A. Saunders, 2008. Forecasting U.S. Insured Hurricane Losses. In H. F. Diaz and R. J. Murnane, editors, *Climate Extremes and Society*, pages 189–208. Cambridge University Press, New York.

Laas, D., 2016. Bayesian Estimation of Extreme Value Mixture Models: Enhancements and Simplifications, Institute of Insurance Economics, University of St. Gallen, Working Paper.

Lee, D., W. K. Li, and T. S. T. Wong, 2012. Modeling Insurance Claims via a Mixture Exponential Model Combined with Peaks-Over-Threshold Approach. *Insurance: Mathematics and Economics*, 51(3):538–550.

Lindsey, J. K., 1974a. Comparison of Probability Distributions. *Journal of the Royal Statistical Society. Series B*, 36(1):38–47.

Lindsey, J. K., 1974b. Construction and Comparison of Statistical Models. *Journal of the Royal Statistical Society. Series B*, 36(3):418–425.

MacDonald, A., 2011. Extreme Value Mixture Modelling with Medical and Industrial Applications, University of Canterbury, PhD Thesis.

MacDonald, A., C. Scarrott, D. Lee, B. Darlow, M. Reale, and G. Russell, 2011. A Flexible Extreme Value Mixture Model. *Computational Statistics & Data Analysis*, 55(6):2137–2157.

MacDonald, A., C. J. Scarrott, and D. Lee, 2013. Boundary Correction, Consistency and Robustness of Kernel Densities Using Extreme Value Theory, University of Canterbury, Working Paper.

Marín, J. M., M. T. Rodríguez-Bernal, and M. Wiper, 2005. Using Weibull Mixture Distributions to Model Heterogeneous Survival Data. *Communications in Statistics - Simulation and Computation*, 34:673–684.

McNeil, A. J., 1997. Estimating the Tails of Loss Severity Distributions Using Extreme Value Theory. *ASTIN Bulletin*, 27(1):117–137.

McNeil, A. J., R. Frey, and P. Embrechts, 2005. *Quantitative Risk Management, Concepts, Techniques and Tools*. Princton University Press, Princton, New Jersey.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, 1953. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21(6):1087–1092.

Milidonis, A. and M. F. Grace, 2008. Tax-Deductible Pre-event Catastrophe Loss Reserves: The Case of Florida. *ASTIN Bulletin*, 38(1):13–51.

Moscadelli, M., 2004. The Modelling of Operational Risk: Experience with the Analysis of the Data Collected by the Basel Committee, Banca D'Italia, Temi di Discussione, No. 517.

Murphy, S. A. and A. W. van der Vaart, 2000. On Profile Likelihood. *Journal of the American Statistical Association*, 95(450):449–465.

Northrop, P. J. and N. Attalides, 2014. Posterior Propriety in Objective Bayesian Extreme Value Analyses, University College London, Department of Statistical Science, Working Paper No. 323.

Oumow, B., M. D. Carvalho, and A. C. Davison, 2012. Bayesian P-spline Mixture Modeling of Extreme Forest Temperatures, Working Paper.

Pickands, J., 1975. Statistical Inference Using Extreme Order Statistics. *The Annals of Statistics*, 3(1):119–131.

Reiss, R. D. and M. Thomas, 2007. *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology, and Other Fields*. Birkhäuser Verlag, Basel, Switzerland, 3rd edition.

Richardson, S. and P. J. Green, 1997. On Bayesian Analysis of Mixtures with an Unknown Number of Components. *Journal of the Royal Statistical Society. Series B*, 59(4):731–792.

Roberts, G. O., A. Gelman, and W. R. Gilks, 1997. Weak Convergence and Optimal Scaling of Random Walk Metropolis Algorithms. *The Annals of Applied Probability*, 7(1):110–120.

Roberts, G. O. and J. S. Rosenthal, 2001. Optimal Scaling for Various Metropolis Hastings Algorithms. *Statistical Science*, 16(4):351–367.

Scarrott, C. and Y. Hu, 2015. evmix: Extreme Value Mixture Modelling, Threshold Estimation and Boundary Corrected Kernel Density Estimation. (Available on CRAN. http://www.math.canterbury.ac.nz/∼c.scarrott/evmix/).

Scarrott, C. and A. MacDonald, 2012. A Review of Extreme Value Threshold Estimation and Uncertainty Quantification. *REVSTAT - Statististical Journal*, 10(1):33–60.

Scarrott, C. J., 2016. Univariate Extreme Value Mixture Modelling. In J. Yan and D. K. Dey, editors, *Extreme Value Modeling and Risk Analysis: Methods and Applications*, pages 41–67. Chapman & Hall, Boca Raton, Florida.

Schmutz, M. and R. R. Doerr, 1998. *The Pareto Model in Property Reinsurance, Formulas and Applications*. Zurich, Switzerland.

Scollnik, D. P. M., 2007. On Composite Lognormal-Pareto Models. *Scandinavian Actuarial Journal*, (1):20–33.

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

Sinha, S. K., 1986. Bayes Estimation of the Reliability Function and Hazard Rate of a Weibull Failure Time Distribution. *Trabajos de Estadistica*, 1(2):47–56.

So, M. K. P. and R. K. S. Chan, 2014. Bayesian Analysis of Tail Asymmetry Based on a Threshold Extreme Value Model. *Computational Statistics and Data Analysis*, 71(1997):568–587.

Stephenson, A. and M. Ribatet, 2006. A User's Guide to the evdbayes Package (Version 1.1).

Tancredi, A., C. Anderson, and A. O'Hagan, 2006. Accounting for Threshold Uncertainty in Extreme Value Estimation. *Extremes*, 9:87–106.

Tsionas, E. G., 2002. Bayesian Analysis of Finite Mixtures of Weibull Distributions. *Communications in Statistics - Theory and Methods*, 31(1):37–48.

Villa, C., 2015. Bayesian Estimation of the Threshold of a Generalised Pareto Distribution for Heavy-Tailed Observations, School of Mathematics, Statistics and Actuarial Science, University of Kent, Working Paper.

Wiper, M. and D. Rios, 2001. Mixtures of Gamma Distributions with Applications. *Journal of Computational and Graphical Statistics*, 10(3):440–454.

Wong, T. S. T. and W. K. Li, 2010. A Threshold Approach for Peaks-Over-Threshold Modeling Using Maximum Product of Spacings. *Statistica Sinica*, 20:1257–1272.

Zhang, X., M. L. King, and R. J. Hyndman, 2006. A Bayesian Approach to Bandwidth Selection for Multivariate Kernel Density Estimation. *Computational Statistics & Data Analysis*, 50(11):3009–3031.

Zhao, X., C. Scarrott, L. Oxley, and M. Reale, 2010. Extreme Value Modelling for Forecasting Market Crisis Impacts. *Applied Financial Economics*, 20(1-2):63–72.

Zhao, X., C. J. Scarrott, L. Oxley, and M. Reale, 2011. GARCH Dependence in Extreme Value Models with Bayesian Inference. *Mathematics and Computers in Simulation*, 81(7):1430–1440.