BAYESIAN ESTIMATION OF EXTREME VALUE MIXTURE MODELS: SIMPLIFICATIONS AND ENHANCEMENTS

DANIELA LAAS

WORKING PAPERS ON RISK MANAGEMENT AND INSURANCE NO. 173

EDITED BY HATO SCHMEISER
CHAIR FOR RISK MANAGEMENT AND INSURANCE

APRIL 2016
Bayesian Estimation of Extreme Value Mixture Models: Simplifications and Enhancements

Daniela Laas

This version: April 4th 2016

Abstract

This paper focuses on various computational aspects in the Markov chain Monte Carlo simulation of the Bayesian posterior distributions for extreme value mixture models. Using a comprehensive simulation study and empirical application to historical insurance losses, we consider several enhancements and examine their benefits in comparison to the commonly used standard method. Our analysis shows that parallel tempering can substantially improve the convergence and mixing behaviour of the Markov chains. In contrast, a reparameterisation of the generalized pareto distribution to overcome the threshold dependence of the scale parameter does not seem to be necessary in a real world setting with limited samples size. The analysis further supports the use of a simplified sampling algorithm with maximum likelihood approximation of the tail fraction compared to a full Bayesian approach.

Keywords: Extreme Value Mixture Models, Bayesian Analysis, Markov Chain Monte Carlo Simulation, Parallel Tempering, Reparameterisation, Tail Fraction Estimation, Insurance Losses
1 Introduction

The tradeoff of high costs of capital versus a great impact of extreme insurance losses on an insurer’s solvency requires an exact estimation of the risk of large claims. Due to its applicability to many distributions, the probability of extreme losses is frequently estimated via the peaks-over-threshold (POT) approach (see, e.g., McNeil, 1997, Rootzén and Tajvidi, 1997, Cebrían et al., 2003, Jagger et al., 2008, Brodin and Rootzén, 2009). Other fields of application include, among others, finance (see, e.g., Chavez-Demoulin and Embrechts, 2004), operational risk management (see, e.g., Chavez-Demoulin et al., 2006), and hydrology (see, e.g., Engeland et al., 2005). In order to overcome the threshold selection problem, various extreme value mixture models (EVMMs) have been proposed in the last years. They approximate the whole unknown distribution by splicing the generalized pareto distribution (GPD) at the threshold with some parametric distribution or density estimator for the central part of the support (also called bulk, see, e.g., Scarrott and MacDonald, 2012). In this way, the threshold becomes an additional parameter that can be estimated together with the parameters of the GPD and bulk model.

As the likelihood function frequently has several local maxima, the estimation of the EVMM parameters via the maximum likelihood (ML) method may be very difficult (see, e.g., Scarrott, 2016). Moreover, the ML method and related approaches (e.g., profile likelihood estimation) all provide one single estimator for the threshold. In this way, the plausibility of various thresholds, which is typical for the POT approach and EVMMs, is negated (see, e.g., Scarrott and MacDonald, 2012). The Bayesian approach, in contrast, does not require a direct optimization of the likelihood and leads to a whole distribution for each parameter (see, e.g., Scarrott and MacDonald, 2012). Inferences (e.g., predictions about future observations) are therefore not based on one specific threshold, but on various thresholds that are compatible with the data. In addition, if available, the Bayesian framework permits the consideration of prior expert knowledge on the parameters or other quantities of interest (see, e.g., Behrens et al., 2004, for an application of the method of Coles and Tawn, 1996). This may be especially valid for estimations in the tail, where only few observations are available (see, e.g., Behrens et al., 2004).

In view of these advantages, various EVMMs have been proposed in the Bayesian context (for a detailed overview of the whole class of EVMMs, see Scarrott, 2016). In their seminal paper, Behrens et al. (2004) suggest the combination of the GPD and a gamma bulk model. They further note the possibility of several other parametric distributions for the non-extreme part, such as the weibull or normal distribution. In the Bayesian framework, a normal bulk model is used by Cabras and Castellanos (2011) and Oumow et al. (2012), among others. Cabras and Castellanos (2011) also consider weibull-GPD and lognormal-GPD mixtures and develop a partial Bayesian approach for the estimation of a semiparametric EVMM. Do Nascimento et al. (2012) and Fúquene Patiño (2015) extend the model of Behrens et al. (2004) and approximate the central part of the distribution by means of a finite mixture and a Dirichlet process mixture of gamma distributions, respectively. Moreover, Tancredi et al. (2006), MacDonald et al. (2011), and MacDonald et al. (2013) propose the Bayesian estimation of EVMMs with nonparametric bulk components in form of a mixture of uniform distributions, a kernel density, and a boundary corrected kernel density, respectively.

The posterior distribution cannot be calculated analytically and is usually approximated via Markov chain Monte Carlo (MCMC) sampling. For this, most papers use the Metropolis Hastings (MH) algorithm. If applied to EVMMs, various difficulties can arise. These include problems of convergence or bad
mixing as a result of a multimodal posterior distribution and sampling inefficiency due to the threshold dependence of the GPD scale parameter (see, e.g., Scarrott and MacDonald, 2012). In addition, various papers with an additional parameter for the tail fraction approximate this parameter via the maximum likelihood estimate and apply MH sampling only to the remaining model parameters (see, e.g., MacDonald et al., 2011, and Oumow et al., 2012).

In this paper, we comprehensively analyse these computational issues in form of a simulation study and an empirical application to insurance losses. In particular, we examine the benefits of parallel tempering for the simulation of the Markov chains. The method of parallel tempering extends the MH algorithm in order to improve the convergence and mixing behaviour in the case of multimodal target distributions (see, e.g., Gill and Casella, 2004). Furthermore, the study investigates the need of a reparameterisation of the GPD to overcome the parameter dependence. Finally, we contrast the results of a full Bayesian approach for all parameters (i.e., with MH sampling also for the tail fraction) with those under the standard method that approximates the tail proportion via the ML estimator. The analysis comprises three EVMMs with different bulk distributions (gamma, lognormal, and weibull density). The algorithms are evaluated by means of various criteria, such as the convergence behaviour and autocorrelations of the Markov chains and the approximation of the underlying distribution.

The analysis of the simulated and empirical data sets indicates that the parameter dependence does not have major negative effects in applications with limited sample sizes or samples that do not come from an EVMM distribution. In addition, the use of the simplified tail fraction sampling method only marginally influences the estimation of the extreme quantiles. The parallel tempering algorithm proves to be very useful for the simulation of the Markov chains of the EVMM parameters, as it substantially lowers the rate of non-convergence and autocorrelations (the lag-50 autocorrelations mostly by at least 40%).

Our contribution is relevant for researchers and practitioners. The threshold dependence of the GPD scale parameter is frequently assumed to be a non-negligible problem (see, in particular, Scarrott and MacDonald, 2012), but, to the knowledge of the author, the effects have not been comprehensively examined, so far. Similarly, although the adequacy of the simplified tail fraction estimation method has not been investigated via a comparison with the full Bayesian approach, various papers are based on the simplified method. Our findings justify the procedure and results in previous papers where the parameter dependence is not taken into account (see, e.g., Behrens et al., 2004, and Cabras and Castellanos, 2011), or the tail fraction is approximated via the maximum likelihood estimator (see, e.g., MacDonald et al., 2011, and Oumow et al., 2012). Moreover, our results simplify the MCMC simulation in future analyses. The method of parallel tempering is regularly applied in natural sciences (e.g., physics, see, e.g., Earl and Deem, 2005), but has not been considered for the estimation of EVMMs. In light of the frequent multimodality of the posterior distributions of EVMMs (see, e.g., Cabras and Castellanos, 2011, and Scarrott and MacDonald, 2012), the results obtained in other fields of application, and the sampling scheme, we may expect that the method improves the MCMC simulation of the EVMM parameters. However, as the method requires higher computational efforts, its benefits have to be examined thoroughly, especially if the “temperatures” are chosen in a simple way. Our results finding substantial benefits encourage the use of the method in the context of EVMMs and can therefore help to avoid convergence or mixing problems in future applications.
The paper is structured as follows: Section 2 briefly introduces the EVMM approach. The third section describes the standard Bayesian estimation of EVMMs and the proposed alternatives and enhancements. In Sections 4 and 5, we present the results of the simulation study and empirical application, respectively. The last section presents our conclusions.

2 Extreme Value Mixture Models

In order to approximate different types of distributions and improve the model fit, a large variety of EVMMs has been proposed (for a comprehensive overview, see, e.g., Scarrott, 2016). In many cases, the density has the form (see, e.g., Scarrott, 2016):

$$f_{\text{MM}}(x| \theta, u, \sigma_u, \xi) = \begin{cases} (1 - \phi_u) \cdot h(x| \theta), & x \leq u \\ \phi_u \cdot g(x| u, \sigma_u, \xi), & x > u. \end{cases}$$

(1)

Here, $h(\cdot| \theta)$ denotes the density for the non-extreme part of the distribution (the bulk) and $\theta = (\theta_1, ..., \theta_K)$ the corresponding parameter vector. $g(x| u, \sigma_u, \xi)$ is the density of the GPD distribution with shape $\xi \in \mathbb{R}$, location $u$ (the threshold), and scale $\sigma_u > 0$, i.e. (see, e.g., Do Nascimento et al., 2012):

$$g(x| u, \sigma_u, \xi) = \begin{cases} \frac{1}{\sigma_u} \cdot \left(1 + \xi \cdot \frac{x-u}{\sigma_u}\right)^{-\frac{1+\xi}{\xi}}, & \xi > 0, x > u \\ \frac{1}{\sigma_u} \cdot \left(1 + \xi \cdot \frac{x-u}{\sigma_u}\right)^{-\frac{1+\xi}{\xi}}, & \xi < 0, u < x < u - \sigma_u/\xi \\ \frac{1}{\sigma_u} \cdot \exp\left(-\frac{x-u}{\sigma_u}\right), & \xi = 0, x > u. \end{cases}$$

(2)

The use of the GPD for the tail of the distribution is theoretically justified by Balkema and de Haan (1974) and Pickands (1975). Their work shows that, for a broad class of distributions $F$ of a random variable $X$, the conditional distribution of the excesses $X - u$ over a threshold $u$ converges to a GPD distribution as $u$ is raised towards the upper endpoint of $F$ (see, e.g., McNeil et al., 2005). The bulk model $h$ extends the density to the whole support and permits the estimation of $u$ from the data. It may be a parametric distribution, a semiparametric density estimator, or a nonparametric model (see, e.g., Scarrott and MacDonald, 2012, and Scarrott, 2016). In this paper, EVMMs with the following three bulk models are considered: a gamma distribution (see Behrens et al., 2004), a lognormal distribution (see, e.g., Scollnik, 2007, and Cabras and Castellanos, 2011), and a weibull distribution (see, e.g., Cabras and Castellanos, 2011, and Scollnik and Sun, 2012). In line with the model names in the R-package evmix (see Scarrott and Hu, 2015), the models are referred to as gammagpd, lognormgpd, and weibullgpd mixtures.

The parameter $\phi_u$ in (1) determines the probability mass in the tail. It is either treated as an extra parameter (parameterised tail fraction approach) or specified by the remaining parameters (see, e.g., Scarrott, 2016). For example, under the bulk model based tail fraction approach, $\phi_u$ is set at $1 - H(u| \theta)$ (see, e.g., Scarrott, 2016).
3 Bayesian Inference for Extreme Value Mixture Models

3.1 Standard Method

The Bayesian approach combines the information from the data set and prior knowledge on the parameters in order to derive a whole distribution (the posterior) for the parameters and quantities of interest (for details, see, e.g., Gelman et al., 2013). In the case of EVMMs, the typical procedure (here called standard method) and relevant elements are as follows:

**Likelihood** Let \( x = (x_1, ..., x_n) \) be an independent sample from \( f_{MM} \). For \( \xi \neq 0 \) (the case \( \xi = 0 \) is analogous), the likelihood has the form (see, e.g., Scarrott, 2016):

\[
L_{MM}(\theta, u, \sigma_u, \xi, \phi_u|x) = \prod_{i=1}^{n} \left( 1 - \phi_u \right) \frac{h(x_i|\theta)}{H(u|\theta)} \prod_{i=1}^{n} \phi_u \cdot g(x_i|u, \sigma_u, \xi)
\]  

(3)

Under the bulk model based tail fraction approach, \( \phi_u \) is given by \( \phi_u = 1 - H(u|\theta) \). Moreover, if the parameterised tail fraction approach is used, \( \phi_u \) is typically replaced by the ML estimate for \( \phi_u \), i.e., the proportion of observations above \( u \) (see, e.g., MacDonald et al., 2011, Oumow et al., 2012). Thus, under both approaches, \( \phi_u \) is determined by the remaining parameters and we write \( L_{MM}(\theta, u, \sigma_u, \xi|x) \).

**Prior Distributions** The priors \( \pi_B(\theta|\eta) \) for the parameter vectors \( \theta \) of the different bulk models are summarized in the third column of Table A1 in Appendix A (\( \eta \) denotes the prior parameters). In order to ensure that the posterior distribution is proper (i.e., the density integrates to one, see, e.g., Gelman et al., 2013), we select proper prior distributions commonly used in the literature.\(^1\)

With regard to the prior \( \pi_u \) for the threshold, many papers choose a truncated normal distribution over the range \((\epsilon_1; \infty)\), with \( \epsilon_1 \) set at some low data value (see, e.g., Behrens et al., 2004, and MacDonald et al., 2011). As recommended by Cabras and Castellanos (2011), to ensure the existence of \( K \) data points below and 2 above \( u \) (here, \( K = 2 \)), we use a truncated normal distribution \( N(m_u, s_u^2) \) with support \([x_{(3)}; x_{(n-2)}]\) (the third and \((n-2)\)th order statistics).

If no prior information is available, typical priors for the GPD parameters are a bivariate normal distribution for \((\log(\sigma_u), \xi)\) (see, e.g., MacDonald et al., 2011) or Jeffrey’s prior (see, e.g., Do Nascimento et al., 2012, and Fúquene Patiño, 2015). The latter was derived by Castellanos and Cabras (2007) and implies a proper posterior density. It is given by (see Castellanos and Cabras, 2007):

\[
\pi_T(\sigma_u, \xi) \propto \sigma_u^{-1} \cdot (1 + \xi)^{-1} \cdot (1 + 2\xi)^{-1/2}, \quad \xi > -0.5; \quad \sigma_u > 0.
\]  

(4)

**Posterior Distribution** Assuming independence between \( \theta, u, \sigma_u, \xi \) (as done by Behrens et al., 2004, and Cabras and Castellanos, 2011, among others) and applying Bayes’ rule, we obtain the following structure of the posterior \( \pi_{MM} \) of the mixture model approach (see, e.g., Cabras and Castellanos, 2011):

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

(5)

\(^1\)As described in Section 4.1, we choose prior parameters \( \eta \) that lead to very flat densities. In this case, the choice of the specific distribution family has no or only marginal influence on the posterior (see Gelman et al., 2013, p. 54).
The posterior cannot be computed analytically and is usually approximated via Markov chain Monte Carlo simulation (MCMC) (for details, see, e.g., Gamerman and Lopes, 2006). For this, many papers use the Metropolis-Hastings (MH) algorithm developed by Metropolis et al. (1953) and Hastings (1970) (see, e.g., Behrens et al., 2004, Do Nascimento et al., 2012, MacDonald et al., 2011). The detailed algorithm is given in Appendix B (see Algorithm 1).

3.2 GPD Reparameterisation

If the conditional excesses \( X_u^+ = X - u | X > u \) over a threshold \( u \) follow a GPD distribution with shape parameter \( \xi \) and scale \( \sigma_u \), then for each higher threshold \( v > u \) the excesses \( X_v^+ = X - v | X > v \) are GPD distributed with the same shape \( \xi \) and scale \( \sigma_v = \sigma_u + \xi(v - u) \) (see, e.g., Coles, 2001). This threshold dependence of the scale parameter is neglected under the standard method and may make the MCMC sampler inefficient (see Scarrott and MacDonald, 2012). Moreover, the assumption of independent priors \( \pi_u \) and \( \pi_T \) is inconsistent with this relationship between \( \sigma_u \) and \( u \) (see Scarrott and MacDonald, 2012). In order to resolve the dependence, MacDonald et al. (2011) and Oumow et al. (2012), among others, pass on to the Poisson point process representation of the GPD (see, e.g., MacDonald et al., 2011). However, according to Scarrott (2016), some analyses indicate that this approach might lead to an overestimation of \( u \). Alternatively, the GPD can be reparameterised and expressed in terms of the modified scale parameter (see, Coles, 2001, and Scarrott and MacDonald, 2012):

\[
\sigma^* = \sigma_u - \xi u. \tag{6}
\]

This modified scale is independent of the threshold, i.e., if \( X_u^+ \) has a GPD with modified scale \( \sigma^* \), than also \( X_v^+ (v \geq u) \). The joint posterior \( \pi_{\sigma M}^* \) for the mixture model with modified scale \( \sigma^* \) is given by:

\[
\pi_{\sigma M}^*(\theta, u, \sigma^*, \xi | m_u, s_u^2, x) \propto L_{\sigma M}(\theta, u, \sigma^* + \xi u, \xi | x) \cdot \pi_B(\theta | \eta) \cdot \pi_u(u | m_u, s_u^2) \cdot \pi_T^*(\sigma^*, \xi). \tag{7}
\]

Due to the invariance of Jeffrey’s prior under monotone transformations (see, e.g., Held and Bové, 2014), Jeffrey’s prior \( \pi_T^*(\sigma^*, \xi) \) can be calculated by deriving the density of \( (\sigma^*, \xi) \) from Jeffrey’s prior for \( (\sigma_u, \xi) \). In doing so, we obtain \( \pi_T^*(\sigma^*, \xi) = \pi_T(\sigma^* + \xi u, \xi) \).

3.3 Tail Fraction Estimation

As indicated above, previous papers that consider the tail fraction approach typically approximate the parameter \( \phi_u \) via the ML estimator, i.e., the proportion of observations above \( u \). Thus, the chain for the tail fraction is derived from the Markov chain \((u^{(t)})_t\) for \( u \) (see, e.g., MacDonald et al., 2011, Oumow et al., 2012). In a full Bayesian framework, a prior \( \pi_\phi \) for \( \phi_u \) has to be specified (e.g., a beta distribution with parameters \( a_{\phi_u} \) and \( b_{\phi_u} \)), and the joint posterior becomes:

\[
\pi_{\phi M}^*(\theta, u, \sigma_u, \xi, \phi_u | \eta, m_u, s_u^2, a_{\phi_u}, b_{\phi_u}, x) \propto L_{\phi M}(\theta, u, \sigma_u, \xi, \phi_u | x) \cdot \pi_B(\theta | \eta) \cdot \pi_u(u | m_u, s_u^2) \cdot \pi_T^*(\sigma^*, \xi) \cdot \pi_\phi(\phi_u | a_{\phi_u}, b_{\phi_u}). \tag{8}
\]

The MH algorithm (see Algorithm 2 in Appendix B) for this full model requires the choice of a proposal distribution for the tail fraction and the sampling of \( \phi_u \) from this proposal. We update \( \phi_u \) together with the threshold \( u \) in block and sample both parameters from truncated normal distributions. The

---

2This reparameterisation is used, e.g., by Fawcett and Walshaw, 2006, and Kennedy et al., 2011, for their hierarchical POT models.
parameters of the proposals are based on the maximum likelihood estimates for $u$ and $\phi_u$ given the remaining parameters. The MH steps for the remaining parameters $\theta$, $\sigma_u$, and $\xi$ are the same as those for the posterior (5).

### 3.4 Parallel Tempering

According to previous applications, the posterior distribution for the threshold is frequently multimodal (see, e.g., Cabras and Castellanos, 2011, MacDonald et al., 2013, Scarrott, 2016). If the modes are well separated, standard MCMC sampler like the MH algorithm may generate slow-mixing chains that are trapped in one mode for a large number of iterations (see, e.g., Gill and Casella, 2004). In consequence, the approximation of the posterior may be unsatisfactory and posterior estimates may be biased (see, e.g., Gill and Casella, 2004).

In order to cope with multimodal posterior distributions, various algorithms have been proposed (e.g., simulated annealing, parallel tempering, and simulated tempering, see, e.g., Gill and Casella, 2004). Here, we consider the method of parallel tempering (PT), as it can be implemented in a straightforward way and there is no need to specify a cooling schedule (as required in the case of simulated annealing, see Gill and Casella, 2004).

Let $\pi$ denote the posterior of interest (i.e., $\pi_{MM}$, $\pi_{MM}^\phi$, or $\pi_{MM}^\phi$) and $\psi$ the parameter vector of the considered model (e.g., $\psi = (\theta, u, \sigma_u, \xi)$ in the case of $\pi_{MM}$; for the following description, refer to Gill and Casella, 2004, and Craiu and Rosenthal, 2014). The principle of the PT algorithm is the simulation of $P$ Markov chains $(\psi_1^{(t)})_{t=1,...,T}, \ldots, (\psi_P^{(t)})_{t=1,...,T}$ with stationary distributions $\pi_1 = (\pi)^{1/\tau_1}, \ldots, \pi_P = (\pi)^{1/\tau_P}$, respectively. The parameters $1 = \tau_1 < \tau_2 < \ldots < \tau_P$ are increasing temperatures that heat up the posterior to densities with less pronounced modes. After every $s$ steps, a switch of the values of two randomly selected chains $p_1$ and $p_2$ is proposed and accepted with the corresponding MH acceptance probability. The chain for $\tau_1 = 1$ then describes the posterior $\pi$ of interest (for details, see Algorithm 3 in Appendix B).

### 4 Simulation Study

#### 4.1 Description of the Analysis

The different simulation approaches are compared in a comprehensive simulation study. We focus on the parameterised tail fraction approach and only consider distributions with $\xi > 0$ and positive support (as typical for distributions of insurance or operational losses, see, e.g., McNeil et al., 2005).

**Data Sets** To examine the ability of the methods to reproduce the original parameters, for each of the three considered EVMMs (the gammagpd, lognormgpd, and weibullgpd mixtures), we first draw samples from the models themselves. The threshold is always set at the 90% quantile of the bulk distribution and $\sigma_u$ is fixed at 5. As previous analyses show that the scale of the GPD does not influence the results (see, e.g., Behrens et al., 2004, and Villa, 2015), we do not vary this parameter. For the shape parameter of the GPD, we consider the two cases $\xi = 0.5$ and $\xi = 1.25$. The tail fraction $\phi_u$ is set at 0.1 (in line with the choice of $u$), and at 0.05. The parameters of the bulk models are chosen in order to imply different

---

3I am very grateful to Carl Scarrott and Nuttanan Wichitaksorn for their suggestions regarding the implementation of the full Bayesian Approach.
shapes of the bulk density \( h(x|\theta) \rightarrow 0 \) or \( h(x|\theta) \rightarrow \infty \) for \( x \rightarrow 0 \) and varying degrees of discontinuity of \( f_{MM} \) at the threshold. For the gammagpd mixture, we use a rate \( \beta = 0.2 \) and consider two shapes \( \alpha \in \{0.5, 2\} \). The parameters \( \mu \) and \( \sigma \) of the lognormal distribution in the lognormgpd model are fixed at 1.5 and 0.75, respectively. The simulations for the weibullgpd mixture are performed for a scale of 5 and shapes of 0.5 and 1.25 of the weibull density.

All EVMMs are further applied to samples from the following parametric distributions: A loglogistic distribution with shape 0.85 and scale 3 (density without mode and \( \xi = 1/0.85 \)), a loglogistic distribution with shape 2 and scale 3 (density with mode and \( \xi = 0.5 \)), and an inverse gamma distribution with shape 2 and scale 4 (density with mode and \( \xi = 0.5 \); see, e.g., Klugman et al., 2008).

We consider two sample sizes, \( n = 1,000 \) and \( n = 10,000 \), and draw \( N = 50 \) samples from each mixture, loglogistic, and inverse gamma distribution.

**Calibration** Assuming that prior information is not available, we choose prior parameters that lead to very flat prior distributions with large variances (as done, e.g., by MacDonald et al., 2011). The parameters of the bulk priors are shown in Table A1 in Appendix A. The parameters \( m_u \) and \( s_u \) are set at the 90% data quantile (see, e.g., Behrens et al., 2004, and MacDonald et al., 2013) and at 10,000, respectively. For the full Bayesian approach, we further use \( a_{\phi_u} = b_{\phi_u} = 1 \).

The variances of the proposal distributions used under the MH algorithm are calibrated in some trial simulations based on the acceptance rates of the proposed values (as done also by MacDonald et al., 2011, and Oumow et al., 2012). Gelman et al. (1996) show that, under certain assumptions, the optimal asymptotic acceptance rate is 0.23. However, Roberts and Rosenthal (2001) note that high sampling efficiency can also be obtained for acceptance probabilities below or above the optimal value. Moreover, 0.23 is not necessarily optimal if the target distribution is low-dimensional or multimodal (see Gelman, 1996, and Roberts et al., 1997). We therefore allow acceptance rates between 0.15 and 0.30. If two parameters are updated at once, the range is extended to \([0.10; 0.30]\).\(^4\)

For the PT method, we have to determine the temperatures \( \tau_1, ..., \tau_P \). The choice may influence the performance of the PT algorithm, and various selection methods have been proposed (see, e.g., Earl and Deem, 2005, and Vousden et al., 2015). In our analysis, we do not apply any of these methods and use \( P = 4 \) temperatures \( \tau_1 = 1, \tau_2 = 1.5, \tau_3 = 2, \) and \( \tau_4 = 2.5 \). Thus, we consider a relatively simple version of the PT method. Swaps between two chains are proposed after each \( s = 10 \) iterations.

In our final calculations, for each sample, two Markov chains are generated both by means of the standard MH algorithm and using PT. After the convergence analysis, the two chains are merged, leading to one long chain with the MH steps and one chain with the PT steps. As some trial runs show substantial autocorrelations of the Markov chains for some samples, the chain lengths are set at relatively high values.\(^5\)

\(^4\)For the full Bayesian tail fraction approach, the proposal variances calibrated for the simplified method are used. This sometimes leads to different acceptance rates for \( u \) and \( \phi_u \).

\(^5\)Depending on the algorithm and sample size, different chain lengths are chosen. For example, for the samples of size 1,000 from the loglogistic and inverse gamma distributions and the simplified tail fraction approach, each of the two MH chains consists of 75,000 iterations (after burn-in). The two PT chains have lengths of 50,000 steps.
Convergence Assessment  The convergence of the Markov chains to the stationary distribution is examined by means of a visual comparison of various chains with different starting vectors (as proposed by Gelman, 1996) as well as the Gelman-Rubin statistic (see Gelman and Rubin, 1992, and Brooks and Gelman, 1998). Furthermore, as the PT algorithm seems to have generated converging chains in nearly all cases, we contrast the estimated posterior densities from PT with those from standard MH in order to assess the convergence of the standard MH chains. We then remove all data sets for which the chains have not converged under at least one of the considered approaches. In addition, we discard some samples for which some chains have converged but mix extremely badly and therefore lead to a biased estimation of the posterior. Eventually, we also exclude a few data sets where some of the chains did not reach the range of the posterior within the burn-in-period (but eventually converged later).

Evaluation Criteria  The number of data sets that have to be discarded is a first indicator of the convergence and mixing behaviour of the Markov chains under the different simulation approaches. In order to investigate the mixing behaviour for the \( N' \) not removed samples, we calculate the lag-50 autocorrelations of the Markov chains and take the averages across all \( N' \) samples. As more autocorrelated chains need more time to move along the parameter space, the autocorrelation function is a commonly applied measure to evaluate the mixing behaviour (see, e.g., Jackman, 2009, and Greenberg, 2013). It is also the basis for several measures of sampling inefficiency (see, e.g., Jackman, 2009, and Greenberg, 2013). Following MacDonald et al. (2011) and MacDonald et al. (2013), the accuracy of the estimates is assessed by means of the coverage rates of the 95\% posterior intervals, the mean posterior interval lengths, as well as the averages of the integrated squared errors (ISE) of the posterior predictive densities \( \hat{f}_i \). The 95\% posterior interval corresponds to the range that comprises 95\% of the posterior probability (see, e.g., Gelman et al., 2013), and the coverage rate is the proportion of the \( N' \) posterior intervals which cover the true parameter (see, e.g., MacDonald et al., 2011). For sample \( x^{(i)} \), the posterior predictive density and ISE are given by (see, e.g., MacDonald et al., 2013):\(^7\)

\[
\hat{f}_i(y) = \int_{\Psi} f_{MM}(y|\psi)\pi(\psi|x^{(i)})d\psi
\]

\[
ISE(\hat{f}_i) = ||f - \hat{f}_i||_2^2 = \int_0^\infty (f(y) - \hat{f}_i(y))^2dy.
\]

Here, \( \pi \) denotes the posterior under the considered approach and \( \psi \) the corresponding parameter vector. Moreover, \( f \) is the true underlying density.

4.2 MH and PT Results for Different Model Parameterisations

4.2.1 Large Samples from Extreme Value Mixture Distributions

In the first part of the analysis, we restrict on the simplified tail fraction approach and compare the results for the different model parameterisations and MCMC algorithms. Tables 1 and 2 (see pp. 12/13) summarize the results for the samples of size \( n = 10,000 \) that are drawn from an EVMM density. As the results are similar for \( \xi = 0.5 \) and \( \xi = 1.25 \), only the output for \( \xi = 0.5 \) is given.

The first data column shows the number \( N \) of data sets that have to be removed due to non-convergence or extremely bad mixing. While the PT algorithm generates converging and adequately mixing Markov

\(^6\)Please note that in various cases it is very difficult to decide whether the mixing is tolerable. Thus, the decision whether to exclude a sample is rather subjective. We tried to take consistent decisions for the two model parameterisations for a given data set and not to loose too many data sets.

\(^7\)MacDonald et al. (2013) use the median of the integrated squared errors. In this paper, we always calculate the mean, as the median only accounts for the performance of the algorithms in the 50\% best cases. If one approach is good for all data sets and another for only half of the samples, the median leads to wrong conclusions.
chains for almost all data sets and both model parameterisations, the performance of the standard MH algorithm varies substantially. If the MH sampler is applied to the mixture with GPD scale $\sigma^*$, nearly all chains converge and mix sufficiently, even if the posterior is multimodal. In contrast, if the original model parameterisation (i.e., with $\sigma_u$) is considered and the samples are drawn from an almost continuous density (i.e., a density with only a small “jump” at the threshold), the standard algorithm cannot cope with the multimodality of the posterior for various samples (see, e.g., D1, D3, D5).

In line with these results, the PT algorithm leads to equally well or substantially better mixing chains than the standard MH also for the remaining (i.e., not removed) data sets (see data columns 2 to 7). If the autocorrelations under the standard MH algorithm are 0.01 or lower, the PT method achieves the same high mixing performance. In the remaining cases, the autocorrelations of the parameter chains can always be reduced by 40% or more if PT is used instead of the standard MH algorithm. With regard to the GPD parameterisation, sampling of $\sigma^*$ helps to reduce the high autocorrelations of the chains for $u$ and $\phi_u$ if the posterior for $u$ is not highly concentrated around one point, i.e., if the discontinuity of the original density is relatively small (see, e.g., D1, D3, D5, D10). The autocorrelations of the chains for $\theta_1$ and $\theta_2$ are also somewhat reduced, but they are rather low under both model parameterisations. In the case of a strongly discontinuous original density, the autocorrelations of the chains for $u$ are higher under the reparameterised model, but remain at an acceptable level (see, e.g., D6, D7, D9). The mixing of the Markov chains for $\xi$ is always substantially worse under the mixture with $\sigma^*$.

Figure 1 illustrates the results for one sample from (D1). Under the standard MH algorithm, the chain for $u$ in the original model only rarely reaches the second mode of the posterior and then stays there for a rather long period (see Subfigure a). For the reparameterised model, in contrast, the chain mixes relatively well. In the case of PT, sampling of $\sigma^*$ also enhances the mixing behaviour of the Markov chains for the threshold. However, the PT chains generated under the original model also move relatively well around the parameter space. If the PT algorithm is used, the reparameterisation is therefore not indispensable (especially as it leads to higher autocorrelated chains for $\xi$).

The posterior interval coverage rates, interval lengths, and the mean integrated squared errors are very similar for the two model parameterisations and MCMC methods (see the remaining columns 8 to 20). This may also be expected, as all chains have finally converged to the stationary distribution. The mean integrated squared errors are of order $10^{-5}$ or lower. Thus, the original densities are well reproduced. Similarly, the posterior interval coverage rates are mostly very high.\(^8\) Only in a few cases, low coverage rates can be observed, but this does not imply a bad fit. For example, in the case of distribution (D5), a good fit can also be achieved with higher thresholds, as the lognormal bulk distribution is able to model quite high losses. Moreover, the very low coverage rates for the parameter $\phi_u$ for data sets from highly discontinuous densities are the result of extremely narrow posterior intervals and not of a poor estimation of this parameter. Indeed, in the respective cases, the mean deviations of the posterior means of $\phi_u$ from the true value of 0.1 or 0.05 are only of order $10^{-3}$ or lower.

\(^8\)This corresponds to the results of Behrens et al. (2004), and Do Nascimento et al. (2012), for the bulk model based tail fraction approach (we use the parameterised tail fraction approach) and their EVMMs with a single or several mixed gamma distributions for the bulk.
Figure 1: Chains for $u$ under Different Sampling Approaches and Model Parameterisations

This figure shows the traceplots for the threshold $u$ generated for one sample from (D1) under the different approaches. The chains for Subfigures (a) and (b) are generated by means of the standard MH algorithm for the original gammagpd mixture and the reparameterised model with scale $\sigma^*$, respectively. Subfigures (c) and (d) provide the Markov chains generated via PT. In order to improve the visual comparability, only the first 50,000 iterations of the MH chains are shown.
This table summarizes the results for the first part of data sets of size $n = 10,000$ that are drawn from EVMM densities. The first three columns indicate the density the samples are simulated from, the estimated EVMM and GPD parameterisation (* refers to the parameterisation with $\alpha$), as well as the used sampling algorithm (MH or PT). The first data column (i.e., column 4) shows the number $N$ of data sets that have to be removed due to non-convergence or substantial mixing problems of the Markov chains. The remaining results are only based on those samples that are not discarded under none of the four approaches. Data columns 2 to 7 (entitled “ACR”) contain the average lag-50 autocorrelations of the parameter chains across the not removed data sets. The mean integrated squared error (ISE) of the posterior predictive densities is given in data column 8. The block of columns denoted “Posterior Interval Coverage” shows the coverage rates of the 95% posterior intervals of the EVMM parameters. The average lengths of the 95% posterior intervals are provided in the last six columns.
<table>
<thead>
<tr>
<th>Data</th>
<th>EVMM</th>
<th>N</th>
<th>ACR</th>
<th>Mean</th>
<th>Posterior Interval Coverage</th>
<th>Posterior Interval Length</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>θ₁</td>
<td>θ₂</td>
</tr>
<tr>
<td>(D7)</td>
<td>Weibullgpd</td>
<td>2</td>
<td></td>
<td></td>
<td>5.58×10⁻⁶</td>
<td>0.96</td>
</tr>
<tr>
<td>τ = 0.5, φ = 5,</td>
<td>Weibullgpd</td>
<td>0</td>
<td></td>
<td></td>
<td>5.52×10⁻⁶</td>
<td>0.96</td>
</tr>
<tr>
<td>u = 6.77, φ = 0.1</td>
<td>Weibullgpd*</td>
<td>1</td>
<td></td>
<td></td>
<td>5.51×10⁻⁶</td>
<td>0.96</td>
</tr>
<tr>
<td>σₓ = 5, ξ = 0.5</td>
<td>Weibullgpd*</td>
<td>0</td>
<td></td>
<td></td>
<td>5.63×10⁻⁶</td>
<td>0.96</td>
</tr>
<tr>
<td>(D8)</td>
<td>Weibullgpd</td>
<td>1</td>
<td></td>
<td></td>
<td>5.40×10⁻⁶</td>
<td>0.90</td>
</tr>
<tr>
<td>τ = 0.5, φ = 5,</td>
<td>Weibullgpd</td>
<td>0</td>
<td></td>
<td></td>
<td>5.38×10⁻⁶</td>
<td>0.90</td>
</tr>
<tr>
<td>u = 6.77, φ = 0.05</td>
<td>Weibullgpd*</td>
<td>0</td>
<td></td>
<td></td>
<td>5.40×10⁻⁶</td>
<td>0.90</td>
</tr>
<tr>
<td>σₓ = 5, ξ = 0.5</td>
<td>Weibullgpd*</td>
<td>0</td>
<td></td>
<td></td>
<td>5.37×10⁻⁶</td>
<td>0.90</td>
</tr>
<tr>
<td>(D9)</td>
<td>Weibullgpd</td>
<td>2</td>
<td></td>
<td></td>
<td>1.93×10⁻⁵</td>
<td>0.96</td>
</tr>
<tr>
<td>τ = 1.25, φ = 5,</td>
<td>Weibullgpd</td>
<td>0</td>
<td></td>
<td></td>
<td>1.93×10⁻⁵</td>
<td>0.93</td>
</tr>
<tr>
<td>u = 13.62, φ = 0.1</td>
<td>Weibullgpd*</td>
<td>3</td>
<td></td>
<td></td>
<td>1.94×10⁻⁵</td>
<td>0.96</td>
</tr>
<tr>
<td>σₓ = 5, ξ = 0.5</td>
<td>Weibullgpd*</td>
<td>0</td>
<td></td>
<td></td>
<td>1.93×10⁻⁵</td>
<td>0.93</td>
</tr>
<tr>
<td>(D10)</td>
<td>Weibullgpd</td>
<td>0</td>
<td></td>
<td></td>
<td>1.74×10⁻⁵</td>
<td>0.92</td>
</tr>
<tr>
<td>τ = 1.25, φ = 5,</td>
<td>Weibullgpd</td>
<td>0</td>
<td></td>
<td></td>
<td>1.74×10⁻⁵</td>
<td>0.94</td>
</tr>
<tr>
<td>u = 13.62, φ = 0.05</td>
<td>Weibullgpd*</td>
<td>0</td>
<td></td>
<td></td>
<td>1.74×10⁻⁵</td>
<td>0.94</td>
</tr>
<tr>
<td>σₓ = 5, ξ = 0.5</td>
<td>Weibullgpd*</td>
<td>0</td>
<td></td>
<td></td>
<td>1.72×10⁻⁵</td>
<td>0.92</td>
</tr>
</tbody>
</table>

Table 2: Samples from EVMMs, n = 10,000, Part II

This table summarizes the output for the second part of data sets of size n = 10,000 that are drawn from EVMM densities. The first three columns indicate the density the samples are simulated from, the estimated EVMM and GPD parameterisation (* refers to the parameterisation with σ*), as well as the used sampling algorithm (MH or PT). The first data column (i.e., column 4) shows the number N of data sets that have to be removed due to non-convergence or substantial mixing problems of the Markov chains. The remaining results are only based on those samples that are not discarded under none of the four approaches. Data columns 2 to 7 (entitled “ACR”) contain the average lag-50 autocorrelations of the parameter chains across the not removed data sets. The mean integrated squared error (ISE) of the posterior predictive densities is given in data column 8. The block of columns denoted “Posterior Interval Coverage” shows the coverage rates of the 95% posterior intervals of the EVMM parameters. The average lengths of the 95% posterior intervals are provided in the last six columns.
This table summarizes the results for the data sets of size $n = 1,000$ that are drawn from EVMM densities. The first three columns indicate the density the samples are simulated from, the estimated EVMM and GPD parameterisation (* refers to the parameterisation with $\sigma^*$), as well as the used sampling algorithm (MH or PT). The first data column (i.e., column 4) shows the number $N$ of data sets that have to be removed due to non-convergence or substantial mixing problems of the Markov chains. The remaining results are only based on those samples that are not discarded under none of the four approaches. Data columns 2 to 7 (entitled “ACR”) contain the average lag-50 autocorrelations of the parameter chains across the not removed data sets. The mean integrated squared error (ISE) of the posterior predictive densities is given in data column 8. The block of columns denoted “Posterior Interval Coverage” shows the coverage rates of the 95% posterior intervals of the EVMM parameters. The average lengths of the 95% posterior intervals are provided in the last six columns.
<table>
<thead>
<tr>
<th>Data</th>
<th>EVMM</th>
<th>EVMM</th>
<th>ACR</th>
<th>Mean</th>
<th>ISE</th>
<th>EVMM</th>
<th>ACR</th>
<th>Mean</th>
<th>ISE</th>
</tr>
</thead>
<tbody>
<tr>
<td>(D11) Loglogistic</td>
<td>Gammagpd*</td>
<td>9</td>
<td>0.19</td>
<td>0.39</td>
<td>0.51</td>
<td>0.82</td>
<td>0.12</td>
<td>0.84</td>
<td>1.06</td>
</tr>
<tr>
<td></td>
<td>Gammagpd</td>
<td>9</td>
<td>0.13</td>
<td>0.30</td>
<td>0.16</td>
<td>0.18</td>
<td>0.08</td>
<td>0.18</td>
<td>1.06</td>
</tr>
<tr>
<td></td>
<td>PT</td>
<td>0</td>
<td>0.05</td>
<td>0.09</td>
<td>0.05</td>
<td>0.02</td>
<td>0.05</td>
<td>1.07</td>
<td>10^{-4}</td>
</tr>
<tr>
<td>(c) Weibullgpd</td>
<td>MH</td>
<td>11</td>
<td>0.22</td>
<td>0.43</td>
<td>0.83</td>
<td>0.84</td>
<td>0.11</td>
<td>0.85</td>
<td>8.56</td>
</tr>
<tr>
<td></td>
<td>PT</td>
<td>3</td>
<td>0.07</td>
<td>0.08</td>
<td>0.27</td>
<td>0.27</td>
<td>0.04</td>
<td>0.28</td>
<td>8.43</td>
</tr>
<tr>
<td></td>
<td>PT</td>
<td>0</td>
<td>0.05</td>
<td>0.09</td>
<td>0.05</td>
<td>0.02</td>
<td>0.05</td>
<td>1.07</td>
<td>10^{-4}</td>
</tr>
<tr>
<td>(D12) Loglogistic</td>
<td>Gammagpd</td>
<td>0</td>
<td>0.12</td>
<td>0.19</td>
<td>0.60</td>
<td>0.54</td>
<td>0.09</td>
<td>0.61</td>
<td>6.66</td>
</tr>
<tr>
<td></td>
<td>Gammagpd*</td>
<td>0</td>
<td>0.07</td>
<td>0.10</td>
<td>0.21</td>
<td>0.10</td>
<td>0.44</td>
<td>0.21</td>
<td>6.60</td>
</tr>
<tr>
<td></td>
<td>PT</td>
<td>0</td>
<td>0.02</td>
<td>0.03</td>
<td>0.07</td>
<td>0.04</td>
<td>0.14</td>
<td>0.07</td>
<td>6.64</td>
</tr>
<tr>
<td>(b) Lognormgpd</td>
<td>MH</td>
<td>0</td>
<td>0.62</td>
<td>0.60</td>
<td>0.29</td>
<td>0.05</td>
<td>0.09</td>
<td>0.29</td>
<td>4.44</td>
</tr>
<tr>
<td></td>
<td>Lognormgpd*</td>
<td>0</td>
<td>0.69</td>
<td>0.67</td>
<td>0.34</td>
<td>0.07</td>
<td>0.32</td>
<td>0.33</td>
<td>4.44</td>
</tr>
<tr>
<td></td>
<td>PT</td>
<td>0</td>
<td>0.20</td>
<td>0.21</td>
<td>0.11</td>
<td>0.02</td>
<td>0.11</td>
<td>0.11</td>
<td>4.44</td>
</tr>
<tr>
<td>(c) Weibullgpd</td>
<td>MH</td>
<td>0</td>
<td>0.18</td>
<td>0.29</td>
<td>0.42</td>
<td>0.23</td>
<td>0.10</td>
<td>0.41</td>
<td>1.11</td>
</tr>
<tr>
<td></td>
<td>PT</td>
<td>0</td>
<td>0.06</td>
<td>0.10</td>
<td>0.14</td>
<td>0.08</td>
<td>0.04</td>
<td>0.14</td>
<td>1.11</td>
</tr>
<tr>
<td></td>
<td>PT</td>
<td>0</td>
<td>0.06</td>
<td>0.09</td>
<td>0.13</td>
<td>0.06</td>
<td>0.13</td>
<td>0.13</td>
<td>1.11</td>
</tr>
<tr>
<td>(D13) Inverse Gamma</td>
<td>Gammagpd*</td>
<td>3</td>
<td>0.24</td>
<td>0.27</td>
<td>0.31</td>
<td>0.04</td>
<td>0.06</td>
<td>0.31</td>
<td>4.99</td>
</tr>
<tr>
<td></td>
<td>Gammagpd</td>
<td>3</td>
<td>0.27</td>
<td>0.29</td>
<td>0.36</td>
<td>0.12</td>
<td>0.23</td>
<td>0.36</td>
<td>4.96</td>
</tr>
<tr>
<td></td>
<td>PT</td>
<td>0</td>
<td>0.09</td>
<td>0.10</td>
<td>0.11</td>
<td>0.02</td>
<td>0.02</td>
<td>0.11</td>
<td>4.95</td>
</tr>
<tr>
<td>(b) Lognormgpd</td>
<td>MH</td>
<td>1</td>
<td>0.39</td>
<td>0.33</td>
<td>0.51</td>
<td>0.37</td>
<td>0.09</td>
<td>0.50</td>
<td>1.99</td>
</tr>
<tr>
<td></td>
<td>Lognormgpd*</td>
<td>1</td>
<td>0.34</td>
<td>0.29</td>
<td>0.36</td>
<td>0.21</td>
<td>0.30</td>
<td>0.36</td>
<td>1.99</td>
</tr>
<tr>
<td></td>
<td>PT</td>
<td>0</td>
<td>0.12</td>
<td>0.10</td>
<td>0.12</td>
<td>0.08</td>
<td>0.10</td>
<td>0.12</td>
<td>1.98</td>
</tr>
<tr>
<td>(c) Weibullgpd</td>
<td>MH</td>
<td>11</td>
<td>0.22</td>
<td>0.30</td>
<td>0.42</td>
<td>0.01</td>
<td>0.05</td>
<td>0.43</td>
<td>1.06</td>
</tr>
<tr>
<td></td>
<td>PT</td>
<td>3</td>
<td>0.07</td>
<td>0.09</td>
<td>0.12</td>
<td>0.00</td>
<td>0.01</td>
<td>0.12</td>
<td>1.06</td>
</tr>
<tr>
<td></td>
<td>PT</td>
<td>0</td>
<td>0.08</td>
<td>0.10</td>
<td>0.13</td>
<td>0.02</td>
<td>0.05</td>
<td>0.13</td>
<td>1.06</td>
</tr>
</tbody>
</table>

**Table 4: Samples from Parametric Distributions**

This table gives an overview of the results for the samples from various loglogistic and inverse gamma distributions. The first three columns indicate the original density ($\alpha$ and $\theta$ are the shape and scale parameters, respectively), the estimated EVM and GPD parameterisation (* refers to the parameterisation with $\sigma^*$), as well as the used sampling method (MH or PT). The block of data columns in dark gray summarizes the output for the samples of size $n = 10,000$. The results for $n = 1,000$ are shown in the block in light gray. In each block, the first column indicates the number of samples with convergence or substantial mixing problems of the Markov chains. Moreover, columns 2 to 7 show the mean lag-50 autocorrelations of the parameter chains across all samples. The last column gives the mean integrated squared error (ISE) of the posterior predictive density.
4.2.2 Small Samples from EVMMs and Samples from Other Parametric Distributions

The simulation output for the data sets of size \( n = 1,000 \) from the EVMM densities and for the samples from the loglogistic and inverse gamma distributions is summarized in Tables 3 and 4, respectively (see pp. 14/15).\(^9\) Due to the incompatibility of the lognormal bulk model with a density without a mode, we have not fitted the lognormgpd mixture to the samples from the loglogistic distribution with shape 0.85.

The results confirm the findings in the previous section that the PT algorithm substantially enhances the MCMC simulation in terms of convergence and mixing. Except for (D5), the number \( N \) of data sets that has to be removed is similar or considerably lower under PT than the MH algorithm. In addition, the autocorrelations decrease significantly through parallel tempering. With a few exceptions, the reduction is 40% or more. For the samples of size \( n = 10,000 \) from the loglogistic and inverse gamma distributions (D11) to (D13), the mean lag-50 autocorrelations decrease by even 60% or more. There are also a few examples where convergence problems arise under PT, but in the respective cases, the MH algorithm fails as well (see in particular the output for D5 in Table 3 and for D12b and \( n = 1,000 \) in Table 4). With regard to the model parameterisation, sampling of \( \sigma^* \) reduces the number \( N \) of unacceptable sampling results for some distributions (see, e.g., D11c in Table 4), but leads to increased problems for others (see, e.g., D1, D3, D5 in Table 3 and D11a and D13c in Table 4). Similarly, the autocorrelations of the Markov chains can only sometimes be clearly reduced by means of the model reparameterisation (see, e.g., D11a and D11c). For many samples, the transition to the model with \( \sigma^* \) does not only induce an enormous increase in the autocorrelations of the chains for \( \xi \), but also no significant improvement or even a slight deterioration of the mixing of the chains for the bulk model parameters and \( u \) (see, e.g., D3 and D5 in Table 3, and D12b, D13a, D13c in Table 4). Thus, the dominance of the reparameterised model seems to hold no longer if the sample size is rather small or the data originates from a distribution that does not have the form of an EVMM and for which the tail approximation via the GPD only holds asymptotically.

Due to the exclusion of the samples with non-convergent or insufficiently mixing Markov chains, the mean ISEs and posterior interval lengths are similar under the different approaches.\(^10\) Analogously, the posterior interval coverage rates for the parameters of the mixture densities (D1) to (D9) differ only marginally. As the sample characteristics tend to deviate more from the characteristics of the true underlying distribution for small data sets, the mean integrated squared errors are higher for \( n = 1,000 \) than \( n = 10,000 \). For example, for the samples from the mixture densities, the mean ISEs are increased by a factor of approximately 10. In addition, a higher degree of uncertainty leads to longer posterior intervals for \( n = 1,000 \) (see also the results of Behrens et al., 2004). The results in Table 3 further show relatively low coverage rates for the threshold for (D1), (D3), and (D5). This is in line with the findings of MacDonald et al. (2011) and Do Nascimento et al. (2012) and shows that the replication of the true threshold may be difficult for small sample sizes and high prior variances (see Do Nascimento et al., 2012).

4.3 Results for the Sampling Method for the Tail Fraction

Our results for the samples of size \( n = 10,000 \) generated from the mixture distributions indicate that the simplified sampling algorithm with ML approximation of the tail fraction \( \phi_u \) is able to reproduce the original parameters. The posterior interval coverage rates for \( \theta_1, \theta_2, u, \sigma_u, \) and \( \xi \) are high and the

\(^9\)For small sample sizes, the location of the threshold is relatively uncertain independent on the degree of discontinuity of the underlying density. We therefore restrict on the distributions with \( \phi_u = 0.1 \).

\(^10\)For the samples from the loglogistic and inverse gamma distributions, the posterior interval lengths are not shown. They are available upon request from the author.
posterior mean estimates for $\phi_u$ only deviate marginally from the true values (see Section 4.2.1). In order to further check the adequacy of the algorithm, we compare the estimates obtained under the simplified algorithm with the estimates under the full Bayesian approach presented in Section 3.3. We consider both an idealized situation with large samples ($n = 10,000$) from six extreme value mixture distributions (with different values of $\phi_u$ and therefore varying degree of discontinuity) and the more realistic setting of relatively small data sets from non-EVMM densities. In view of the results in the previous section, the Markov chains are generated via parallel tempering and for the original model parameterisation.

The full Bayesian approach requires substantially higher computational efforts for the MCMC simulation. In addition, we observe some tendency towards slightly higher convergence and mixing problems.

Table 5 contrasts the estimation results under the two approaches. For each mixture model and data set, the parameters and quantiles are estimated via the posterior medians (as done, e.g., by Cabras and Castellanos, 2011) and the table shows the mean values above all samples. The application of the full Bayesian approach is indicated by the Greek letter $\phi$ at the end of the model name.

The parameter estimates (see data columns one to six) for large samples from the six extreme value mixture distributions are very similar under the two approaches. Only in the case of a mixture density with a rather small jump at the threshold (and therefore some uncertainty in the estimation of $u$), some small deviations are possible for the threshold estimates (see D5 and D10). For small samples from the parametric distributions (D11) to (D13), the estimates of $\xi$ are almost identical, as well. The (average) posterior medians for the remaining parameters differ, but for the most part, the deviations are relatively small. In most cases, the relative differences are lower than 5%. In the remaining cases, the estimates mostly differ by around 10%. Substantial discrepancies can only be observed for the application of the weibullgpd mixture to the samples from (D11). For (D11), especially the posterior medians for $u$ differ enormously under the simplified and full Bayesian approach. The consistency of the estimates of $\xi$ in this example can be explained by the threshold stability of the GPD, i.e., if the excesses above a certain threshold $u_0$ are GPD distributed with shape $\xi_0$, than the excesses above each higher threshold $v > u_0$ are also GPD distributed, with the same shape (see, e.g., McNeil et al., 2005).

A comparison of the traceplots of the parameter chains shows a high conformity of the Markov chains under the two approaches for the samples from the mixture distributions (D1), (D2), and (D6). Thus, the whole posterior distributions are very similar, not only the medians. The same applies to the chains for the gammagpd and lognormgpd parameters and the data sets from (D11) to (D13). For (D5), (D7), and (D10) and the weibullgpd approximation to (D11) to (D13), the posterior ranges differ for several samples. This discrepancy may be either the result of indeed unequal posteriors or indicate the existence of mixing problems. We varied the parameters of the MH and PT algorithms (e.g., the proposal variances) and obtained the same results as before, but some other sampling method might be necessary.

The model parameters are typically not the main quantities of interest. More important is usually the approximation of the value at risk or return levels and thus the quantiles in the tail. As shown by the last five columns in Table 5, the quantile estimates are almost identical under the simplified and full Bayesian approaches, even if the parameter posteriors do not match perfectly. For the samples from extreme

11For this reason, a lower number $N$ of samples is used for a sample size of $n = 10,000$. We choose $N$ such that we retain 25 samples after the omission of the data sets with convergence or substantial mixing problems.
<table>
<thead>
<tr>
<th>Data</th>
<th>EVMM</th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
<th>$u$</th>
<th>$\sigma_u$</th>
<th>$\xi$</th>
<th>$\phi_u$</th>
<th>$Q_{0.90}$</th>
<th>$Q_{0.95}$</th>
<th>$Q_{0.975}$</th>
<th>$Q_{0.99}$</th>
<th>$Q_{0.999}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(D1) Gammagpd</td>
<td>Gammagpd</td>
<td>0.50</td>
<td>0.20</td>
<td>6.09</td>
<td>4.99</td>
<td>0.50</td>
<td>0.11</td>
<td>6.83</td>
<td>10.94</td>
<td>16.85</td>
<td>28.55</td>
<td>98.26</td>
</tr>
<tr>
<td>$\alpha = 0.5$, $\beta = 0.2$, $u = 6.76$, $\phi_u = 0.1$, $n = 10,000$</td>
<td>Gammagpd$^a$</td>
<td>0.50</td>
<td>0.20</td>
<td>6.09</td>
<td>4.93</td>
<td>0.50</td>
<td>0.11</td>
<td>6.84</td>
<td>10.94</td>
<td>16.84</td>
<td>28.55</td>
<td>98.56</td>
</tr>
<tr>
<td>(D2) Gammagpd</td>
<td>Gammagpd</td>
<td>0.50</td>
<td>0.20</td>
<td>6.76</td>
<td>5.07</td>
<td>0.50</td>
<td>0.05</td>
<td>5.26</td>
<td>6.84</td>
<td>10.97</td>
<td>19.29</td>
<td>68.41</td>
</tr>
<tr>
<td>$\alpha = 0.5$, $\beta = 0.2$, $u = 6.76$, $\phi_u = 0.05$, $n = 10,000$</td>
<td>Gammagpd$^a$</td>
<td>0.50</td>
<td>0.20</td>
<td>6.76</td>
<td>5.08</td>
<td>0.50</td>
<td>0.05</td>
<td>5.26</td>
<td>6.84</td>
<td>10.97</td>
<td>19.29</td>
<td>68.36</td>
</tr>
<tr>
<td>(D5) Lognormgpd</td>
<td>Lognormgpd</td>
<td>1.50</td>
<td>0.75</td>
<td>18.14</td>
<td>8.02</td>
<td>0.50</td>
<td>0.06</td>
<td>11.77</td>
<td>15.92</td>
<td>21.75</td>
<td>33.75</td>
<td>103.43</td>
</tr>
<tr>
<td>$\mu = 1.5$, $\sigma = 0.75$, $u = 11.72$, $\phi_u = 0.1$, $n = 10,000$</td>
<td>Lognormgpd$^a$</td>
<td>1.50</td>
<td>0.75</td>
<td>16.50</td>
<td>7.46</td>
<td>0.50</td>
<td>0.07</td>
<td>11.78</td>
<td>15.93</td>
<td>21.78</td>
<td>33.81</td>
<td>103.62</td>
</tr>
<tr>
<td>(D6) Lognormgpd</td>
<td>Lognormgpd</td>
<td>1.50</td>
<td>0.75</td>
<td>11.72</td>
<td>5.08</td>
<td>0.50</td>
<td>0.05</td>
<td>9.85</td>
<td>11.79</td>
<td>15.93</td>
<td>24.25</td>
<td>73.34</td>
</tr>
<tr>
<td>$\mu = 1.5$, $\sigma = 0.75$, $u = 11.72$, $\phi_u = 0.05$, $n = 10,000$</td>
<td>Lognormgpd$^a$</td>
<td>1.50</td>
<td>0.75</td>
<td>11.72</td>
<td>5.08</td>
<td>0.49</td>
<td>0.05</td>
<td>9.86</td>
<td>11.79</td>
<td>15.93</td>
<td>24.25</td>
<td>73.30</td>
</tr>
<tr>
<td>(D7) Weibullgpd</td>
<td>Weibullgpd</td>
<td>0.50</td>
<td>5.20</td>
<td>6.75</td>
<td>4.99</td>
<td>0.50</td>
<td>0.10</td>
<td>6.80</td>
<td>10.92</td>
<td>16.78</td>
<td>28.41</td>
<td>97.27</td>
</tr>
<tr>
<td>$\tau = 0.5$, $\phi = 5$, $u = 6.77$, $\phi_u = 0.1$, $n = 10,000$</td>
<td>Weibullgpd$^a$</td>
<td>0.50</td>
<td>5.13</td>
<td>6.74</td>
<td>4.93</td>
<td>0.51</td>
<td>0.10</td>
<td>6.78</td>
<td>10.80</td>
<td>16.59</td>
<td>28.16</td>
<td>97.40</td>
</tr>
<tr>
<td>(D10) Weibullgpd</td>
<td>Weibullgpd</td>
<td>1.25</td>
<td>5.03</td>
<td>14.76</td>
<td>5.61</td>
<td>0.48</td>
<td>0.05</td>
<td>10.46</td>
<td>13.60</td>
<td>17.84</td>
<td>26.16</td>
<td>73.85</td>
</tr>
<tr>
<td>$\tau = 1.25$, $\phi = 5$, $u = 13.62$, $\phi_u = 0.05$, $n = 10,000$</td>
<td>Weibullgpd$^a$</td>
<td>1.24</td>
<td>5.04</td>
<td>15.19</td>
<td>5.87</td>
<td>0.48</td>
<td>0.04</td>
<td>10.48</td>
<td>13.62</td>
<td>17.89</td>
<td>26.33</td>
<td>74.26</td>
</tr>
<tr>
<td>(D11) Loglogistic</td>
<td>Gammagpd</td>
<td>0.77</td>
<td>0.30</td>
<td>2.66</td>
<td>7.31</td>
<td>1.22</td>
<td>0.57</td>
<td>41.25</td>
<td>100.81</td>
<td>241.07</td>
<td>756.88</td>
<td>13.580</td>
</tr>
<tr>
<td>$\alpha = 0.85$, $\beta = 0$, $\theta = 2$, $n = 1,000$</td>
<td>Gammagpd$^a$</td>
<td>0.77</td>
<td>0.30</td>
<td>2.95</td>
<td>7.67</td>
<td>1.22</td>
<td>0.54</td>
<td>41.23</td>
<td>100.58</td>
<td>239.99</td>
<td>751.20</td>
<td>13.370</td>
</tr>
<tr>
<td>(D12) Loglogistic</td>
<td>Gammagpd</td>
<td>2.18</td>
<td>0.68</td>
<td>4.96</td>
<td>3.28</td>
<td>0.44</td>
<td>0.31</td>
<td>9.26</td>
<td>13.54</td>
<td>19.35</td>
<td>30.48</td>
<td>92.75</td>
</tr>
<tr>
<td>$\alpha = 2$, $\beta = 0$, $\theta = 3$, $n = 1,000$</td>
<td>Gammagpd$^a$</td>
<td>2.19</td>
<td>0.68</td>
<td>5.21</td>
<td>3.33</td>
<td>0.45</td>
<td>0.28</td>
<td>9.26</td>
<td>13.53</td>
<td>19.34</td>
<td>30.56</td>
<td>94.57</td>
</tr>
<tr>
<td>(D13) Inverse Gamma</td>
<td>Gammagpd</td>
<td>7.72</td>
<td>5.35</td>
<td>1.67</td>
<td>1.94</td>
<td>0.45</td>
<td>0.70</td>
<td>7.63</td>
<td>11.40</td>
<td>16.58</td>
<td>26.53</td>
<td>81.94</td>
</tr>
<tr>
<td>$\alpha = 2$, $\beta = 4$, $\theta = 1$, $n = 1,000$</td>
<td>Gammagpd$^a$</td>
<td>7.60</td>
<td>5.21</td>
<td>1.73</td>
<td>1.94</td>
<td>0.45</td>
<td>0.68</td>
<td>7.62</td>
<td>11.39</td>
<td>16.58</td>
<td>26.61</td>
<td>82.74</td>
</tr>
</tbody>
</table>

Table 5: Comparison of the Simplified and Full Bayesian Tail Fraction Estimation

This table contrasts the results of the parameter and quantile estimation under the simplified sampling algorithm (see Algorithm 1 in Appendix B) and the full Bayesian approach (see Algorithm 2 in Appendix B). The use of the latter is indicated by a superscript of the Greek letter $\phi$ at the end of the model name. Under both approaches, the Markov chains are generated via parallel tempering. For each EVMM and sample, the parameters and quantiles $Q_u$ are estimated via the posterior medians and the columns show the means above all data sets.
value mixture densities, the (average) posterior medians differ by around 1% or less at the quantile levels considered in this analysis. The maximum relative difference for the data from (D11) to (D13) is 5.69%, but in most cases, the deviations are substantially lower. In view of the high uncertainty associated with the quantile estimates and existence of various other risk sources, these differences are negligible both from a statistical and economic perspective. The high uncertainty of the quantile estimates themselves is shown by large 95% posterior interval ranges (not included in the table). For example, in the case of the gammagpd mixture and (D1), the average posterior interval lengths of the posterior distributions for $Q_{0.975}$, $Q_{0.99}$, and $Q_{0.999}$ are 11%, 17%, and 49% of the size of the true quantile values. Due to the lower sample size, the uncertainty is even higher for (D11) to (D13). Additional risk sources that are relevant in practical applications include, among others, the model risk that the EVMM approach is inappropriate for the considered application (i.e., the underlying distribution cannot be approximated via an EVMM density with low estimation errors), the risk that the sample of past losses is not representative for the future loss distribution (e.g., due to a structural change), and the risk of an incorrect trending of claims in order to account for cost and price changes (see, e.g., Actuarial Standards Board, 2007, and Venter and Sahasrabuddhe, 2012).

5 Empirical Analysis

In our empirical application, we analyse a sample of automobile insurance claims of the Secura Belgian Re. The data set is provided by Beirlant et al. (2004) as an add-on to their book on extreme value theory. In the context of EVMMs, the sample is considered by Wong and Li (2010), but they focus on the tail approximation under different non-Bayesian estimation methods. The data set comprises 371 inflation-adjusted automobile insurance losses above 1.2 million EUR that were recorded by the Secura Belgian Re between 1988 and 2001 (see Beirlant et al., 2004). For our analysis, we subtract 1.2 million EUR from each loss and divide the losses by 1 million. Thus, we consider a reinsurer’s claims in million EUR resulting from a portfolio of excess-of-loss reinsurance contracts with priority 1.2 million EUR.

Figure 2 shows the histogram of the data. In addition, Table 6 summarizes the descriptive statistics. The claims range from 0.01 to 6.70, the average loss is 1.03, and the standard deviation 1.01. The skewness is 2.43, i.e., the distribution is skewed to the right.

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>371</th>
<th>Minimum</th>
<th>0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>1.03</td>
<td>Median</td>
<td>0.74</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>1.01</td>
<td>90% Quantile</td>
<td>2.12</td>
</tr>
<tr>
<td>Skewness</td>
<td>2.43</td>
<td>99% Quantile</td>
<td>5.72</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>11.18</td>
<td>Maximum</td>
<td>6.70</td>
</tr>
</tbody>
</table>

Figure 2: Histogram Secura Data

Table 6: Descriptive Statistics Secura Data

\[\text{The data set is available online at http://lstat.kuleuven.be/Wiley/ (accessed 24/11/2015).}\]
The Markov chains for the gammagpd mixture have converged under all approaches. For the models with a lognormal or weibull bulk component, the chains also seem to have converged and the parameter estimates are similar under the MH and PT algorithms. However, the behaviour of some chains generated under the PT algorithm for higher temperatures $\tau$ indicates that there might be some problems for these two mixtures.

<table>
<thead>
<tr>
<th>EVMM</th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
<th>$u$</th>
<th>$\sigma_u$</th>
<th>$\xi$</th>
<th>$\phi_u$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gammagpd</td>
<td>MH</td>
<td>0.27</td>
<td>0.54</td>
<td>0.36</td>
<td>0.15</td>
<td>0.22</td>
</tr>
<tr>
<td>Gammagpd</td>
<td>PT</td>
<td>0.09</td>
<td>0.17</td>
<td>0.15</td>
<td>0.08</td>
<td>0.10</td>
</tr>
<tr>
<td>Gammagpd*</td>
<td>MH</td>
<td>0.25</td>
<td>0.44</td>
<td>0.38</td>
<td>0.20</td>
<td>0.93</td>
</tr>
<tr>
<td>Gammagpd*</td>
<td>PT</td>
<td>0.11</td>
<td>0.18</td>
<td>0.16</td>
<td>0.12</td>
<td>0.37</td>
</tr>
<tr>
<td>Gammagpd*</td>
<td>MH</td>
<td>0.41</td>
<td>0.69</td>
<td>0.63</td>
<td>0.29</td>
<td>0.20</td>
</tr>
<tr>
<td>Gammagpd*</td>
<td>PT</td>
<td>0.16</td>
<td>0.26</td>
<td>0.31</td>
<td>0.16</td>
<td>0.09</td>
</tr>
<tr>
<td>Lognormgpd</td>
<td>MH</td>
<td>0.72</td>
<td>0.67</td>
<td>0.33</td>
<td>0.29</td>
<td>0.15</td>
</tr>
<tr>
<td>Lognormgpd</td>
<td>PT</td>
<td>0.29</td>
<td>0.33</td>
<td>0.21</td>
<td>0.18</td>
<td>0.09</td>
</tr>
<tr>
<td>Lognormgpd</td>
<td>MH</td>
<td>0.75</td>
<td>0.70</td>
<td>0.41</td>
<td>0.60</td>
<td>0.68</td>
</tr>
<tr>
<td>Lognormgpd</td>
<td>PT</td>
<td>0.15</td>
<td>0.21</td>
<td>0.18</td>
<td>0.32</td>
<td>0.33</td>
</tr>
<tr>
<td>Lognormgpd</td>
<td>MH</td>
<td>0.91</td>
<td>0.83</td>
<td>0.72</td>
<td>0.40</td>
<td>0.22</td>
</tr>
<tr>
<td>Lognormgpd</td>
<td>PT</td>
<td>0.31</td>
<td>0.35</td>
<td>0.49</td>
<td>0.25</td>
<td>0.14</td>
</tr>
<tr>
<td>Weibullgpd</td>
<td>MH</td>
<td>0.07</td>
<td>0.31</td>
<td>0.16</td>
<td>0.14</td>
<td>0.11</td>
</tr>
<tr>
<td>Weibullgpd</td>
<td>PT</td>
<td>0.10</td>
<td>0.08</td>
<td>0.21</td>
<td>0.15</td>
<td>0.18</td>
</tr>
<tr>
<td>Weibullgpd*</td>
<td>MH</td>
<td>0.17</td>
<td>0.80</td>
<td>0.45</td>
<td>0.34</td>
<td>0.97</td>
</tr>
<tr>
<td>Weibullgpd*</td>
<td>PT</td>
<td>0.17</td>
<td>0.15</td>
<td>0.27</td>
<td>0.26</td>
<td>0.81</td>
</tr>
<tr>
<td>Weibullgpd*</td>
<td>MH</td>
<td>0.18</td>
<td>0.61</td>
<td>0.37</td>
<td>0.23</td>
<td>0.22</td>
</tr>
<tr>
<td>Weibullgpd*</td>
<td>PT</td>
<td>0.10</td>
<td>0.25</td>
<td>0.34</td>
<td>0.11</td>
<td>0.16</td>
</tr>
</tbody>
</table>

Table 7: Mixing Results for the Secura Belgian Re Data

This table shows the mixing of the Markov chains in terms of the lag-50 autocorrelations for the Secura Belgian Re data. The first column shows the estimated EVMM and modeling approach (original model, reparameterised model with scale $\sigma^*$, and original model with full Bayesian tail fraction estimation). The second column indicates the sampling algorithm (PT or MH).

Table 7 summarizes the mixing behaviour of the Markov chains. For the gammagpd and lognormgpd mixtures, PT leads to a reduction of the lag-50 autocorrelations by at least one third. In the case of the weibullgpd mixture, the benefits of the parallel tempering algorithm are lower and we even observe some increases in the autocorrelations under the standard model. The empirical analysis further confirms the result that the GPD reparameterisation does not bring major benefits or has even adverse effects in the case of a small sample. For all three mixtures, sampling of $\sigma^*$ leads to higher autocorrelations of the Markov chains for the tail parameters $u$, $\sigma_u$, $\xi$, and $\phi_u$. If a weibull bulk model is used, the autocorrelations are even more than doubled under the reparameterised model and the chains are temporarily stuck in a remote mode at the edge of the posterior range. These problems are illustrated in Figure 3, which contrasts the traceplots for the threshold and shape parameter $\xi$ under the two model parameterisations.

In Subfigure (b), the chain for $u$ is trapped at a very low value for around 5,000 iterations and, for a shorter while, also at a very high value (see approximately iteration 65,000). In addition, the Markov chains for $\xi$ are highly autocorrelated under the reparameterised model, substantially more than under the original mixture (see Subfigures c and d).

The table also shows the autocorrelations under the full Bayesian approach (which have not been considered, so far). The values mainly exceed the autocorrelations under the simplified standard approach, especially if the standard MH algorithm is used.
Figure 3: Weibullgpd and Weibullgpd* Traceplots for the Secura Belgian Re Data

This figure shows the traceplots for the threshold $u$ and shape parameter $\xi$ in the application of the original and reparameterised weibullgpd mixtures to the Secura Belgian Re data set. The chains are generated via the MH algorithm. The subfigures on the left (a and c) illustrate the chains for the original model. The traceplots generated under the reparameterised model with $\sigma^*$ are shown on the right (Subfigures b and d).

Table 8 contrasts the results (posterior medians and 95% posterior intervals) of the quantile and parameter estimation under the the simplified tail fraction approach and the full Bayesian method (the original model parameterisation is used in both cases). For the lognormgpd mixture, both methods lead to almost the same parameter and quantile estimates. The 95% posterior intervals are also consistent (except for $\theta_1$). For the mixtures with gamma and weibull bulk components, the estimates for the bulk model parameters $\theta_1$ and $\theta_2$ are very similar, too. The posterior medians for the remaining parameters differ, but the posterior intervals overlap and always include the median from the alternative approach. Due to these differences in the posteriors of the tail parameters, the estimates of the tail quantiles are also different. However, the deviations are lower than for the parameters and largely negligible in light of the high uncertainty in the estimates (see the posterior interval lengths). If a gamma bulk model is used, the quantiles estimated under the simplified approach differ by 5% or less from the quantiles under the full Bayesian approach (for the probabilities considered here). The relative differences for the weibullgpd mixture range between 3% and 12%.

Finally, we note that the estimated quantiles mainly lie very close to the empirical quantiles (given in the first row of Table 8). In empirical applications, this is usually considered as an indicator for a good tail fit (see, e.g., MacDonald et al., 2011, Do Nascimento et al., 2012, Lee et al., 2012). However, as the empirical quantiles can differ from the true quantiles of the (unknown) underlying distribution, a firm judgement of the approximation would only be possible if we knew the true distribution.\textsuperscript{13}

\textsuperscript{13}For this paper, this is less relevant, as the objective is the identification of potential differences between the simplified and full Bayesian tail fraction approach, not the evaluation of the tail approximation.
Table 8: Parameter and Quantile Estimates for the Secura Belgian Re Data

<table>
<thead>
<tr>
<th>EVMM</th>
<th>( \theta_1 )</th>
<th>( \theta_2 )</th>
<th>( \sigma_\alpha )</th>
<th>( \xi )</th>
<th>( \phi_\alpha )</th>
<th>( Q_{0.95} )</th>
<th>( Q_{0.975} )</th>
<th>( Q_{0.99} )</th>
<th>( Q_{0.999} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empirical</td>
<td>2.90</td>
<td>3.90</td>
<td>5.72</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gammadpd</td>
<td>1.28</td>
<td>1.3</td>
<td>5.19</td>
<td>0.99</td>
<td>-0.25</td>
<td>0.01</td>
<td>2.90</td>
<td>3.76</td>
<td>5.37</td>
</tr>
<tr>
<td>Gammadpd( \phi )</td>
<td>1.27</td>
<td>1.3</td>
<td>3.35</td>
<td>1.28</td>
<td>-0.20</td>
<td>0.04</td>
<td>2.97</td>
<td>3.97</td>
<td>5.45</td>
</tr>
<tr>
<td>Lognormgd</td>
<td>1.35</td>
<td>1.87</td>
<td>1.72</td>
<td>0.95</td>
<td>0.11</td>
<td>0.15</td>
<td>2.98</td>
<td>3.85</td>
<td>5.08</td>
</tr>
<tr>
<td>Lognormgd( \phi )</td>
<td>1.32</td>
<td>1.85</td>
<td>1.70</td>
<td>0.93</td>
<td>0.12</td>
<td>0.17</td>
<td>2.96</td>
<td>3.83</td>
<td>5.09</td>
</tr>
<tr>
<td>Weibullgd</td>
<td>1.19</td>
<td>1.01</td>
<td>3.84</td>
<td>1.28</td>
<td>-0.28</td>
<td>0.03</td>
<td>2.86</td>
<td>3.85</td>
<td>5.39</td>
</tr>
<tr>
<td>Weibullgd( \phi )</td>
<td>1.19</td>
<td>1.02</td>
<td>2.57</td>
<td>1.39</td>
<td>-0.10</td>
<td>0.10</td>
<td>3.26</td>
<td>4.31</td>
<td>5.56</td>
</tr>
</tbody>
</table>

This table shows the parameter and quantile estimates for the Secura Belgian Re data set under the original EVMM parameterisation with simplified tail fraction estimation and the full Bayesian approach. The use of the latter is indicated by the Greek letter \( \phi \) at the end of the model name. The table shows the results for the Markov chains generated via parallel tempering (the MH results are similar). The parameters and quantiles are estimated via the posterior medians. In addition, the 95% HPD intervals are given in brackets.

6 Conclusions

The Bayesian estimation of EVMMs has various advantages compared to the maximum likelihood method. In this paper, we address several computational issues in the simulation of the Markov chains for the posterior distributions. We describe three possible enhancements of the standard simulation method and present the detailed sampling algorithms. In order to evaluate the adequacy of commonly used simplifications and the benefits of the proposed enhancements, we perform a comprehensive simulation study with different sample sizes and distribution types and apply the approaches to empirical automobile insurance losses. The approaches are compared with respect to the convergence and mixing behaviour of the Markov chains and their effects on the parameter and quantile estimation.

The first problem considered in our analysis is the threshold dependence of the GPD scale parameter. We resolve this dependence by means of a GPD reparameterisation. The comparison of the results for the original and reparameterised models shows similar coverage rates and lengths of the posterior intervals. Moreover, in the case of relatively small samples from EVMMs and data from other parametric distributions, the reparameterisation does not generally improve the convergence and mixing behaviour of the Markov chains. We even observe some tendency that the chains are getting stuck in a remote mode more frequently. For large samples from extreme value mixture model distributions (which is rather unlikely in real applications), sampling of \( \sigma^* \) may increase the sampling efficiency, but the benefits are rather small if parallel tempering is applied. Our results therefore indicate that the threshold dependence of the GPD scale does not necessarily need to be taken into account.

Second, we examine the consequences of the simplified tail fraction approximation via the maximum likelihood estimator. For this, we develop a full Bayesian approach and compare the parameter and quantile estimates. The simulation study and empirical application support the simplified approach. The latter requires substantially less computational effort, generates better mixing chains, and, most importantly,
provides adequate parameter estimates for the large majority of samples. There are also some data sets with some differences in the parameter estimates under the simplified and full Bayesian method. However, the estimates of the tail quantiles, which are typically more relevant than the parameters themselves, are similar in all our examples.

Finally, we propose to enhance the simulation by means of parallel tempering. The method is not a panacea, and in some cases both the MH sampler and the PT method fail. However, for nearly all data sets (and the original model with simplified tail fraction estimation), parallel tempering leads to equally good or better convergence results than the standard MH algorithm and substantially improves the mixing behaviour. Even in the case of a rather ad-hoc choice of the temperature, the autocorrelations are reduced considerably and the Markov chains jump back and forth between different modes. Due to the simulation of several parallel chains, parallel tempering requires higher computational efforts than the standard MH algorithm, but the additional burden can in parts be compensated through the need of shorter burn-in periods and chain lengths.

Our analyses focus on the gammagpd, lognormgpd, and weibullgpd mixtures. We assume that the results also hold for EVMMs with other bulk distributions, for example a kernel density (as proposed by MacDonald et al., 2011, and MacDonald et al., 2013) or a mixture of gamma distributions (see Do Nascimento et al., 2012). Moreover, similar results are expected under the bulk model based tail fraction approach and in the case of a continuity constraint for the density (see, e.g., Scarrott, 2016). These assumptions could be checked in future studies. In addition, the benefits of a parameter orthogonalization could be evaluated (see also Scarrott and MacDonald, 2012). As shown by Davison (2003), orthogonal GPD parameters in the sense of Cox and Reid (1987) can be obtained by expressing the GPD in terms of $\xi$ and $\sigma^* = \sigma_a(1 + \xi)$. Finally, future work could examine whether the convergence problems under the PT algorithm for some samples can be resolved through an advanced technique for the selection of the temperatures (see, e.g., Earl and Deem, 2005, or Vousden et al., 2015) or by means of an alternative sampling approach designed for multimodal posterior distributions (see, e.g., Gill and Casella, 2004). For example, Ergashev et al. (2013) use simulated annealing.

### Appendix A

<table>
<thead>
<tr>
<th>Bulk Model</th>
<th>Density $h$</th>
<th>Priors</th>
<th>Prior Parameters</th>
<th>Proposals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gamma</td>
<td>$h(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \cdot \exp(-\beta x)$</td>
<td>$\alpha \sim Ga(a, b)$</td>
<td>$a = 1; b = 0.01$</td>
<td>$\alpha$: Gamma*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\mu = \alpha/\beta \sim IG(c, d)$</td>
<td>$c = 1.5; d = 5$</td>
<td>$\mu$: Gamma*</td>
</tr>
<tr>
<td>Lognormal</td>
<td>$h(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right)$</td>
<td>$\mu \sim N(m, s^2)$</td>
<td>$m = 1; s = 1,000$</td>
<td>$\mu$: Normal</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma^2 \sim IG(c, d)$</td>
<td>$c = 2.5; d = 5$</td>
<td>$\sigma^2$: Gamma</td>
</tr>
<tr>
<td>Weibull</td>
<td>$h(x) = \frac{\tau}{\xi} \left(\frac{x}{\xi}\right)^{\tau-1} \cdot \exp\left(-\left(\frac{x}{\xi}\right)^\tau\right)$</td>
<td>$\tau \sim Ga(a, b)$</td>
<td>$a = 1; b = 0.01$</td>
<td>$\tau$: Gamma</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\log(\phi) \sim N(m, s^2)$</td>
<td>$m = 0; s = 2$</td>
<td>$\log(\phi)$: Normal $Z(-100; 100)$</td>
</tr>
</tbody>
</table>

Table A1: Bulk Model Specifications

This table provides some details on the considered bulk models and their Bayesian estimation. The first column shows the density, the second and third column display the prior distributions and parameters, and the last column gives the proposals used for the MH algorithm. The priors for the gammagpd mixture are chosen based on Wiper and Rios (2001) and Do Nascimento et al. (2012). For the priors for the lognormal bulk component, we refer to Richardson and Green (1997) and de Alba (2006). A gamma prior for the weibull shape parameter is also used by Tsionas (2002) and Marin et al. (2005).  

* In order to prevent the chain from getting stuck, the gamma proposals for $\alpha$ and $\mu$ are truncated at 0.001 and 0.00001, respectively.
Appendix B

This section outlines the algorithms for the MCMC simulations of the EVMM posterior distributions under the various approaches. The algorithms are implemented in R, where the likelihood functions $L_{MM}$ are available in the evmix-package by Scarrott and Hu (2015).

The first algorithm describes the Metropolis Hastings sampler for the two EVMM parameterisations and the standard case that the tail fraction is determined by the remaining parameters. The approach for the original model (model 0 below) mainly corresponds to the sampling method used by Behrens et al. (2004) and Do Nascimento et al. (2012) and, in parts, also to the procedure of MacDonald (2011). The steps for the remaining models are adapted accordingly.

In the following, let $m = 0$ denote the original EVMM form and $m = 1$ the model with GPD scale $\sigma^*$. Similarly, set $\pi[0] = \pi_{MM}$ and $\pi[1] = \pi_{M_M}$, as well as $\sigma_0 = \sigma_u$ and $\sigma_1 = \sigma^*$. We further write $M$ for the maximum of a sample $x = \{x_1, \ldots, x_m\}$ and $x(i)$ for its $i$th order statistic. The densities of the truncated normal distribution with mean $\mu$, variance $V$, and range $A$, and the truncated gamma distribution with shape $a$, rate $b$, and range $A$, are denoted with $f_{TM}(x; \mu, V, A)$ and $f_{TG}(x; a, b, A)$, respectively. $V_\xi$, $\sigma_m$, $V_u$, and $V_\sigma$ are the proposal variances and calibrated so as to obtain appropriate acceptance rates (see Section 4.1).

In line with Behrens et al. (2004) and Do Nascimento et al. (2012), the ranges $A$ for the proposals of $u$, $\sigma_0$, and $\xi$ are chosen in order to ensure $(1 + \xi \cdot (x - u)/\sigma_0) > 0$ for all $x > u$. Analogously, we require $(1 + \xi \cdot (x - u)/(\sigma_1 + \xi u)) > 0$ in the case of the reparameterised models. We further ensure $\sigma_u = \sigma_1 + \xi u > 0$. The specific ranges are given in Table B1.

Under the tail fraction approach, the parameter $\phi_u$ is determined via the proportion of threshold exceedances. Thus, the chains for $\phi_u$ are derived from the chains for $u$ and each new threshold $\hat{u}$ also implies a new tail fraction $\hat{\phi}_u$. Although not explicitly shown in the following algorithm, this is taken into account in the calculation of the acceptance probabilities $\alpha_u$.

<table>
<thead>
<tr>
<th>$m$</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_\xi$</td>
<td>$(a; \infty)$ with $a = \max \left{ \frac{x(t-1) - u(t-1)}{\xi(t-1)}, -0.495 \right}$</td>
<td>$(a; \infty)$ with $a = \max \left{ -\frac{x(t-1) - u(t-1)}{\xi(t-1)}, -0.495 \right}$</td>
</tr>
<tr>
<td>$A_{\sigma_m}$</td>
<td>$(-\xi(t) \cdot u(t-1); \infty)$, if $\xi(t) &gt; 0$</td>
<td>$(-\xi(t) \cdot M; \infty)$, if $\xi(t) &gt; 0$</td>
</tr>
<tr>
<td>$A_u$</td>
<td>$(a; x(n-2))$ with $a = x(3)$, if $\xi(t) &gt; 0$</td>
<td>$[a_1; a_2]$ with $a_1 = \max \left{ -\frac{x(t)}{\xi(1); x(3)}, x(n-2) \right}$, if $\xi(t) &gt; 0$</td>
</tr>
<tr>
<td>&amp; $a = \max \left{ M + \frac{\sigma(t)}{\xi(t); x(3)}, x(3) \right}$, else</td>
<td>$a_2 = x(n-2)$, $\xi(t) &gt; 0$</td>
<td></td>
</tr>
</tbody>
</table>

Table B1: Ranges of the Proposal Distributions in Algorithm 1

This table specifies the ranges $A_\xi$, $A_{\sigma_m}$, and $A_u$ of the proposal distributions in Algorithm 1 for the two model parameterisations $m = 0, 1$. The ranges are derived based on the requirements $(1 + \xi \cdot (x - u)/\sigma_0) > 0$ for all $x > u$ and $\sigma_0 > 0$ in the case of $m = 0$ and $(1 + \xi \cdot (x - u)/(\sigma_1 + \xi u)) > 0$ and $\sigma_1 + \xi u > 0$ if $m = 1$.
Algorithm 1: MH-Sampling for the Posterior $\pi^{[m]}$  

Initialization: Choose starting values $\xi^{(0)}$, $\sigma_m^{(0)}$, $u^{(0)}$, and $\theta_1^{(0)}, ..., \theta_K^{(0)}$.  

Simulation: For $t = 1, ..., T$ do  

1. Sampling $\xi$:  

   1.1 Sample $\tilde{\xi} \sim N(\xi^{(t-1)}, V_{\xi})$; 
   1.2 Calculate 
   \[
   \alpha_{\xi} = \min \left\{ \frac{\pi^{[m]}(\theta^{(t-1)}, u^{(t-1)}, \sigma_m^{(t-1)}, \tilde{\xi}|x) \cdot f_{TN}(\xi^{(t-1)}|\tilde{\xi}, V_{\xi}, A_{\xi})}{\pi^{[m]}(\theta^{(t-1)}, u^{(t-1)}, \sigma_m^{(t-1)}, \xi^{(t-1)}|x) \cdot f_{TN}(\xi^{(t-1)}|\xi^{(t-1)}, V_{\xi}, A_{\xi})}; 1 \right\};
   \]
   1.3 Set $\xi^{(t)} = \tilde{\xi}$ with probability $\alpha_{\xi}$ and $\xi^{(t)} = \xi^{(t-1)}$ otherwise.  

2. Sampling $\sigma_m$:  

   2.1 Draw $\tilde{\sigma}_m \sim N(\sigma_m^{(t-1)}, V_{\sigma_m})$;  
   2.2 Compute 
   \[
   \alpha_{\sigma_m} = \min \left\{ \frac{\pi^{[m]}(\theta^{(t-1)}, u^{(t-1)}, \tilde{\sigma}_m, \xi^{(t)}|x) \cdot f_{TN}(\sigma_m^{(t-1)}|\tilde{\sigma}_m, V_{\sigma_m}, A_{\sigma_m})}{\pi^{[m]}(\theta^{(t-1)}, u^{(t-1)}, \sigma_m^{(t-1)}, \xi^{(t)}|x) \cdot f_{TN}(\sigma_m^{(t-1)}|\sigma_m^{(t-1)}, V_{\sigma_m}, A_{\sigma_m})}; 1 \right\};
   \]
   2.3 Set $\sigma_m^{(t)} = \tilde{\sigma}_m$ with probability $\alpha_{\sigma_m}$ and $\sigma_m^{(t)} = \sigma_m^{(t-1)}$ otherwise.  

3. Sampling $u$:  

   3.1 Draw $\tilde{u} \sim N(u^{(t-1)}, V_u)$;  
   3.2 Determine 
   \[
   \alpha_u = \min \left\{ \frac{\pi^{[m]}(\theta^{(t-1)}, \tilde{u}, \sigma_m^{(t-1)}, \xi^{(t)}|x) \cdot f_{TN}(u^{(t-1)}|\tilde{u}, V_u, A_u)}{\pi^{[m]}(\theta^{(t-1)}, u^{(t-1)}, \sigma_m^{(t-1)}, \xi^{(t)}|x) \cdot f_{TN}(u^{(t-1)}|u^{(t-1)}, V_u, A_u)}; 1 \right\};
   \]
   3.3 Set $u^{(t)} = \tilde{u}$ with probability $\alpha_u$ and $u^{(t)} = u^{(t-1)}$ otherwise.  

4. Sampling $\theta = (\theta_1, ..., \theta_K)$:  

   4.1 Lognormal bulk: Draw $\tilde{\theta}_1 = \tilde{\mu} \sim N(\mu^{(t-1)}, V_{\mu})$ and $\tilde{\theta}_2 = (\tilde{\sigma})^2 \sim G(a_{\theta_2}^{(t-1)}, b_{\theta_2}^{(t-1)})$;  
   Gamma bulk: Draw $\tilde{\theta}_1 = \tilde{\alpha} \sim G(a_{\theta_1}^{(t-1)}, b_{\theta_1}^{(t-1)})$, $\tilde{\theta}_2 = (\alpha/\beta) \sim G(a_{\theta_2}^{(t-1)}, b_{\theta_2}^{(t-1)})$;  
   Weibull bulk: Draw $\tilde{\theta}_1 = \tilde{\tau} \sim G(a_{\theta_1}^{(t-1)}, b_{\theta_1}^{(t-1)})$ and $\log(\tilde{\theta}_2) = \log(\tilde{\phi}) \sim N(\log(\phi^{(t-1)}); V_\phi)$, $\mathcal{I}(-100; 100)$;  
   The shape $a_{\theta_k}^{(t-1)}$ and rate $b_{\theta_k}^{(t-1)}$ of the Gamma proposals are chosen such that the proposal mean and variance equal $\theta_k^{(t-1)}$ and $V_{\theta_k}$;  
   4.2 Calculate 
   \[
   \alpha_\theta = \min \left\{ \frac{\pi^{[m]}(\tilde{\theta}, u^{(t)}, \sigma_m^{(t-1)}, \xi^{(t)}|x) \cdot \prod_{k=1}^K f_k(\tilde{\theta}_k^{(t-1)}|\tilde{\theta}_k, V_{\theta_k})}{\pi^{[m]}(\theta^{(t-1)}, u^{(t)}, \sigma_m^{(t-1)}, \xi^{(t)}|x) \cdot \prod_{k=1}^K f_k(\theta_k^{(t-1)}|\theta_k, V_{\theta_k})}; 1 \right\};
   \]
   Here, $f_k$ denotes the density of the proposal for $\theta_k$, $k = 1, ..., K$.  
   4.3 Set $\theta^{(t)} = \tilde{\theta}$ with probability $\alpha_\theta$ and $\theta^{(t)} = \theta^{(t-1)}$ otherwise.

\footnote{Behrens et al. (2004) and Do Nascimento et al. (2012) use a gamma proposal for the case $\xi^{(t)} \geq 0$. However, with this proposal, the chain is sometimes getting stuck in our simulations.}

25
Algorithm 2: MH-Sampling for the Extended Model with Posterior $\pi_{MM}^\phi$

Initialization: Choose starting values $\xi^{(0)}$, $\sigma_u^{(0)}$, $u^{(0)}$, $\phi_u^{(0)}$, $\theta_1^{(0)}, \ldots, \theta_K^{(0)}$.

Simulation: For $t = 1, \ldots, T$ do

1. Sampling $\xi$ and $\sigma_u$: As Steps 1 and 2 in Algorithm 1 for $m = 0$ with sampled tail fraction $\phi_u^{(t-1)}$ and $\pi_{MM}^\phi$ instead of $\pi^{(0)}$. 

2. Sampling $u$ and $\phi_u$:
   
   2.1 Calculate the ML estimators $\hat{u}$ and $\hat{\phi}_u$ for $u$ and $\phi_u$ given $\xi = \xi^{(t)}$, $\sigma_u = \sigma_u^{(t)}$, and $\theta = \theta^{(t-1)}$. In addition, calculate the standard deviation $\hat{\phi}_u = \sqrt{\hat{\phi}_u \cdot (1 - \hat{\phi}_u) / n}$ of $\hat{\phi}_u$.

   2.2 Draw $\tilde{u} \sim N(\hat{u}, V_{\phi_u}) \cdot \mathcal{I}(A_u)$ with $A_u$ specified in Table B1 and $\hat{\phi}_u \sim N(\hat{\phi}_u, V_{\phi_u}) \cdot \mathcal{I}(0; 1)$ with $V_{\phi_u} = v \cdot \hat{\phi}_u^2$. The parameter $v$ is calibrated ex-ante based on the acceptance rates for the proposed values.

   2.3 Determine $a_{u,\phi} = \frac{\pi_{MM}^\phi(\theta^{(t-1)}, \tilde{u}, \sigma_u^{(t)}, \xi^{(t)}, \tilde{\phi}_u | x) \cdot f_{TN}(u^{(t-1)} | \hat{u}, V_{\phi_u}, A_u) \cdot f_{TN}(\phi_u^{(t-1)} | \hat{\phi}_u, V_{\phi_u}, (0; 1))}{\pi_{MM}^\phi(\theta^{(t-1)}, u^{(t-1)}, \sigma_u^{(t)}, \xi^{(t)}, \phi_u^{(t-1)} | x) \cdot f_{TN}(\tilde{u} | \hat{u}, V_{\phi_u}, A_u) \cdot f_{TN}(\tilde{\phi}_u | \hat{\phi}_u, V_{\phi_u}, (0; 1))}$ and set $\alpha_{u,\phi} = \min\{a_{u,\phi}; 1\}$.

   2.4 Set $u^{(t)} = \tilde{u}$ and $\phi_u^{(t)} = \tilde{\phi}_u$ with probability $\alpha_{u,\phi}$. Otherwise, $u^{(t)} = u^{(t-1)}$ and $\phi_u^{(t)} = \phi_u^{(t-1)}$.

3. Sampling $\theta$: As Step 4 in Algorithm 1 for $m = 0$ with sampled tail fraction $\phi_u^{(t)}$ and $\pi_{MM}^\phi$ instead of $\pi^{(0)}$.

Algorithm 3: Parallel Tempering

Let $\pi = \pi_{MM} = \pi^{(0)}$ be the target density and $\pi_p = \pi^{1/\tau_p}$ for temperatures $1 = \tau_1 < \tau_2 < \ldots < \tau_P$. The procedure for the posteriors of the reparameterised models and for $\pi_{MM}^\phi$ is analogous. The following algorithm is an adaption of the general PT algorithm described in Gill and Casella (2004) and Craiu and Rosenthal (2014) to EVMMs.

Initialization: Choose $\xi^{(0)}$, $\sigma_u^{(0)}$, $u^{(0)}$, $\theta^{(0)}$ and set $\Psi_p^{(0)} = (\theta^{(0)}, u^{(0)}, \sigma_u^{(0)}, \xi^{(0)})$, $p = 1, \ldots, P$.

Simulation: For $t = 1, \ldots, T$ do

1. For $p = 1, \ldots, P$ update the chain for the posterior $\pi_p$ (i.e., determine $\Psi_p^{(t)} = (\theta_p^{(t)}, u_p^{(t)}, \sigma_u^{(t)}, \xi^{(t)})$) according to Algorithm 1 (with $\pi_{MM}$ being replaced by $\pi_p$).

2. If $t/s$ is an integer (i.e., after every $s$ steps) select randomly $p_1, p_2 \in \{1, \ldots, P\}$. Swap $\Psi_{p_1}^{(t)}$ and $\Psi_{p_2}^{(t)}$ with probability

   $$\alpha_T = \min \left(1, \frac{\pi_{p_1}^{(t)}(\Psi_{p_2}^{(t)}) \cdot \pi_{p_2}^{(t)}(\Psi_{p_1}^{(t)})}{\pi_{p_1}^{(t)}(\Psi_{p_1}^{(t)}) \cdot \pi_{p_2}^{(t)}(\Psi_{p_2}^{(t)})} \right).$$

Finalization: Keep the chains for $\tau_1 = 1$ for the approximation of the posterior $\pi_{MM}$.
References


