Insurance: Mathematics and Economics 88 (2019) 77–92

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

journal homepage: www.elsevier.com/locate/ime

Severity modeling of extreme insurance claims for tariffication

Christian Laudagé [a,*], Sascha Desmettre [b,c], Jörg Wenzel [a]

[a] Department of Financial Mathematics, Fraunhofer ITWM, Fraunhofer-Platz 1, 67663 Kaiserslautern, Germany
[b] Department of Mathematics, TU Kaiserslautern (TUK), Erwin-Schrödinger Straße, 67663, Kaiserslautern Germany
[c] Institute for Mathematics and Scientific Computing, University of Graz, Heinrichstraße 36, AT-8010 Graz, Austria

ARTICLE INFO

Article history:
Received August 2018
Received in revised form May 2019
Accepted 11 June 2019
Available online 19 June 2019

JEL classification:
C24
G22

Keywords:
Extreme claims
Generalized linear model
Truncated gamma distribution
Extreme value theory
Peaks-over-threshold
Generalized Pareto distribution

ABSTRACT

Generalized linear models are common instruments for the pricing of non-life insurance contracts. They are used to estimate the expected frequency and severity of insurance claims. However, these models do not work adequately for extreme claim sizes. To accommodate for these extreme claim sizes, we develop the threshold severity model, that splits the claim size distribution in areas below and above a given threshold. More specifically, the extreme insurance claims above the threshold are modeled in the sense of the peaks-over-threshold methodology from extreme value theory using the generalized Pareto distribution for the excess distribution, and the claims below the threshold are captured by a generalized linear model based on the truncated gamma distribution. Subsequently, we develop the corresponding concrete log-likelihood functions above and below the threshold. Moreover, in the presence of simulated extreme claim sizes following a log-normal as well as Burr Type XII distribution, we demonstrate the superiority of the threshold severity model compared to the commonly used generalized linear model based on the gamma distribution.

1. Introduction

Property–casualty insurers rely on statistical methods to price their insurance contracts adequately. A common assumption is that the claim frequency and severity are independent. In that regard, a widespread practice is to model the claim severity using a generalized linear model (GLM) based on a gamma distribution, see e.g. Ohlsson and Johansson (2010), Wüthrich (2017) and Garrido et al. (2016). However, due to the light-tailed of the gamma distribution this approach may treat extreme claim sizes inappropriately. To model these extreme claim sizes alternative approaches are required.

For instance, the work of Shi (2014) presents regression models that are able to incorporate fat tails. Among other things, Shi (2014) mentions the use of a GLM with an inverse Gaussian distribution for long-tail claims. In a similar fashion, Shi et al. (2015) make use of a fat-tailed regression with a generalized gamma distribution in order to include heavy-tailed distributions for modeling extreme claim severity. Also in this context, focusing on binary rare events data, Calabrese and Osmetti (2011) propose the so-called generalized extreme value (GEV) regression model that uses the quantile function of the GEV distribution as link function.

Another obvious way to handle extreme claim sizes over a certain threshold is extreme value theory and in particular the peaks-over-threshold method (POT); see e.g. the classical monograph (Embrechts et al., 1997) for a comprehensive overview with a focus on finance and insurance. Notably, these methods can be used for the risk management of property–casualty insurances as it is for instance exemplified in McNeil (1997) and Embrechts et al. (1999). Furthermore, for risk analysis it is a usual procedure to combine two distributions in a so-called splicing or composite model, see e.g. Lee et al. (2012), Reynkens et al. (2017), Li et al. (2016) and the references therein. In doing so, the claim size distribution is modeled with a distribution for the body that covers light and moderate claim sizes and a distribution for the tail that covers extreme claim sizes. Nevertheless, all these approaches do not allow for a sufficient classification of the policyholders, which is in line with the pricing policy and which we aim at in this work – think e.g. of a household insurance where the total insured value and the future damages depend on each other. Another example is a car liability insurance where the premium possibly depends on the car brand. This implicitly assumes that the choice of the car brand affects the drivers risk attitude or the likelihood of theft. For simplicity, we refer to such characteristics as tariff features.

In order to consider extreme claim sizes in the premium calculation and at the same time allowing for a classification by tariff

* Corresponding author.
E-mail addresses: christian.laudage@itwm.fraunhofer.de (C. Laudagé), desmettre@mathematik.uni-kl.de (S. Desmettre), joerg.wenzel@itwm.fraunhofer.de (J. Wenzel).

https://doi.org/10.1016/j.insmatheco.2019.06.002

features, most notably the insured sum, we develop the so-called *threshold severity model*. This model transfers the idea of a splicing model for risk management into the context of tariffication by taking specific tariff features into account. This means that we apply the POT method from extreme value theory and split the claim severity at a certain large threshold. For the data below the threshold we develop a new GLM for the truncated gamma distribution. The extreme claim size data is then modeled by the generalized Pareto distribution (GPD). We estimate the parameters of these distributions by taking right-censored data points into account. This is a typical situation for primary insurers, because of the limitation by insured sums. Moreover, we give concrete estimators for the expected claim severity, above and below the threshold. To the best of our knowledge, the threshold severity model for the first time combines the POT modeling for extreme claim sizes with GLMs based on the truncated gamma distribution. For further analysis, we wish to stress that we work under the assumption of independence between claim frequency and claim severity. For models that allow for dependence – however not in the context of our threshold severity model – we refer to Shi et al. (2015) and Garrido et al. (2016) among others.

Finally, we illustrate and evaluate the behavior of the standard gamma GLM and the threshold severity model in the presence of extreme claim size data in a simulation study. Among other things, we fit the threshold severity model and the gamma GLM against log-normal and Burr Type XII distributed claim sizes. To do so, we use a finite number of insured sums which is e.g. common for personal and car liability insurances. We find that the standard gamma GLM yields misleading relations between the tariff features and the expected claim severity, whereas the threshold severity model is able to explain these relations significantly better.

The remainder of the paper is structured as follows. In Section 2 we introduce the modeling framework and define the threshold severity model. In Section 3 we list the likelihood-function for the splicing distribution of the threshold severity model. In Section 4 we use concrete distributions to model the body and the tail of the splicing distribution. To this end, we develop a GLM for the truncated gamma distribution in Section 4.2, followed by a description of modeling extreme claims using the GPD in Section 4.3. The advantages of the threshold severity model compared to a gamma GLM in the presence of extreme claim sizes are illustrated by a simulation study in Section 5. Therefore, in Section 5.1 we explain the simulated (heavy-tailed) claim sizes. Then, in Section 5.2 we demonstrate the results of fitting the threshold severity model and the gamma GLM to the simulated data. Afterwards, in Section 5.3 we examine effects of different thresholds on the threshold severity model.

2. Modeling framework

Notation and Set-Up. By $\mathcal{B}(U)$ we denote the smallest σ-field containing all open sets of a topological space U. The power set of a topological space U is denoted by $\mathcal{P}(U)$.

To model insurance claims in a given time period for groups of policyholders we work on a probability space denoted by $(\Omega, \mathcal{F}, P)$. The independent random variables $X_1, X_2, \ldots$ on $(\Omega, \mathcal{F})$ with state space $(\mathbb{R}_{>0}, \mathcal{B}(\mathbb{R}_{>0}))$ model the claim severity. The claim severity is the cost to the insurance underwriter for an event of damage. The claim frequency for the given time period is a random variable N on $(\Omega, \mathcal{F})$ with state space $(\mathbb{N}_0, \mathcal{P}(\mathbb{N}_0))$. The random variable $S := \sum_{i=1}^{N} X_i$ describes the total damage in the given time period.

Distributions for risk management describe the entire insurance portfolio. For this purpose, Lee and Lin (2010) and Willmot and Woo (2007) use mixed Erlang distributions for the modeling of insurance claims. We transfer this idea from risk theory into the tariffication context and use the fact that the inclusion of a weight in the mixed Erlang distribution can also be interpreted as the probability that a policyholder corresponds to a certain group of policyholders. Thereby such a group can be characterized by the values of tariff features. In order to build a connection to the described situation in risk management, we introduce tariff features as non-negative, discrete random variables $R_1, \ldots, R_d$ on $(\Omega, \mathcal{F})$. Let $R = (R_1, \ldots, R_d)$ denote the vector of tariff features. The random variable R_i represents the ith tariff feature.

We impose the following standing assumptions:

Assumption 2.1.

(i) Given the information of the tariff features, all claim sizes have the same distribution, i.e. there is a random variable X such that for each $i \in \{1, \ldots, d\}$ and for each $A \in \mathcal{B}(\mathbb{R}_{>0})$ we have:

$$P(X_i \in A|R) = P(X \in A|R), \quad P\text{-a.s.} \tag{2.1}$$

(ii) Given the information of the tariff features the claim frequency and severity are independent, i.e. X and N are conditionally independent given R.

Remark 2.2.

1. Assumption 2.1 (i) states that claim sizes of the same tariff features have to follow the same distribution, whereas Assumption 2.1 (ii) is an analogue of the common assumption that claim frequency and severity are independent.
2. Under Assumption 2.1 (i) and (ii) we obtain the following conditional version of Wald's identity for a concrete combination of tariff features (i.e. a vector of covariates) $r \in \mathbb{R}^d_{\geq 0}$:

$$E(S|R = r) = E(N|R = r)\, E(X|R = r). \tag{2.2}$$

This model is also known as independent aggregate claims model or two-part model, compare e.g. Garrido et al. (2016) and Shi et al. (2015).

In the following we call a concrete combination of tariff features a *tariff cell*. Our aim is to estimate the conditional expectation $E(X|R = r)$ w.r.t. a concrete tariff cell r. To secure themselves against unaffordable claims, primary insurers only pay for damages up to a specified amount. This amount constitutes a right-censoring point which we call the insured sum and consider it as a tariff feature. Let Y denote the random variable of the actual damage, which may be larger than the insured sum, i.e. Y describes the relevant damage for the policyholder. Without loss of generality, let I denote the index of the tariff feature corresponding to the insured sum. Obviously, the claim severity is then given by

$$X := \min(Y, R_I). \tag{2.3}$$

The insurer only observes realizations for X, i.e. right-censored data. In what follows, we need to determine the distribution of Y based on this censored data.

The Threshold Severity Model. We split the distribution of Y at a certain threshold $u \in \mathbb{R}_{>0}$. This allows to model the body and the tail of the claim size distribution separately. Therefore, for a given tariff cell r let H_r and G_r be distribution functions with parameter vectors Θ_H and Θ_G. To obtain a continuous distribution function for the actual damage Y at the splicing point u, assume that $H_r(u; \Theta_H) > 0$ and $G_r(u; \Theta_G) = 0$ hold. Also let $q_r(\Theta_q)$ be the probability of exceeding the given threshold u with parameter vector Θ_q. Then the distribution function of Y with parameter

vector $\Theta = \left(\Theta_H, \Theta_G, \Theta_q\right)$ is given by

$$F_r\left(y;\Theta\right) = \begin{cases} 0 & ,y \leq 0, \\ \left(1-q_r\left(\Theta_q\right)\right)\dfrac{H_r\left(y;\Theta_H\right)}{H_r\left(u;\Theta_H\right)} & ,0 < y \leq u, \\ \left(1-q_r\left(\Theta_q\right)\right) + q_r\left(\Theta_q\right) G_r\left(y;\Theta_G\right) & ,y > u. \end{cases} \tag{2.4}$$

Let h_r and g_r denote the density functions of H_r and G_r, respectively. The density function of Y is then given by

$$f_r\left(y;\Theta\right) = \begin{cases} 0 & ,y \leq 0, \\ \left(1-q_r\left(\Theta_q\right)\right)\dfrac{h_r\left(y;\Theta_H\right)}{H_r\left(u;\Theta_H\right)} & ,0 < y \leq u, \\ q_r\left(\Theta_q\right) g_r\left(y;\Theta_G\right) & ,y > u. \end{cases} \tag{2.5}$$

To the best of our knowledge it is the first time that splicing distributions are used to determine the expected claim severity for the premium calculation of composite insurance contracts. We refer to this model as the *threshold severity model* (TSM).

Remark 2.3 (*Properties of the TSM*)**.**

1. We wish to stress that we use the same threshold u for each tariff cell within our model, i.e. u is independent of r. To motivate this modeling decision consider the following facts:
   - In *risk management* of insurance companies so-called splicing models are applied (compare e.g. Reynkens et al., 2017; Lee et al., 2012 and the references therein). A key feature of these models is that they are based on a single threshold which is estimated for the entire dataset. This allows for a separate modeling of extreme damages as a form of contaminations. As we aim at a consistent modeling of both fields, risk management *and* premium calculation/tariffication, we opted to use the same single threshold also in the threshold severity model.
   - The introduction of a threshold for extreme claim sizes is associated with a sparse number of extreme events. On the other hand reliable estimators require a sufficient number of these rare events, see e.g. Section 6.5 in Embrechts et al. (1997) or Desmettre and Deege (2016). Since extreme events are already rare over the entire insurance portfolio, different thresholds for different tariff cells would complicate the estimation of the thresholds significantly.

   However, our model should possess the feature that the threshold can be exceeded more or less often, depending on the insured sum. We achieve this by allowing the threshold to be exceeded with different probabilities depending on the tariff cells. In particular, in our model the probability $q_r\left(\Theta_q\right)$ can depend on the tariff cell, and above all on the insured sum.
2. The threshold selection itself is a delicate issue, which has been dealt with in a separate strand of literature. However, the focus of this paper is not the threshold selection and we work under the assumption that the threshold is either externally given (e.g. from a risk management context as described above) or has been determined in advance. For the determination of the threshold we refer e.g. to Desmettre and Deege (2016) and Section 6.5 in Embrechts et al. (1997) as well as Gissibl et al. (2017) and the references therein. In the following we thus assume a given threshold $u \in \mathbb{R}_{>0}$. Later, in Section 5.2 we will also demonstrate the use of the mean excess plot for the determination of the threshold.

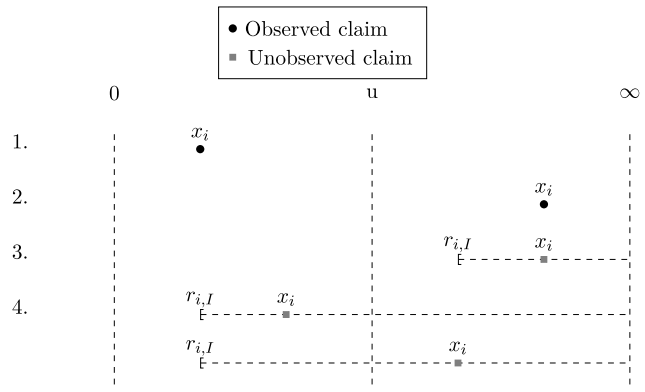

Fig. 1. The four cases of censored and uncensored claims.

3. Maximum-likelihood estimation

The Log-Likelihood Function. The parameter vector Θ can be estimated via the maximum-likelihood method. To formalize the method, we assume m claim sizes and denote the m-dimensional vector of claim sizes by $\mathbf{x}$ as well as the matrix of the corresponding tariff cells by $\mathbf{r}$, such that $r_{i,j}$ denotes the jth tariff feature corresponding to the ith damage. Especially, remember that $r_{i,I}$ is the insured sum of observation i. Further, let r_i be the tariff cell of claim x_i, i.e. the ith row of $\mathbf{r}$. As mentioned earlier, the claim size is limited by the insured sum. Therefore, the insurer only observes right-censored data. In accordance to Section 3.1 in Reynkens et al. (2017) we obtain the following cases of censoring:

1. Uncensored with $x_i \leq u$,
2. Uncensored with $u < x_i$,
3. Censored with $u < r_{i,I}$,
4. Censored with $r_{i,I} \leq u$.

These four cases are illustrated in Fig. 1, which is orientated on Figure 1 in Reynkens et al. (2017).

Also in the spirit of Reynkens et al. (2017) we denote by S_k the indices of data points corresponding to the censoring case $k \in \{1, \ldots, 4\}$. The log-likelihood function is then given by

$$\begin{aligned} l\left(\Theta;\mathbf{x},\mathbf{r}\right) = &\sum_{i\in S_1}\left(\ln\left(1-q_{r_i}\left(\Theta_q\right)\right) + \ln\left(\frac{h_{r_i}\left(x_i;\Theta_H\right)}{H_{r_i}\left(u;\Theta_H\right)}\right)\right) \\ &+ \sum_{i\in S_2}\left(\ln\left(q_{r_i}\left(\Theta_q\right)\right) + \ln\left(g_{r_i}\left(x_i;\Theta_G\right)\right)\right) \\ &+ \sum_{i\in S_3}\left(\ln\left(q_{r_i}\left(\Theta_q\right)\right) + \ln\left(1-G_{r_i}\left(r_{i,I};\Theta_G\right)\right)\right) \\ &+ \sum_{i\in S_4}\ln\left(1-\left(1-q_{r_i}\left(\Theta_q\right)\right)\frac{H_{r_i}\left(r_{i,I};\Theta_H\right)}{H_{r_i}\left(u;\Theta_H\right)}\right). \end{aligned} \tag{3.1}$$

Maximization of the Log-Likelihood. The parameter vectors Θ_H, Θ_G, Θ_q are disjoint for the concrete distributions which we introduce in Section 4. In that case, the last term of (3.1) does not allow for a separate maximization w.r.t. the parameter vectors. This technical optimization problem can be addressed with the help of an EM-algorithm in the spirit of Reynkens et al. (2017). In the situation that all possible insured sums are larger than the threshold, i.e. the right censoring points lie above the threshold, we end up with a log-likelihood function for the *whole* claim size

distribution which *separates* into parts for claim sizes above and below the threshold.

This situation is typical for considerably many common non-life insurance contracts. For instance, typical insured sums corresponding to German personal liability insurances are blanket 5, 10 or 15 million Euro, or even up to 50 million Euro and higher. Moreover, in German car liability insurance contracts, in most cases insured sums are larger than required by law in the environment of roughly 100 million Euro. These contract types are the ones we aim at within the setting of the threshold severity model. The log-likelihood function for these types of contracts is given by

$$\begin{aligned} l(\Theta; \mathbf{x}, \mathbf{r}) = &\sum_{i \in S_1} \left(\ln\left(1 - q_{r_i}\left(\Theta_q\right)\right) + \ln\left(\frac{h_{r_i}\left(x_i; \Theta_H\right)}{H_{r_i}\left(u; \Theta_H\right)}\right)\right) \\ &+ \sum_{i \in S_2} \left(\ln\left(q_{r_i}\left(\Theta_q\right)\right) + \ln\left(g_{r_i}\left(x_i; \Theta_G\right)\right)\right) \\ &+ \sum_{i \in S_3} \left(\ln\left(q_{r_i}\left(\Theta_q\right)\right) + \ln\left(1 - G_{r_i}\left(r_{i,I}; \Theta_G\right)\right)\right). \end{aligned} \tag{3.2}$$

After having maximized the log-likelihood function, we obtain an estimator $\hat{\Theta}$ for the parameter vector and are able to calculate the desired expectation for a tariff cell $r \in \mathbb{R}^d_{\geq 0}$ by

$$E_{\hat{\Theta}}\left(\min(Y, R_I) | R = r\right) = \int_0^{r_I} y f_r\left(y; \hat{\Theta}\right) dy + r_I\left(1 - F_r\left(r_I; \hat{\Theta}\right)\right). \tag{3.3}$$

4. Estimators below and above a given threshold

In this section we want to use concrete distributions for the conditional distribution functions below and above the threshold for a tariff cell r. First, in Section 4.1 we explain the estimation of $q_r\left(\Theta_q\right)$ via a regression model. Subsequently, in Section 4.2, we describe the modeling of the claim severity *below* the given threshold using a GLM and assuming a gamma distribution for H_r. In particular, the conditional distribution function

$$P\left(Y \leq y | Y \leq u, R = r\right) = \frac{H_r\left(y; \Theta_H\right)}{H_r\left(u; \Theta_H\right)}, \quad 0 < y \leq u, \tag{4.1}$$

describes a truncated gamma distribution. The starting point for a new regression model based on the truncated gamma distribution is the classical gamma GLM as described in Section 2 in Ohlsson and Johansson (2010) and Section 7 in Wüthrich (2017).

Afterwards, in Section 4.3, we apply methods from extreme value theory to estimate the claim severity *above* the given threshold. In accordance with Embrechts et al. (1997) the conditional distribution function

$$P\left(Y \leq y | Y > u, R = r\right) = G_r\left(y; \Theta_G\right), \quad y > u, \tag{4.2}$$

is approximated by the GPD. In Section 4.4 we specify the estimator for the expected claim size in which we take into account that the claim size is limited by the insured sum.

4.1. Estimation of the probability of exceeding the threshold

Due to the fact that binary regression models have significant drawbacks for rare events data (compare e.g. Calabrese and Osmetti, 2011), in our model, the probability of exceeding the threshold only depends on the insured sum. The probability $q_r\left(\Theta_q\right)$ can then be estimated using a GLM like e.g. a probit- or logit-model. For the simulation study in Section 5 we use a logit-model. The parameter vector w.r.t. the insured sum as the only regressor is denoted by δ. Hence, we have $\Theta_q = \delta$ and

$$q_r\left(\Theta_q\right) = \frac{1}{1 + e^{-(\delta_0 + \delta_I r_I)}}. \tag{4.3}$$

Remark 4.1. Using Eq. (4.3) it is possible to estimate Θ_q by maximizing (3.1). Note, that in this general setting we have to estimate Θ_q together with Θ_H. In contrast, in the setting of (3.2), where all insured sums lie above the threshold, we can fit the logit-model separately from the parameter vectors Θ_H and Θ_G.

4.2. Estimation below a given threshold using GLMs

Below the given threshold there are usually enough data to calculate adjusted premiums with regression models. The main idea of GLMs is to map the covariates to the expectation of the underlying distribution. The mapping is described by the so called link function. Usually, a logarithmic link function is used for the claim severity. In particular, we want to achieve a multiplicative structure of premium predictions for light and moderate losses. This means, that for the given threshold u there exists a parameter vector $\alpha \in \mathbb{R}^{d+1}$ such that for all $r \in \mathbb{R}^d_{\geq 0}$

$$E\left(Y | Y \leq u, R = r\right) = \exp\left(\alpha_0 + \sum_{j=1}^{d} \alpha_j r_j\right). \tag{4.4}$$

Now we want to focus on the underlying distribution of the GLM. Commonly, the claim severity is modeled by a gamma distribution, as it is motivated in Section 2.1.2 in Ohlsson and Johansson (2010). Since we want to separate contaminations in the form of extreme claim sizes from the remaining data points, we use a *truncated* gamma distribution to model the conditional distribution function below the threshold. To formulate this assumption and to clarify the parametrization we work with, we briefly recall the definition of the gamma and the truncated gamma distribution:

Definition 4.2 (*Gamma Distributions*)**.** Let Γ and γ denote the complete and incomplete gamma function, respectively. For parameters $\alpha, \beta > 0$ and the parametrization[1] $\phi = 1/\alpha > 0$ and $\theta = -\beta/\alpha < 0$, we call a random variable X with density function

$$f_G\left(x; \phi, \theta\right) := \frac{\beta^\alpha x^{\alpha - 1} e^{-\beta x}}{\Gamma\left(\alpha\right)}, \quad x \geq 0, \tag{4.5}$$

gamma distributed. Further, we write $X \sim \mathcal{G}\left(\phi, \theta\right)$ with cumulative distribution function

$$F_G\left(x; \phi, \theta\right) := \frac{\gamma\left(\alpha, \beta x\right)}{\Gamma\left(\alpha\right)}, \quad x \geq 0, \tag{4.6}$$

and call the parameter ϕ the dispersion parameter. For a given threshold $u \in \mathbb{R}_{>0}$, a random variable X with density function

$$f_{TG}\left(x; \phi, \theta, u\right) := \frac{f_G\left(x; \phi, \theta\right)}{F_G\left(u; \phi, \theta\right)} 1_{(0,u]}\left(x\right), \quad x \geq 0, \tag{4.7}$$

is said to be distributed according to the truncated gamma distribution, and we denote this by $X \sim \mathcal{G}\left(\phi, \theta, u\right)$.

In line with Definition 4.2 we assume that for all $r \in \mathbb{R}^d_{\geq 0}$ we have

$$\left(Y | Y \leq u, R = r\right) \sim \mathcal{G}\left(\phi, \theta_r, u\right), \tag{4.8}$$

with $\phi > 0$ and $\theta_r < 0$.

Remark 4.3. Note that the scale parameter θ_r of the truncated gamma distribution depends on the specific tariff cell r.

[1] This parametrization is also in line with the parametrization of the exponential dispersion family with parameters $\phi > 0$ and $\theta < 0$ as e.g. given in Section 2.1 in Ohlsson and Johansson (2010) and Section 7.3 in Wüthrich (2017).

Classic gamma GLMs are based on the connection between the scale parameter and the expectation, which can be expressed by the derivative of the cumulant function from the exponential dispersion family. The cumulant function for the gamma distribution is given by $b(\theta) := -\ln(-\theta)$ with derivative $b'(\theta) = -1/\theta$. In an analogous way, we specify the cumulant function in the truncated gamma GLM by calculating the expectation of a truncated gamma distributed random variable:

Proposition 4.4. *For $\phi > 0, \theta < 0$ and $u \in \mathbb{R}_{>0}$ let $Z \sim \mathcal{G}(\phi, \theta, u)$. Then we have that*

$$E(Z) = -\frac{1}{\theta} + \frac{-u\left(-\frac{\theta u}{\phi}\right)^{\frac{1}{\phi}-1} e^{\frac{\theta u}{\phi}}}{\gamma\left(\frac{1}{\phi}, -\frac{\theta u}{\phi}\right)}. \tag{4.9}$$

Proof. Following the proof of Lemma 7.8 in Wüthrich (2017), the moment generating function M_Z of the truncated gamma distributed random variable Z is verified to be

$$\begin{aligned} M_Z(t) &= \int_0^u e^{tx} \frac{f_G(x;\phi,\theta)}{F_G(u;\phi,\theta)} dx \\ &= \exp\left(\frac{b(\theta+t\phi) - b(\theta)}{\phi}\right) \frac{F_G(u;\phi,\theta+t\phi)}{F_G(u;\phi,\theta)}. \end{aligned}$$

The expectation is then obtained by calculating $\frac{\mathrm{d}}{\mathrm{d}t} M_Z(t)\Big|_{t=0}$, where we have used that

$$\frac{\mathrm{d}}{\mathrm{d}t}\gamma\left(\frac{1}{\phi}, -\left(\frac{\theta}{\phi}+t\right)u\right) = -u\left(-\left(\frac{\theta}{\phi}+t\right)u\right)^{\frac{1}{\phi}-1} e^{\left(\frac{\theta}{\phi}+t\right)u}. \quad \square$$

Remark 4.5.

1. Thus, the expectation in (4.9) is represented by the derivative of the cumulant function of a gamma distributed random variable and an additional term, which originates from truncation at the given threshold u.
2. Note that the expectation in (4.9) as function of ϕ and θ is not bijective. However, we aim at a bijective connection between the conditional expectation and at least the scaling parameter θ in the sense of a classical gamma GLM. Hence, we need a prior estimation for the dispersion parameter. We do this with an approximation argument. Therefore, note that the truncated gamma distribution for a specific tariff cell r *converges uniformly* to a gamma distribution, i.e. it holds

$$\lim_{u\to\infty} \sup_{0<x<\infty} \left| \frac{H_r(x;\Theta_H)}{H_r(u;\Theta_H)} 1_{(0,u]}(x) + 1_{(u,\infty)}(x) - H_r(x;\Theta_H) \right| = 0. \tag{4.10}$$

This result justifies the approximation of the conditional distribution by the corresponding unconditional distribution for large thresholds, i.e. we are allowed to approximate the truncated gamma distribution by a gamma distribution given that the threshold is large enough. Since the threshold represents the separation between the basic claim sizes and the extreme events, we can suppose that threshold estimations are sufficiently large. Furthermore, we only aim at a suitable approximation for the dispersion parameter. Therefore, we suggest to use common estimators for a gamma GLM that do not take censoring into account. Due to the uniform convergence, for this estimation we can use all claim sizes below the *extreme* threshold. For possible dispersion estimators we refer to Section 3.1.1 in Ohlsson and Johansson (2010), Section 7.3.3 in Wüthrich (2017) and Section 2.2 in Garrido et al. (2016).

Proposition 4.4 is an essential ingredient for the estimation of the parameter vector α of the GLM in (4.4). For ease of notation we define the function $f : \mathbb{R}_{<0} \to \mathbb{R}$ by

$$\begin{aligned} f(\theta) &:= -\frac{1}{\theta} + \frac{-u\left(-\frac{\theta u}{\phi}\right)^{\frac{1}{\phi}-1} e^{\frac{\theta u}{\phi}}}{\gamma\left(\frac{1}{\phi}, -\frac{\theta u}{\phi}\right)} \\ &= \frac{\mathrm{d}}{\mathrm{d}\theta} \ln\left(-\frac{\left(\gamma\left(\frac{1}{\phi}, -\frac{\theta u}{\phi}\right)\right)^{\phi}}{\theta} \right), \end{aligned} \tag{4.11}$$

and suppress the dependence on the parameters ϕ and u in what follows.

Furthermore, it is a prerequisite for the maximum-likelihood estimation of the parameter vector that the function f given by (4.11) is injective. We address this in the following proposition whose proof is quite technical and is delegated to the Appendix:

Proposition 4.6. *The function f given by (4.11) for $\phi > 0, \theta < 0$ and $u \in \mathbb{R}_{>0}$ is continuous, strictly increasing and takes values in $\left(0, \frac{u}{1+\phi}\right)$. Therefore, $f : \mathbb{R}_{<0} \to \left(0, \frac{u}{1+\phi}\right)$ is bijective and has a continuously differentiable inverse.*

Thus, when using the data vectors $\mathbf{x}$ and $\mathbf{r}$ as in Section 3, the maximization of (3.1) has to be executed alongside the following steps:

1. Find an estimator $\hat{\phi}$ for the dispersion parameter as described in Remark 4.5.
2. Set $\Theta_H = \left(\alpha, \hat{\phi}\right)$ and
   $H_{r_i}(y;\Theta_H) = F_G\left(y;\hat{\phi}, f^{-1}\left(\exp\left(\alpha_0 + \sum_{j=1}^d \alpha_j r_{i,j}\right)\right)\right)$.
3. Find an estimator for α by maximizing (3.1) w.r.t. the linear constraints that for all $i \in \{1,\dots,m\}$ the following inequalities hold:

$$\alpha_0 + \sum_{j=1}^d \alpha_j r_{i,j} < \ln\left(\frac{u}{1+\hat{\phi}}\right). \tag{4.12}$$

Remark 4.7 (*Solving the Nonlinear Programming Problem*)**.**

1. Consider that the regressors are non-negative. Therefore, the admissible set of the constrained optimization problem is non-empty.
2. The inverse of the function f has to be determined numerically.
3. In the setting of (3.2), i.e. when all insured sums lie above the threshold, we can determine α by maximizing the following part of the log-likelihood function corresponding to the truncated gamma GLM and the *uncensored* claim sizes *below* the threshold:

$$\operatorname*{argmax}_{\alpha\in\mathbb{R}^{d+1}} \sum_{i\in S_1} \ln\left(f_{TG}\left(x_i;\hat{\phi}, f^{-1}\left(\exp\left(\alpha_0 + \sum_{j=1}^d \alpha_j r_{i,j}\right)\right)\right)\right), \tag{4.13}$$

w.r.t. the linear constraints (4.12). For this purpose we use an adaptive barrier algorithm, see e.g. Section 16.3 in Lange (2010) and Section 11.2 in Lange (2013).

4.3. Estimation above a given threshold using extreme value theory

In what follows, we want to examine the distribution function G_r in (2.4) more closely. Since extreme value data are rare, the

application of GLMs for G_r is problematic and we opt to model G_r using results from extreme value theory. To properly describe the distribution above a certain threshold, we introduce the excess distribution function:

Definition 4.8 (*Excess Distribution Function*)**.** Let Z be a random variable on $\mathbb{R}_{>0}$ with distribution function F. For $u \in \mathbb{R}_{>0}$ and for all $z \in \mathbb{R}_{>0}$ we call

$$F_u(z) := P(Z - u \leq z | Z > u) = \frac{F(u+z) - F(u)}{1 - F(u)}, \tag{4.14}$$

the excess distribution function of Z.

Following the POT approach (see e.g. Section 6.5 in Embrechts et al. (1997)), we want to determine the excess distribution function of Y. Therefore, we recall the definition of a GPD:

Definition 4.9 (*GPD*)**.** For $\xi, u \in \mathbb{R}$ and $\beta \in \mathbb{R}_{>0}$ we define the distribution function $G_{\xi,\beta;u}$ by

$$G_{\xi,\beta;u}(x) = \begin{cases} 1 - \left(1 + \xi \frac{x-u}{\beta}\right)^{-\frac{1}{\xi}} & , \xi \neq 0, \\ 1 - e^{-\frac{x-u}{\beta}} & , \xi = 0, \end{cases} \tag{4.15}$$

where $x \geq u$ if $\xi \geq 0$ and $x \in \left[u, u - \frac{\beta}{\xi}\right]$ if $\xi < 0$. $G_{\xi,\beta;u}$ is called a generalized Pareto distribution (GPD). We denote the density of a GPD by $g_{\xi,\beta;u}$ and set $G_{\xi,\beta} := G_{\xi,\beta;0}$.

The Theorem of Pickands, Balkema and de Haan (see e.g. Theorem 3.4.13 (b) in Embrechts et al. (1997)) states that for a large class of underlying distributions,[2] there exists a positive function $\beta(u)$ such that

$$\lim_{u \uparrow x_F} \sup_{0 < x < x_F - u} \left| F_u(x) - G_{\xi,\beta(u)}(x) \right| = 0, \tag{4.16}$$

where $G_{\xi,\beta(u)}$ is the generalized Pareto distribution as given in Definition 4.9, and x_F is the right endpoint of F. The POT method uses this *uniform convergence* to justify the approximation of the excess distribution function by a GPD above a given, large threshold u. For an application of this methodology see e.g. Section 2.2 and 5.2 in Singh et al. (2013). We follow this procedure and approximate the excess distribution function of the actual damage Y by the GPD. In line with the arguments of Remark 2.3, the following reasons justify that we not only use the same threshold, but also the same GPD independent of the tariff cell, especially independent of the insured sum:

- Extreme claim sizes can be considered as accidents and the choice of the insured sum only reflects the personal assessment of the policyholder about his risk potential. Of course, this personal assessment can also be a misjudgment. Consider e.g. a typical German personal liability insurance with insured sums higher than 5 million Euro. These amounts are beyond perception of one's daily life, such that an underestimation of the risk potential seems possible. Therefore, in this regard, extreme damages are independent of the insured sum.
- The usage of the same GPD can handle a sparse number of contaminations in the form of outliers in the data. This leads to a single surcharge for extreme actual damages. Limiting the actual damage with the insured sum then leads to proportionately fairly distributed premium supplements.
- Lastly, the GPD is also known to be a useful risk management tool for insurance companies (compare e.g. Embrechts et al., 1999). Hence, our procedure enables the insurer to use the same GPD for a *consistent* pricing and risk management.

[2] More explicitly the distribution function is an element of the maximum domain of attraction of the generalized extreme value distribution, see e.g. Section 3.3 and 3.4 in Embrechts et al. (1997).

Remark 4.10. Nevertheless, there are several possible ways to incorporate additional tariff features in the estimation above the threshold u. Among them are the application of multivariate peaks-over-thresholds methods along the lines of e.g. Kiriliouk et al. (2019) and Rootzén et al. (2018) or the design of suitable shape-scale regression models for the GPD in the spirit of e.g. Pupashenko et al. (2015). However, a thorough investigation of these methods would go far beyond the scope of this paper, and we thus leave this delicate issue for future research.

Following the previous explanations, we approximate the excess distribution functions of the tariff cells by a single GPD, i.e. $\Theta_G = (\xi, \beta)$ and $G_r(y; \Theta_G) = G_{\xi,\beta;u}(y) = G_{\xi,\beta}(y - u)$. Accordingly, we have to maximize the following part of the log-likelihood function (3.1):

$$\underset{\xi \in \mathbb{R}, \beta > 0}{\operatorname{argmax}} \left(\sum_{i \in S_2} \ln\left(g_{\xi,\beta;u}(x_i)\right) + \sum_{i \in S_3} \ln\left(1 - G_{\xi,\beta;u}(r_{i,I})\right) \right) \tag{4.17}$$

Remark 4.11. For the simulation study in Section 5 we maximize (4.17) with the Nelder–Mead method. We determine the starting values for this optimization algorithm by the probability weighted moment approach as described in Section 6.5 in Embrechts et al. (1997).

4.4. Estimator for the expected claim size

After having found an estimator $\hat{\Theta}$ for the parameter vector Θ, we are able to specify the estimator for the expected claim size

$$\begin{aligned} E(X|R=r) = {} & P(X \leq u | R = r)\, E(X | X \leq u, R = r) \\ & + P(X > u | R = r)\, E(X | X > u, R = r)\,. \end{aligned} \tag{4.18}$$

To accomplish this, we have to distinguish if the insured sum is below or above the threshold, because the actual damage Y can be significantly larger than the insured sum r_I, i.e. there are scenarios $\omega \in \Omega$ such that $X(\omega) = r_I < Y(\omega)$. We thus obtain the following form of our estimator:

$$E_{\hat{\Theta}}(X|R=r) = \begin{cases} \left(1 - q_r\left(\hat{\delta}\right)\right) \exp\left(\hat{\alpha}_0 + \sum_{j=1}^{d} \hat{\alpha}_j r_j\right) & \\ \quad + q_r\left(\hat{\delta}\right) \left(\int_u^{r_I} y\, g_{\xi,\beta;u}(y)\, dy \right. & \\ \quad \left. + r_I \left(1 - G_{\xi,\beta;u}(r_I)\right) \right) & , r_I \geq u, \\ \left(1 - q_r\left(\hat{\delta}\right)\right) \int_0^{r_I} y \dfrac{f_G\left(y; \hat{\phi}, g\left(\hat{\alpha}, r\right)\right)}{F_G\left(u; \hat{\phi}, g\left(\hat{\alpha}, r\right)\right)} dy & \\ \quad + r_I \left(1 - \left(1 - q_r\left(\hat{\delta}\right)\right) \dfrac{F_G\left(r_I; \hat{\phi}, g\left(\hat{\alpha}, r\right)\right)}{F_G\left(u; \hat{\phi}, g\left(\hat{\alpha}, r\right)\right)}\right) & , r_I < u, \end{cases} \tag{4.19}$$

where

$$g\left(\hat{\alpha}, r\right) := f^{-1}\left(\exp\left(\hat{\alpha}_0 + \sum_{j=1}^{d} \hat{\alpha}_j r_j\right)\right). \tag{4.20}$$

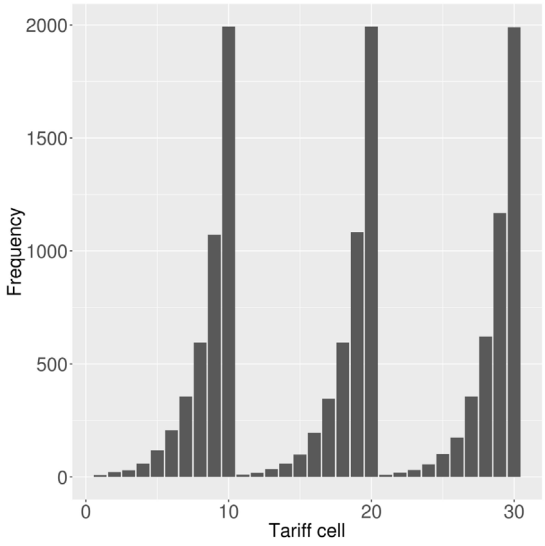

Fig. 2. Histogram of simulated claim sizes.

Remark 4.12.

1. We observe that in the case $r_I \geq u$ we can easily calculate the conditional expectation below the threshold in the sense of a classical GLM. Otherwise, we have to solve the above integral numerically.
2. Note that, e.g. for personal liability insurances, the censoring point (insured sum) can depend on the kind of damage. In this case, there are more than one censoring point within an insurance contract possible. The insurer can use the distribution of Y as approximation for the distribution of X. This is e.g. the case if the insurer determines the premium using the widespread gamma GLM, see among others Section 2.1.2 in Ohlsson and Johansson (2010) and Section 7.3.2 in Wüthrich (2017).

5. Simulation study

In this section, we show that the threshold severity model outperforms the classical gamma GLM when fitting to simulated claim sizes from other regression models. To this end, in Section 5.1 we present the simulated claims corresponding to these models. First of all, for the sake of completeness, in Section 5.1.1, we simulate claims from the TSM (*Scenario 1*). Then, in Section 5.1.2 and Section 5.1.3, we use heavy-tailed regression models based on the log-normal (*Scenario 2*) and Burr Type XII (*Scenario 3*) distributions to generate claim sizes. Afterwards, in Sections 5.2.1, 5.2.2 and 5.2.3, we present and compare the predictions stemming from the gamma GLM and the TSM w.r.t. these three scenarios. In Section 5.3, we use the sample from Section 5.1.1 to demonstrate the effect of changing the threshold when fitting the TSM. All simulations and calculations are performed with the help of R version 3.5.2 (The R Foundation, 2018).

5.1. Simulation of claims

In what follows, for the sake of brevity, we set the index of the insured sum to 1 and denote it by $v\,(= r_1 = r_I)$. Inspired by typical German personal liability insurances, we use insured sums of 5 million, 20 million and 50 million. Furthermore, we allow for a second tariff feature taking integer values from 1 to 10. This corresponds to e.g. a classification based on the mileage or on the car's power. This additional tariff feature is denoted by $w\,(= r_2)$. This results in a total number of 30 tariff cells. To reflect these tariff cells in the following illustrations, we display all values of w clustered w.r.t. the ascendingly ordered possible insured sums, i.e. we use the lexicographic order for (v, w).

Table 1
Excess probabilities for the simulation w.r.t. insured sums of 5, 20, 50 million.

	5 million	20 million	50 million
Excess probability	0.64%	0.93%	1.96%

Note, that the claim frequency of observed claims can vary significantly w.r.t. the tariff cells. Hence, we aim at a variation of the expected claim frequency between the tariff cells from round about 10 up to 2000 claims. Further, we assume that the claim frequency is independent from the insured sum v. To achieve the desired claim frequencies, we use a GLM based on a Poisson distribution. By trial and error we then end up with the following mean for the GLM:

$$E\,(N|w) \approx e^{1.714+0.589w}. \tag{5.1}$$

This leads to 13410 simulated claims. Fig. 2 presents the histogram of claims w.r.t. the tariff cells.

5.1.1. Scenario 1

For the simulation of claim sizes from the TSM we use the following parameters:

$$u = 10^6, \delta_0 = \ln(1/175), \delta_1 = 2.5 \times 10^{-8}, \delta_2 = 0, \phi = 0.5,$$
$$\alpha_0 = 10, \alpha_1 = 0, \alpha_2 = 1/5, \xi = 0.4, \beta = 2\,400\,000. \tag{5.2}$$

Note, that for the logit-model and the truncated gamma GLM we assume that only one of the tariff features v and w has an influence on the expectation. We incorporate this fact by setting the other regression parameter to zero. In line with the explanations in Section 4.1, we only incorporate the insured sum for the logit-model. Since we model extreme claim sizes, only a few claims in a tariff cell should be larger than the threshold u of 1 million. On the other hand, we aim at significant differences in the excess probabilities w.r.t. the insured sum v. Hence, the intercept and the regression parameter of v are chosen such that we end up with the excess probabilities as displayed in Table 1.

Note as well that the threshold u of 1 million is significantly smaller than the possible insured sums. Thus, the insured sum especially influences extreme claim sizes. Hence, we let only the second tariff feature w influence the expectation of small and moderate claims. As a result, the corresponding conditional expectation from the truncated gamma GLM is given by

$$E\,(X|X \leq u, R_2 = w) = e^{10+\frac{1}{5}w}. \tag{5.3}$$

As GPD parameters we use a shape of $\xi = 0.4$ and a scale of $\beta = 2\,400\,000$. Without limitation at the insured sum, this results in an expectation of 5 million, which is significantly larger than the threshold of 1 million. Together with the fact that the considered GPD has a standard deviation of roughly 9 million, this implies that an insured sum of 5 million is exceeded with a considerable probability. Thus, the data for calibration of the standard GLM will be contaminated by a notable amount of extremes.

Fig. 3 displays the structure of the simulated claims. In this particular case, 168 claim sizes lie above the threshold. Moreover, the mean excess plot (see e.g. Section 6.2.2 in Embrechts et al. (1997)) in Fig. 3 illustrates the breaking point of the splicing distribution at the threshold of 1 million.

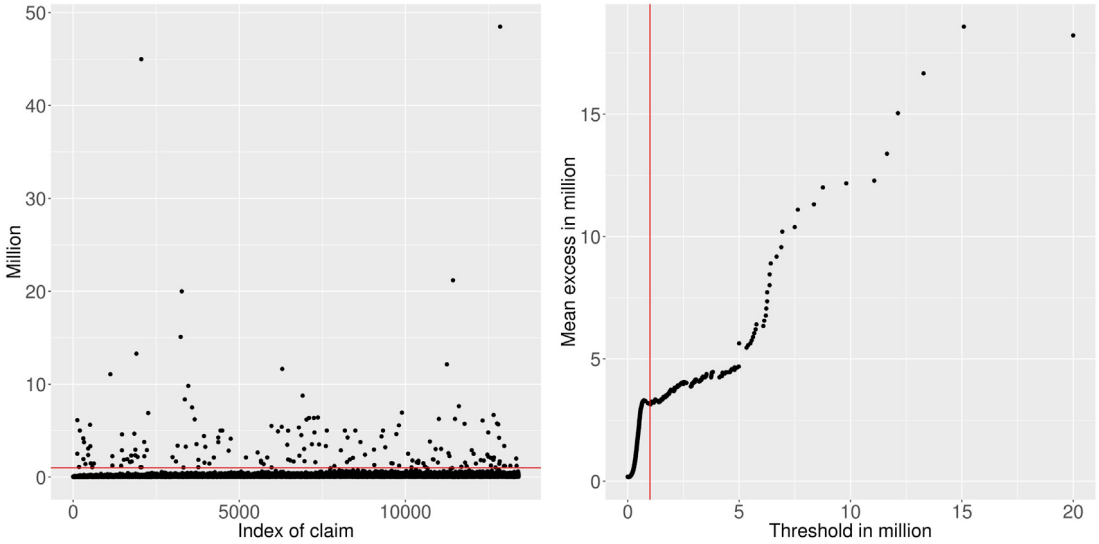

Fig. 3. Simulated claim sizes (left) and mean excess plot of the simulated claims (right). The red lines correspond to the threshold of 1 million. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5.1.2. Scenario 2

For the second scenario we simulate claims from a heavy-tailed log-normal regression. For a tariff cell r, a claim size is simulated as follows:

1. Simulate a normal random variable $Z \sim N(\mu, \sigma)$ with mean $\mu = \alpha_0 + \alpha_1 v + \alpha_1 w$ and standard deviation $\sigma > 0$.
2. Obtain the log-normal random variable by $X = e^Z$.

In order to obtain a significant influence of the insured sum, we use the following parameters in this scenario:

$$\alpha_0 = 5.5, \alpha_1 = 4 \times 10^{-8}, \alpha_2 = 0.02, \sigma = 2.75. \tag{5.4}$$

The simulated claim sizes are illustrated in Fig. 4. To stress the calibration of the threshold severity model we aim at a sufficient number of claims close to the insured sums. This is also reflected by the plot.

The first step in order to fit the threshold severity model is the choice of the threshold u. Using techniques from extreme value theory, we consider the mean excess plot in Fig. 4. To obtain a good approximation via the GPD, it is recommended to choose a threshold such that the mean excess plot is *roughly* linear for higher values. Following this criterion, there are several possible choices for the threshold. One choice would be at the kink of 2.25 million. This would only lead to 42 exceedances. In order to obtain a bit more data points above the threshold we use a lower value. Additionally, to obtain a *roughly* linear behavior above the threshold (disregard extreme outliers), we decided to choose a threshold of 1.55 million. This leads to 59 exceedances.

5.1.3. Scenario 3

For the last scenario we simulate claim sizes from a Burr Type XII distribution, i.e. we focus on random variables with density function

$$f_B(y; \beta, \lambda, \tau) = \frac{\lambda \beta^\lambda \tau y^{\tau-1}}{(\beta + y^\tau)^{\lambda+1}}, \quad y > 0, \ \beta, \lambda, \tau > 0. \tag{5.5}$$

A random variable Y with such a density function is denoted by $Y \sim \text{Burr}(\beta, \lambda, \tau)$. To incorporate tariff cells, we follow Beirlant et al. (1998) and use a regression for the parameter β. Hence, we obtain the conditional distribution

$$(Y|R = r) \sim \text{Burr}(\beta(r), \lambda, \tau), \tag{5.6}$$

where $\beta(r) := \exp(\tau(\alpha_0 + \alpha_1 v + \alpha_1 w))$. For this Burr regression we choose the following parameter values:

$$\alpha_0 = 8, \alpha_1 = 4 \times 10^{-8}, \alpha_2 = 0.02, \lambda = 1.5, \tau = 0.7. \tag{5.7}$$

The motivation for the parameter values of λ and τ is as follows: As described in Section 2.1.2 in Ohlsson and Johansson (2010), a typical distribution of claim sizes is right-skewed. To obtain this, we choose $\lambda = 1.5$ (compare Section 10.4.1 in Shi (2014)). Additionally, to obtain a heavy-tailed distribution, we choose $\tau = 0.7$. Note that this leads to a distribution, for which only the first moment exists (see also Section 10.4.1 in Shi (2014)). The simulated claim sizes are shown in Fig. 5. Using as before the mean excess plot, we choose a threshold at the kink of 1.75 million, which results in 79 exceedances.

5.2. Results

In order to evaluate the performance of the threshold severity model, we fit the gamma GLM and the TSM to the simulated data from Section 5.1.

In the following, we denote the true mean, the predictive mean, and the number of claims of the ith tariff cell by μ_i, $\hat{\mu}_i$, and m_i, respectively. Moreover, the total number of claims is given by m. The vectors of observed claims and corresponding estimated means below the threshold are $\mathbf{x}^{\leq u}$ and $\hat{\mu}^{\leq u}$ and the number of claims less or equal the threshold u is given by $m^{\leq u}$. Finally, $r^{\leq u}$ denotes the number of estimated regression parameters of the truncated gamma GLM.

Calibration of the gamma GLM: For the gamma GLM we always use the insured sum v and the second tariff feature w as regressors. In this context, the regression parameters of the intercept, of the insured sum, and the second tariff feature are denoted by α_0^G, α_1^G, and α_2^G, respectively. When fitting the gamma GLM to the censored data, we denote the dispersion parameter of the gamma GLM by ϕ^G. Thus, our aim is to maximize the following log-likelihood function:

$$\begin{aligned} l(\alpha^G, \phi^G; \mathbf{x}, \mathbf{r}) = &\sum_{i \in S_1 \cup S_2} \ln\left(f_G\left(x_i; \phi^G, \alpha_0^G + \alpha_1^G v_i + \alpha_2^G w_i\right)\right) \\ &+ \sum_{i \in S_3} \ln\left(1 - F_G\left(v_i; \phi^G, \right.\right. \\ &\left.\left. \exp\left(\alpha_0^G + \alpha_1^G v_i + \alpha_2^G w_i\right)\right)\right). \end{aligned} \tag{5.8}$$

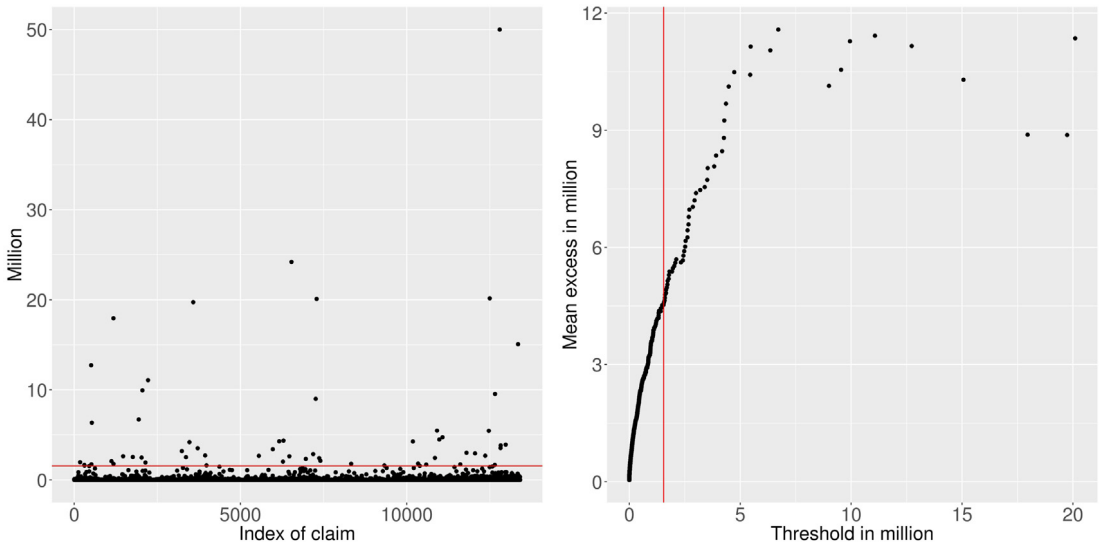

Fig. 4. Simulated claim sizes (left) and mean excess plot of the simulated claims (right) w.r.t. the log-normal regression. The red lines correspond to a threshold of 1.55 million. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

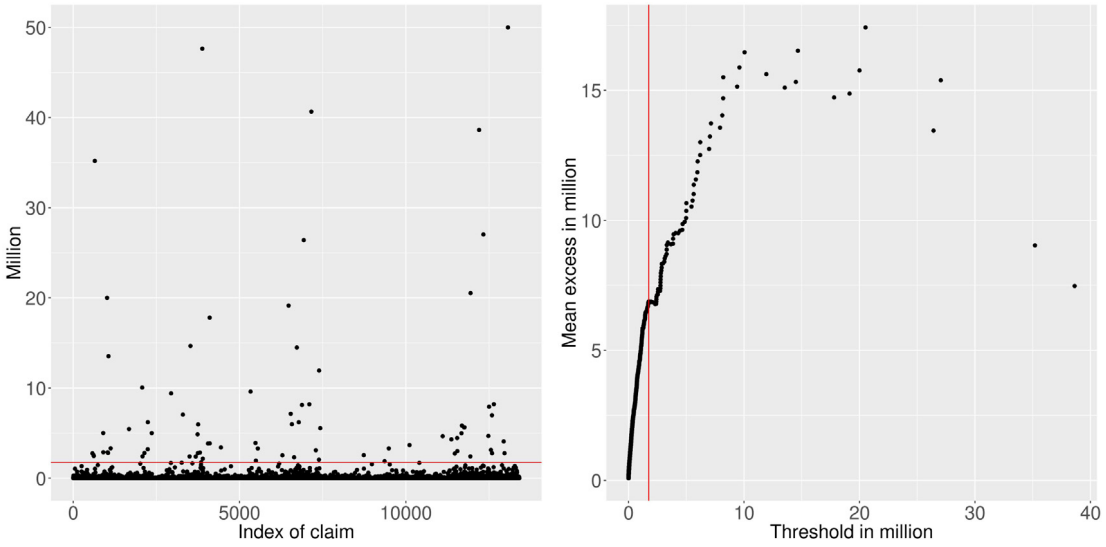

Fig. 5. Simulated claim sizes (left) and mean excess plot of the simulated claims (right) w.r.t. the Burr regression. The red lines correspond to a threshold of 1.75 million. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

To achieve this, we use the `optim` function of the `stats`-package with the default *Nelder–Mead* method.

Calibration of the TSM: A crucial point for the calibration of the truncated gamma GLM is the approximation of the dispersion parameter ϕ. As described in Remark 4.5, we use all claim sizes below the threshold to fit a standard gamma GLM. For this we exploit the `glm.fit` function. Within this gamma GLM, several possible dispersion estimators are available. In our simulation study we compare the following three standard estimators (see Section 3.1.1 in Ohlsson and Johansson (2010)):

- **Pearson's chi-square statistic**: Given the variance function $v(\mu) = \mu^2$ of the standard gamma GLM we obtain this estimator by:

$$\hat{\phi}_P := \frac{1}{m^{\leq u} - r^{\leq u}} \sum_{i \in S_1 \cup S_2} \frac{\left(x_i - \hat{\mu}_i\right)^2}{v\left(\hat{\mu}_i\right)}. \tag{5.9}$$

- **Deviance statistic**: We obtain this statistic from the (unscaled) deviance statistic $D(\cdot, \cdot)$ (see Section 3.1 in Ohlsson and Johansson (2010)) by

$$\hat{\phi}_D := \frac{D\left(\mathbf{x}^{\leq u}, \hat{\mu}^{\leq u}\right)}{m^{\leq u} - r^{\leq u}}. \tag{5.10}$$

- **ML-estimator**: Given the regression parameters from the output of the `glm.fit` function, we maximize the corresponding log-likelihood function w.r.t. the dispersion parameter. We denote this estimator by $\hat{\phi}_{ML}$.

The optimization of the truncated gamma GLM is performed using the function `constrOptim` of the `stats`-package with the default *Nelder–Mead* method. To ensure a robust calibration, we use different initial values, chosen from a neighborhood of the outputs of the `glm.fit` function. Furthermore, we fit the GPD parameters with the classical `optim` function which also stems from the `stats`-package. The initial values are determined by the `gpd` method of the `evir`-package.

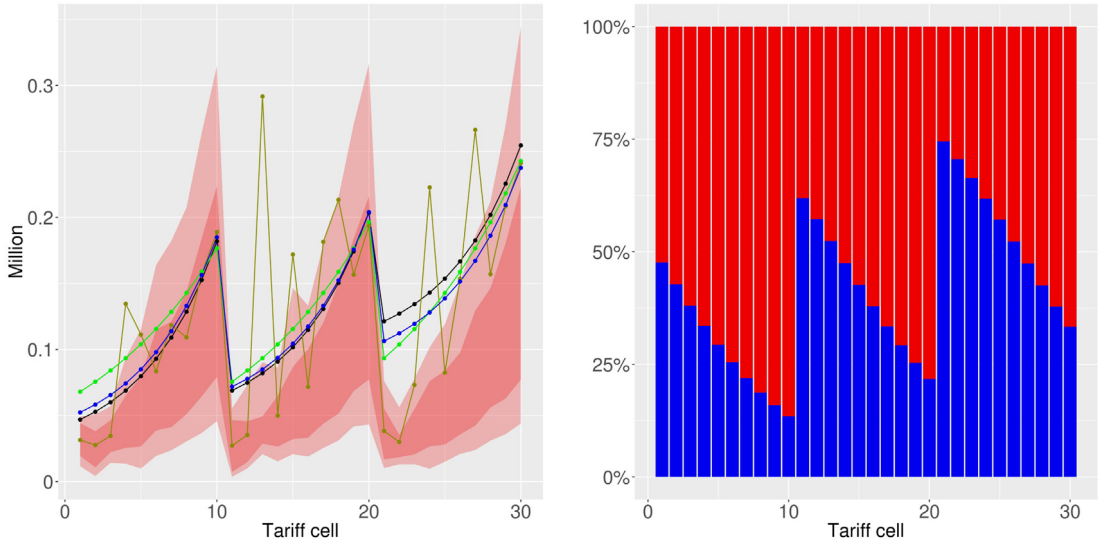

Fig. 6. Left: True mean (black) and empirical mean (yellow) of the simulated data and predictions stemming from the standard gamma GLM (green) and the TSM (blue). The different shades of red represent the intervals from the 25% to the 75% and from the 10% to the 90% sample quantile. Right: Proportions between the predictions from the truncated gamma GLM (red) and the predictions from the GPD (blue) in the TSM. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Observed statistics: In what follows, one of our main priorities is to quantify the relative deviation between the true and predictive mean of a specific tariff cell. Therefore, we calculate (weighted) averages of the relative differences for every scenario w.r.t. all tariff cells. These statistics are given by

$$\bar{z}_1 := \frac{1}{30}\sum_{i=1}^{30}\frac{\left|\hat{\mu}_i - \mu_i\right|}{\mu_i}, \qquad \bar{z}_2 := \sum_{i=1}^{30}\frac{m_i}{m}\frac{\left|\hat{\mu}_i - \mu_i\right|}{\mu_i}. \tag{5.11}$$

Moreover, we calculate the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). Here, the AIC and the BIC are based on 4 estimated parameters for the gamma GLM, 7 estimated parameters for the TSM in Section 5.2.1, and 8 estimated parameters for the TSM in Section 5.2.2 and Section 5.2.3.

5.2.1. Results for scenario 1

We use the true threshold of 1 million to fit the threshold severity model as given in Scenario 1, i.e. we use the true model to fit the simulated data in this scenario. We perform this assessment for the sake of completeness and for demonstrating the inappropriateness of the gamma GLM in the presence of extreme claim sizes. In that regard, Section 5.3 also investigates the dependence of the fitted parameters on the chosen threshold.

The statistics in Table 2 w.r.t. the different dispersion estimators for the TSM indicate that the ML-estimator yields the best log-likelihood value for the TSM. Moreover, all possible dispersion estimators lead to better distributional fits than the gamma GLM. Since our chosen optimization criterion is the log-likelihood function, we calibrate the TSM w.r.t. the ML-dispersion estimator in what follows. Hence, using (5.11), for the threshold severity model the average of the relative differences between the true and predictive means given by $\bar{z}_1$ is 8.65% lower. Using $\bar{z}_2$, the improvement is only 1.92%. Additionally, compared to the gamma GLM, the TSM shows a decrease of 2.69% for the AIC as well as 2.68% for the BIC.

Table 2
Log-likelihood function, AIC, BIC, $\bar{z}_1$ and $\bar{z}_2$ for the predictions of the gamma GLM and the TSM.

	Log-likelihood	AIC	BIC	$\bar{z}_1$	$\bar{z}_2$
Gamma GLM	−175 491.54	350 991.09	351 021.10	14.69%	5.63%
TSM $\left(\hat{\phi}_P \approx 0.500\right)$	−170 773.39	341 560.78	341 613.31	6.04%	3.71%
TSM $\left(\hat{\phi}_D \approx 0.542\right)$	−170 796.33	341 606.66	341 659.18	6.04%	3.71%
TSM $\left(\hat{\phi}_{ML} \approx 0.501\right)$	−170 773.38	341 560.76	341 613.29	6.04%	3.71%

We obtain the following parameter estimates for the TSM w.r.t. $\hat{\phi}_{ML}$:

$$\begin{aligned}&\hat{u} = 1 \times 10^6, \hat{\delta}_0 = -4.863, \hat{\delta}_1 = 1.758 \times 10^{-8},\\ &\hat{\delta}_2 = 0, \hat{\phi}_{ML} = 0.501,\\ &\hat{\alpha}_0 = 10.031, \hat{\alpha}_1 = 0, \hat{\alpha}_2 = 0.196, \hat{\xi} = 0.404, \hat{\beta} = 2050787.\end{aligned} \tag{5.12}$$

The parameter estimates for the gamma GLM are as follows:

$$\hat{\alpha}_0^G = 10.985, \hat{\alpha}_1^G = 7.059 \times 10^{-9}, \hat{\alpha}_2^G = 0.106, \hat{\phi}^G = 0.990. \tag{5.13}$$

In Fig. 6 (left) we present the empirical and true mean of the simulated data. The intervals between the sample quantiles in Fig. 6 (left) show that the true mean of the claim severity of a specific tariff cell is significantly larger than the corresponding sample median. This is due to the fact, that the GPD is the excess distribution function of the underlying data. We deduce from Fig. 6 that the predictions from the threshold severity model in most cases are closer to the true mean than the predictions of the gamma GLM. Taking the previous values for the statistics $\bar{z}_1$ and $\bar{z}_2$ into account, this is an expected behavior. More interestingly, we observe that in the case of an insured sum of 50 million the slope of the predictions of the gamma GLM in the direction of the additional tariff feature w is too high. On the other hand, in the case of an insured sum of 5 million, we observe a slope which is too weak.

Using the true threshold, Fig. 6 (right) illustrates the proportion between the predictions for the expected claim severity from

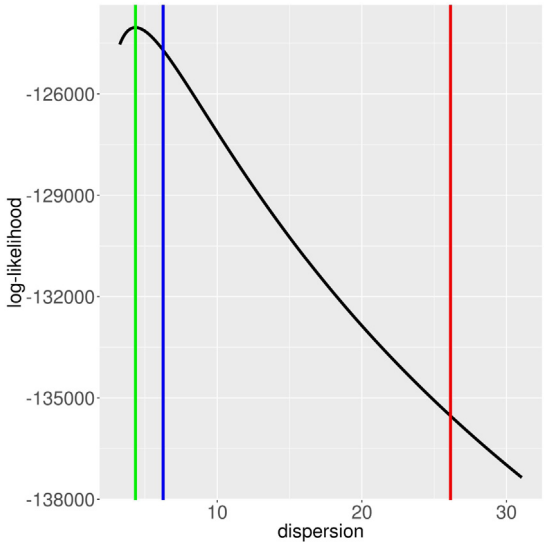

Fig. 7. Log-likelihood in the gamma GLM for fitting the dispersion of the TSM in dependence of the dispersion parameter; $\hat{\phi}_{ML}$ (green), $\hat{\phi}_D$ (blue), $\hat{\phi}_P$ (red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the truncated gamma GLM and the predictions for the expected claim severity from the GPD in the TSM, taking the excess probability into account. More specifically, for each tariff cell, the estimated values of $P(X \leq u|R_1 = v)\, E(X|X \leq u, R_2 = w)$ and $P(X > u|R_1 = v)\, E(X|X > u, R_1 = v)$ are presented as a proportion of the corresponding total sum. Remember that for the TSM the second tariff feature w is only contained in the truncated gamma GLM. Thereby, we deduce from Fig. 6 (right) that the truncated gamma GLM explains the positive impact of the predictor w on the conditional expectation for each insured sum. Furthermore, some predictions are mostly determined by the GPD (model above the threshold).

5.2.2. Results for scenario 2

We fit the threshold severity model within the same setting as in Section 5.2.1 to the sample from Scenario 2, i.e. we only use the second tariff feature w as regressor for the truncated gamma GLM. In this setting, we obtain values of 44.57% and 42.86% for $\bar{z}_1$ and $\bar{z}_2$ (w.r.t. the best TSM in terms of the log-likelihood value given for $\hat{\phi}_{ML} \approx 4.470$). This undesirable result is caused by the fact that for the log-normal regression both tariff features influence the whole support of the distribution, i.e. small and moderate as well as extreme claim sizes depend on the insured sum v and the second tariff feature w.

Hence, in the following we use the insured sum v *as well as* the second tariff feature w as regressors for the truncated gamma GLM. In Table 3 we display the same statistics as in the previous section. As before, for the TSM we obtain the best overall distributional fit using the ML-dispersion estimator. In contrast to Scenario 1, the log-likelihood for the different dispersion estimators as well as the dispersion estimators themselves, vary significantly. For instance, Pearson's chi-square estimator $\hat{\phi}_P$ is significantly larger than $\hat{\phi}_{ML}$. Moreover, $\hat{\phi}_P$ leads to a significantly smaller log-likelihood value than $\hat{\phi}_{ML}$. This can be explained as follows: Recall that $\hat{\phi}_{ML}$ maximizes the log-likelihood of the used (approximating) gamma GLM w.r.t. the data below the threshold. This is illustrated in Fig. 7. In addition, we depict from this illustration that $\hat{\phi}_P$ is far away from this optimal value, which implies that the log-likelihood of the gamma GLM is noticeably smaller. Since the gamma GLM can be regarded as a limiting case of the truncated gamma GLM, $\hat{\phi}_P$ also leads to an undesirable value for the log-likelihood of the truncated gamma GLM of the TSM. In contrast, $\hat{\phi}_{ML}$ maximizes the log-likelihood of the gamma GLM and is thus also a good approximation for the dispersion parameter in terms of the log-likelihood of the TSM. This example illustrates that the log-likelihood of the TSM can significantly depend on the dispersion estimator.

On the other hand, $\hat{\phi}_P$ leads to better values for $\bar{z}_1$ and $\bar{z}_2$ than $\hat{\phi}_{ML}$. This means that the predictions of $\hat{\phi}_P$ (on average) are closer to the true values. Hence, we recommend to test different dispersion estimators and to use the one which best suits the desired requirements. As our main concern at this point is a good

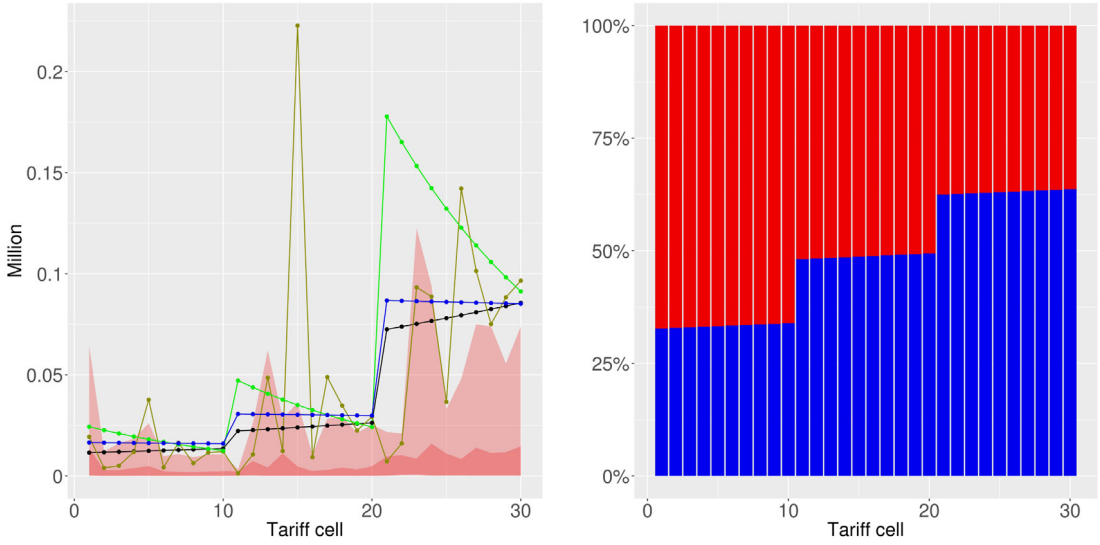

Fig. 8. Left: True mean (black) and empirical mean (yellow) of the simulated data and predictions stemming from the standard gamma GLM (green) and the TSM (blue) w.r.t. the log-normal regression. The different shades of red represent the intervals from the 25% to the 75% sample quantile and the 10% to the 90% sample quantile. Right: Proportions between the predictions from the truncated gamma GLM (red) and the predictions from the GPD (blue) in the TSM w.r.t. the log-normal regression. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

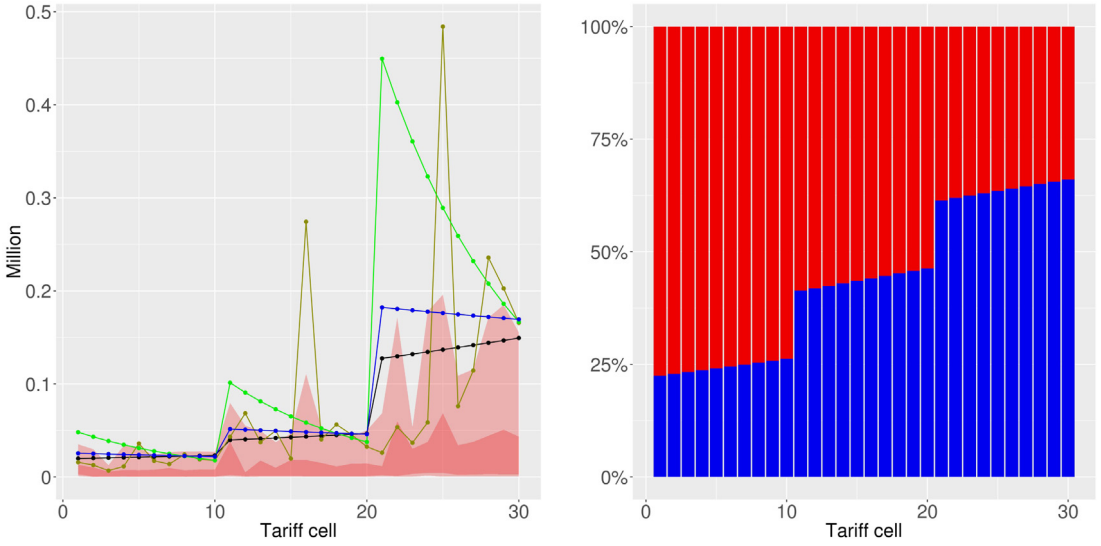

Fig. 9. Left: True mean (black) and empirical mean (yellow) of the simulated data and predictions stemming from the standard gamma GLM (green) and the TSM (blue) w.r.t. the Burr regression. The different shades of red represent the intervals from the 25% to the 75% and the 10% to the 90% sample quantile. Right: Proportions between the predictions from the truncated gamma GLM (red) and the predictions from the GPD (blue) in the TSM w.r.t. the Burr regression. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
Log-likelihood function, AIC, BIC, $\bar{z}_1$ and $\bar{z}_2$ for the predictions of the gamma GLM and the TSM.

	Log-likelihood	AIC	BIC	$\bar{z}_1$	$\bar{z}_2$
Gamma GLM	−126 838.62	253 685.24	253 715.25	53.31%	14.58%
TSM $\left(\hat{\phi}_P \approx 26.129\right)$	−136 814.10	273 644.19	273 704.22	20.42%	13.33%
TSM $\left(\hat{\phi}_D \approx 6.261\right)$	−126 002.48	252 020.96	252 080.99	21.67%	13.35%
TSM $\left(\hat{\phi}_{ML} \approx 4.347\right)$	−125 332.80	250 681.61	250 741.64	21.67%	13.35%

distributional fit, we focus in the following on the calibration of the TSM corresponding to $\hat{\phi}_{ML}$.

The more robust predictions of the threshold severity model are further substantiated by a decrease of $\bar{z}_1$ by 31.64%. This means that the threshold severity model estimates (on average) the true expectation over all tariff cells significantly better than the gamma GLM. Hence, the threshold severity model is more robust against extreme claim sizes of log-normal distributed type. We also obtain a decrease of 1.23% for $\bar{z}_2$. Hence, we end up with a slight improvement for the weighted differences.

The parameters of the TSM w.r.t. $\hat{\phi}_{ML}$ are given by

$$\begin{aligned} &\hat{u} = 1.55 \times 10^6, \hat{\delta}_0 = -6.604, \hat{\delta}_1 = 3.725 \times 10^{-8}, \\ &\hat{\delta}_2 = 0, \hat{\phi}_{ML} = 4.347, \\ &\hat{\alpha}_0 = 9.201, \hat{\alpha}_1 = 2.407 \times 10^{-8}, \hat{\alpha}_2 = -0.006, \\ &\hat{\xi} = 0.889, \hat{\beta} = 1524561. \end{aligned} \tag{5.14}$$

The parameter estimates for the gamma GLM are

$$\hat{\alpha}_0^G = 9.952, \hat{\alpha}_1^G = 4.421 \times 10^{-8}, \hat{\alpha}_2^G = -0.074, \hat{\phi}^G = 5.082. \tag{5.15}$$

Fig. 8 (left) demonstrates the inferior predictions of the gamma GLM in comparison to the TSM. The deviations from the true expectations are particularly pronounced for the insured sum of 50 million. It appears that for each insured sum the gamma GLM explains a decrease w.r.t. the second tariff feature w. This can be explained by the different number of claims for the tariff cells. The gamma GLM predicts the expectations for tariff cells with a huge number of claims much better than for other tariff cells. In contrast, the TSM is nearly unaffected by the second tariff feature w.

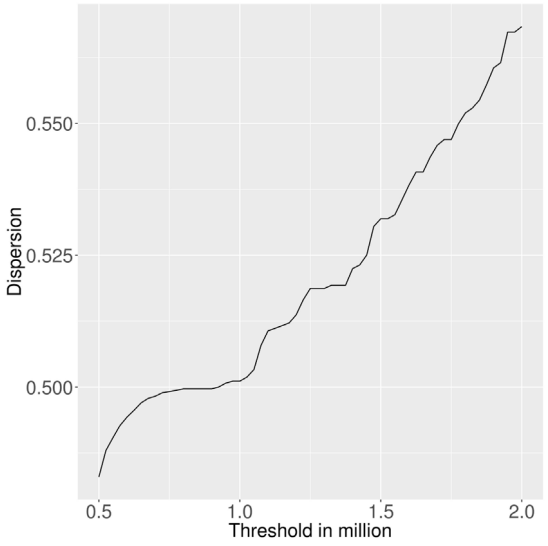

Fig. 10. Estimates for the dispersion parameter w.r.t. the threshold.

This also explains the larger reduction of $\bar{z}_1$ in comparison to $\bar{z}_2$. The better fit of the gamma GLM for tariff cells which contain larger numbers of simulated claims is more strongly represented in the (weighted) $\bar{z}_2$-statistic than in $\bar{z}_1$. Hence, we can conclude that – additionally to a better distributional fit – the TSM leads to an improvement which stems from better predictions for tariff cells with a small number of claims.

For the sake of completeness, Fig. 8 (right) contains the same plot of the proportions as in the previous section. Here, we

Table 4
Log-likelihood function, AIC, BIC, $\bar{z}_1$ and $\bar{z}_2$ for the predictions of the gamma GLM and the TSM.

	Log-likelihood	AIC	BIC	$\bar{z}_1$	$\bar{z}_2$
Gamma GLM	−145 235.16	290 478.31	290 508.33	74.82%	23.51%
TSM $\left(\hat{\phi}_P \approx 13.313\right)$	−151 304.90	302 625.79	302 685.82	16.53%	8.24%
TSM $\left(\hat{\phi}_D \approx 4.074\right)$	−143 658.35	287 332.71	287 392.74	17.78%	8.59%
TSM $\left(\hat{\phi}_{ML} \approx 3.004\right)$	−143 208.98	286 433.96	286 493.99	17.78%	8.59%

deduce that the tariff feature w has no significant influence on the predictions of the TSM.

5.2.3. Results for scenario 3

As in Section 5.2.2, if we only use the second tariff feature w as regressor for the truncated gamma GLM, we obtain 40.26% and 32.47% for $\bar{z}_1$ and $\bar{z}_2$ (w.r.t. $\hat{\phi}_{ML} \approx 3.185$). Therefore and because of the influence of the insured sum over the whole support of the Burr Type XII distribution, we use the second tariff feature w as well as the insured sum v as regressors for the truncated gamma GLM.

Table 4 allows for an analog interpretation as in Scenario 2. Hence, we focus again on the TSM w.r.t. $\hat{\phi}_{ML}$. A decrease of 57.04% as well as 14.92% for $\bar{z}_1$ and $\bar{z}_2$ confirms that the TSM is more robust than the gamma GLM for the simulated Burr Type XII claim sizes. Due to the better log-likelihood value we obtain smaller AIC and BIC values for the TSM in comparison with the gamma GLM. This underlines the better distributional fit of the TSM compared to the gamma GLM.

For the TSM w.r.t. $\hat{\phi}_{ML}$ we obtain the following parameter values:

$$\begin{aligned} &\hat{u} = 1.75 \times 10^6, \hat{\delta}_0 = -6.786, \hat{\delta}_1 = 4.927 \times 10^{-8}, \\ &\hat{\delta}_2 = 0, \hat{\phi}_{ML} = 3.004, \\ &\hat{\alpha}_0 = 9.768, \hat{\alpha}_1 = 2.857 \times 10^{-8}, \hat{\alpha}_2 = -0.022, \\ &\hat{\xi} = 0.653, \hat{\beta} = 3\,307\,626. \end{aligned} \tag{5.16}$$

For the gamma GLM we obtain the following parameter estimates:

$$\hat{\alpha}_0^G = 10.642, \hat{\alpha}_1^G = 4.968 \times 10^{-8}, \hat{\alpha}_2^G = -0.110, \hat{\phi}^G = 3.744. \tag{5.17}$$

In Fig. 9 (left) we obtain an analog behavior as in Fig. 8 (left). The predictions using the TSM are robust w.r.t. heavy-tailed claim sizes. In contrast to the TSM, the gamma GLM leads to a significant negative impact of the second tariff feature w. Hence, especially for the insured sum of 50 million, we observe large deviations of the gamma GLM predictions from the true expectations. As the threshold severity model does not lead to such large differences, we conclude that the POT-approximation for extreme claim sizes is sufficiently accurate. Additionally, in contrast to the previous evaluations, the proportions in Fig. 9 (right) indicate that the TSM explains a slight decrease of the expectation w.r.t. the tariff feature w. This decrease stems from the truncated gamma GLM, because it is the only model component which takes the second tariff feature w into account.

5.3. Impact of the chosen threshold

In order to analyze the behavior of the threshold severity model w.r.t. changes of the threshold, we fit the model repeatedly against the sample of Scenario 1 for different thresholds in the interval [0.5, 2] million. For each fit in this interval, we choose the dispersion estimator that leads to the best log-likelihood value.

Fig. 10 illustrates the behavior of the estimates of the dispersion parameter. We deduce the following: For thresholds lower than the true threshold of 1 million, the estimation admits a stable area. In contrast, the estimation for the dispersion parameter increases significantly for thresholds larger than the true threshold. We therefore conclude that the data which has been generated according to the GPD, has as expected a substantial impact on the estimation of the dispersion parameter. Due to the described behavior, the plot of the dispersion parameter may also serve as a tool to determine a threshold for the TSM.

Fig. 11 displays the estimates for the remaining parameters of the threshold severity model. Taking the ordinate-axes' scaling into account, the parameter estimates of the truncated gamma GLM fluctuate weakly over all thresholds. This is caused by the large number of claims below the threshold and by the fact that we contaminate the estimation only by a moderate number of extreme claims. In contrast to that, the estimation of the GPD parameters breaks down close to a value of roughly 0.5 million. This is again reasonable, since with a decreasing threshold, the estimation is contaminated by an increasing number of gamma distributed random variables.

Moreover, as measured by $\bar{z}_1$ and the AIC, for almost all considered thresholds, the estimation based on the TSM is superior to the estimation based on the gamma GLM, which is depicted in Fig. 12. As mentioned before, the substantial increase of $\bar{z}_1$ close to a value of 0.5 million is caused by poor parameter estimates. A reason for the increasing values of the AIC is that for larger thresholds the truncated gamma GLM behaves more similarly to the classical gamma GLM.

6. Conclusion and outlook

It is a common procedure in practice to model the claim severity of insurance claims by generalized linear models that are based on the gamma distribution. However, such a framework may lead to an underestimation of the expected claim size if the data set contains extreme insurance claims. To overcome this issue, in this paper we have developed a procedure to model the tail of the distribution separately by dividing the claim size distribution by a given threshold. To handle claim sizes below this threshold, a GLM based on the truncated gamma distribution is implemented. Above the threshold, we employ the peaks-over-threshold method by approximating the extreme insurance claims by the generalized Pareto distribution. In the resulting threshold severity model, we obtain explicit estimators for the claim severity below and above a given threshold. Moreover, we simulate log-normal and Burr Type XII distributed claim sizes. Subsequently, we demonstrate that the extreme value modeling renders predictions that are significantly closer to the true expectations than predictions from the standard GLM based on the gamma distribution. Furthermore, we note that the threshold severity model is sufficiently robust against threshold changes. Thus, the threshold severity model seems to be a convenient tool for the modeling of extreme insurance claims.

Several possibly relevant aspects are left as directions for future research: Since we focused on the insured sum as the dominating tariff feature in order to estimate extreme claim sizes, a possible future extension of the threshold severity model is the inclusion of further tariff features. To address this issue, multivariate peaks-over-thresholds methods (see e.g. Kiriliouk et al., 2019; Rootzén et al., 2018) or suitable shape-scale regression models for the GPD (see e.g. Pupashenko et al., 2015) seem to be promising approaches. Furthermore, we determine the dispersion parameter of the truncated gamma GLM by an approximation

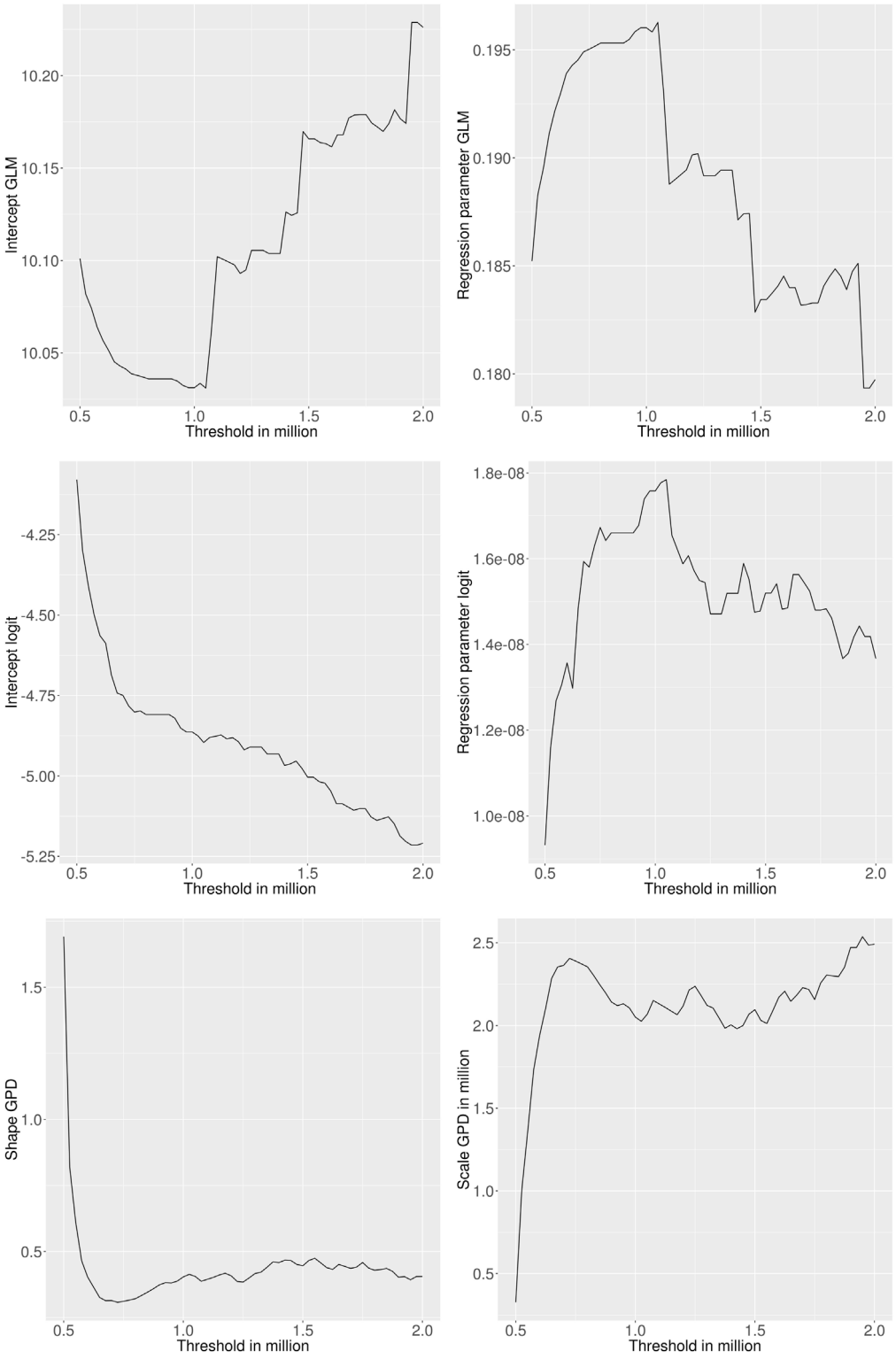

Fig. 11. Parameter estimates in the TSM w.r.t. the threshold.

using a gamma GLM. Thus, an investigation of methods for the estimation of the dispersion parameter that do not require this asymptotic argument might be interesting as well.

Acknowledgments

We wish to thank Wieger Hinderks as well as the participants of the TU Kaiserslautern Seminar in Mathematical Finance and the European Actuarial Journal Conference 2018 in Leuven for fruitful comments and suggestions. We also thank an anonymous referee for helpful comments which improved the contents and the readability of the paper significantly. S. Desmettre is grateful for funding within the DFG, Germany Research Training Group 1932 *Stochastic Models for Innovations in the Engineering Sciences*. S. Desmettre is also supported by the Austrian Science Fund (FWF) project F5508-N26, which is part of the Special Research Program

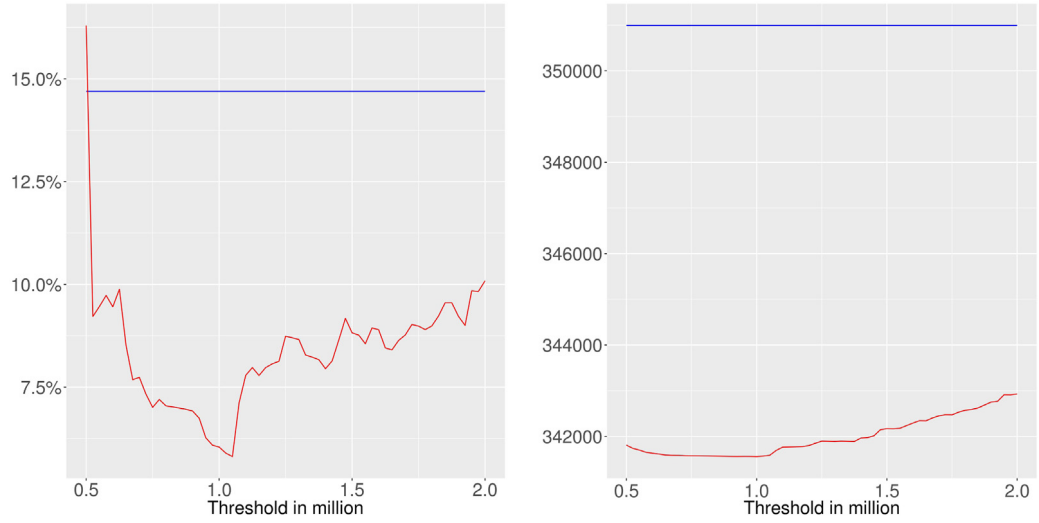

Fig. 12. $\bar{z}_1$ (left) and AIC (right) of the gamma GLM (blue) and the TSM (red) as a function of the threshold. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$$f'(\theta) = \frac{u}{\phi} \cdot \frac{\sum_{n,m=1}^{\infty} \theta^{n+m-1} (-u/\phi)^{n+m} \gamma_\phi(n+1)\gamma_\phi(m) \cdot (n-m)}{\left(\sum_{n=1}^{\infty} \theta^n (-u/\phi)^n \gamma_\phi(n)\right)^2}$$
$$= \frac{u}{\phi} \cdot \frac{\sum_{n>m} \theta^{n+m-1} (-u/\phi)^{n+m} (n-m)\left(\gamma_\phi(n+1)\gamma_\phi(m) - \gamma_\phi(m+1)\gamma_\phi(n)\right)}{\left(\sum_{n=1}^{\infty} \theta^n (-u/\phi)^n \gamma_\phi(n)\right)^2}.$$

Box I.

Quasi-Monte Carlo Methods: Theory and Applications. Moreover, the authors gratefully acknowledge support from NAWI Graz, Austria.

Appendix

Proof of Proposition 4.6. We find an equivalent representation of the function f using the following series expansion of the lower incomplete gamma function (see e.g. Abramowitz and Stegun, 1984):

$$\gamma(a,x) = \int_0^x t^{a-1} \exp(-a)\, dt$$
$$= \Gamma(a)\, x^a \exp(-x) \sum_{n=0}^{\infty} \frac{x^n}{\Gamma(a+n+1)}.$$

In order to obtain shorter expressions, let us also set $\gamma_{1/a}(n) := 1/\Gamma(a+n)$. Using the expansion above in (4.11), we can write f as the quotient of two power series:

$$f(\theta) = \frac{u}{\phi} \cdot \frac{\sum_{n=1}^{\infty} \theta^n (-u/\phi)^n \gamma_\phi(n+1)}{\sum_{n=1}^{\infty} \theta^n (-u/\phi)^n \gamma_\phi(n)}.$$

Note that for $\theta < 0, u > 0, \phi > 0$ both power series have all positive coefficients and hence the quotient is a positive analytical function. Noting that $\gamma_\phi(n+1) = \gamma_\phi(n)/(1/\phi + n)$,

$$\lim_{\theta \to 0} f(\theta) = \frac{u}{\phi} \cdot \frac{\gamma_\phi(2)}{\gamma_\phi(1)} = \frac{u}{1+\phi},$$

is easily obtained for the limit. On the other hand it is easy to see that

$$\lim_{\theta \to -\infty} f(\theta) = 0.$$

Using the quotient rule, we can also calculate the derivative of f as given in Box I. Note that the rearrangement in the last line is possible, since both series are absolutely convergent. Using again $\gamma_\phi(n+1) = \gamma_\phi(n)/(1/\phi + n)$, we see that the last term in parenthesis in the numerator is always negative, while $(n-m)$ is always positive and $\theta^{n+m-1}(-u/\phi)^{n+m}$ is also negative for $\theta < 0$. Thus, $f'(\theta) > 0$ which shows that the function is strictly increasing. Finally the monotonicity implies the existence of an inverse on the whole range $(0, u/(1+\phi))$ and since f is analytic its inverse is too. □

References

Abramowitz, M., Stegun, I.A. (Eds.), 1984. Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. In: A Wiley-Interscience Publication. Selected Government Publications, John Wiley & Sons, Inc; National Bureau of Standards, New York, Washington, D.C.

Beirlant, J., Goegebeur, Y., Verlaak, R., Vynckier, P., 1998. Burr regression and portfolio segmentation. Insurance Math. Econom. 23, 231–250.

Calabrese, R., Osmetti, S.A., 2011. Generalized extreme value regression for binary rare events data: An application to credit defaults, UCD Geary Institute Discussion Paper Series, WP2011/20.

Desmettre, S., Deege, M., 2016. Modeling redemption risks of mutual funds using extreme value theory. J. Risk 18 (6), 1–37.

Embrechts, P., Klüppelberg, C., Mikosch, T., 1997. Modelling Extremal Events for Insurance and Finance, third ed. Springer-Verlag Berlin Heidelberg New York.

Embrechts, P., Resnick, S.I., Samorodnitsky, G., 1999. Extreme value theory as a risk management tool. N. Am. Actuar. J. 3 (2), 30–41.

Garrido, J., Genest, C., Schulz, J., 2016. Generalized linear models for dependent frequency and severity of insurance claims. Insurance Math. Econom. 70, 205–215.

Gissibl, N., Klüppelberg, C., Mager, J., 2017. Big data: Progress in automating extreme risk analysis. pp. 171–189.
Kiriliouk, A., Rootzén, H., Segers, J., Wadsworth, J., 2019. Peaks over thresholds modelling with multivariate generalized pareto distributions. Technometrics 61 (1), 123–135.
Lange, K., 2010. Numerical Analysis for Statisticians, second ed. Springer-Verlag New York.
Lange, K., 2013. Optimization, second ed. Springer-Verlag New York.
Lee, D., Li, W.K., Wong, T.S.T., 2012. Modeling insurance claims via a mixture exponential model combined with peaks-over-threshold approach. Insurance Math. Econom. 51, 538–550.
Lee, S.C.K., Lin, X.S., 2010. Modeling and evaluating insurance losses via mixtures of erlang distributions. N. Am. Actuar. J. 14, 107–130.
Li, Y., Tang, N., Jiang, X., 2016. Bayesian approaches for analyzing earthquake catastrophic risk. Insurance Math. Econom. 68, 110–119.
McNeil, A.J., 1997. Estimating the tails of loss severity distributions using extreme value theory. Astin Bull. 27 (1), 117–137.
Ohlsson, E., Johansson, B., 2010. Non-Life Insurance Pricing with Generalized Linear Models, first ed. Springer-Verlag Berlin Heidelberg.
Pupashenko, D., Ruckdeschel, P., Kohl, M., 2015. L_2-Differentiability of generalized linear models. Statist. Probab. Lett. 97, 155–164.
Reynkens, T., et al., 2017. Modelling censored losses using splicing: A global fit strategy with mixed erlang and extreme value distributions. Insurance Math. Econom. 77, 65–77.
Rootzén, H., Segers, J., Wadsworth, J., 2018. Multivariate peaks over thresholds models. Extremes 21, 115–145.
Shi, P., 2014. Fat-tailed regression models. Predictive Model. Appl. Actuar. Sci. 1, 236–259.
Shi, P., Feng, X., Ivantsova, A., 2015. Dependent frequency–severity modeling of insurance claims. Insurance Math. Econom. 64, 417–428.
Singh, A.K., Allen, D.E., Robert, P.J., 2013. Extreme market risk and extreme value theory. Math. Comput. Simulation 94, 310–328.
The R Foundation for Statistical Computing. R version 3.5.2 (2018-12-20). 2018.
Willmot, G.E., Woo, J.K., 2007. On the class of erlang mixtures with risk theoretic applications. N. Am. Actuar. J. 11 (2), 99–115.
Wüthrich, M., Non-Life Insurance: Mathematics & Statistics, March 14, 2017). Available at SSRN: https://ssrn.com/abstract=2319328.