An extreme value approach for modeling Operational Risk losses depending on covariates

Valérie Chavez-Demoulin[1], Paul Embrechts[2], Marius Hofert[3]

2013-11-26

Abstract

A general methodology for modeling loss data depending on covariates is developed. The parameters of the frequency and severity distributions of the losses may depend on covariates. The loss frequency over time is modeled with a non-homogeneous Poisson process with rate function depending on the covariates. This corresponds to a generalized additive model which can be estimated with spline smoothing via penalized maximum likelihood estimation. The loss severity over time is modeled with a non-stationary generalized Pareto distribution (alternatively, a generalized extreme value distribution) depending on the covariates. Since spline smoothing can not directly be applied in this case, an efficient algorithm based on orthogonal parameters is suggested. The methodology is applied both to simulated loss data and a database of real operational risk losses. Estimates, including confidence intervals, for risk measures such as Value-at-Risk as required by the Basel II/III framework are computed. Furthermore, an implementation of the statistical methodology in R is provided.

Keywords

Operational risk, Value-at-Risk, extreme value theory, covariates, spline smoothing, penalized maximum likelihood.

MSC2010

60G70, 62-07, 62P05.

1 Introduction

The aim of the paper is threefold: first, we present a statistical approach for the modeling of business loss data as a function of covariates; second, this methodology is exemplified in the context of an Operational Risk dataset to be detailed later in the paper; third, a publicly available software implementation (including a simulated data example) is developed to apply the presented methodology.

The fact that we apply the new statistical tools to business "loss" data is not really essential but rather reflects the properties of the Operational Risk dataset at hand (and data of a similar kind). "Losses" can without any problem be changed into "gains"; relevant is that we concentrate our

---

[1] Faculty of Business and Economics, University of Lausanne, Switzerland, `valerie.chavez@unil.ch`

[2] RiskLab, Department of Mathematics and Swiss Finance Institute, ETH Zurich, 8092 Zurich, Switzerland, `embrechts@math.ethz.ch`

[3] RiskLab, Department of Mathematics, ETH Zurich, 8092 Zurich, Switzerland, `marius.hofert@math.ethz.ch`. The author (Willis Research Fellow) thanks Willis Re for financial support while this work was being completed.

analysis on either the left or the right tail of an underlying performance distribution function. This more general interpretation will become clear from the sequel. Slightly more precise, the typical data to which our methodology applies is of the marked point process type, that is, random losses occur at random time points and one is interested in estimating the aggregate loss distribution dynamically over time. Key features will be the existence of extreme (rare) events, the availability of covariate information, and a dynamic modeling of the underlying parameters as a function of the covariates. Operational Risk data typically exhibits such features; see later references. Our concentration on an example from the financial services industry also highlights the recent interest shown in more stringent capital buffers for banks (under the Basel guidelines) and insurance (referring to Solvency 2); for some background on these regularity frameworks, see for instance McNeil et al. (2005) and the references therein.

The methodology presented in this paper is applied to a database of OpRisk losses collected from public media. We are aware that other databases are available. In particular, it would have been interesting to get further explanatory variables such as firm size (not present in our database) which may have an impact on the loss severity and frequency; see, for instance Ganegoda and Evans (2013), Shih et al. (2000) and Cope and Labbi (2008). The database at our disposal is, however, original, rather challenging to model (mainly due to data scarcity), and shows stylized features any OpRisk losses can show. Our findings regarding the estimated parameters, are in accordance with Moscadelli (2004) (infinite-mean models), the latter being based on a much larger database. We also provide an implementation including a reproducible simulation study in a realistic OpRisk context; it shows that even under these difficult features, the methodology provides a convincing fit.

Recall that, under the capital adequacy guidelines of the Basel Committee on Banking Supervision (see `http://www.bis.org/bcbs`, shortened throughout the paper as *Basel* or *the Basel Committee*), *Operational Risk* (OpRisk) is defined as “The risk of a loss resulting from inadequate or failed internal processes, people and systems or from external events. This definition includes legal risk, but excludes strategic and reputational risk.”; see BIS (2006, p. 144). By nature, this risk category, as opposed to Market and Credit Risk, is much more akin to non-life insurance risk or loss experience from industrial quality control. OpRisk came under regulatory scrutiny in the wake of Basel II in the late nineties; see BIS (2006). This is relevant as data was only systematically collected fairly recently, leading to reporting bias in all OpRisk datasets. We will come back to this issue later in the paper. An important aspect of the Basel framework is industry’s freedom of choice of the internal model. Of course, industry has to show to the regulators that the model fits well; on the latter, Dutta and Perry (2006) contains a list of basic principles an internal model has to satisfy. Recent events like Société Générale (rogue trading), UBS (rogue trading), the May 6, 2010 Flash Crash (algorithmic trading), the Libor scandal (fraud involving several banks), and litigation coming out of the subprime crisis have catapulted OpRisk very high on the regulators’ and politicians’ agenda.

Basel II allows banks to opt for an increasingly sophisticated approach, starting from the Basic Indicator Approach and the Standardized Approach to the Advanced Measurement Approach. The latter is commonly realized via the Loss Distribution Approach (LDA) which we will consider in this paper. All LDA models in use aim for a (loss-frequency, loss-severity) approach. Regulators prescribe the use of the risk measure Value-at-Risk (VaR) over a one-year horizon at the 99.9% confidence level, denoted by $\mathrm{VaR}_{0.999}$. Note that in this notation we stick to the use of “confidence level” in order to refer to what statisticians prefer to call “percentile”; this unfortunate choice is by now universal in the regulatory as well as more practical academic literature. For the latter, see for instance Jorion (2007, Section 16.1.2).

The wisdom of the choice of $\mathrm{VaR}_{0.999}$ is highly contested; see for instance Daníelsson et al. (2001). We will see later that Expected Shortfall (ES) as another risk measure is not always a viable alternative, mainly due to the extreme (even infinite mean) heavy-tailedness of OpRisk data. In this respect, we would like to point to the current discussion triggered by the Basel document BIS (2012), see in particular Question 8 on page 41, and the relevant paper on risk measure forecasting Gneiting (2011). In summary, BIS (2012) asks for a possible transition for market risk from VaR to ES as the underlying risk measure. We know that in general VaR is not subadditive, whereas ES is. However, Gneiting (2011) implies that, whereas VaR in general is statistically backtestable (VaR is elicitable in the terminology of Gneiting (2011)), ES is not. Furthermore, whereas VaR in general has certain robustness properties, ES does not, at least as discussed in Cont et al. (2010). We therefore strongly believe that VaR, despite all its shortcomings, will remain in force for a further while. As a consequence of this, and the extreme heavy-tailedness of OpRisk data, we will focus on VaR in this paper. For further details on this, see Embrechts et al. (2013a).

Key publications on OpRisk from the regulatory front are de Fontnouvelle et al. (2004), de Fontnouvelle et al. (2005), Dutta and Perry (2006), Mori et al. (2007), and Moscadelli (2004). The latter paper uses Extreme Value Theory (EVT) to analyze data from the second Quantitative Impact Study with over 47 000 observations. Interesting contributions from industry include Baud et al. (2002), Baud et al. (2003) and Frachot et al. (2004) for Crédit Lyonnais, Aue and Kalkbrener (2006) for Deutsche Bank, and Soprano et al. (2009) for UniCredit Group. More methodological papers relevant for our approach are Chavez-Demoulin and Embrechts (2004), El-Gamal et al. (2007), and Böcker and Klüppelberg (2010). For the analysis of economic business factors influencing OpRisk, as well as the resulting reputational risk consequences, see for instances Cummins et al. (2006), Jarrow (2008), Jarrow et al. (2010), and Chernobai et al. (2011). Unfortunately, publically available OpRisk data is hard (if not impossible) to come by. An excellent source of statistical information on consortium data (based on the ORX data) can be found on the ORX website: `www.orx.org/orx-research`. Questions researched include data homogeneity, scaling, correlation, and capital modeling. For another study on real OpRisk data (from several Italian banking groups), explicitly modeling dependence between weekly aggregated losses of business lines or event types via a copula approach, see Brechmann et al. (2013). Finally, more recent textbook treatments are Cruz (2002), Akkizidis and Bouchereau (2006), Panjer (2006), Böcker (2010), Shevchenko (2011), and Bolancé et al. (2012); see also McNeil et al. (2005, Chapter 10).

For the LDA, the Basel Committee has decomposed a bank's activities into business lines (the standard is 8) and event types (typically 7) so that it is possible to model along substructures of this matrix; often aggregation to business line level is chosen. Some methodological issues underlying the analysis of loss data in such a matrix structure, which is also known as Unit of Measure (UOM) in the regulatory jargon, are discussed in Embrechts and Puccetti (2008). Papers analyzing data at the business line and event type levels include Moscadelli (2004), de Fontnouvelle et al. (2004) at the individual bank level, Dutta and Perry (2006) at the enterprise level. The latter two also attempt a modeling by time. Papers on modeling operational risk data using covariates include Ganegoda and Evans (2013), Shih et al. (2000), and Cope and Labbi (2008) where the authors investigate whether the size of operational risk losses can be correlated with firm size and geographical region. Na et al. (2006) supposes that the operational loss can be broken down into a component common to all banks and idiosyncratic components specific to each loss. Dahen and Dionnne (2010) investigates how severity and frequencies of external losses can be scaled for integration with internal data.

Going down to more granular levels of modeling increases the variance of the resulting estimators due to data scarcity. The methodology presented in the following sections allows for "pooling" of the data (explained later), introducing business line, event type, and time as covariates. This approach allows for a greater flexibility in analyzing the data, as well as making model comparisons possible.

The paper is organized as follows. Section 2 provides a brief introduction to EVT in terms of the block maxima and the Peaks-over-Threshold (POT) approaches. In Section 3 we extend these classical approaches, allowing the (constant) parameters of the model to (dynamically) vary with covariates. This constitutes the main statistical methodology used in the paper and focus is put on the dynamic POT approach. A fitting method is developed in Section 3.3 and it is applied to simulated data in Section A.2. In Section 4 we model a dataset of publicly available OpRisk losses with the presented dynamic, EVT-based approaches. Section 5 provides an in-depth discussion of the presented methodology.

2 Classical EVT approach for modeling losses

The standard approaches for describing the extreme events of a stationary time series are the block maxima approach (which models the maxima of a set of blocks dividing the series) and the POT approach (which focuses on exceedances over a fixed high threshold). The POT method has the advantage of being more flexible in modeling data, because more data points are incorporated. The method we use is an extension of the POT method to a non-stationary setup; an extension of the block maxima method to the non-stationary case is also proposed and serves as another possibility to evaluate risk measures. We begin with the latter.

2.1 The Block Maxima Method

Consider the maximum of a sequence of independent and identically distributed random variables $X_1, \dots, X_q$ from a continuous distribution F. Suppose there exists a point x_0 (possibly ∞) such that $\lim_{x\to x_0} F(x) = 1$. For any fixed $x < x_0$ we have

$$\mathbb{P}(\max\{X_1, \dots, X_q\} \le x) = \mathbb{P}(X_i \le x,\ i = 1, \dots, q) = F^q(x),$$

which tends to 0 as $q \to \infty$. To obtain a non-degenerate limiting distribution for the maximum, the X_i's must be rescaled by sequences (a_q) (positive) and (b_q) leading to $W_q = a_q^{-1}(\max(X_1, \dots, X_q) - b_q)$. As $q \to \infty$, a possible limiting distribution of W_q is such that $P(W_q \le w) = F^q(b_q + a_q w)$ has a limit. It can be shown that the latter expression, rewritten as $\left(1 - \frac{q(1-F(b_q+a_q w))}{q}\right)^q$, has a limit if and only if $\lim_{q\to\infty} q(1 - F(b_q + a_q w))$ exists. If so, the only possible limit is

$$\lim_{q\to\infty} q(1 - F(b_q + a_q w)) = \begin{cases} \left(1 + \xi \frac{w-\mu}{\sigma}\right)_+^{-1/\xi}, & \text{if } \xi \neq 0, \\ \exp(-\frac{w-\mu}{\sigma}), & \text{if } \xi = 0, \end{cases} \tag{1}$$

where $(x)_+ = \max\{x, 0\}$ with $\xi, \mu \in \mathbb{R}$ and $\sigma > 0$. We shall not go into details about (1). In theory, it requires that F satisfies some properties of regular variation but in the vast majority of applications, simpler conditions are sufficient. This leads to the remarkable result that given suitable sequences (a_q) and (b_q), the non-degenerate limiting distribution must be a *generalized extreme value (GEV) distribution*, given by

$$H_{\xi,\mu,\sigma}(w) = \begin{cases} \exp\Big(-\big(1+\xi\frac{w-\mu}{\sigma}\big)_+^{-1/\xi}\Big), & \text{if } \xi \neq 0, \\ \exp(-\exp(-(\frac{w-\mu}{\sigma}))), & \text{if } \xi = 0. \end{cases} \tag{2}$$

The parameters $\mu \in \mathbb{R}$ and $\sigma > 0$ are respectively the *location* and *scale* parameters of the distribution. The parameter $\xi \in \mathbb{R}$ controls the *shape* of the distribution.

In applications, we consider $X_{t'_1}, \dots, X_{t'_{n'}}$ following a non-degenerate, continuous distribution function F. We interpret $X_{t'_1}, \dots, X_{t'_{n'}}$ (≥ 0) as losses in some monetary unit over a time period $[0, T]$ with $0 \leq t'_1 \leq \dots \leq t'_{n'} \leq T$. We fit the GEV distribution (2) to the series of (typically) annual maxima. Taking q to be the number of observations in a year and considering m' years observed over the period $[0, T]$ (such that $m'q = n'$), we have *block maxima* $M_q^{(1)}, \dots, M_q^{(m')}$ from m' blocks of size q. Assuming these block maxima to be independent, the log-likelihood is given by

$$\ell(\xi, \mu, \sigma; M_q^{(1)}, \dots, M_q^{(m')}) = \log\Big(\prod_{i=1}^{m'} h_{\xi,\mu,\sigma}(M_q^{(i)}) \mathbb{1}_{\{1+\xi(M_q^{(i)}-\mu)/\sigma>0\}}\Big),$$

where $h_{\xi,\mu,\sigma}(w)$ denotes the density of $H_{\xi,\mu,\sigma}(w)$. By maximizing the log-likelihood with respect to ξ, μ, σ, one obtains the maximum likelihood estimators $\hat{\xi}$, $\hat{\mu}$, $\hat{\sigma}$. The fitted distribution is used to estimate the $1/p$*-year return level*, that is, the level exceeded once every $1/p$ years on average, $p \in (0, 1)$. This is equivalent to using VaR_α at confidence level $\alpha = 1 - p$ as risk measure. Based on the estimates $\hat{\xi}$, $\hat{\mu}$, $\hat{\sigma}$, VaR_α is estimated by

$$\widehat{\mathrm{VaR}}_\alpha = \begin{cases} \hat{\mu} + \hat{\sigma}\big((-\log(1-\alpha))^{-\hat{\xi}} - 1\big)/\hat{\xi}, & \text{if } \hat{\xi} \neq 0, \\ \hat{\mu} + \hat{\sigma}\big(-\log(-\log(1-\alpha))\big), & \text{if } \hat{\xi} = 0. \end{cases} \tag{3}$$

The latter estimation is an approximation in the limit of a large quantile level, α being typically close to 1.

2.2 The Peaks-over-Threshold approach

Given a suitably large threshold $u \geq 0$, let $\{t_1, \dots, t_n\} \subseteq \{t'_1, \dots, t'_{n'}\}$ denote those time points (in increasing order) for which $X_{t'_1}, \dots, X_{t'_{n'}}$ exceed u, that is, let $X_{t_1}, \dots, X_{t_n}$ denote the *exceedances over* u with corresponding *excesses* $Y_{t_i} = X_{t_i} - u$, $i \in \{1, \dots, n\}$. It follows from Embrechts et al. (1997, pp. 166) that

1) the number of exceedances N_t approximately follows a Poisson process with intensity λ, that is, $N_t \sim \mathrm{Poi}(\Lambda(t))$ with *integrated rate function* $\Lambda(t) = \lambda t$;
2) the excesses $Y_{t_1}, \dots, Y_{t_{N_t}}$ over u approximately follow (independently of N_t) a *generalized Pareto distribution (GPD)*, denoted by $\mathrm{GPD}(\xi, \beta)$ for $\xi \in \mathbb{R}$, $\beta > 0$, with distribution function

$$G_{\xi,\beta}(x) = \begin{cases} 1 - \big(1 + \xi x/\beta\big)^{-1/\xi}, & \text{if } \xi \neq 0, \\ 1 - \exp(-x/\beta), & \text{if } \xi = 0, \end{cases}$$

for $x \geq 0$, if $\xi \geq 0$, and $x \in [0, -\beta/\xi]$, if $\xi < 0$.

For a precise formulation and the full details of the proof, see Leadbetter (1991); note that ξ is the same as in (1).

In the following, we assume $\xi > -1$. If $\xi > 0$ (which is in accordance with most OpRisk loss models), the mathematical condition needed in order for the above asymptotics to hold is known as *regular variation*, that is,

$$\bar{F}(x) = 1 - F(x) = x^{-1/\xi} L(x) \tag{4}$$

for some slowly varying function $L : (0, \infty) \to (0, \infty)$ measurable so that

$$\lim_{x \to \infty} \frac{L(tx)}{L(x)} = 1 \quad \text{for all } t > 0.$$

The underlying theorems are known under the names of Gnedenko and Pickands–Balkema–de Haan; see McNeil et al. (2005, Theorems 7.8 and 7.20). Equation (4) implies that the tail of the loss distribution is of power- (or Pareto-) type, a property which typically holds for OpRisk data. One can show that a model satisfying (4) with $\xi \in (0, 1)$ has at least a finite first moment whereas for $\xi > 1$, F has infinite first moment; for examples of the latter within an OpRisk context, see Moscadelli (2004). More methodological consequences are for instance to be found in Nešlehová et al. (2006) and the references therein; see also Section 5. The asymptotic independence condition between Poisson exceedance times and GPD excesses yields an approximate likelihood function of the form

$$L(\lambda, \xi, \beta; \boldsymbol{Y}) = \frac{(\lambda T)^n}{n!} \exp(-\lambda T) \prod_{i=1}^{n} g_{\xi,\beta}(Y_{t_i}),$$

where $\boldsymbol{Y} = (Y_{t_1}, \dots, Y_{t_n})$ and $g_{\xi,\beta}$ is the density of $G_{\xi,\beta}$. It follows that the log-likelihood splits into two parts

$$\ell(\lambda, \xi, \beta; \boldsymbol{Y}) = \ell(\lambda; \boldsymbol{Y}) + \ell(\xi, \beta; \boldsymbol{Y}),$$

where

$$\ell(\lambda; \boldsymbol{Y}) = -\lambda T + n \log(\lambda) + \log(T^n/n!) \quad \text{and} \quad \ell(\xi, \beta; \boldsymbol{Y}) = \sum_{i=1}^{n} \ell(\xi, \beta; Y_{t_i})$$

with

$$\ell(\xi, \beta; y) = \begin{cases} -\log(\beta) - (1 + 1/\xi) \log(1 + \xi y/\beta), & \text{if } \xi > 0,\ y \geq 0 \text{ or } \xi < 0,\ y \in [0, -\beta/\xi), \\ -\log(\beta) - y/\beta, & \text{if } \xi = 0, \\ -\infty, & \text{otherwise.} \end{cases}$$

The maximization for the two estimation problems related to 1) and 2) can thus be carried out separately. One can further show that this standard EVT (POT) methodology also holds for certain classes of stationary models beyond the assumption of independence and identical distribution; see Embrechts et al. (1997, Section 4.4).

3 A dynamic EVT approach

In practice, it is often the case that stationarity assumptions (such as independence and identical distribution) for time series extremes are violated. For example, OpRisk losses might depend on covariates, that is, on additional variables which are possibly predictive of the outcome. Covariates can be economic factors, business lines and/or event types, or also time.

We now extend the classical approaches described above to more dynamic ones, in which we let the model parameters depend on covariates. The new methodology allows the dependence (on covariates) to be parametric, non-parametric, or semi-parametric and can also include interactions. In this work, we focus on two covariates: a factor x (in Section 4: business line) and time t (in Section 4: year). The model presented can easily be extended to more covariates. Let $\boldsymbol{\theta} \in \mathbb{R}^p$ be the vector of p EVT model parameters ($p = 3$ for the GEV distribution and $p = 2$ for the GPD). A general model for $\boldsymbol{\theta}$ can then be built via

$$g_k(\theta_k) = f_k(x) + h_k(t), \quad k \in \{1, \dots, p\}, \tag{5}$$

where g_k denotes a link function (constrained to the parameter space), f_k maps the factor levels to correspondingly many constants, and $h_k(t)$ is either a linear (parametric) function or, in most of the cases, a smoothed (non-parametric) function of $t \in \mathcal{A} \subseteq \mathbb{R}$.

The idea of letting GEV parameters depend on covariates in a parametric way has been developed in Coles (2001, Chapter 6). Its generalization to a semi-parametric form using smoothing splines is new. This approach allows any functional form for the dependence of the (annual, say) maxima over time which may be useful in various kinds of applications where there is a significant evolution through time (in the form of a trend or seasonality). Assuming all h_k in (5) to be smoothed functions, the parameter vector $\boldsymbol{\theta}$ of the EVT model can be estimated by maximizing the penalized log-likelihood

$$\ell(\boldsymbol{\theta}; \cdot\,) - \sum_{k=1}^{p} \Big(\gamma_k \int_{\mathcal{A}} h_k''(t)^2 \, dt \Big), \tag{6}$$

where the first term $\ell(\boldsymbol{\theta}; \cdot\,)$ is the log-likelihood based on either the block-maxima or the EVT POT-model. The introduction of the penalty terms is a standard technique to avoid over-fitting when one is interested in fitting smooth functions h_k (see Hastie and Tibshirani (1990) or Green and Silverman (2000)). Intuitively, the penalty functions $\int_{\mathcal{A}} h_k''(t)^2 \, dt$ measure the roughness of twice-differentiable curves and the parameters γ_k are chosen to regulate the smoothness of the estimates $\hat{h}_k$. Large values of γ_k produce smoother curves, while smaller values produce rougher curves.

In this paper, we mainly consider the methodology for the POT approach, which is more relevant in our case because it allows one to incorporate more data. However, as a further assessment of the results, we also apply the semi-parametric GEV-based approach to yearly maxima of OpRisk losses; this is done in Section 4.2.3.

Focusing on the POT approach, we follow Chavez-Demoulin and Davison (2005) and let the intensity λ, as well as the GPD parameters ξ and β depend on covariates (x and t). This is done in Sections 3.1 and 3.2. Section 3.3 then presents the details about the fitting procedure.

3.1 Loss frequency

The number of exceedances is assumed to follow a non-homogeneous Poisson process with rate function

$$\lambda = \lambda(x,t) = \exp(f_\lambda(x) + h_\lambda(t)) \tag{7}$$

where f_λ denotes a function mapping the factor levels of the covariate x to correspondingly many constants and $h_\lambda : [0,T] \to \mathbb{R}$ a general measurable function not depending on specific parameters.

This model is a standard *generalized additive model* which leads to an estimate for λ; see, for example, Wood (2006). In the statistical software R for example, such models can be fit with the function `gam(..., family=poisson)` from the package `mgcv`. Pointwise asymptotic confidence intervals can be constructed from the estimated standard errors obtained by `predict(..., se.fit=TRUE)`. For how to use these and other functions we implemented (in the R package `QRM` ($\geq$ 0.4-10)), see the simulation examples in Section A.2.

3.2 Loss severity

We use a similar form as (7) for each of the parameters ξ and β of the approximate GPD of the excesses. However, for convergence of the simultaneous fitting procedure for ξ and β, it is crucial that these parameters are orthogonal with respect to the Fisher information metric; see Chavez-Demoulin (1999, p. 96) for a counter-example otherwise. We therefore reparameterize the GPD parameter β by

$$\nu = \log((1+\xi)\beta) \tag{8}$$

which is orthogonal to ξ in this sense; see Cox and Reid (1987) for the general theory and Chavez-Demoulin and Davison (2005) for a similar reparameterization in a EVT POT-context. Note that this reparameterization is only valid for $\xi > -1$ (which is a rather weak assumption for the applications we consider). The corresponding reparameterized log-likelihood ℓ^r for the excesses is thus

$$\ell^r(\xi,\nu;\boldsymbol{Y}) = \ell(\xi, \exp(\nu)/(1+\xi); \boldsymbol{Y}).$$

We assume that ξ and ν are of the form

$$\xi = \xi(x,t) = f_\xi(x) + h_\xi(t), \tag{9}$$
$$\nu = \nu(x,t) = f_\nu(x) + h_\nu(t), \tag{10}$$

where f_ξ, f_ν denote functions in the factor levels of the covariate x as in (7) and $h_\xi, h_\nu : [0,T] \to \mathbb{R}$ are general measurable functions. Concerning the excesses, our goal is to estimate ξ and ν, and thus

$$\beta = \beta(x,t) = \frac{\exp(\nu(x,t))}{1+\xi(x,t)} \tag{11}$$

as functions of the covariate x and of time t based on estimators $\hat{f}_\xi, \hat{h}_\xi, \hat{f}_\nu, \hat{h}_\nu$ of $f_\xi, h_\xi, f_\nu, h_\nu$, respectively. The latter estimates are obtained from the observed vectors $\boldsymbol{z}_i = (t_i, x_i, y_{t_i})$, $i \in$

$\{1, \dots, n\}$, where $0 \le t_1 \le \dots \le t_n \le T$ denote the exceedance times, x_i, $i \in \{1, \dots, n\}$, the corresponding observed covariates, and y_{t_i} the corresponding realization of the (random) excess Y_{t_i} over the threshold u, $i \in \{1, \dots, n\}$. In contrast to (7), simultaneously fitting (9) and (10) does not lie within the scope of a standard generalized additive modeling procedure. We therefore develop a suitable backfitting algorithm (including a bootstrap for computing confidence intervals) for this task in the following section.

Before doing so, let us briefly mention a graphical goodness-of-fit test for the GPD model. If the excesses Y_{t_i}, $i \in \{1, \dots, n\}$, (approximately independently) follow $\mathrm{GPD}(\xi_i, \beta_i)$ distributions, then $R_i = 1 - G_{\xi_i, \beta_i}(Y_{t_i})$, $i \in \{1, \dots, n\}$, (approximately) forms a random sample from a standard uniform distribution. We can thus graphically check (for example, with a Q-Q plot) whether, approximately,

$$r_i = -\log(1 - G_{\hat{\xi}_i, \hat{\beta}_i}(y_{t_i})), \quad i \in \{1, \dots, n\}, \tag{12}$$

are distributed as independent standard exponential variables.

3.3 The penalized maximum likelihood estimator and its computation

In this section, we present the penalized maximum likelihood estimator and a backfitting algorithm for simultaneously estimating the parameters ξ and ν (thus β) associated with the approximate GPD for the excesses; see (9) and (10) above. In our application in Section 4, covariates will be business line x and year t (in which the loss happened) of the OpRisk loss under consideration.

In order to fit reasonably smooth functions h_ξ, h_ν to the observations $\boldsymbol{z}_i$, $i \in \{1, \dots, n\}$, we use the penalized likelihood (6). The *penalized log-likelihood* ℓ^p corresponding to the observations $\boldsymbol{z}_i$, $i \in \{1, \dots, n\}$, is given by

$$\ell^p(f_\xi, h_\xi, f_\nu, h_\nu; \boldsymbol{z}_1, \dots, \boldsymbol{z}_n) = \ell^r(\boldsymbol{\xi}, \boldsymbol{\nu}; \boldsymbol{y}) - \gamma_\xi \int_0^T h_\xi''(t)^2 \, dt - \gamma_\nu \int_0^T h_\nu''(t)^2 \, dt \tag{13}$$

where $\gamma_\xi, \gamma_\nu \ge 0$ denote *smoothing parameters*, $\boldsymbol{y} = (y_{t_1}, \dots, y_{t_n})$, and

$$\ell^r(\boldsymbol{\xi}, \boldsymbol{\nu}; \boldsymbol{y}) = \sum_{i=1}^n \ell^r(\xi_i, \nu_i; y_{t_i}) \tag{14}$$

for $\ell^r(\xi_i, \nu_i; y_{t_i}) = \ell(\xi_i, \exp(\nu_i)/(1 + \xi_i); y_{t_i})$. Larger values of the smoothing parameters lead to smoother fitted curves.

Let $0 = s_0 < s_1 < \dots < s_m < s_{m+1} = T$ denote the ordered and distinct values among $\{t_1, \dots, t_n\}$. A function h defined on $[0, T]$ is a cubic spline with the above knots if the two following conditions are satisfied: 1) on each interval $[s_i, s_{i+1}]$, $i \in \{1, \dots, m\}$, h is a cubic polynomial; 2) at each knot s_i, $i \in \{1, \dots, m\}$, h and its first and second derivatives are continuous, hence the same is true on the entire domain $[0, T]$. A cubic spline on $[0, T]$ is a natural cubic spline if in addition to the two latter conditions it satisfies the natural boundary condition: 3) the second and third derivatives of h at 0 and T are zero. It follows from Green and Silverman (2000, p. 13) that for a natural cubic spline h with knots $s_1, \dots, s_m$, one has

$$\int_0^T h''(t)^2 \, dt = \boldsymbol{h}^\top K \boldsymbol{h},$$

where $\boldsymbol{h} = (h_{s_1}, \dots, h_{s_m}) = (h(s_1), \dots, h(s_m))$ and K is a symmetric (m, m)-matrix of rank $m-2$ only depending on the knots $s_1, \dots, s_m$. The penalized log-likelihood (13) can thus be written as

$$\ell^p(f_\xi, h_\xi, f_\nu, h_\nu; \boldsymbol{z}_1, \dots, \boldsymbol{z}_n) = \ell^r(\boldsymbol{\xi}, \boldsymbol{\nu}; \boldsymbol{y}) - \gamma_\xi \boldsymbol{h}_\xi^\top K \boldsymbol{h}_\xi - \gamma_\nu \boldsymbol{h}_\nu^\top K \boldsymbol{h}_\nu \tag{15}$$

with $\boldsymbol{h}_\xi = (h_\xi(s_1), \dots, h_\xi(s_m))$ and $\boldsymbol{h}_\nu = (h_\nu(s_1), \dots, h_\nu(s_m))$.

With these formulas, it is possible to develop a backfitting algorithm for estimating the GPD parameters ξ and ν (thus β). The basic idea is an iterative weighted least squares procedure (see Algorithm 3.2 below) that alternates between Newton steps for ξ and ν (see Algorithm 3.1 below). We construct bootstrapped pointwise two-sided confidence intervals with a post-blackend bootstrap (see Algorithm 3.3 below).

Algorithm 3.1 computes one Newton step (for both parameters) for solving the likelihood equations. The quantities $f_\xi^{(1)}, h_\xi^{(1)}, f_\nu^{(1)}, h_\nu^{(1)}$ denote initial values for $f_\xi, h_\xi, f_\nu, h_\nu$, respectively. They are computed from the classical (non-dynamic) approach described in Section 2; see Algorithm 3.2 below for this step. The formulas `xi.formula` and `nu.formula` refer to the (parametric, non-parametric, or semi-parametric) model specification; see, for example, (24)–(29) in Section 4. The derivatives of the reparameterized log-likelihood appearing in the following algorithm can be found in Section A.1.

Algorithm 3.1 (Newton step; `QRM:::gamGPDfitUp()`)

Let $k \in \mathbb{N}$ and n-dimensional parameter vectors $\boldsymbol{\xi}^{(k)} = (\xi_1^{(k)}, \dots, \xi_n^{(k)})$ and $\boldsymbol{\nu}^{(k)} = (\nu_1^{(k)}, \dots, \nu_n^{(k)})$ be given. Furthermore, let $\boldsymbol{z}_i = (t_i, x_i, y_{t_i})$, $i \in \{1, \dots, n\}$, be given.

1) Setup: Specify formulas `xi.formula` and `nu.formula` for the calls to `gam()` in Steps 2.2) and 3.2) below for fitting (9) and (10), respectively.

2) Update $\boldsymbol{\xi}^{(k)}$:

    2.1) Newton step for the score component: Compute (componentwise)

    $$\boldsymbol{\xi}^{\text{Newton}} = \boldsymbol{\xi}^{(k)} - \frac{\ell_\xi^r(\boldsymbol{\xi}^{(k)}, \boldsymbol{\nu}^{(k)}; \boldsymbol{y})}{\ell_{\xi\xi}^r(\boldsymbol{\xi}^{(k)}, \boldsymbol{\nu}^{(k)}; \boldsymbol{y})}.$$

    2.2) Fitting: Compute $\boldsymbol{\xi}^{(k+1)}$ via calling `fitted(xiObj)` for

    `xiObj <- gam(`$\boldsymbol{\xi}^{\text{Newton}}$`~xi.formula,...,weights=`$-\ell_{\xi\xi}^r$`)`

    for the specified formula `xi.formula`.

3) Given $\boldsymbol{\xi}^{(k+1)}$, update $\boldsymbol{\nu}^{(k)}$:

    3.1) Newton step for the score component: Compute (componentwise)

    $$\boldsymbol{\nu}^{\text{Newton}} = \boldsymbol{\nu}^{(k)} - \frac{\ell_\nu^r(\boldsymbol{\xi}^{(k+1)}, \boldsymbol{\nu}^{(k)}; \boldsymbol{y})}{\ell_{\nu\nu}^r(\boldsymbol{\xi}^{(k+1)}, \boldsymbol{\nu}^{(k)}; \boldsymbol{y})}.$$

    3.2) Fitting: Compute $\boldsymbol{\nu}^{(k+1)}$ via calling `fitted(nuObj)` for

    `nuObj <- gam(`$\boldsymbol{\nu}^{\text{Newton}}$`~nu.formula,...,weights=`$-\ell_{\nu\nu}^r$`)`

    for the specified formula `nu.formula`.

4) Return $\boldsymbol{\xi}^{(k+1)}$ and $\boldsymbol{\nu}^{(k+1)}$.

Based on Algorithm 3.1, the following backfitting algorithm computes the estimators

$$\begin{aligned}\hat{\boldsymbol{\xi}} &= (\hat{\xi}(x_1,t_1),\dots,\hat{\xi}(x_n,t_n)),\\ \hat{\boldsymbol{\beta}} &= \Bigl(\frac{\exp(\hat{\nu}(x_1,t_1))}{1+\hat{\xi}(x_1,t_1)},\dots,\frac{\exp(\hat{\nu}(x_n,t_n))}{1+\hat{\xi}(x_n,t_n)}\Bigr)\end{aligned}$$

of the GPD parameters at each of the observed vectors $\boldsymbol{z}_i$, $i \in \{1,\dots,n\}$, where entries corresponding to the same covariates and time points are equal.

Algorithm 3.2 (Fitting the GPD parameters; `QRM::gamGPDfit()`)

1) Setup: Fix $\varepsilon_\xi, \varepsilon_\nu > 0$ sufficiently small (for example, $\varepsilon_\xi = \varepsilon_\nu = 10^{-5}$).
2) Initialization step (use `gpd.fit()` from the R package `ismev`):
    2.1) Compute the (classical) maximum likelihood estimators ξ^{MLE} and β^{MLE} of the GPD parameters based on all the data y_{t_i}, $i \in \{1,\dots,n\}$, as described in Section 2 2).
    2.2) Compute $\nu^{\text{MLE}} = \log((1+\xi^{\text{MLE}})\beta^{\text{MLE}})$.
    2.3) Set $k=1$, $\boldsymbol{\xi}^{(k)} = (\xi^{\text{MLE}},\dots,\xi^{\text{MLE}})$, and $\boldsymbol{\nu}^{(k)} = (\nu^{\text{MLE}},\dots,\nu^{\text{MLE}})$.
3) Iteration:
    3.1) Based on $\boldsymbol{\xi}^{(k)}$ and $\boldsymbol{\nu}^{(k)}$, compute the new, updated parameter vectors $\boldsymbol{\xi}^{(k+1)}$ and $\boldsymbol{\nu}^{(k+1)}$ with Algorithm 3.1 (Newton step).
    3.2) Check convergence: If the mean relative differences satisfy

$$\frac{1}{n}\sum_{i=1}^n\biggl|\frac{\xi_i^{(k)}-\xi_i^{(k+1)}}{\xi_i^{(k)}}\biggr| \le \varepsilon_\xi \quad\text{and}\quad \frac{1}{n}\sum_{i=1}^n\biggl|\frac{\nu_i^{(k)}-\nu_i^{(k+1)}}{\nu_i^{(k)}}\biggr| \le \varepsilon_\nu,$$

stop. In this case, compute

$$\boldsymbol{\beta}^{(k+1)} = \Biggl(\frac{\exp(\nu_1^{(k+1)})}{1+\xi_1^{(k+1)}},\dots,\frac{\exp(\nu_n^{(k+1)})}{1+\xi_n^{(k+1)}}\Biggr)$$

and return the estimates $\hat{\boldsymbol{\xi}} = \boldsymbol{\xi}^{(k+1)}$ and $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}^{(k+1)}$. Otherwise, set k to $k+1$ and continue with Step 3.1).

The following algorithm wraps around Algorithm 3.2 to compute a list of estimates (as returned by Algorithm 3.2) of length $B+1$ with the post-blackend bootstrap of Chavez-Demoulin and Davison (2005), where B denotes the number of bootstrap replications. This list can then be used to compute bootstrapped pointwise two-sided $1-\alpha$ confidence intervals for the GPD parameters ξ and ν (or β) for each combination of covariate x and time point t. Predicted values can also be computed; see the function `GPD.predict()` in `QRM`.

Algorithm 3.3 (Post-blackend bootstrap; `QRM::gamGPDboot()`)

1) Compute estimates $\hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\nu}}$ (and $\hat{\boldsymbol{\beta}}$) with Algorithm 3.2 and corresponding residuals r_i, $i \in \{1,\dots,n\}$, as in (12).

2) For b from 1 to B do:
    2.1) Within each group of covariates, randomly sample (with replacement) r_i, $i \in \{1, \dots, n\}$, to obtain $r_i^{(b)}$, $i \in \{1, \dots, n\}$.
    2.2) Compute the corresponding excesses $y_{t_i}^{(b)} = G^{-1}_{\hat{\xi}_i, \hat{\beta}_i}(1 - \exp(-r_i^{(b)}))$, $i \in \{1, \dots, n\}$, where, for all $p \in [0, 1]$,

$$G^{-1}_{\xi,\beta}(p) = \begin{cases} \beta((1-p)^{-\xi} - 1)/\xi, & \text{if } \xi \neq 0, \\ -\beta \log(1-p), & \text{if } \xi = 0. \end{cases}$$

    2.3) Compute estimates $\hat{\boldsymbol{\xi}}^{(b)}, \hat{\boldsymbol{\nu}}^{(b)}$ (and $\hat{\boldsymbol{\beta}}^{(b)}$) with Algorithm 3.2.

3) Return a list of length $B+1$ containing all estimated objects (including $\hat{\boldsymbol{\xi}}$, $\hat{\boldsymbol{\xi}}^{(b)}$, $\hat{\boldsymbol{\nu}}$, $\hat{\boldsymbol{\nu}}^{(b)}$, and $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\beta}}^{(b)}$, $b \in \{1, \dots, B\}$).

Based on the estimates $\hat{\lambda}$, $\hat{\xi}$, and $\hat{\beta}$ for a fixed covariate x and time point t, one can compute estimates of the risk measures VaR and ES (depending on x and t) as

$$\widehat{\mathrm{VaR}}_\alpha = u + \frac{\hat{\beta}}{\hat{\xi}}\Bigg(\bigg(\frac{1-\alpha}{\hat{\lambda}}\bigg)^{-\hat{\xi}} - 1\Bigg), \tag{16}$$

$$\widehat{\mathrm{ES}}_\alpha = \begin{cases} \frac{\widehat{\mathrm{VaR}}_\alpha + \hat{\beta} - \hat{\xi}u}{1-\hat{\xi}}, & \text{if } \hat{\xi} \in (0,1), \\ \infty, & \hat{\xi} > 1. \end{cases} \tag{17}$$

Bootstrapped pointwise two-sided confidence intervals for VaR_α and ES_α can be constructed from the fitted values of λ and (list of) bootstrapped estimates of ξ and β by computing the corresponding empirical quantities.

3.4 Smoothing parameter selection

For simplicity, we consider Model (5) for θ with $p = 1$ and with link function g equal to the identity (it is straightforward to extend the following argument to the general form with penalized log-likelihood as given in (6)). In generalized additive models, asymptotic distribution theory provides tools for inference and assessment of model fit based on the deviance and its associated number of parameters used to estimate the model. A quantity related to the smoothing parameter γ in (6) is the *degrees of freedom*; see Green and Silverman (2000, p. 110). This gives an indication of the effective number of parameters used in the model for a specific value of the smoothing parameter. It is defined as

$$\mathrm{Df} = m - \mathrm{tr}(S) - \mathrm{tr}\Big((D^\top A(I-S)D)^{-1} D^\top A(I-S)^2 D\Big),$$

where $S = E(E^\top AE + \gamma K)^{-1}E^\top A$, $A = E[-\frac{\partial^2}{\partial\theta^2}\ell(\theta; \cdot)]$, $D = \frac{\partial\theta}{\partial f}$, and $E = \frac{\partial\theta}{\partial h}$. Very often, the term "degrees of freedom" substitutes the term "smoothing parameter" for simplicity, the degrees of freedom 1 corresponding to linearity; in R the function `s()` is used, where, as smoothing parameter, the input is Df (with Df = 1 representing linearity).

There are two different approaches of choosing the smoothing parameter (degrees of freedom), in practice. When the aim is a pure exploration of data on different scales, varying the smoothing parameter is an efficient way to proceed. The other approach follows the paradigm that the data itself must choose the smoothing parameter and thus an automatic procedure is recommended. Following Hastie and Tibshirani (1990, pp. 158), we suggest the use of Akaike's Information Criterion (AIC). The derivation of AIC is intended to create an approximation to the Kullback–Leibler divergence

$$\mathbb{E}_{f_{\mathrm{T}}}\left[\log \frac{f_{\mathrm{T}}(x)}{f_{\mathrm{C}}(x)}\right],$$

where f_{T} denotes the density of the true model and f_{C} the density of the candidate model. The AIC criterion is readily adapted to a wide range of statistical models. In the context of semiparametric models its form is given by

$$\mathrm{AIC} \propto -2\ell + 2\,\mathrm{Df}, \tag{18}$$

see Simonoff and Tsai (1999).

4 Application to an OpRisk loss database

In this section, we apply the methodology presented before to a database of OpRisk losses collected[4] from public media. The database consists of 1413 OpRisk events reported in the public media since 1970. For our analysis, we consider the 1387 events since 1980. For each event, the following information is given: a reference number; the organization affected (one or 0.07 % missing); the country of head office; the country where the event happened (one or 0.07 % missing); the business line; the event type; the type of insurance; the year of the event; the gross loss amount in GBP (437 or 31.51 % missing); the net loss amount in GBP (1353 or 97.55 % missing); the regulator involved (838 or 60.42 % missing), the source from which the information was drawn (29 or 2.09 % missing); and a loss description. The data has been sourced from various webpages, newspapers, press releases, regulator announcements, and databases (Bloomberg, Yahoo Finance etc.). In later years, webpages were used predominantly (including, for example, Reuters, Financial Times, and BBC). The reported OpRisk loss events happened in 62 different countries. The countries with the largest number of events in the considered time period are the USA (615 or 44.34 % of the reported events), the UK (361 or 26.03 %; additionally, there is one reported event on the Cayman Islands and two reported events on the Isle of Man), Japan (70 or 5.05 %), Australia (32 or 2.31 %), and India (28 or 2.02 %); 15 events (1.08 %) are reported under "Various" and not associated with a specific country. With regards to insurance, 887 of the events (63.95 %) were (partially) insured, 483 (34.82 %) were not insured (for the remaining 17 losses, 13 were missing and there was no information available yet for the other 4).

4.1 Descriptive analysis of the loss data

For our analysis of this database, we work with gross losses and adjust the loss amounts for inflation to the niveau of 2013 with the consumer price index as given in Table 50 in the document *Consumer*

---

[4] by Willis Professional Risks

Price Inflation Reference Tables, August 2013 obtained from the webpage `http://www.ons.gov.uk/ons/rel/cpi/consumer-price-indices/august-2013/index.html`. To obtain an inflation adjusted value (to the niveau of 2013) of a loss in a certain year, the loss is multiplied by the composite price index of 2013 divided by the composite price index of the loss' year. Table 1 shows a summary of selected information about the ten largest losses, including the corresponding business line (BL) which is Agency Services (AS), Asset Management (AM), Commercial Banking (CB), Corporate Finance (CF), Insurance (I), Payment and Settlement (PS), Retail Banking (RBa), Retail Brokerage (RBr), Trading and Sales (TS), or an unallocated business line (UBL). It also includes the event type (ET) which is Internal Fraud (IF), External Fraud (EF), Employment Practices and Workplace Safety (EPWS), Clients, Products, and Business Practices (CPBP), Damage to Physical Assets (DPA), Business Disruption and System Failures (BDSF), and Execution, Delivery, and Process Management (EDPM); see BIS (2006, p. 302 and p. 305) for more details on this classification. Due to space limitation, the description was largely shortened. Insurance coverage was present for all but the largest, third-largest, and fifth-largest of these losses. However, the database does not give information about the amount of insurance coverage.

Organization	Loss	BL	ET	Year	Description
Madoff and investors	40 819	AM	EF	2008	B. Madoff's Ponzi scheme
Parmalat	14 608	PS	EF	2003	Dubious transactions with funds on Cayman Islands
Bank of America et al.	12 500	CB	CPBP	2013	Fannie Mae claims regarding sold mortgage loans
Bank of America	9268	UBL	CPBP	2011	Settle Countrywide Financial Corp.'s residential mortgage-backed security repurchase exposure
BTA Bank	8518	CB	IF	2010	Fraud by chairman, diverting funds to companies he owned
Agricult. Bank of China	6173	CB	IF	2004	Financial crimes and bookkeeping irregularities
T. Imar Bankasi T.A.S.	5528	RBr	IF	2003	Fraudulent computer program
Société Générale	4548	UBL	IF	2008	A trader entered futures positions circumventing internal regulations
Banca Naz. del Lavoro	4407	RBr	IF	1989	Four counts of fraud
J.P. Morgan	3760	UBL	CPBP	2013	US authorities demand money due to mis-sold securities to Fannie Mae and Freddie Mac

Table 1 Summary of the ten largest losses (adjusted for inflation to the level of 2013; rounded to M GBP).

In what follows, we consider the 950 available gross losses in M GBP. The left-hand side of Figure 1 provides a graphical summary of these (inflation adjusted) available losses over time. On the right-hand side of Figure 1, the number of losses over time is given per business line. There are obvious concerns about changes in measurement. Furthermore, the increasing frequency of events is probably also due to improved record-keeping (reporting bias), a feature that our model may take into account; see Section 4.2.

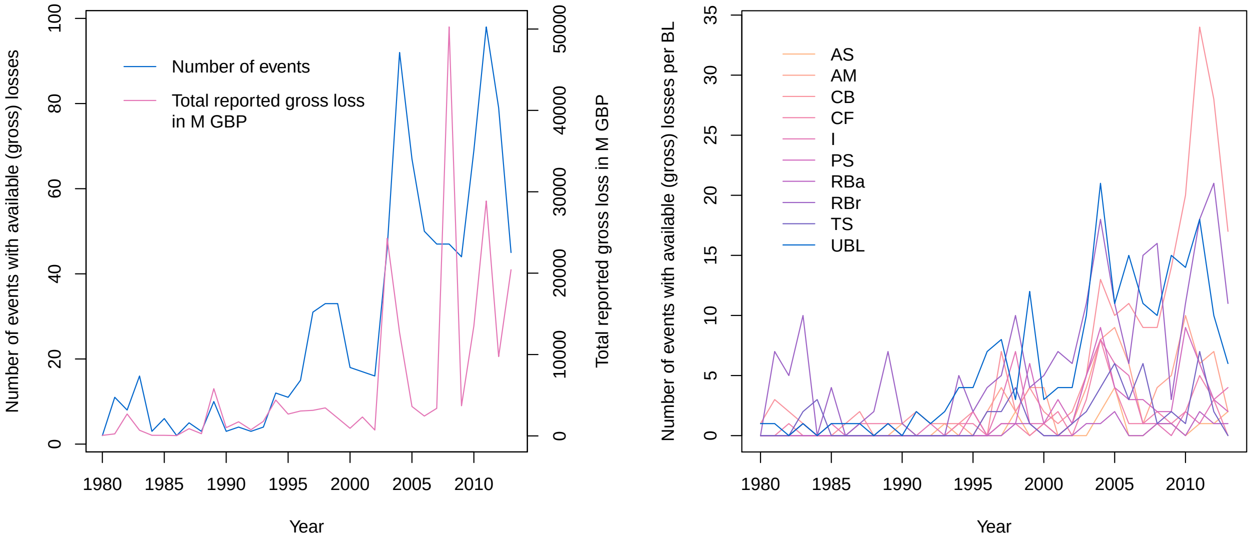

Figure 1 OpRisk data. Number of available losses and total available gross losses (aggregated per year) over time (left). Number of available losses for each business line over time (right).

Figure 2 shows the available, positive gross losses in M GBP per business line on a logarithmic scale. It seems obvious from this figure that the losses in different business lines are not identically distributed and thus can not be modeled with the same parameters. Our model will take this into account by interpreting the business lines (or event types) as covariates. One particular feature of our approach is that the model can be fitted to all the data simultaneously. This allows us to "pool" the data, that is, all the data is used to fit the model, not just the data available for a certain business line (or event type); the latter would not be feasible due to the data being sparse (especially over time).

Remark 4.1

The OpRisk database under study is clearly fairly limited with respect to sample size. Compare for instance with the sample sizes in the simulated example in Section A.2. Our analysis hence should be viewed as a first step towards an LDA model. Before its introduction into practice, more extensive datasets are needed. On the other hand, as we shall see, our analysis clearly is able to capture main features of OpRisk (and similar) data.

The *Basel matrix* B contains the number of available gross losses divided into *business lines* (rows) and *event types* (columns). In our case, this is a $(10,7)$-matrix. By summing, for each business line, over the different event types, one obtains the corresponding *Basel vector* b. Based on the 950

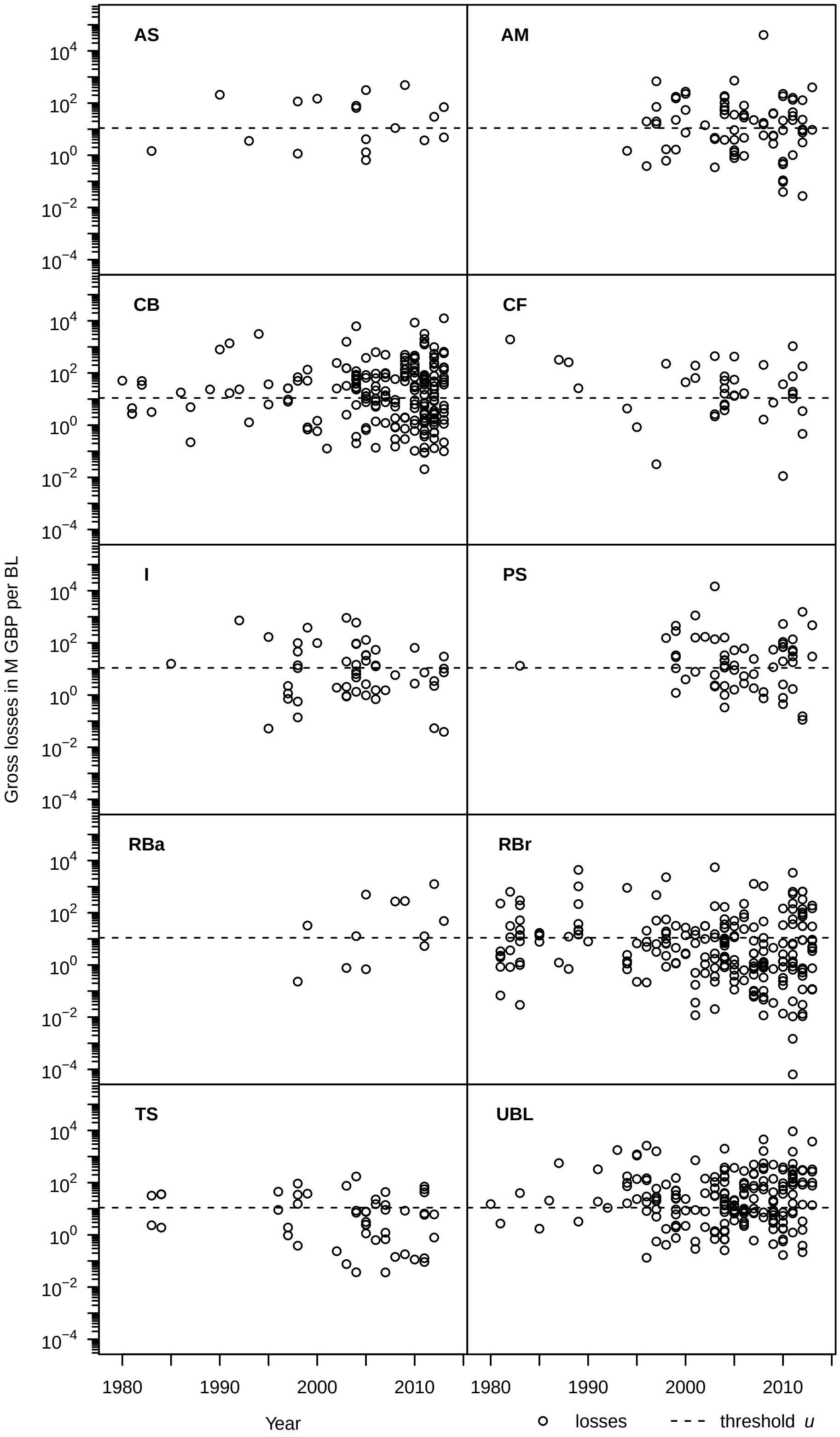

Figure 2 OpRisk data. Available positive gross losses in M GBP per business line including the threshold u chosen in Section 4.2.1 (median of all losses).

available gross losses since 1980, the Basel matrix B and the Basel vector b are given by

$$B = \begin{array}{c} \begin{array}{ccccccc} \text{IF} & \text{EF} & \text{EPWS} & \text{CPBP} & \text{DPA} & \text{BDSF} & \text{EDPM} \end{array} \\ \left(\begin{array}{rrrrrrr} 2 & 1 & 0 & 12 & 0 & 0 & 3 \\ 12 & 3 & 4 & 55 & 0 & 0 & 6 \\ 60 & 54 & 4 & 77 & 1 & 0 & 11 \\ 12 & 4 & 0 & 23 & 0 & 0 & 2 \\ 13 & 2 & 2 & 32 & 0 & 0 & 4 \\ 10 & 3 & 0 & 38 & 2 & 0 & 9 \\ 3 & 0 & 0 & 4 & 0 & 2 & 3 \\ 71 & 62 & 5 & 73 & 1 & 0 & 14 \\ 13 & 3 & 2 & 28 & 0 & 1 & 2 \\ 60 & 2 & 20 & 107 & 0 & 3 & 10 \end{array}\right) \end{array} \begin{array}{l} \\ \text{AS} \\ \text{AM} \\ \text{CB} \\ \text{CF} \\ \text{I} \\ \text{PS} \\ \text{RBa} \\ \text{RBr} \\ \text{TS} \\ \text{UBL} \end{array} \quad \text{and} \quad b = \left(\begin{array}{r} 18 \\ 80 \\ 207 \\ 41 \\ 53 \\ 62 \\ 12 \\ 226 \\ 49 \\ 202 \end{array}\right) \begin{array}{l} \text{AS} \\ \text{AM} \\ \text{CB} \\ \text{CF} \\ \text{I} \\ \text{PS} \\ \text{RBa} \\ \text{RBr} \\ \text{TS} \\ \text{UBL} \end{array} .$$

Note that the right-hand side of Figure 1 indeed shows the evolution of the Basel vectors over time, so $(b_t)_{t\in\{1980,\dots,2013\}}$, where $b = \sum_{t=1980}^{2013} b_t$.

4.2 Dynamic EVT modeling

We now apply the methodology presented in Section 3 to estimate the different parameters entering the calculation of the annual $\mathrm{VaR}_{0.999}$, including pointwise two-sided 95% confidence intervals. Due to the data being sparse, we "aggregate", for each business line, the losses over all event types; in other words, we proceed as if the label "event type" of the losses was not given. This is common practice in OpRisk modeling and justified by the fact that business lines are often operated separately and thus require a separate consideration. In Sections 4.2.1 and 4.2.2 we focus on the dynamic POT approach, in Section 4.2.3 we present the dynamic block maxima approach.

4.2.1 Dynamic POT analysis

In this section, we consider the median of all losses as threshold u (that is, 11.02 M GBP) and the 475 gross losses from 1980 to 2013 which exceed u. The reason for this choice and the sensitivity of our analysis on the threshold are addressed in Section 4.2.2.

The choice of the smoothing parameters (γ_λ for λ and γ_ξ, γ_ν for the GPD related parameters ξ, ν) or, equivalently, the degrees of freedom, depends on the aim of the analysis. If the purpose is purely explorative, the smoothers can be made suitably large to suit the situation by interpolating the data more. We automatically select the smoothing parameters with the AIC criterion (18). Another option would be cross-validation, but this can become computationally demanding when the number of losses is large. We also adopt informal inference based on likelihood-ratio statistics to assess model validation and comparison.

In applications dealing with OpRisk losses, it is not rare that ES does not exist because the estimated shape parameter ξ is larger than 1. This turns out to be the case here for several business lines; see Figure 5. Hence, for the remaining part of the paper, we can only present results about VaR.

Loss frequency
We fit the following models for λ:

$$\log \lambda(x,t) = c_\lambda, \tag{19}$$
$$\log \lambda(x,t) = f_\lambda(x), \tag{20}$$
$$\log \lambda(x,t) = f_\lambda(x) + c_\lambda t. \tag{21}$$

Model (19) is the constant model, neither depending on business lines nor on time. Model (20) depends on business lines (x is a `factor` running in AS, AM, CB, CF, I, PS, RBa, RBr, TS, UBL) but not on time. Model (21) depends on business lines x and, parametrically, on time t (year in 1980 to 2013) as covariates.

Comparing Models (19) with (20) and (20) with (21) via likelihood-ratio tests leads to both business line and time being significant. With the AIC criterion, we then compare Model (21) to models of the form

$$\log \lambda(x,t) = f_\lambda(x) + h_\lambda^{(\mathrm{Df})}(t), \tag{22}$$

for natural cubic splines $h_\lambda^{(\mathrm{Df})}$ with degrees of freedom $\mathrm{Df} \in \{1,\dots,8\}$; see Figure 3. Note that the case $\mathrm{Df} = 1$ corresponds to Model (21). It is clear from this curve, that a non-parametric dependence on time is suggested. We use the “elbow criterion” to choose the optimal degrees of freedom as 3; to be more precise, in Figure 3 we choose the smallest degrees of freedom so that adding another degree does not lead to a smaller AIC.

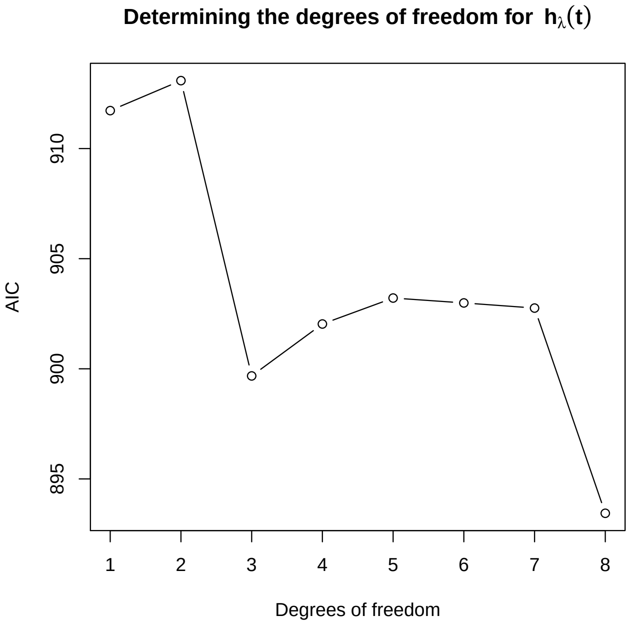

Figure 3 OpRisk data. AIC curve for fitted models of the form $\log \lambda(x,t) = f_\lambda(x) + h_\lambda^{(\mathrm{Df})}(t)$ with different degrees of freedom $\mathrm{Df} \in \{1,\dots,8\}$.

We therefore select the estimated model

$$\log \hat{\lambda}(x,t) = \hat{f}_\lambda(x) + \hat{h}_\lambda^{(3)}(t) \tag{23}$$

for λ, where x denotes the corresponding business line and $\hat{h}_\lambda^{(3)}(t)$ is a natural cubic spline with 3 degrees of freedom. In particular, the selected model shows that considering a homogeneous Poisson process for the occurrence of losses is not adequate. Figure 4 shows $\hat{\lambda}$ for each business line from 1980 to 2013 (solid lines: predicted values; filled dots: fitted values; dashed lines: pointwise asymptotic 95% confidence intervals). Overall, the curves show an increasing pattern through time for all business lines.

Loss severity

We now use the methodology explained in Section 3 to fit dynamic models for the GPD parameters (ξ, ν); recall the reparameterization (8). We fit the following models for (ξ, ν):

$$\xi(x,t) = c_\xi, \qquad \nu(x,t) = c_\nu, \tag{24}$$

$$\xi(x,t) = f_\xi(x), \qquad \nu(x,t) = c_\nu, \tag{25}$$

$$\xi(x,t) = f_\xi(x) + c_\xi t, \qquad \nu(x,t) = c_\nu, \tag{26}$$

$$\xi(x,t) = f_\xi(x), \qquad \nu(x,t) = f_\nu(x), \tag{27}$$

$$\xi(x,t) = f_\xi(x), \qquad \nu(x,t) = f_\nu(x) + c_\nu t, \tag{28}$$

$$\xi(x,t) = f_\xi(x), \qquad \nu(x,t) = f_\nu(x) + h_\nu(t), \tag{29}$$

where h_ν is a smoothed function of t (with variable degrees of freedom). The likelihood-ratio test based on Models (24) and (25) reveals that business line has a significant effect on ξ. Comparing Models (25) with (26) indicates that time does not have a significant effect on ξ. We therefore use Model (25) for ξ. As the test shows for Models (25) and (27), business line is significant for ν. Comparing Models (28) and (27) with the likelihood-ratio test indicates that time also has a significant effect on ν. Finally, Model (29) does not lead to a significant improvement of (28), the estimated degrees of freedom are very close to one. Therefore, ν in (28) is linear in time. Finally, we select the estimated models

$$\hat{\xi}(x,t) = \hat{f}_\xi(x), \qquad \hat{\nu}(x,t) = \hat{f}_\nu(x) + \hat{c}_\nu t. \tag{30}$$

The left-hand side of Figure 5 shows the estimates $\hat{\xi}$ for every business line and corresponding bootstrapped two-sided 95% confidence intervals for our dataset. The right-hand side shows the results of Moscadelli (2004, p. 40) for a broad comparison. Note that, with two exceptions, all estimates (left-hand side) are larger than 1, leading to infinite-mean models (roughly in line with the estimates found by Moscadelli (2004, p. 40) based on a much larger database). However, the confidence intervals for all except two business lines also cover values below 1. Figure 2 justifies somehow that ξ is negative for the business lines AS and TS for which the choice of the threshold is rather "high" for these two (sub)datasets, leading to low chance of sudden appearance of high losses, that are, furthermore, not so high above the threshold. The contrary is observed for the other (sub)datasets for which ξ is bigger than one: Figure 2 shows that for these business lines, there is a high chance (a high proportion of data above the threshold) of observing sudden high losses (with

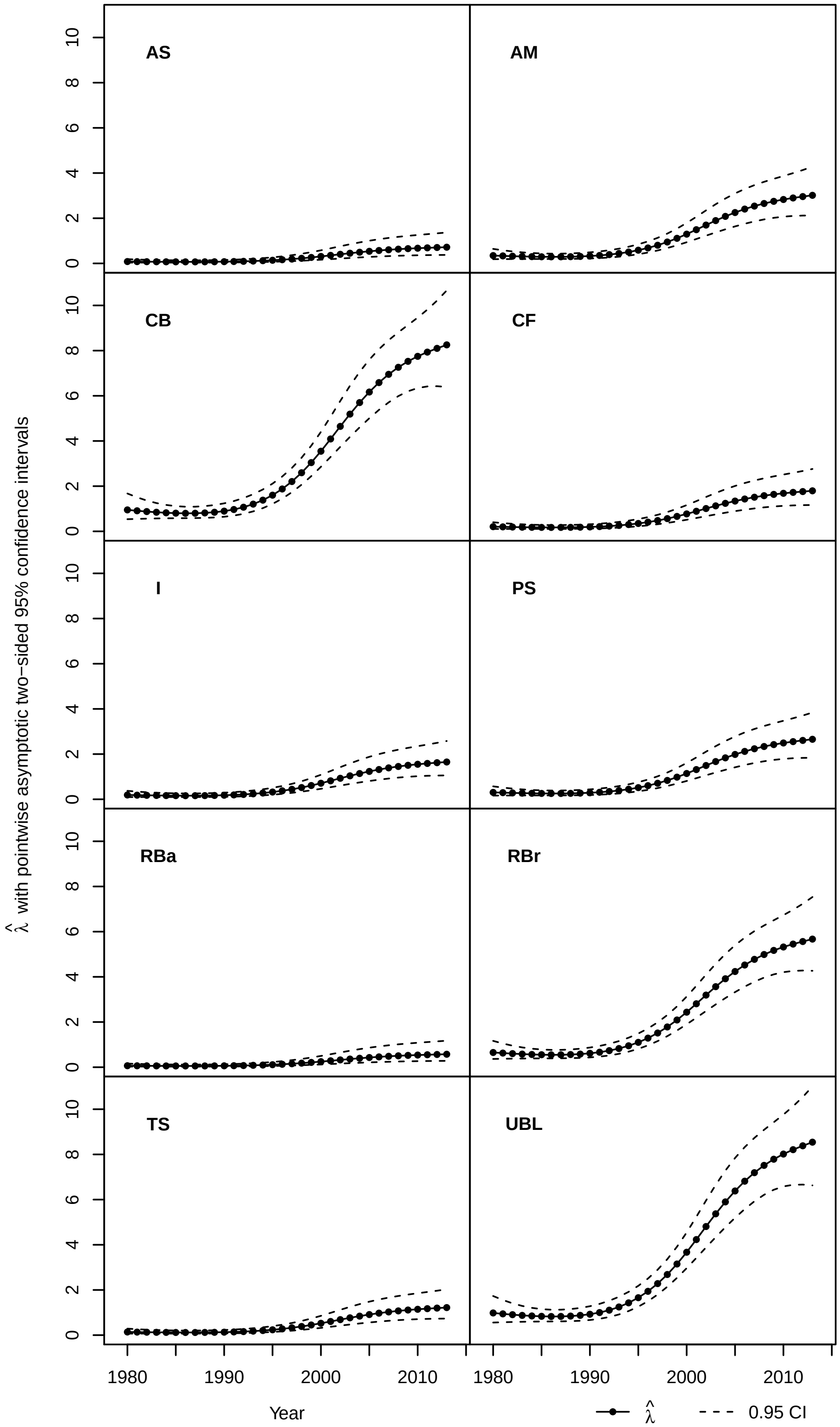

Figure 4 OpRisk data. Estimates $\hat{\lambda}$ including 95% confidence intervals depending on time and business line.

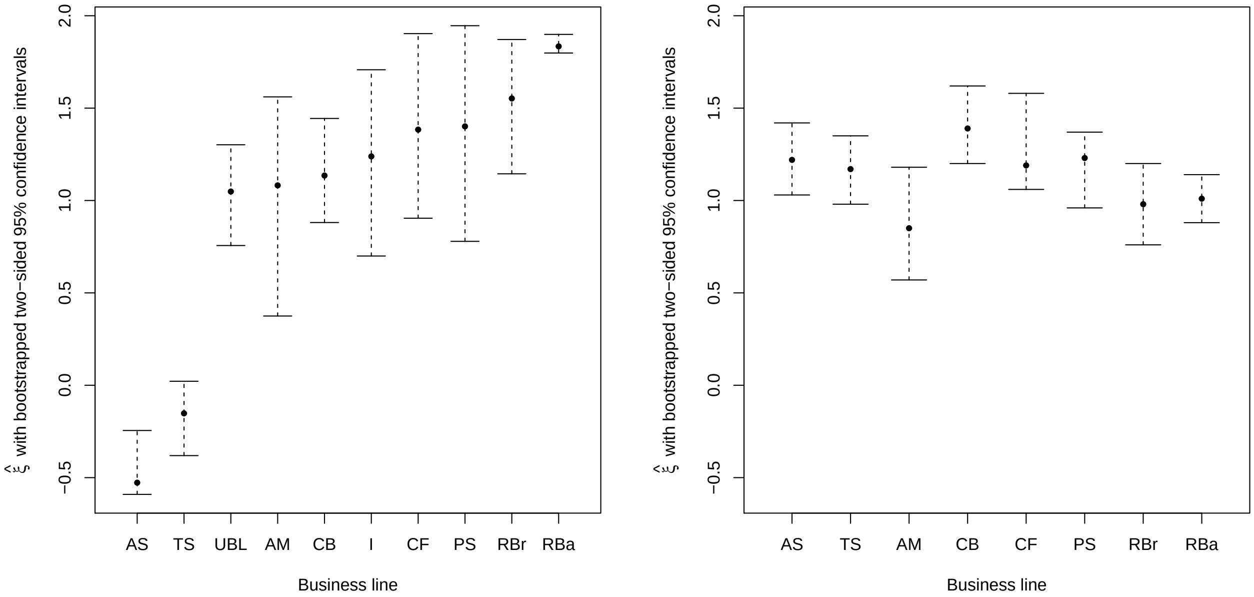

Figure 5 OpRisk data. Estimates $\hat{\xi}$ (left: our dataset; right: Moscadelli (2004)) including (bootstrapped) 95% confidence intervals depending on business lines.

values far above the threshold). The overall larger size of the confidence intervals in comparison to Moscadelli (2004) is due to the lack of data for each business line; see also Figures 1 and 2.

Figure 6 shows the corresponding estimates $\hat{\beta}$ including bootstrapped pointwise two-sided 95% confidence intervals depending on time and on business lines. The linear effect of time on ν is apparent and results in a slight increase of β over the years; this means more variability over time.

Figure 7 shows a Q-Q plot of the residuals (12) computed from the selected Model (30) for the GPD parameters against the quantiles of the standard exponential distribution including pointwise asymptotic 95% confidence intervals. It follows that there is no reason to reject the model based on the given data.

Figure 8 displays the effect of the selected models on $\mathrm{VaR}_{0.999}$. Both the effect of business line and the increase in occurrence of losses are visible. The confidence intervals are rather large for most of the business lines, indicating considerable uncertainty. We refrain from (over)interpreting the loss amounts estimated. As the data is gathered across several institutions, it is difficult to put them in perspective. As mentioned earlier, the main purpose of the paper is to show how the new methodology presented can be used for modeling OpRisk-type data; see also Embrechts and Puccetti (2008) for specifics of this kind of data from a QRM point of view. An ideal application would concentrate on loss data within one company for which then a common total capital denominator would be available.

4.2.2 Sensitivity with respect to the choice of the threshold and selected model

The outcome of a POT analysis is of course affected by the choice of the threshold u above which losses are modeled; see, for instance, de Fontnouvelle et al. (2004) and Shevchenko and Temnov

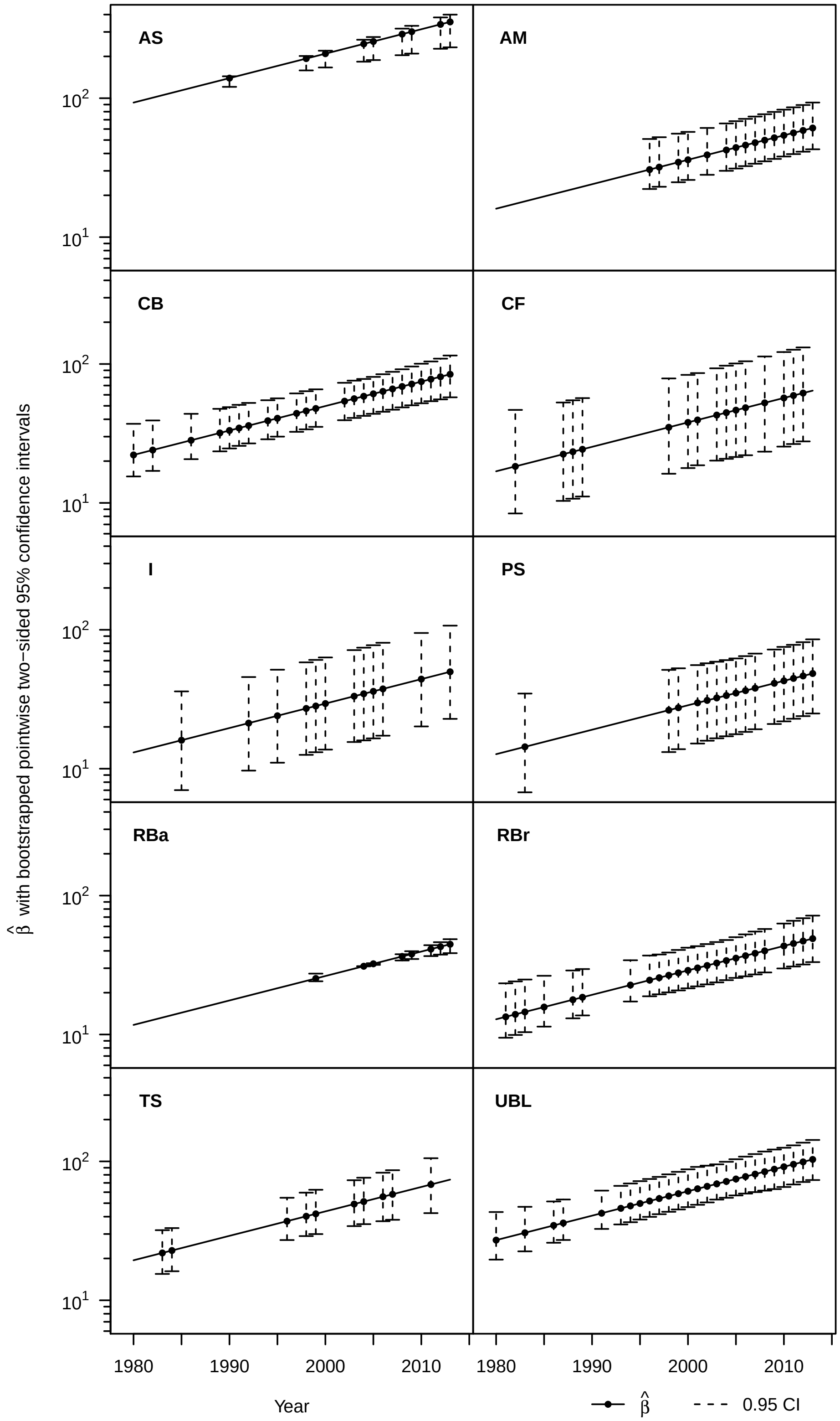

Figure 6 OpRisk data. Estimates $\hat{\beta}$ including bootstrapped pointwise two-sided 95% confidence intervals depending on time and business line.

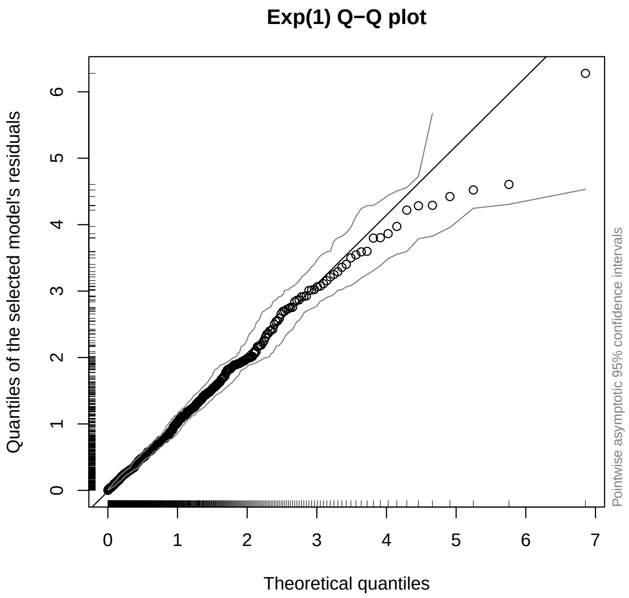

Figure 7 OpRisk data. Q-Q plot of the model's residuals depending on time and business lines.

(2009) for time varying thresholds. The key question is "by how much"? If the threshold is chosen too large, the number of losses is small which increases the variance in the statistical estimation; if the threshold is chosen too small, the asymptotic results by Pickands–Balkema–de Haan from Section 2.2 and Leadbetter (1991) might not provide a valid approximation and thus an estimation bias results. To assess the sensitivity of our modeling approach on the choice of the threshold, we conduct the same analysis with the threshold u being chosen as the 0-quantile (taking all data into account), 0.2-quantile, 0.3-quantile, and 0.4-quantile; recall that we chose the threshold u as the 0.5-quantile (median) above. The models selected for the loss severity are given in Table 2. Residual plots (not included) for smaller values of u showed clear departures from the model. We have indicated the increasing deterioration with decreasing u by the non-technical ↯-symbols. The residuals for the threshold $u = 0.5$ are given in Figure 7 and show a good overall behavior; whence we decided on this level of u.

Threshold u	Model for ξ	Model for ν	Q-Q
0-quantile	$\hat{\xi}(x,t) = \hat{f}_\xi(x) + \hat{c}_\xi t$	$\hat{\nu}(x,t) = \hat{f}_\nu(x) + \hat{c}_\nu t$	↯↯↯
0.2-quantile	$\hat{\xi}(x,t) = \hat{f}_\xi(x)$	$\hat{\nu}(x,t) = \hat{f}_\nu(x)$	↯↯
0.3-quantile	$\hat{\xi}(x,t) = \hat{f}_\xi(x)$	$\hat{\nu}(x,t) = \hat{f}_\nu(x)$	↯↯
0.4-quantile	$\hat{\xi}(x,t) = \hat{f}_\xi(x)$	$\hat{\nu}(x,t) = \hat{f}_\nu(x)$	↯
0.5-quantile	$\hat{\xi}(x,t) = \hat{f}_\xi(x)$	$\hat{\nu}(x,t) = \hat{f}_\nu(x) + \hat{c}_\nu t$	✓

Table 2 Selected models for the loss severity depending on the choice of threshold.

Another question is how sensitive the selected models (23) and (30) are to the modeled time period. To address this, we repeated the whole analysis conducted in Section 4.2.1 but only on

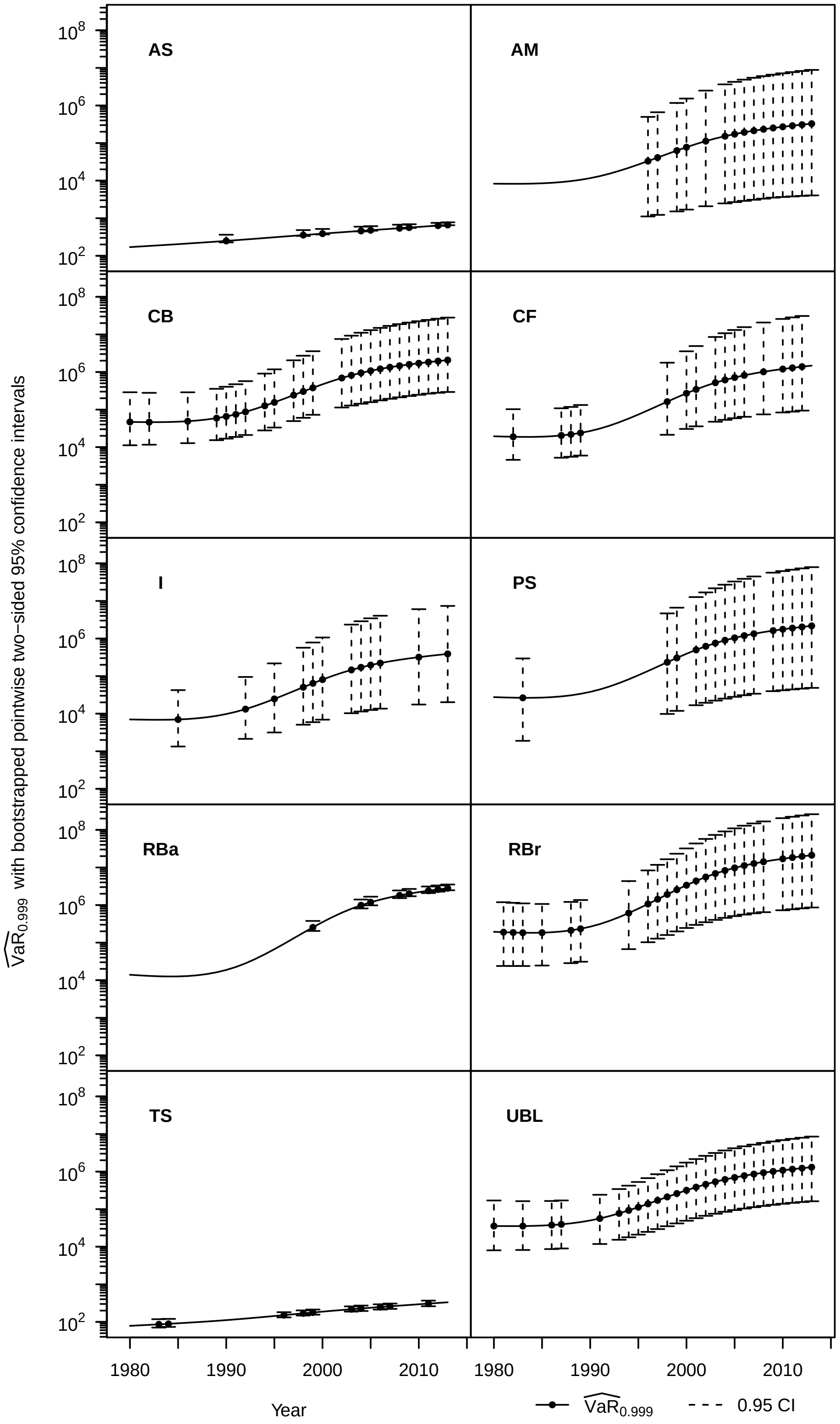

Figure 8 OpRisk data. Estimates $\widehat{\mathrm{VaR}}_{0.999}$ including bootstrapped pointwise two-sided 95% confidence intervals depending on time and business line.

the data from 1980 to 2011. This leaves us with 1192 losses, for 826 the loss amount is available. As before, the threshold u is chosen as the median of the data. We have not included the explicit results of this analysis but only summarize our main findings. The selected loss frequency model is the same as (23). Also, the selected model for the loss severity parameter ξ is the same as before, see (30). Concerning the loss severity parameter ν, a slightly simpler model than (30) is selected, not depending on time. The residual Q-Q plot is also fine for this model, similar to Figure 7. The impact on $\widehat{\mathrm{VaR}}_{0.999}$ is barely visible, the estimates including confidence intervals indeed look similar to Figure 8.

4.2.3 Dynamic block maxima method

In this section, we apply the dynamic block maxima method to the yearly maxima of the logarithmic gross losses observed from 1980 to 2013; note that in Section A.2.2, a small simulation study is presented to show how the dynamic block maxima methodology works. The analysis is motivated by a clear increasing trend of the log-transformed maxima over years. Because there are only 34 yearly maxima available, we do not consider other covariates (business lines or event types) than year.

If one is interested in VaR_α, it can sometimes be useful to work with the log-transformed data. This is allowed since for $0 < X \sim F_X$ and $\log X \sim F_{\log X}$,

$$\mathrm{VaR}_\alpha(X) = F_X^{-1}(\alpha) = \exp(F_{\log X}^{-1}(\alpha)) = \exp(\mathrm{VaR}_\alpha(\log X)). \tag{31}$$

A well-known property from EVT states that the log-transform of a random variable (distribution function) in the maximum domain of attraction (MDA) of a Fréchet ($\xi > 0$ in (1) and (2)) belongs to the MDA of the Gumbel ($\xi = 0$). A special case concerns exact Pareto distribution functions, the log-transform of which is exponential. See Embrechts et al. (1997, p. 148, Example 3.3.33) and further examples therein. Besides the relevant (31), log-transformation moves data from the heavy-tailed (even infinite-mean) case to a more standard short-tailed environment. For the OpRisk data, the left-hand side of Figure 9 shows for the log-transformed data an increasing trend over the years; the symbols represent different business lines. The business lines AS and RBa are never yearly maxima for this period of time. Because we only have a small amount of yearly maxima and because the different business lines all seem to follow the increasing trend in time, in a first analysis, it seems reasonable to let only the location parameter μ of the GEV distribution (2) depend on time (and not on business lines nor event type). We use the penalized log-likelihood of the form (6) for the block maxima approach (GEV), penalized for $\mu = \mu(t)$. The resulting dynamic model is a fully non-parametric model

$$\hat{\mu} = \hat{\mu}(t) = \hat{h}_\mu^{(2)}(t),$$

that is, a natural cubic spline with two degrees of freedom. The scale and shape parameters are set to be constant and their estimates are $\hat{\sigma} = 1.52$ (0.25) and $\hat{\xi} = -0.39$ (0.10) (standard errors in parentheses). Note that the model suggests a short tail for the log-losses ($\hat{\xi} < 0$). Using (3), where $\hat{\mu}$ is replaced by $\hat{h}_\mu^{(2)}(t)$ and the two other parameters by their estimates, we obtain an estimated curve for $\mathrm{VaR}_{0.999}$ including bootstrapped pointwise two-sided 95% confidence intervals for the yearly log-loss maxima and for the original yearly loss maxima using the relation (31); see the right-hand side of Figure 9. The confidence intervals are rather large because of the small number of maxima, highlighting strong uncertainty of the model. Nevertheless, the estimated $\mathrm{VaR}_{0.999}$

curve captures the increasing trend of the yearly maxima over time (all business lines and event types superimposed). The results are not comparable but complementary to the ones obtained from the dynamic POT model in Section 4.2.1. First, (much) less data and therefore less information is available. Second, it has not been possible to incorporate the additional information about the business lines. Third, the model is applied to yearly maxima and the corresponding VaR_α at level α (3) is the value of the (log-)loss maxima expected to be exceeded once every $1/(1-\alpha)$ years. The curve $\widehat{\mathrm{VaR}}_{0.999}$ on the right-hand side of Figure 9, therefore shows the estimated loss maxima that can be exceeded only “once every 1000 years on average”. Over our period under study, the resulting $\widehat{\mathrm{VaR}}_{0.999}$ has not been exceeded by the observed losses.

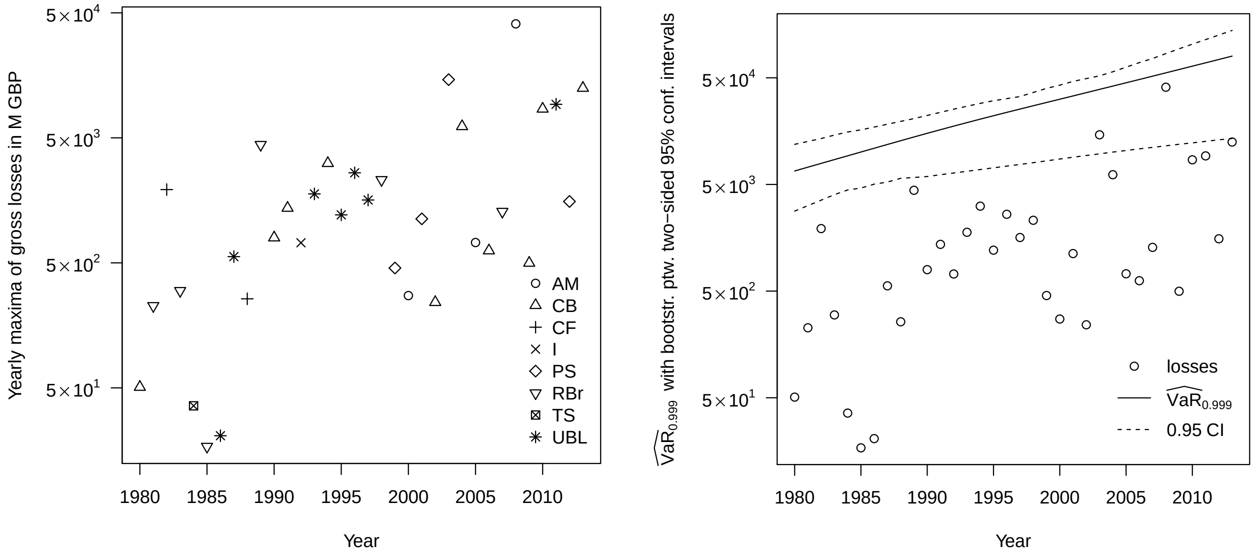

Figure 9 OpRisk data. Yearly loss maxima in M GBP over the period 1980–2013 (left); symbols indicate business lines. Estimates $\widehat{\mathrm{VaR}}_{0.999}$ including bootstrapped pointwise two-sided 95% confidence intervals for the yearly loss maxima over time (right).

5 Discussion

In recent years, banks and insurance companies have paid increasing attention to OpRisk and there is a pressing need for a flexible statistical methodology for the modeling of extreme losses in this context. Recent actions by international regulation show ample evidence of this, especially in the aftermath of the subprime crisis (legal risk). Further events like the LIBOR-case and the worries about fixes in FX-markets will no doubt imply more pressure for risk capital increases. For instance, the Swiss regulation FINMA (October 2013) ordered a 50% increase of the regulatory capital for UBS, a large Swiss bank. Similar actions throughout the industry worldwide will no doubt follow. Based on currently (especially publicly) available OpRisk data, yearly high-quantile-type risk measures (like $\mathrm{VaR}_{0.999}$) are statistically hard to come by. It is to be hoped that more and more complete and reliable data will eventually become available. In anticipation of this, we developed the new methodology in this paper. We offer a EVT-based statistical approach that allows one to choose either appropriate loss frequency and loss severity distributions, or loss maximum distribution, by taking into account dependence of the parameters on covariates and time. The methodology is applied both to simulated data (see Section A.2) and to a database of publicly available OpRisk losses (see Section 4) which shows several stylized features of industry-wide OpRisk data.

Some of the assumptions and consequences of the presented EVT-based methodology require a discussion:

- It is important to say something on infinite mean models resulting from fitted $\hat{\xi}$'s strictly larger than 1 as we partly saw in Section 4. In the context of OpRisk, several analyses based on unbounded severity models have resulted in such $\hat{\xi}$'s. One can, and should question such models; industry has used various techniques avoiding these models, such as tapering or truncation from above; see for instance Kagan and Schoenberg (2001) in the context of seismology. For an interesting overview of these techniques together with applications to wildfire-sizes, earthquake inter-event times and stock-price data, see Patel (2011). Our experience is that for data of such very heavy-tailed type, whatever fitting adjustment one tries, capital charges are highly depending on the tapering or truncation mechanism used. A fully unconditioned POT analysis (possibly resulting in $\hat{\xi} > 1$) points into a direction of extreme uncertainty of high quantile (VaR) estimates. A relevant, thought provoking paper on the economics of climate change in the presence of (very) heavy-tailed risks is Weitzman (2009) and the various reactions on that paper, like Nordhaus (2009). The catch phrase reflecting on uncertainty coming from infinite mean models is "The Dismal Theorem". We quote from Weitzman (2009) as it very much reflects our own thinking: "The economics of fat-tailed catastrophes raises difficult conceptual issues which cause the analysis to appear less scientifically conclusive and to look more contentiously subjective than what comes out of [...] more usual thin-tailed situations. But if this is the way things are with fat tails, then this is the way things are, and it is an inconvenient truth to be lived with rather than a fact to be evaded just because it looks less scientifically objective in cost-benefit applications". Further relevant papers on the topic include Embrechts et al. (2013b), Ibragimov and Walden (2007) and Ibragimov and Walden (2008). Das et al. (2013) contains a discussion in the broader context of model uncertainty within Quantitative Risk Management. For heavy-tailed data, $\hat{\xi} > 1/3$ or $1/4$, say, the so-called *one big jump heuristic* kicks in: "One large loss dwarfs the aggregate of all other losses over a given time period". This is a well-known phenomenon which has precise mathematical formulations and consequences; see for instance Embrechts et al. (1997, Section 8.3.3). Recent fraud (hence OpRisk) events have unfortunately shown how much this principle holds in the world of banking.
- In theory, the model allows all possible interactions between the covariates, in practice this is only possible for large sample sizes. The simulated example in Section A.2 partly considers interactions between the variables "group" and "year". This has been possible because enough data has been drawn for each combination of interaction level. Furthermore, for sufficiently large datasets, asymptotic distribution theory provides tools for inference and assessment of model fit based on the notion of deviance and its associated degrees of freedom of an appropriate χ^2 statistic; see Nelder and Wedderburn (1972). In the context of small samples, standard approximate distributions for the statistics can be unreliable. For more discussion on these points, see, for instance, Hastie and Tibshirani (1990) and Appendix B. Indeed "large enough" is difficult to define. For the OpRisk data considered in this paper, we found it reasonable not to include the covariate "event type" (in order to gain in sample size). We are aware that a larger dataset would

usefully allow us to fit all possible interactions (combinations) of "business line", "event type" and "year" and also a more reliable comparison of the fitting procedure using likelihood ratio statistics.

- Note that we do not model a specific business line at a given time point. For this one would clearly not have sufficient data for all year–business lines combinations. We fit a model to all available losses simultaneously. This is what we mean by "pooling". Although considering all data (pooling) provides higher global information and significance for the business-line effect, the lack of data (at least for some business lines) and possible heterogeneity across risk types may lead to model misspecification and should be carefully looked at.
- In Section 4.2.3, we use log-transformed data. As mentioned, in practice, it has the effect of reducing the estimated value of the shape parameter ξ. Theoretical proofs of the latter for some distributions are provided in Embrechts et al. (1997, p. 148). The numerical reduction from $\hat{\xi} > 1$ obtained from the original data to $\hat{\xi} < 1$ for the log-transformed data may sometimes be useful; for $\xi > 1$ (infinite-mean) the usual Taylor expansions can be made but do not yield a consistent estimator. In this case, the log-likelihood may be arbitrarily large, so a maximum likelihood estimators may take values which correspond to local maxima of the log-likelihood; see Davison and Smith (1990).
- The fact that we implemented the statistical fitting procedure, the bootstrap, and relevant model functionals in R is not relevant. The reason why we chose R is that it is open source, widely known in the statistical community, provides an implementation for the (non-trivial) fitting of generalized additive models, and graphics. For implementing the model in another programming language, a fitting procedure for generalized additive models has to be provided in addition to the functions we implemented; see Wood (2006) for details.
- The field of applications clearly extends to broader domains of finance, insurance, environmental risk, and any industrial domain where the modeling of extremes or extremal point measures like VaR or ES depending on covariates is of interest.

A Appendix

A.1 Derivatives of the reparameterized log-likelihood

Algorithm 3.1 requires to compute the first two derivatives of the reparameterized log-likelihood with respect to ξ and ν. We now present these ingredients. The reparameterized log-likelihood contribution is

$$\begin{aligned}\ell^r(\xi,\nu;y) &= \ell(\xi,\exp(\nu)/(1+\xi);y)\\ &= \begin{cases}\log(1+\xi)-\nu-(1+1/\xi)\log(1+\xi(1+\xi)\exp(-\nu)y), & \text{case 1,}\\ -(\nu+\exp(-\nu)y), & \text{case 2,}\\ -\infty, & \text{otherwise,}\end{cases}\end{aligned}$$

with case 1 being the set where $\xi > 0$ and $y \geq 0$, or $\xi < 0$ and $y \in [0, -\exp(\nu)/(\xi(1+\xi)))$, and case 2 the one where $\xi = 0$. This implies

$$\begin{aligned}
&\ell^r_\xi(\xi, \nu; y) = \frac{\partial}{\partial \xi} \ell^r(\xi, \nu; y) \\
&= \begin{cases} 1/(1+\xi) + \log(1+\xi(1+\xi)\exp(-\nu)y)/\xi^2 - (1+1/\xi)\frac{(1+2\xi)\exp(-\nu)y}{1+\xi(1+\xi)\exp(-\nu)y}, & \text{case 1,} \\ 0, & \text{otherwise} \end{cases}
\end{aligned}$$

and

$$\begin{aligned}
&\ell^r_{\xi\xi}(\xi, \nu; y) = \frac{\partial^2}{\partial \xi^2} \ell^r(\xi, \nu; y) \\
&= \begin{cases} -1/(1+\xi)^2 - 2\log(1+\xi(1+\xi)\exp(-\nu)y)/\xi^3 + \frac{2(1+2\xi)\exp(-\nu)y}{\xi^2(1+\xi(1+\xi)\exp(-\nu)y)} & \\ -(1+1/\xi)\exp(-\nu)y\frac{2(1+\xi(1+\xi)\exp(-\nu)y)-(1+2\xi)^2\exp(-\nu)y}{(1+\xi(1+\xi)\exp(-\nu)y)^2}, & \text{case 1,} \\ 0, & \text{otherwise.} \end{cases}
\end{aligned}$$

Furthermore,

$$\ell^r_\nu(\xi, \nu; y) = \frac{\partial}{\partial \nu} \ell^r(\xi, \nu; y) = \begin{cases} \frac{-1+(1+\xi)\exp(-\nu)y}{1+\xi(1+\xi)\exp(-\nu)y}, & \text{case 1,} \\ -1 + \exp(-\nu)y, & \text{case 2,} \\ 0, & \text{otherwise} \end{cases}$$

and

$$\ell^r_{\nu\nu}(\xi, \nu; y) = \frac{\partial^2}{\partial \nu^2} \ell^r(\xi, \nu; y) = \begin{cases} \frac{-(1+\xi)^2\exp(-\nu)y}{(1+\xi(1+\xi)\exp(-\nu)y)^2}, & \text{case 1,} \\ -\exp(-\nu)y, & \text{case 2,} \\ 0, & \text{otherwise.} \end{cases}$$

By replacing ξ by ξ_i, ν by ν_i, and y by y_{t_i} above, we obtain the required derivatives of the reparameterized log-likelihood as given in (14).

A.2 Demonstration of the dynamic approaches based on simulated data

In this section, we provide an example based on simulated data (sample size of 2000) to check correctness of our implementation and to indicate how/that our methodology works. We only briefly discuss this example; all details for the POT approach, including more plots, can be accessed via `game(demo)` in `QRM`.

A.2.1 The dynamic POT approach

We generate a dataset of losses over a time period of 10 years for two groups (factor with levels “A” and “B”). The simulated losses are drawn from a (non-stationary) GPD depending on the covariates “year” and “group”. We then fit the (Poisson process) intensity λ and the two GPD parameters ξ and β depending on “year” and “group” and compute bootstrapped confidence intervals using the methodology presented in Sections 3.1 to 3.3. The precise models fitted for the loss frequency and severity are

$$
\begin{aligned}
\log \lambda(x,t) &= f_\lambda(x) + h_\lambda^{(3)}(t,x),\\
\xi(x,t) &= f_\xi(x) + h_\xi^{(3)}(t,x),\\
\nu(x,t) &= f_\nu(x) + h_\nu^{(3)}(t,x),
\end{aligned}
$$

where x denotes the corresponding group and $h_\lambda^{(3)}, h_\xi^{(3)}, h_\nu^{(3)}$ are group-specific (hence the interaction with x) natural cubic splines with 3 degrees of freedom. Finally, we compute dynamic estimates of VaR_α; see Figure 10. Overall, as a comparison with the true parameters and values indicates, our methodology can be applied to fit a model to the entire dataset and captures the different functional forms driven by the covariates.

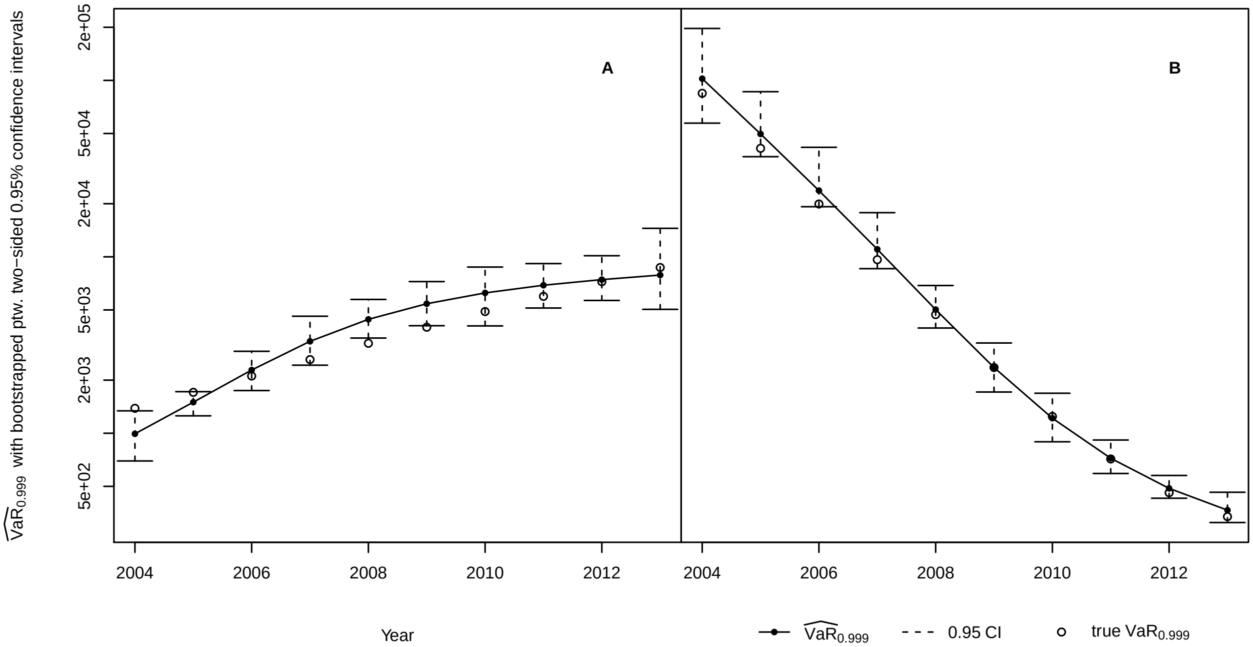

Figure 10 Simulated data. Estimates $\widehat{\mathrm{VaR}}_{0.999}$ including bootstrapped pointwise two-sided 95% confidence intervals depending on time and group, obtained from the dynamic POT method.

A.2.2 The dynamic block maxima approach

We run a similar simulation study to assess the performance of the dynamic GEV approach introduced in Section 4.2.3. We generate a dataset of losses over a time period of 10 years for two groups (factor with levels “A” and “B”). The simulated losses are drawn from a (non-stationary) GEV with location parameter μ depending on the covariates “year” and “group” using non-linear functions. We fit the semi-parametric model

$$\mu(x,t) = f_\mu(x) + h_\mu^{(3)}(t,x)$$

for the location parameter μ, where, again, x denotes the corresponding group and $h_\mu^{(3)}$ is a group-specific natural cubic splines with 3 degrees of freedom. The shape and scale parameters are set constant, like in the application of Section 4.2.3. We then compute dynamic estimates of VaR_α; see Figure 11.

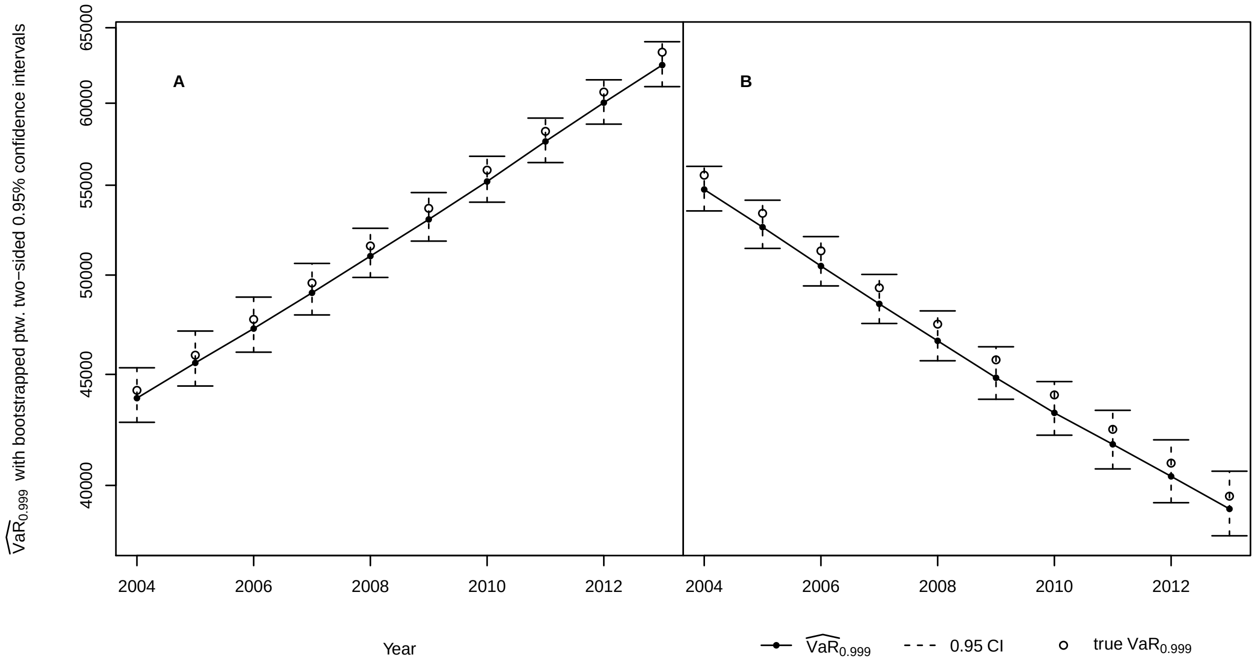

Figure 11 Simulated data. Estimates $\widehat{\text{VaR}}_{0.999}$ including bootstrapped pointwise two-sided 95% confidence intervals depending on time and group, obtained from the dynamic GEV block maxima method.

Acknowledgement

The authors would like to thank the referees and an editor for various comments that led to a much improved version of the paper.

References

Akkizidis, I. S. and Bouchereau, V. (2006), Guide to Optimal Operational Risk and Basel II, Boca Raton, New York: Auerbach Publications.

Aue, F. and Kalkbrener, M. (2006), LDA at work: Deutsche Bank's approach to quantifying operational risk, *Operational Risk*, 1(4), Deutsche Bank AG, Frankfurt, Germany, 49–93.

Baud, N., Frachot, A., and Roncalli, T. (2002), Internal Data, External Data and Consortium Data – How to Mix Them for Measuring Operational Risk, Groupe de Recherche Opérationelle Crédit Lyonnais, `http://www.thierry-roncalli.com/#published` (2013-02-03).

Baud, N., Frachot, A., and Roncalli, T. (2003), How to avoid over-estimating capital charge for operational risk? *risk.net*, Groupe de Recherche Opérationelle Crédit Lyonnais.

BIS (2006), Basel II: International Convergence of Capital Measurement and Capital Standards: A Revised Framework – Comprehensive Version, `http://bis.org/publ/bcbs128.pdf` (2010-05-17).

BIS (2012), Fundamental review of the trading book, `http://www.bis.org/publ/bcbs219.pdf` (2012-02-03).

Böcker, K. (2010), Rethinking Risk Measurement and Reporting: Uncertainty, Bayesian Analysis and Expert Elicitation, ed. by K. Böcker, Risk Books, London.

Böcker, K. and Klüppelberg, C. (2010), Multivariate models for operational risk, *Quantitative Finance*, 10(8), 855–869.

Bolancé, C., Guillén, M., Gustafsson, J., and Nielsen, J. P. (2012), Quantitative Operational Risk Models, London: Chapman & Hall/Crc Finance Series.

Brechmann, E. C., Czado, C., and Paterlini, S. (2013), Flexible Dependence Modeling of Operational Risk Losses and Its Impact on Total Capital Requirements, `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2252481` (2013-10-26).

Chavez-Demoulin, V. (1999), Two Problems in Environmental Statistics: Capture-Recapture Analysis and Smooth Extremal Models, PhD thesis, EPFL.

Chavez-Demoulin, V. and Davison, A. C. (2005), Generalized additive models for sample extremes, *Applied Statistics*, 54(1), 207–222.

Chavez-Demoulin, V. and Embrechts, P. (2004), Smooth extremal models in finance and insurance, *Journal of Risk and Insurance*, 71(2), 183–199.

Chernobai, A., Jorion, P., and Yu, F. (2011), The Determinants of operational risk in U.S. financial institutions, *Journal of Financial and Quantitative Analysis*, 46(6), 1683–1725.

Coles, S. (2001), An Introduction to Statistical Modeling of Extreme Values, Springer.

Cont, R., Deguest, R., and Scandolo, G. (2010), Robustness and sensitivity analysis of risk measurement procedures, *Quantitative Finance*, 10(6), 593–606.

Cope, E. and Labbi, A. (2008), Operational loss scaling by exposure indicators: evidence from the ORX database, *Journal of Operational Risk*, 3(4), 25–46.

Cox, D. R. and Reid, N. (1987), Parameter Orthogonality and Approximate Conditional Inference, *Journal of the Royal Statistical Society: Series B*, 49(1), 1–39.

Cruz, M. G. (2002), Modeling, Measuring and Hedging Operational Risk, Chichester: Wiley.

Cummins, J. D., Lewis, C. M., and Wei, R. (2006), The market value impact of operational loss events for US banks and insurers, *Journal of Banking and Finance*, 30(10), 2605–2634.

Dahen, H. and Dionnne, G. (2010), Scaling models for the severity and frequency of external operational loss data, *Journal of Banking and Finance*, 34(7), 1484–1496.

Daníelsson, J., Embrechts, P., Goodhart, C., Keating, C., Muennich, F., Renault, O., and Shin, H. S. (2001), An Academic Response to Basel II, `http://www.math.ethz.ch/~embrechts/ftp/Responsev3.pdf` (2013-02-03).

Das, B., Embrechts, P., and Fasen, V. (2013), Four Theorems and a Financial Crisis, *International Journal of Approximate Reasoning*, to appear, `http://www.math.ethz.ch/~embrechts/ftp/Four_Theorems_2012.pdf` (2013-02-03).

Davison, A. C. and Smith, R. L. (1990), Models for exceedances over high thresholds (with Discussion), *Journal of the Royal Statistical Society. Series B (Methodological)*, 52(3), 393–442.

De Fontnouvelle, P., Jordan, J. S., DeJesus-Ureff, V., and Rosengren, E. S. (2004), Capital and risk: New evidence on implications of large operational losses, working paper 03-5, `http://ideas.repec.org/p/fip/fedbwp/03-5.html` (2013-02-14).

De Fontnouvelle, P., Rosengren, E. S., and Jordan, J. S. (2005), Implications of alternative operational risk modeling techniques, working paper 11103, `http://www.nber.org/papers/w11103` (2013-02-14).

Dutta, K. and Perry, J. (2006), A tale of tails: An empirical analysis of loss distribution models for estimating operational risk capital, working paper 06-13, `http://www.bos.frb.org/economic/wp/wp2006/wp0613.pdf` (2013-02-14).

El-Gamal, M., Inanoglu, H., and Stengel, M. (2007), Multivariate estimation for operational risk with judicious use of extreme value theory, *Journal of Operational Risk*, 2(1), 21–54.

Embrechts, P. and Puccetti, G. (2008), Aggregating operational risk across matrix structured loss data, *Journal of Operational Risk*, 3(2), 29–44.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997), Modelling Extremal Events for Insurance and Finance, Springer.

Embrechts, P., Puccetti, G., Rüschendorf, L., Wang, R., and Beleraj, A. (2013a), An academic response to Basel 3.5, Preprint, ETH Zurich.

Embrechts, P., Puccetti, G., and Rüschendorf, L. (2013b), Model uncertainty and VaR aggregation, *Journal of Banking and Finance*, 37(8), 2750–2764.

Frachot, A., Moudoulaud, O., and Roncalli, T. (2004), Loss Distribution Approach in Practice, *The Basel Handbook: A Guide for Financial Practitioners*, ed. by M. Ong, Groupe de Recherche Opérationelle Crédit Lyonnais.

Ganegoda, A. and Evans, J. (2013), A scaling model for severity of operational losses using generalized additive models for location scale and shape (GAMLSS), *Annals of Actuarial Science*, 7(1), 61–100.

Gneiting, T. (2011), Making and evaluating point forecasts, *Journal of the Americal Statistical Association*, 106, 746–762.

Green, P. J. and Silverman, B. W. (2000), Nonparametric Regression and Generalized Linear Models – A roughness penalty approach, Chapman & Hall/CRC.

Hastie, T. J. and Tibshirani, R. J. (1990), Generalized Additive Models, London: Chapman and Hall.

Ibragimov, R. and Walden, J. (2007), The limits of diversification when losses may be large, *Journal of Banking and Finance*, 31, 2551–2569.

Ibragimov, R. and Walden, J. (2008), Portfolio diversification under local and moderate deviations from power laws, *Insurance: Mathematics and Economics*, 42, 594–599.

Jarrow, R. A. (2008), Operational risk, *Journal of Banking and Finance*, 32, 870–879.

Jarrow, R. A., Oxman, J., and Yildirim, Y. (2010), The cost of operational risk loss insurance, *Review of Derivatives Research*, 13(3), 273–295.

Jorion, P. (2007), Value at Risk: The New Benchmark for Managing Financial Risk, 3rd ed., New York: McGraw-Hill.

Kagan, Y. Y. and Schoenberg, P. (2001), Estimation of the upper cutoff parameter for the tapered Pareto distribution, *Journal of Applied Probability*, 38, 158–175.

Leadbetter, M. R. (1991), On a basis for "Peaks over Threshold" modeling, *Statistics & Probability Letters*, 12, 357–362.

McNeil, A. J., Frey, R., and Embrechts, P. (2005), Quantitative Risk Management: Concepts, Techniques, Tools, Princeton University Press.

Mori, A., Kimata, T., and Nagafuji, T. (2007), The Effect of the Choice of the Loss Severity Distribution and the Parameter Estimation Method on Operational Risk Measurement: Analysis Using Sample Data, Bank of Japan.

Moscadelli, M. (2004), The modelling of operational risk: Experience with the analysis of the data collected by the Basel Committee, Technical Report 517, Banca d'Italia.

Na, H. S., van den Berg, J., Miranda, L. C., and Leipoldt, M. (2006), An econometric model to scale operational losses, *Journal of Operational Risk*, 1(2), 11–31.

Nelder, J. A. and Wedderburn, R. W. M. (1972), Generalized linear models, *Journal of the Royal Statistical Society: Series A*, 135, 370–384.

Nešlehová, J., Embrechts, P., and Chavez-Demoulin, V. (2006), Infinite mean models and the LDA for operational risk, *Journal of Operational Risk*, 1(1), 3–25.

Nordhaus, W. D. (2009), An analysis of the Dismal Theorem, Cowles Foundation Discussion paper no. 1686, Yale University.

Panjer, H. H. (2006), Operational Risk: Modeling Analytics, Wiley.

Patel, R. D. (2011), Testing Local Self-Similarity in Univariate Heavy-Tailed Data, PhD thesis, Department of Statistics, University of California, Los Angeles.

Shevchenko, P. V. and Temnov, G. (2009), Modeling operational risk data reported above a time-varying threshold, *Journal of Operational Risk*, 4(2), 19–42.

Shevchenko, P. V. (2011), Modelling Operational Risk Using Bayesian Inference, Heidelberg: Springer.

Shih, J., Khan, A., and Medepa, P. (2000), Is the size of an operational loss related to firm size? *Operational Risk Magazine*, 2(1), 1–2.

Simonoff, J. S. and Tsai, C.-L. (1999), Semiparametric and additive model selection using an improved AIC criterion, *Journal of Computational and Graphical Statistics*, 8, 1–19.

Soprano, A., Crielaard, B., Riacenza, F., and Ruspantini, D. (2009), Measuring Operational and Reputational Risk: A Practitioner's Approach, Wiley Finance.

Weitzman, M. L. (2009), On modeling and interpreting the economics of catastrophic climate change, *The Review of Economics and Statistics*, 91(1), 1–19.

Wood, S. N. (2006), Generalized Additive Models: An Introduction with R, Boca Raton: Chapman & Hall/CRC.