The Journal of Risk and Insurance

 Vol. 83, No. 3, 735–776 (2016).
DOI: 10.1111/jori.12059

An Extreme Value Approach for Modeling Operational Risk Losses Depending on Covariates

Valérie Chavez-Demoulin
Paul Embrechts
Marius Hofert

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 nonhomogeneous 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 nonstationary generalized Pareto distribution (alternatively, a generalized extreme value distribution) depending on the covariates. Since spline smoothing cannot 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 operational risk losses collected from public media. 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.

Introduction

The aim of the article 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 (OpRisk) data set to be detailed later in the article; third, a publicly available software implementation (including a simulated data example) is developed to apply the presented methodology.

Valérie Chavez-Demoulin is at the Faculty of Business and Economics, University of Lausanne, Switzerland. Chavez-Demoulin can be contacted via e-mail: valerie.chavez@unil.ch. Paul Embrechts is at the RiskLab, Department of Mathematics and Swiss Finance Institute, ETH Zurich, 8092 Zurich, Switzerland. Embrechts can be contacted via e-mail: embrechts@math.ethz.ch. Marius Hofert is at the Department of Statistics and Actuarial Science, University of Waterloo, 200 University Avenue West, Waterloo, ON, Canada N2L 3G1. Hofert can be contacted via e-mail: marius.hofert@math.ethz.ch. The author (Willis Research Fellow) thanks Willis Re for financial support while this work was being completed. The authors would like to thank the referees and an editor for various comments that led to a much improved version of the article.

The fact that we apply the new statistical tools to business "loss" data is not really essential but rather reflects the properties of the OpRisk data set at hand (and data of a similar kind). "Losses" can, without any problem, be changed into "gains"; relevant is that we concentrate our 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. OpRisk data typically exhibit 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, Frey, and Embrechts (2005) and the references therein.

The methodology presented in this article 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) that may have an impact on the loss severity and frequency; see, for instance, Ganegoda and Evans (2013), Shih, Khan, and Medepa (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. We stress that the (limited) public OpRisk data available to us provided the motivation for developing the new statistical extreme value theory (EVT) methodology of this article. We do not (and indeed cannot) formulate general conclusions on the Loss Distribution Approach (LDA) modeling of real, one-company-based OpRisk data. By providing the R-software used, any industry end-user can apply our techniques to his/her internal data. We very much hope to learn from such experiments in the future so that the method provided can be further enhanced. In the "Discussion" section, we do however make some general comments on the quantitative LDA modeling of OpRisk data.

Recall that under the capital adequacy guidelines of the Basel Committee on Banking Supervision (see http://www.bis.org/bcbs, shortened throughout the article 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 Bank for International Settlements [BIS], 2006, p. 144). By nature, this risk category, as opposed to Market and Credit Risk, is much more akin to nonlife insurance risk or loss experience from industrial quality control. OpRisk came under regulatory scrutiny in the wake of Basel II in the late 90s; see BIS (2006). This is relevant as data were systematically collected only fairly recently, leading to reporting bias in all OpRisk data sets. We will come back to this issue later in the article. 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) provide 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 LDA, which we will consider in this article. 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 1-year horizon at the 99.9 percent 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, sect. 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 on Market Risk measurement triggered by the Basel documents BIS (2012, 2013). See in particular Question 8 on page 41 of the former: "What are the likely constraints with moving from VaR to ES, including any challenges in delivering robust backtesting and how might these be best overcome?" These documents, and the latter question in particular, led to a very extensive debate between regulators, practitioners and academics; see Embrechts et al. (2014) for more details. Ingredients of the wider debate are (1) to what extent can financial or insurance risk be mapped to a number to be used as a basis for capital requirement; (2) which number, that is, risk measure to use; and (3) for a given risk measure, how to statistically estimate it and backtest it on historical data? In this article we mainly look at EVT-based methodology for answering the first part of (3), that is, statistical estimation, and this with focus on VaR, at very high quantile levels.

An important question concerns whether OpRisk can be accurately modeled and statistically estimated as proposed under the Basel guidelines: 1-year VaR at a confidence level of 99.9 percent. Whereas for certain subcategories like soft- and hardware failures this may be possible, for risk classes like fraud this becomes much more complicated, if at all possible with the data available. It is worthwhile to compare and contrast this situation with insurance regulation: under the EU Solvency 2 framework (to come into force on January 1, 2016) the capital requirement for OpRisk is volume based, that is, 4.5 percent of the technical provisions for life insurance obligations or 3 percent of the premiums earned during the past 12 months for nonlife. For the Swiss Solvency Test, in force since January 1, 2011, no quantitative consideration is given. OpRisk is treated qualitatively through the Swiss Quality Assessment (SQA). The SQA guidelines state however that "[t]he undertaking must be organized in such a way that all its significant risks can be captured, limited, and monitored." The approach we develop in this article will also be useful in the analysis of particular subclasses of OpRisk for internal use within the context of SQA-like requirements.

Key publications on OpRisk from the regulatory front are de Fontnouvelle et al. (2004), de Fontnouvelle, Rosengren, and Jordan (2005), Dutta and Perry (2006), Mori, Kimata, and Nagafuji (2007), and Moscadelli (2004). The last article uses EVT to analyze data from the second Quantitative Impact Study with over 47,000 observations. Interesting contributions from industry include Baud, Frachot, and Roncalli (2002, 2003), and Frachot, Moudoulaud, and Roncalli (2004) for Crédit Lyonnais; Aue and Kalkbrener (2006) for Deutsche Bank; and Soprano et al. (2009) for UniCredit Group. More methodological articles relevant for our approach are Chavez-Demoulin and Embrechts (2004), El-Gamal, Inanoglu, and Stengel (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 instance, Cummins, Lewis, and Wei (2006), Jarrow (2008), Jarrow, Oxman, and Yildirim (2010), and Chernobai, Jorion, and Yu (2011). Unfortunately, publicly available OpRisk data are 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 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, Czado, and Paterlini (2014). 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, Frey, and Embrechts (2005, chap. 10).

For the LDA, the Basel Committee has decomposed a bank's activities into business lines (the standard is eight) and event types (typically seven) 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, and Dutta and Perry (2006) at the enterprise level. The last two also attempt a modeling by time. Papers on modeling OpRisk data using covariates include Ganegoda and Evans (2013), Shih, Khan, and Medepa (2000), and Cope and Labbi (2008) where the authors investigate whether the size of OpRisk losses can be correlated with firm size and geographical region. Na et al. (2006) suppose that the operational loss can be broken down into a component common to all banks and idiosyncratic components specific to each loss. Dahen and Dionne (2010) investigate 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 article is organized as follows. The "Classical EVT Approach for Modeling Losses" section provides a brief introduction to EVT in terms of the Block Maxima and the Peaks-Over-Threshold (POT) approaches. In the "A Dynamic EVT Approach" section 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 article and focus is put on the dynamic POT approach. A fitting method is developed in "The Penalized Maximum Likelihood Estimator and Risk Measure Estimators" section and it is applied to simulated data in the "Demonstration of the Dynamic Approaches Based on Simulated Data" section in the Appendix. In the "Application to an OpRisk Loss Database" section we model a data set of publicly available OpRisk losses with the presented dynamic, EVT-based approaches. The "Discussion" section provides an in-depth discussion of the presented methodology and data.

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 nonstationary setup; an extension of the Block Maxima method to the nonstationary case is also proposed and serves as another possibility to evaluate risk measures. We begin with the latter.

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 nondegenerate limiting distribution for the maximum, the X_is 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 $\mathbb{P}(W_q \le w) = F^q(b_q + a_q w)$ 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 limits are

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

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 nondegenerate limiting distribution must be a *generalized extreme value (GEV) distribution*, given by

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

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 $X_{t'_1}, \ldots, X_{t'_{n'}}$ following a nondegenerate, continuous distribution function F. We interpret $X_{t'_1}, \ldots, X_{t'_{n'}}$ (≥ 0) as losses in some monetary unit over a time period $[0, T]$ with $0 \leq t'_1 \leq \ldots \leq t'_{n'} \leq T$. 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* $M_q^{(1)}, \ldots, M_q^{(m')}$ from m' blocks of size q. Assuming these block maxima to be independent, the log-likelihood is given by

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

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 high quantile, α being typically close to 1.

The POT Approach

Given a suitably large threshold $u \geq 0$, let $\{t_1, \ldots, t_n\} \subseteq \{t'_1, \ldots, t'_{n'}\}$ denote those time points (in increasing order) for which $X_{t'_1}, \ldots, X_{t'_{n'}}$ exceed u; that is, let $X_{t_1}, \ldots, X_{t_n}$ denote the *exceedances over* u with corresponding *excesses* $Y_{t_i} = X_{t_i} - u$, $i \in \{1, \ldots, n\}$. It follows from Embrechts, Klüppelberg, and Mikosch (1997, p. 167) 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 $Y_{t_1}, \ldots, Y_{t_{N_t}}$ 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 dis-

tribution function

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

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

For a precise formulation and full details of the proof, see Leadbetter (1991); note that ξ is the same as in (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, Frey, and Embrechts (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 $\xi \geq 1$, 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, Embrechts, and Chavez-Demoulin (2006) and the references therein; see also the "Discussion" section. In the following, we assume $\xi > -1$. 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 $\boldsymbol{Y} = (Y_{t_1}, \ldots, Y_{t_n})$ 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

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

The maximization for the two estimation problems related to 1 and 1 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, Klüppelberg, and Mikosch (1997, sect. 4.4).

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 that are possibly predictive for 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, nonparametric, or semi-parametric and can also include interactions. With the OpRisk example in mind, we focus on two covariates: a factor x (in the "Application to an OpRisk Loss Database" section: business line) and time t (in the "Application to an OpRisk Loss Database" section: 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

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

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 smooth (nonparametric) 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, chap. 6). Its generalization to a semiparametric 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 sufficiently smooth functions, the parameter vector $\boldsymbol{\theta}$ of the EVT model can be estimated by maximizing the penalized log-likelihood

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

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 overfitting when one is interested in fitting smooth functions h_k (see Hastie and Tibshirani, 1990; Green and Silverman, 2000). Intuitively, the penalty functions $\int_{\mathcal{A}} h_k''(t)^2 \, \mathrm{d}t$ 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 article, 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 semiparametric GEV-based approach to yearly maxima of OpRisk losses; this is done in the "Dynamic Block Maxima Method" section.

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 the "Loss Frequency" and "Loss Severity" sections. The "Penalized Maximum Likelihood Estimator and Risk Measure Estimators" section then presents the details about the fitting procedure.

Loss Frequency

The number of exceedances is assumed to follow a nonhomogeneous 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` ($\geq$ 0.4–11), see the simulation examples in the "Demonstration of the Dynamic Approaches Based on Simulated Data" section.

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 counterexample 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 $0 \le t_1 \le \cdots \le t_n \le T$ 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 (e.g., 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.

The Penalized Maximum Likelihood Estimator and Risk Measure Estimators

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 the "Application to an OpRisk Loss Database" section, 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 $\boldsymbol{z}_i = (t_i, x_i, y_{t_i})$, $i \in \{1,\dots,n\}$, we use the penalized likelihood (6). The *penalized log-likelihood* ℓ^p

corresponding to the observations z_i, $i \in \{1, \ldots, n\}$, is given by

$$\ell^p(f_\xi, h_\xi, f_\nu, h_\nu; z_1, \ldots, z_n) = \ell^r(\boldsymbol{\xi}, \boldsymbol{\nu}; \boldsymbol{y}) - \gamma_\xi \int_0^T h_\xi''(t)^2 \, \mathrm{d}t - \gamma_\nu \int_0^T h_\nu''(t)^2 \, \mathrm{d}t, \tag{13}$$

where $\gamma_\xi, \gamma_\nu \geq 0$ denote *smoothing parameters*, $\boldsymbol{y} = (y_{t_1}, \ldots, y_{t_n})$, 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 $0 = s_0 < s_1 < \ldots < s_m < s_{m+1} = T$ denote the ordered and distinct values among $\{t_1, \ldots, t_n\}$. A function h defined on $[0, T]$ is a cubic spline with the above knots if the following conditions are satisfied: (1) on each interval $[s_i, s_{i+1}]$, $i \in \{1, \ldots, m\}$, h is a cubic polynomial; (2) at each knot s_i, $i \in \{1, \ldots, m\}$, 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 $s_1, \ldots, s_m$, one has

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

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

$$\ell^p(f_\xi, h_\xi, f_\nu, h_\nu; z_1, \ldots, z_n) = \ell^r(\boldsymbol{\xi}, \boldsymbol{\nu}; \boldsymbol{y}) - \gamma_\xi \boldsymbol{h}_\xi^\top K \boldsymbol{h}_\xi - \gamma_\nu \boldsymbol{h}_\nu^\top K \boldsymbol{h}_\nu, \tag{15}$$

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

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 the Fitting and Bootstrapping Confidence Intervals Algorithm in the Appendix) that alternates between Newton steps for ξ and ν (see the Derivatives of the Reparameterized Log-Likelihood Algorithm in the Appendix). We construct bootstrapped pointwise two-sided confidence intervals with a post-blackend bootstrap (see the Demonstration of the Dynamic Approaches Based on Simulated Data Algorithm in the Appendix). Based on the estimates $\hat{\lambda}$, $\hat{\xi}$, and $\hat{\beta}$ for a fixed covariate x and time point t, one can then compute estimates of the risk measures VaR and ES (depending on x and t) as

$$\widehat{\mathrm{VaR}}_\alpha = u + \frac{\hat{\beta}}{\hat{\xi}} \left(\left(\frac{1 - \alpha}{\hat{\lambda}/n'} \right)^{-\hat{\xi}} - 1 \right), \tag{16}$$

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

where $n' = n'(x, t)$ is the total number of losses for a fixed covariate x and time point t; see McNeil, Frey, and Embrechts (2005, p. 283) for the basic formulas. In practice and for prediction purposes, we directly model the rate $\rho = \rho(x, t) = \lambda(x, t)/n'(x, t)$ of the number of exceedances over the number of losses using a logistic regression model (`gam(..., link=logit)`) instead of the Poisson regression model for the number of exceedances. Bootstrapped pointwise two-sided confidence intervals for VaR_α and ES_α can be constructed from the fitted values of λ and bootstrapped estimates of ξ and β by computing the corresponding empirical quantities.

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 $\mathrm{Df} = 1$ 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, p. 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

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

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

$$\mathrm{AIC} \propto -2\ell(\theta;\cdot) + 2\,\mathrm{Df}; \tag{18}$$

see Simonoff and Tsai (1999).

Application to an OpRisk Loss Database

In this section, we apply the methodology presented before to a database of OpRisk losses collected[1] from public media; these data may not exhibit all characteristics of company-specific data. The database consists of 1,413 OpRisk events reported in the public media since 1970. For our analysis, we consider the 1,387 events since 1980. For each event, the following information is given: a reference number, the organization affected (1 or 0.07 percent missing), the country of head office, the country where the event happened (1 or 0.07 percent 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 percent missing), the net loss amount in GBP (1,353 or 97.55 percent missing), the regulator involved (838 or 60.42 percent missing), the source from which the information was drawn (29 or 2.09 percent missing), and a loss description. The data have been sourced from various webpages, newspapers, press releases, regulator announcements, and databases (Bloomberg, Yahoo Finance, etc.). In later years, web pages were used predominantly (including, e.g., 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 United States (615 or 44.34 percent of the reported events), the United Kingdom (361 or 26.03 percent; additionally, there is 1 reported event on the Cayman Islands and 2 reported events on the Isle of Man), Japan (70 or 5.05 percent), Australia (32 or 2.31 percent), and India (28 or 2.02 percent); 15 events (1.08 percent) are reported under "Various" and not associated with a specific country. With regard to insurance, 887 of the events (63.95 percent) were (partially) insured and 483 (34.82 percent) were not insured (for the remaining 17 losses, 13 were missing and there was no information available yet for the other 4).

The Covariates Investigated

In this section, we briefly address the choice of covariates for our data analysis below. As will become clear from the "Descriptive Analysis of the Loss Data" section, the dependence on time is an important aspect we would like to incorporate in our model to deal with the inhomogeneity of the data in time. Furthermore, the Basel II guidelines naturally provide the buckets "business line" and "event type." For our analysis we "aggregate," for each business line, the losses over all event types. 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. Another practical aspect of this aggregation is that it mitigates the problem of too small sample sizes; see also the Basel matrix versus the Basel vector in the "Descriptive Analysis of the Loss Data" section below.

Incorporating additional information via covariates is a distinct feature of the new model proposed. A particular advantage hereby is that the model can be fitted to all the data simultaneously. This allows us to "pool" the data; that is, all the data are used to fit the model, not just the data available for a certain business line; the latter would not be feasible due to the data being sparse, especially over time; see, for example, the business line Retail Banking in Figure 2 below.

[1] By Willis Professional Risks.

TABLE 1
Summary of the 10 Largest Losses (Adjusted for Inflation to the Level of 2013; Rounded to M GBP)

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	9,268	UBL	CPBP	2011	Settle Countrywide Financial Corp.'s residential mortgage-backed security repurchase exposure
BTA Bank	8,518	CB	IF	2010	Fraud by chairman, diverting funds to companies he owned
Agricult. Bank of China	6,173	CB	IF	2004	Financial crimes and bookkeeping irregularities
T. Imar Bankasi T.A.S.	5,528	RBr	IF	2003	Fraudulent computer program
Société Générale	4,548	UBL	IF	2008	A trader entered futures positions circumventing internal regulations
Banca Naz. del Lavoro	4,407	RBr	IF	1989	Four counts of fraud
J.P. Morgan	3,760	UBL	CPBP	2013	U.S. authorities demand money due to mis-sold securities to Fannie Mae and Freddie Mac

Let us also remark that if the identification of covariates is not as clear as in our case here (see the "Descriptive Analysis of the Loss Data" section section below) and no covariates can be identified, we are simply in the classical setup of "Classical EVT Approach for Modeling Losses" section, so this is incorporated as a special case in our model.

Descriptive Analysis of the Loss Data

For our data analysis, 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 Price Inflation Reference Tables, August 2013* obtained from the web page 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's year. Table 1 shows a summary of selected information about the 10 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,

Table 2
Summary Statistics (Mean, Standard Deviation, Median, Third Quartile, Maximum, Skewness, Number of Losses Above 1M GBP) About the 950 Losses Split by Business Lines

BL	Mean	Std. Dev.	Median	3rd Quartile	Max	Skewness	# > 1M
AS	86.21	132.92	20.51	106.22	490.10	1.76	17
AM	572.41	4,558.43	15.53	53.87	40818.88	8.60	67
CB	271.55	1,185.12	20.25	81.63	12500.00	7.70	172
CF	142.58	344.75	16.82	74.57	1928.06	3.85	36
I	70.87	180.10	7.38	34.35	910.73	3.35	42
PS	337.82	1,859.65	18.86	70.05	14607.65	7.29	55
RBa	201.80	368.86	22.57	274.15	1257.18	1.91	9
RBr	125.65	561.16	3.97	25.50	5527.96	7.16	153
TS	19.16	31.62	6.57	31.88	172.55	2.77	35
UBL	228.30	831.22	21.20	114.10	9268.25	7.71	181

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, pp. 302, 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.

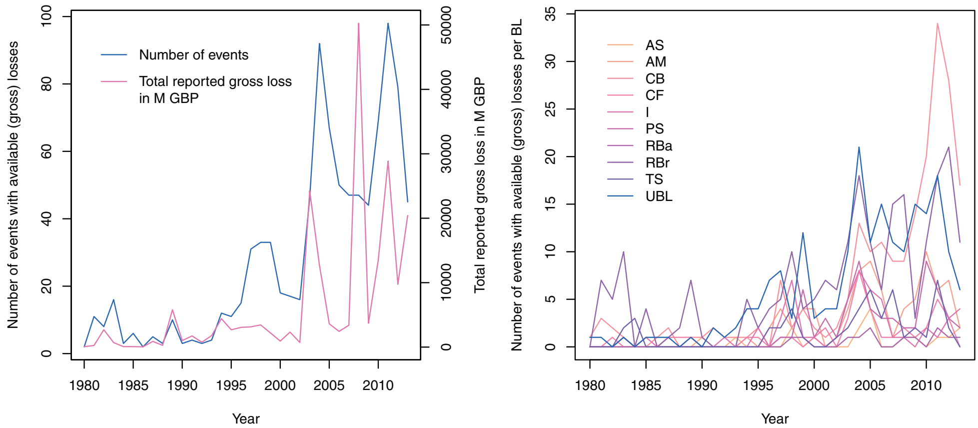
In what follows, we consider the 950 available, inflation-adjusted, gross losses in M GBP (M = million). Table 2 presents summary statistics about the OpRisk losses for each business line; the losses are quite positively skewed.

The left-hand side of Figure 1 provides a graphical summary of the 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. The increasing frequency of events is probably also due to improved record keeping (reporting bias). Furthermore, the frequencies depend on the business lines. Our model takes both of these features into account; see the “Dynamic EVT Modeling” section.

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 cannot be modeled with the same parameters. Our model will take this into account by interpreting the business lines as covariates.

Remark 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 the “Demonstration of the Dynamic Approaches Based on Simulated Data: section in the Appendix. Our analysis hence should be viewed as a first step toward an LDA model. Before its introduction*

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)



into practice, more extensive data sets are needed. On the other hand, as we shall see, our analysis clearly is able to capture some important features of OpRisk (and similar) data.

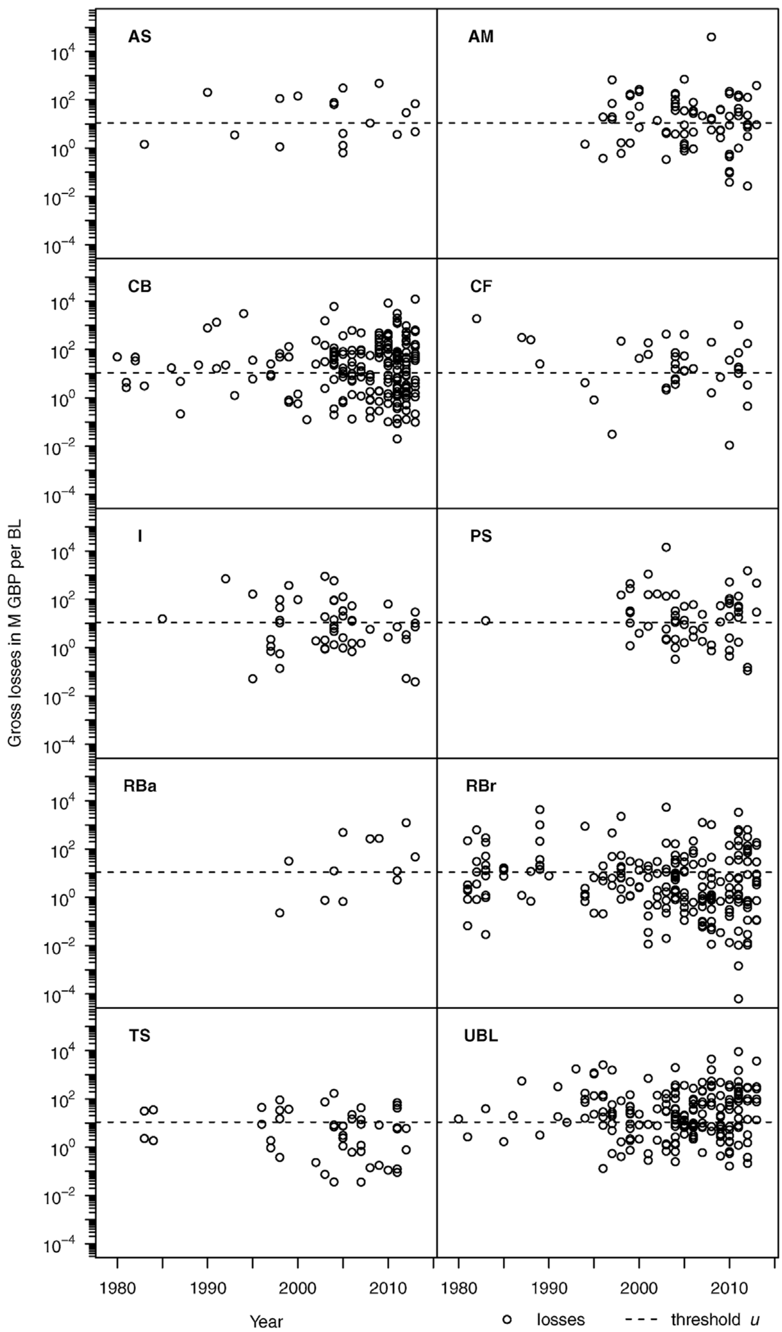
The *Basel matrix B* contains the number of available gross losses divided into *business lines* (rows) and *event types* (columns). In our case, this is a $(10, 7)$ matrix. By summing, for each business line, over the different event types, one obtains the corresponding *Basel vector b*. Based on the 950 available gross losses since 1980, the Basel matrix B and the Basel vector b are given by

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

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 $b = \sum_{t=1980}^{2013} b_t$.

Figure 2

OpRisk Data: Available Positive Gross Losses (on Log Scale) in M GBP per Business Line Including the Threshold u Chosen in the "Dynamic POT Analysis" Section (Median of All Losses)



Dynamic EVT Modeling

We now apply the methodology presented in the "A Dynamic EVT Approach" section to estimate the different parameters entering the calculation of the annual $\mathrm{VaR}_{0.999}$, including pointwise two-sided 95 percent confidence intervals. As mentioned before, due to the data being sparse, we consider the losses aggregated for each business line over all event types. In the "Dynamic POT Analysis," "Sensitivity of the Model Choice With Respect to the Choice of the Threshold," and "Sensitivity of the Model Choice With Respect to the Modeled Rime Period" sections, we focus on the dynamic POT approach; in the "Dynamic Block Maxima Method" section we present the dynamic block maxima approach.

Dynamic POT Analysis. In this section, we consider the median of all losses as threshold u (i.e. 11.02M GBP) and the 475 gross losses from 1980 to 2013 that exceed u. The reason for this choice and the sensitivity of our analysis on the threshold are addressed in the "Sensitivity of the Model Choice With Respect to the Choice of the Threshold" section.

The choice of the smoothing parameters (γ_λ for λ and γ_ξ, γ_ν for the GPD 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 (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 greater than or equal to 1. This turns out to be the case here for several business lines; see Figure 5 below. Hence, for the remaining part of the article, we can only present results about VaR. For further comments on this important issue, see the "Discussion" section.

Loss frequency: We fit the following models for λ:

$$\log \lambda(x,t) = c_\lambda, \tag{19}$$

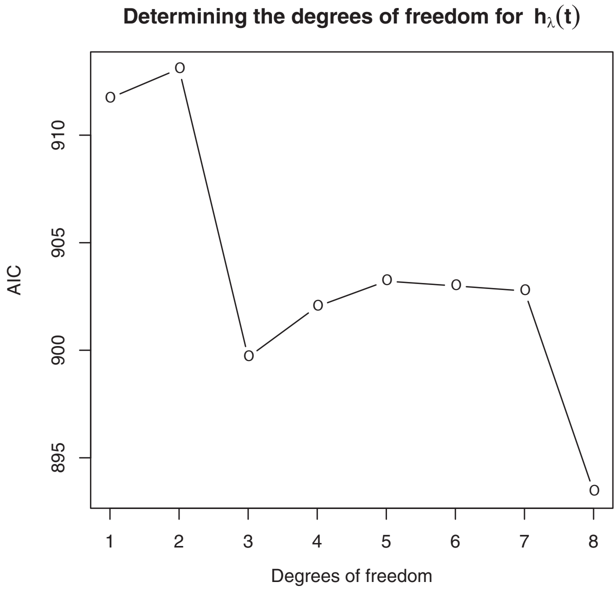
$$\log \lambda(x,t) = f_\lambda(x), \tag{20}$$

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

Model (19) is the classical constant model, depending neither 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, we then compare

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 $\mathrm{Df} \in \{1,\ldots,8\}$



Model (21) with models of the form

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

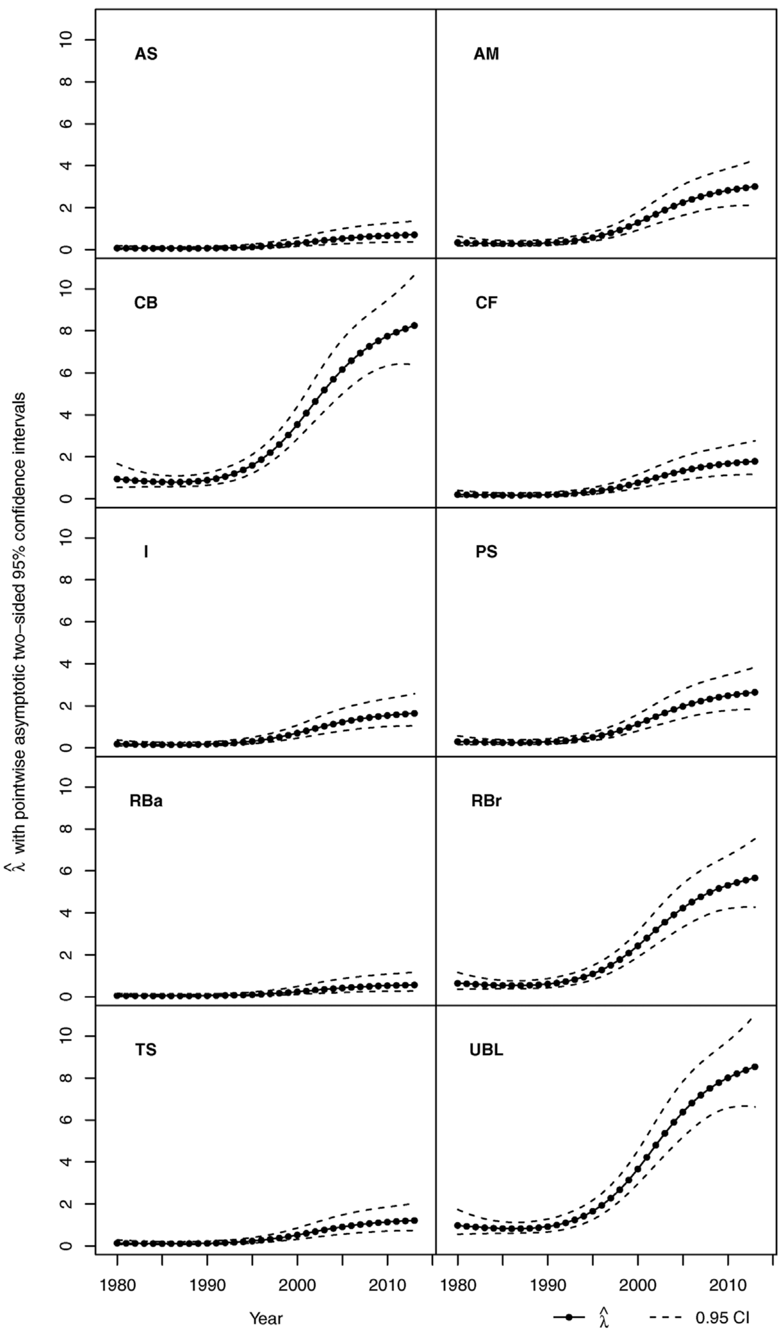
for natural cubic splines $h_\lambda^{(\mathrm{Df})}$ with degrees of freedom $\mathrm{Df} \in \{1,\ldots,8\}$; see Figure 3. Note that the case $\mathrm{Df} = 1$ corresponds to Model (21). It is clear from this curve, that a nonparametric dependence on time is suggested. We use the "elbow criterion" 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.

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 percent confidence intervals). Overall, the curves show an increasing pattern through time for all business lines.

FIGURE 4
OpRisk Data: Estimates $\hat{\lambda}$ Including 95% Confidence Intervals Depending on Time and Business Line



Loss severity: We now use the methodology explained in the "A Dynamic EVT Approach" section to fit dynamic models for the GPD parameters (ξ, ν); recall the reparameterization (8). We fit the following models for (ξ, ν):

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

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

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

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

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

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

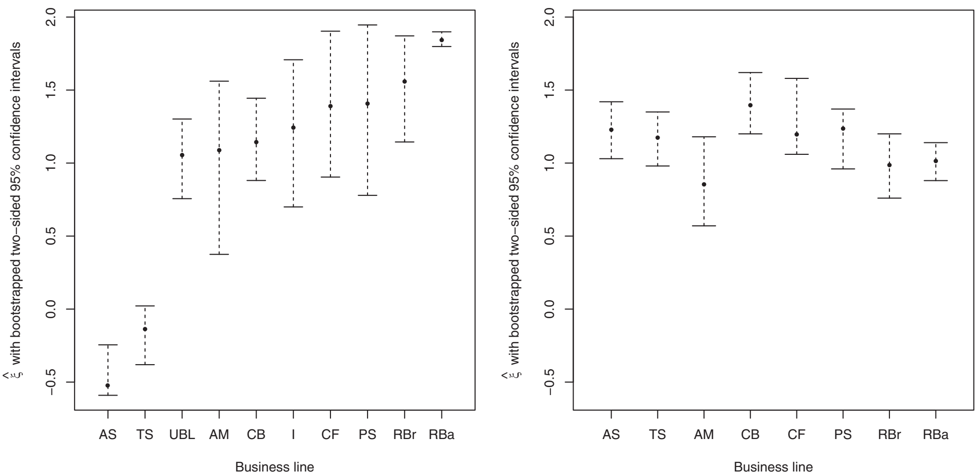
where h_ν is a smooth function of t (with variable degrees of freedom). The likelihood-ratio test based on Models (24) and (25) reveals that business line has a significant effect on ξ. Comparing Models (25) with (26) indicates that time does not have a significant effect on ξ. We therefore use Model (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

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

Note that the more influential or important GPD parameter is ξ, as it determines the tail of the loss severity distribution, see Equation (4), and also whether the loss severity distribution has finite ($\xi \in (0,1)$) or infinite ($\xi \geq 1$) first moment. As our estimate reveals, this important parameter does not depend on time. This is advantageous as we therefore do not switch between the two worlds (finite/infinite first moment) in time. In most applications, it is observed that the shape parameter does not depend on time. A reason for this is that it is likely (and to be expected) that the shape parameter provides evidence in favor of (only) one of the possible underlying loss distributions determining to which maximum domain of attraction (MDA) the loss distribution belongs. Furthermore, considering a fixed threshold u, the number of exceedances above u could vary with time as if the value of the threshold would itself be varying. Theoretically, it can be shown that the shape parameter does not vary for any other threshold value v with $v > u$ given that u is the "smallest" among all thresholds for which the asymptotic result in the first section holds. For more details about this property of the GPD, see McNeil, Frey, and Embrechts (2005, Lemma 7.22).

The left-hand side of Figure 5 shows the estimates $\hat{\xi}$ for every business line and corresponding bootstrapped two-sided 95 percent confidence intervals for our data set. 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

FIGURE 5
OpRisk Data: Estimates $\hat{\xi}$ (Left: Our Data Set; right: Moscadelli, 2004) Including Bootstrapped 95% Confidence Intervals Depending on Business Lines



(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)data sets, 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)data sets for which $\xi \geq 1$: 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 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 percent 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 percent 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 are gathered across several institutions, it is difficult to put them in perspective. As mentioned ear-

Figure 6

OpRisk Data: Estimates $\hat{\beta}$ (on Log Scale) Including Bootstrapped Pointwise Two-Sided 95% Confidence Intervals Depending on Time and Business Line

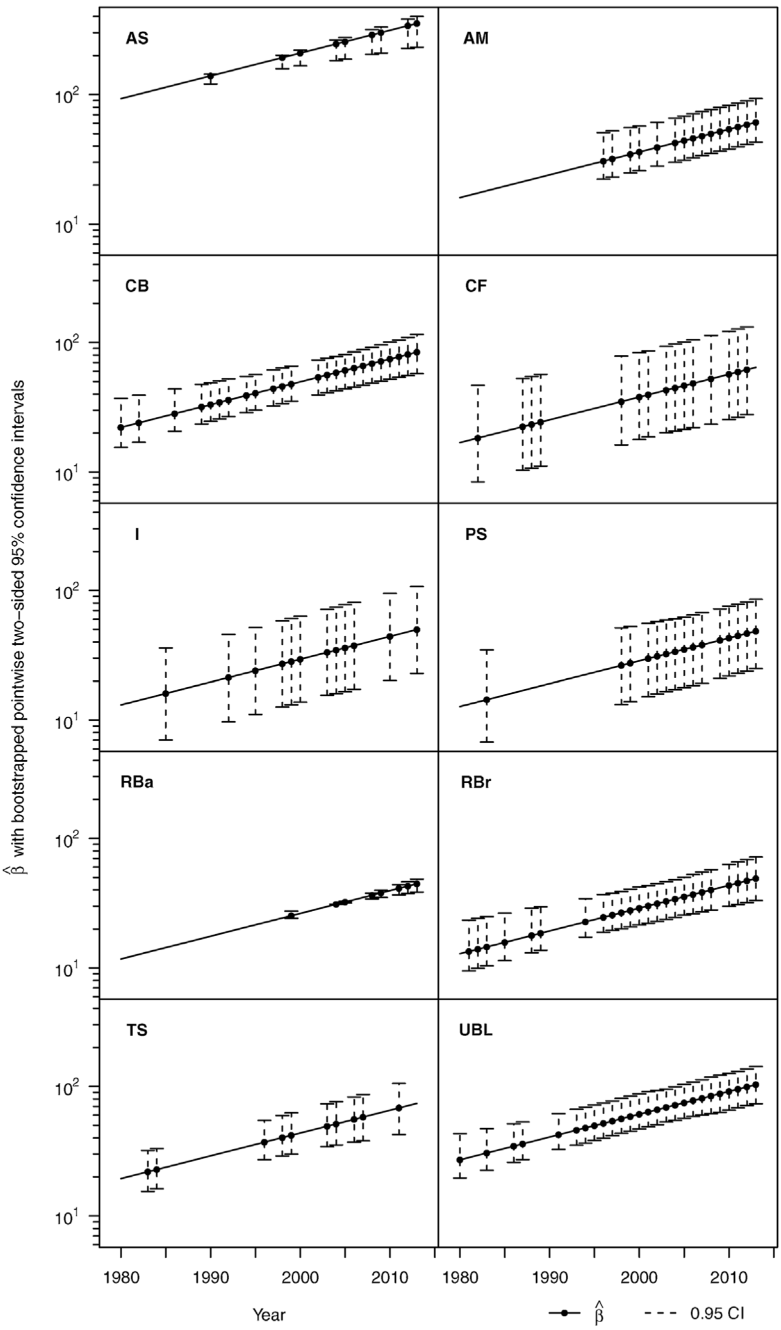
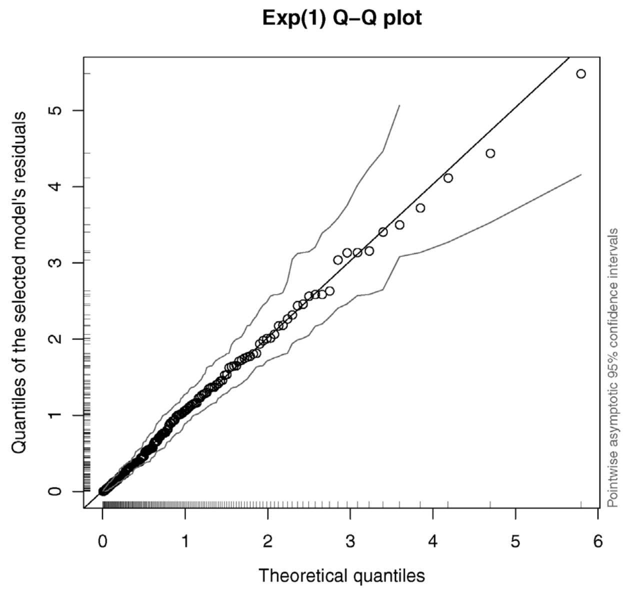


FIGURE 7
OpRisk Data: Q–Q Plot of the Chosen Model's Residuals



lier, the main purpose of the article 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 Quantitative Risk Management point of view. A typical application would concentrate on loss data within one company for which then a common total capital denominator would be available. For an application focusing on institutions belonging to the banking and finance sector and in the U.S. market only, see our study in the "Dynamic POT Analysis Based on the U.S. Banking and Finance Industry" section.

Sensitivity of the Model Choice With Respect to the Choice of the Threshold. 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 (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 "The POT Approach" section 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 3. Residual plots (not included) for smaller values of u showed clear departures from the model. We have indicated the

Figure 8
OpRisk Data: Estimates $\widehat{\mathrm{VaR}}_{0.999}$ (on Log Scale) Including Bootstrapped Pointwise Two-Sided 95% Confidence Intervals Depending on Time and Business Line

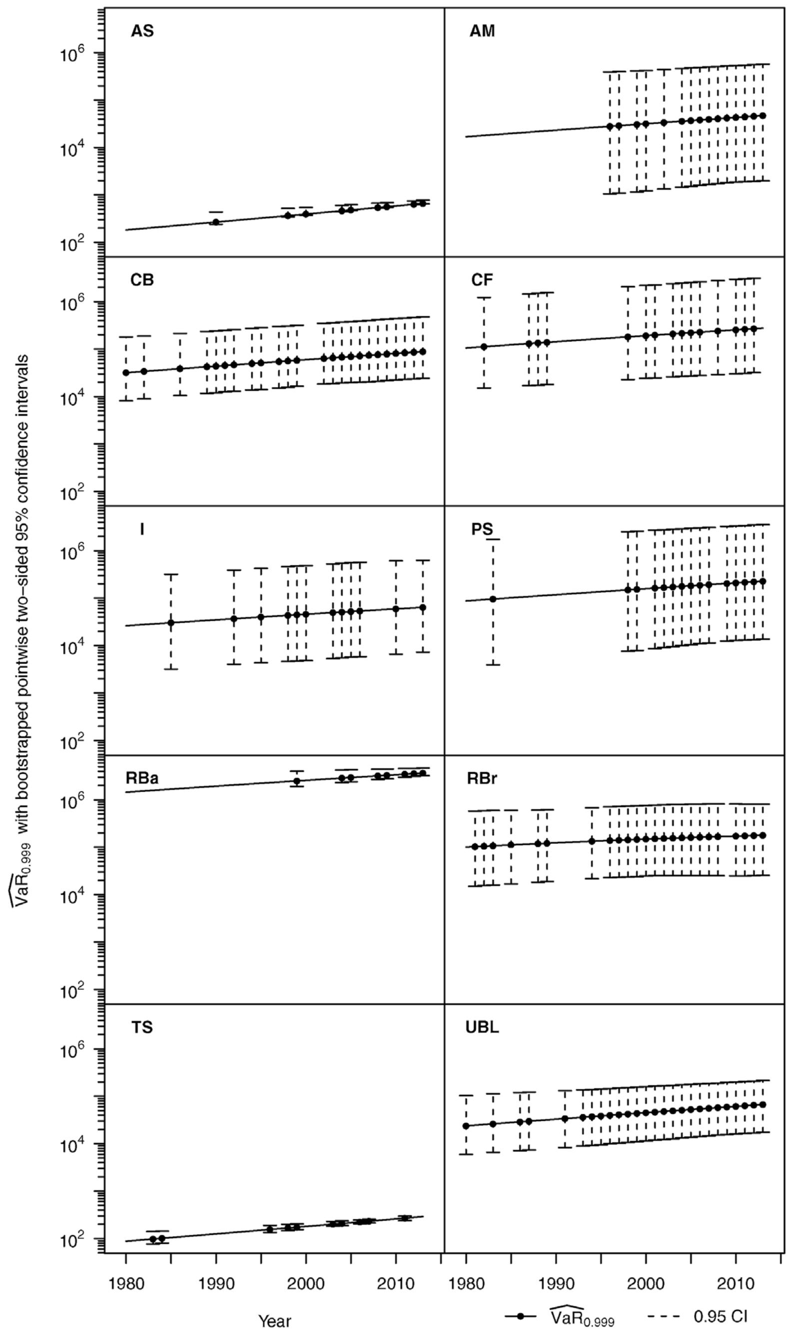


TABLE 3
Selected Models for the Loss Severity Depending on the Choice of Threshold

Threshold u	n	Model for ξ	Model for ν	Q–Q Plot
0-quantile	934	$\hat{\xi}(x,t) = \hat{f}_\xi(x) + \hat{c}_\xi t$	$\hat{\nu}(x,t) = \hat{f}_\nu(x) + \hat{c}_\nu t$	↯↯↯
0.2-quantile	760	$\hat{\xi}(x,t) = \hat{f}_\xi(x)$	$\hat{\nu}(x,t) = \hat{f}_\nu(x)$	↯↯
0.3-quantile	665	$\hat{\xi}(x,t) = \hat{f}_\xi(x)$	$\hat{\nu}(x,t) = \hat{f}_\nu(x)$	↯↯
0.4-quantile	570	$\hat{\xi}(x,t) = \hat{f}_\xi(x)$	$\hat{\nu}(x,t) = \hat{f}_\nu(x)$	↯
0.5-quantile	475	$\hat{\xi}(x,t) = \hat{f}_\xi(x)$	$\hat{\nu}(x,t) = \hat{f}_\nu(x) + \hat{c}_\nu t$	✓

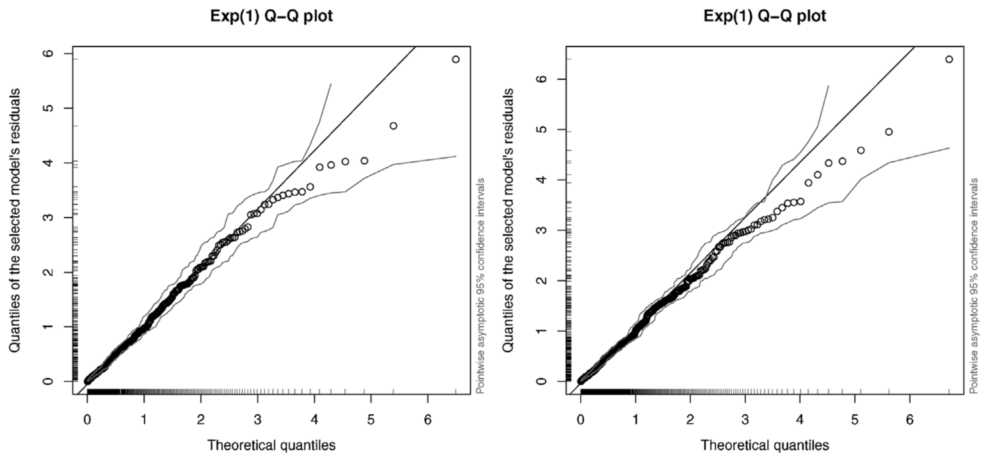
increasing deterioration with decreasing u by the nontechnical ↯symbols. The residuals for the threshold $u = 0.5$ are given in Figure 7 and show a good overall behavior; hence, we decided on this level of u.

Sensitivity of the Model Choice With Respect to the Modeled Time Period. Another question is how sensitive the selected models (23) and (30) are to the modeled time period. To address this, we repeat the whole analysis conducted in the "Dynamic POT Analysis" section for the data from 1980 to 2009, then up to 2010, 2011, and 2012. The results are very similar to the study based on all available data up to 2013 and so we simply summarize our findings in Table 4 (the selected model for the data up to 2013 included for convenience). Besides the end date in the first column, Table 4 contains the number of exceedances n over the threshold u (chosen as the median for all studies). Furthermore, we see that the selected loss frequency model and the selected model for the crucial loss severity parameter ξ is the same for all investigated time periods. The only difference is the chosen model for ν (with no significant dependence on time for the data up to 2009 and up to 2011); however, using the same model as for the data up to 2013 ($\hat{\nu}(x,t) = \hat{f}_\nu(x) + \hat{c}_\nu t$) virtually leads to the same results. The residual Q–Q plots were fine in all cases; see Figure 9 for the plots based on the data up to 2009 and up to 2011. To summarize, the selected model is quite stable over different time periods. The impact on $\widehat{\mathrm{VaR}}_{0.999}$ is also small, the estimates including confidence intervals look similar to Figure 8 for all investigated time periods. As an example, we included the one for the data up to 2011 in Figure 10.

TABLE 4
Selected Models for the Time Periods From 1980 up to 2009, 2010,..., 2013

End Year	n	Model for λ	Model for ξ	Model for ν	Q–Q Plot
2009	329	$\log\hat{\lambda}(x,t) = \hat{f}_\lambda(x) + \hat{h}_\lambda^{(3)}(t)$	$\hat{\xi}(x,t) = \hat{f}_\xi(x)$	$\hat{\nu}(x,t) = \hat{f}_\nu(x)$	✓
2010	364	$\log\hat{\lambda}(x,t) = \hat{f}_\lambda(x) + \hat{h}_\lambda^{(3)}(t)$	$\hat{\xi}(x,t) = \hat{f}_\xi(x)$	$\hat{\nu}(x,t) = \hat{f}_\nu(x) + \hat{c}_\nu t$	✓
2011	413	$\log\hat{\lambda}(x,t) = \hat{f}_\lambda(x) + \hat{h}_\lambda^{(3)}(t)$	$\hat{\xi}(x,t) = \hat{f}_\xi(x)$	$\hat{\nu}(x,t) = \hat{f}_\nu(x)$	✓
2012	452	$\log\hat{\lambda}(x,t) = \hat{f}_\lambda(x) + \hat{h}_\lambda^{(3)}(t)$	$\hat{\xi}(x,t) = \hat{f}_\xi(x)$	$\hat{\nu}(x,t) = \hat{f}_\nu(x) + \hat{c}_\nu t$	✓
2013	475	$\log\hat{\lambda}(x,t) = \hat{f}_\lambda(x) + \hat{h}_\lambda^{(3)}(t)$	$\hat{\xi}(x,t) = \hat{f}_\xi(x)$	$\hat{\nu}(x,t) = \hat{f}_\nu(x) + \hat{c}_\nu t$	✓

FIGURE 9
OpRisk Data Up to and Including the Year 2009 (Left) and 2011 (Right): Q–Q Plots of the Chosen Models' Residuals



Dynamic POT Analysis Based on the U.S. Banking and Finance Industry. As a point of criticism concerning our data study conducted in the "Dynamic POT Analysis" section one might argue that the losses in the database available to us are heterogenous. It is unlikely to assume that an event like Madoff's Ponzi scheme (see Table 1) may occur to a well-supervised bank. In this section, we therefore focus on those losses that appeared in the banking and finance industry in the United States; there are 328 such events with reported losses. As before, the median is chosen as threshold u.

It is clear that this study is based on a very small sample size (and a small amount of bootstrap replications failed due to this fact). Interestingly, though, the selected model for the loss frequency and the selected model for the crucial loss severity parameter ξ are still the same as in the "Dynamic POT Analysis" section; see (23) and (30), respectively. As a model for ν, $\hat{\nu}(x,t) = \hat{f}_\nu(x)$ is selected, so a slightly simpler model than we saw before (not depending on time as a covariate). The residual Q-Q plot is shown in Figure 11; there is no indication of a misspecified model. The estimated $\mathrm{VaR}_{0.999}$ is displayed in Figure 12; as the rate ρ only depends on the business line in this case, so does the estimated $\mathrm{VaR}_{0.999}$. We omit further details.

Dynamic Block Maxima Method. In this section, we apply the dynamic block maxima method to the yearly maxima of all logarithmic gross losses observed from 1980 to 2013; note that in "The Dynamic Block Maxima Approach" section in the Appendix, 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.

FIGURE 10
OpRisk Data up to and Including Year 2011: Estimates $\widehat{\mathrm{VaR}}_{0.999}$ (on log scale) Including Bootstrapped Pointwise Two-Sided 95% Confidence Intervals Depending on Time and Business Line

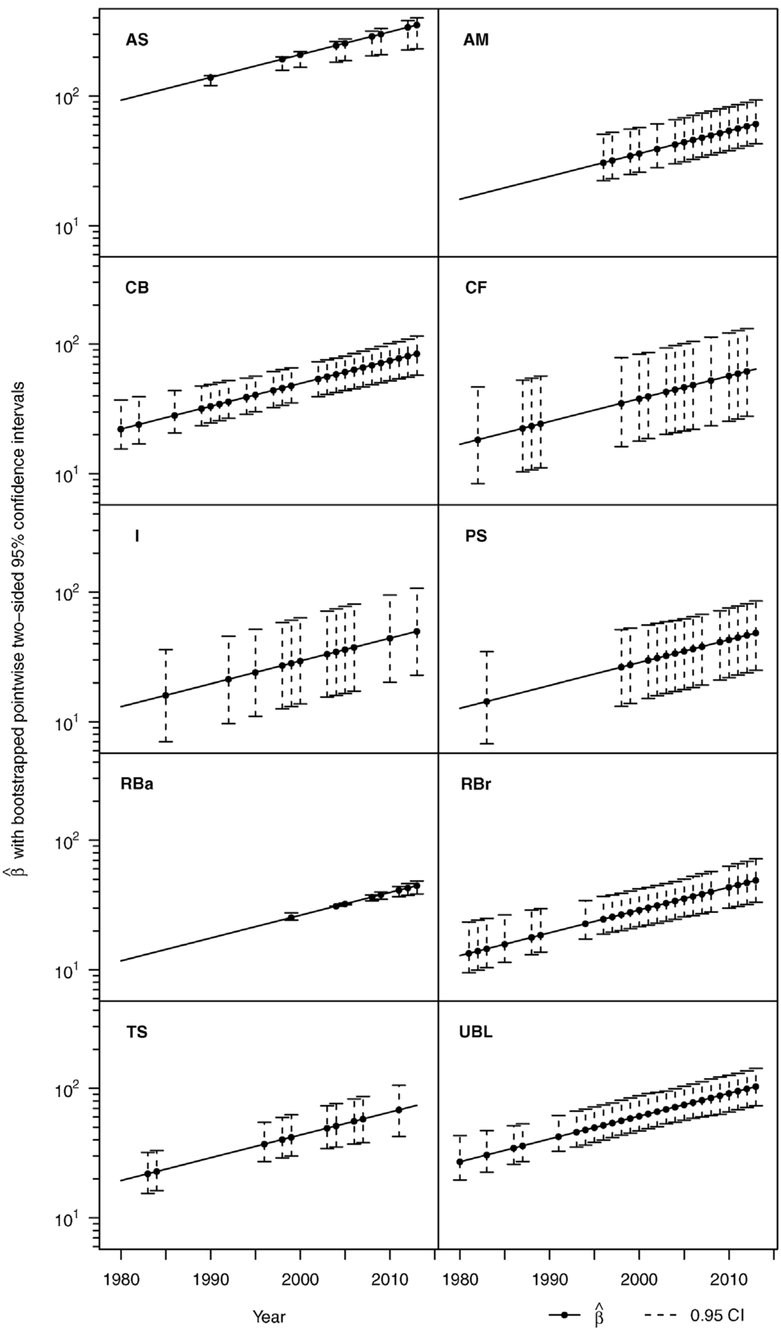
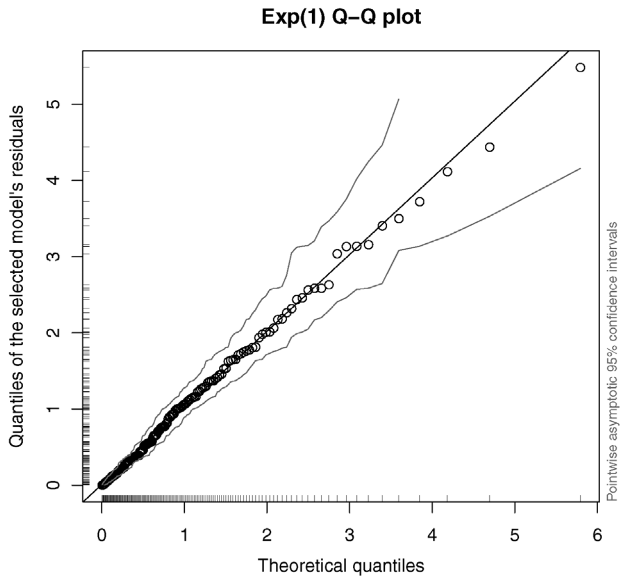


FIGURE 11
OpRisk Data for the U.S. Banking and Finance Industry: Q–Q Plot of the Chosen Model's Residuals



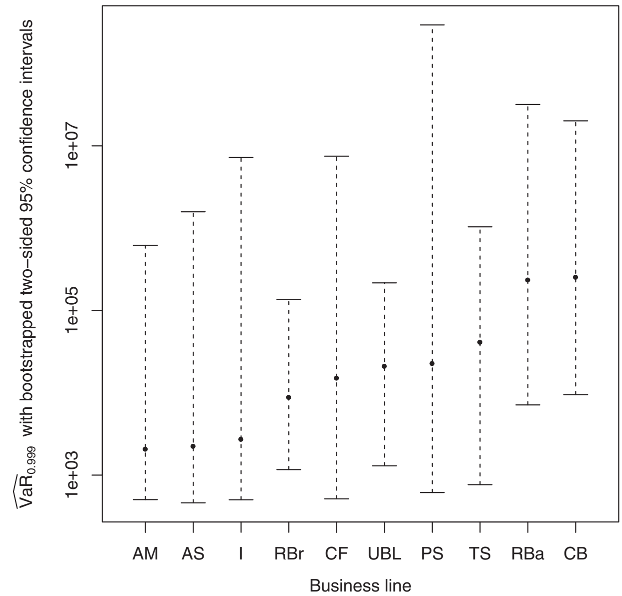
If one is interested in VaR_α, it can sometimes be useful to work with the log-transformed data. This is allowed since for $0 < X \sim F_X$ and $\log X \sim F_{\log X}$,

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

A well-known property from EVT states that the log-transform of a random variable (distribution function) in the maximum domain of attraction 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, Klüppelberg, and Mikosch (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 13 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 nonparametric model

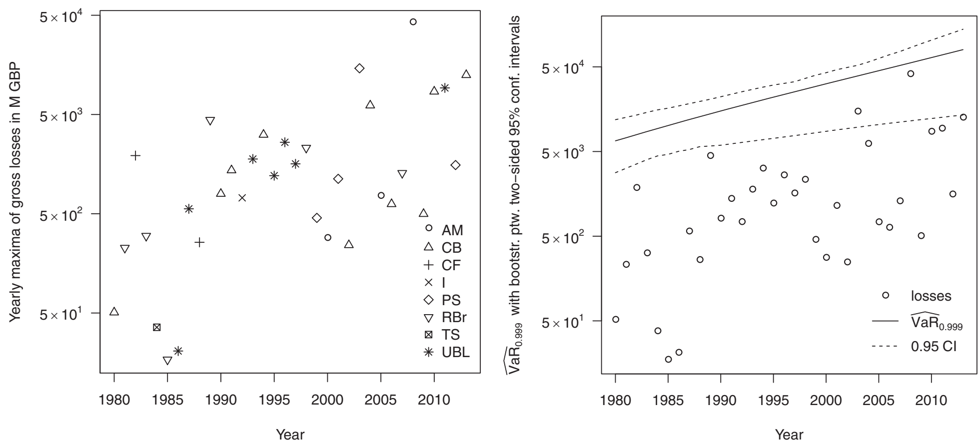
$$\hat{\mu} = \hat{\mu}(t) = \hat{h}_\mu^{(2)}(t),$$

FIGURE 12
OpRisk Data for the U.S. Banking and Finance Industry: Estimates $\widehat{\text{VaR}}_{0.999}$ (on Log Scale) Including Bootstrapped Pointwise Two-Sided 95% Confidence Intervals Depending on Time and Business Line



that is, a natural cubic spline with 2 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 $\text{VaR}_{0.999}$ including bootstrapped pointwise two-sided 95 percent 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 13. The confidence intervals are rather large because of the small number of maxima, highlighting strong uncertainty of the model. Nevertheless, the estimated $\text{VaR}_{0.999}$ 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 the "Dynamic POT Analysis" section. 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 $\widehat{\text{VaR}}_{0.999}$ on the right-hand side of Figure 13 therefore shows the estimated loss maxima that can be exceeded only "once every 1,000 years on average." Over our period under study, the resulting $\widehat{\text{VaR}}_{0.999}$ has not been exceeded by the observed losses.

FIGURE 13
OpRisk Data: Yearly Loss Maxima (on Log Scale) in M GBP Over the Period 1980–2013 (Left) (Symbols Indicate Business Lines) and Estimates $\widehat{\mathrm{VaR}}_{0.999}$ (on log scale) Including Bootstrapped Pointwise Two-Sided 95% Confidence Intervals for the Yearly Loss Maxima Over Time (Right)



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 regulator FINMA (October 2013) ordered a 50 percent 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 a new methodology in this article. We offer an 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 the "Demonstration of the Dynamic Approaches Based on Simulated Data" section in the Appendix) and to a database of publicly available OpRisk losses (see the "Application to an OpRisk Loss Database" section), which shows several stylized features of industry-wide OpRisk data.

Some of the assumptions and consequences, advantages and drawbacks of the presented EVT-based methodology and data application require a discussion:

- Let us start with the database investigated in this article. It is well known to be extremely difficult for academia to get its hands on "real" OpRisk data (internal data, from, say, larger banks, or collected from a consortium such as ORIC or ORX). Studies based on real data are thus rather rare in the scientific OpRisk literature. The database of OpRisk losses we investigate differs from "real" OpRisk losses in that it consists of a range of events (not only of losses from the banking industry, although we specifically address this case in the "Dynamic POT Analysis Based on the U.S. Banking and Finance Industry" section), it has been collected by a single person (instead of a company or consortium), the information is drawn from public media (rather than internally), starting about 30 years before the notion of OpRisk has appeared in the Basel II guidelines (as opposed to around the year 2004). Also, the sample size is quite small. It was thus not possible to conduct a comprehensive data analysis based on our model for "real" internal data for one specific company and our results should be viewed in this light. Whether our findings also apply to complete internal data the industry with corresponding database access has to investigate further (and the best we could do here is to provide an open-source implementation of the model). Nevertheless, we found it quite reassuring that, for example, the important parameter ξ in our studies is comparable to the findings of Moscadelli (2004), whose database consists of such "real" losses (although also not only collected from a specific company) with a considerably larger sample size.

- It is important to say something on infinite mean models resulting from fitted $\hat{\xi} \geq 1$ as we partly saw in the "Application to an OpRisk Loss Database" section. 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 interevent 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 (which is entirely conceivable when using VaR as a risk measure). A fully unconditioned POT analysis (possibly resulting in $\hat{\xi} \geq 1$) points into a direction of extreme uncertainty of high quantile (VaR) estimates and should trigger a warning concerning the possibility for accurate statistical estimation. A relevant, thought-provoking article on the economics of climate change in the presence of (very) heavy-tailed risks is Weitzman (2009) and the various reactions on that article, 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 Ibragimov and Walden (2007, 2008), and Embrechts, Puccetti, and Rüschendorf (2013). Das et al. (2013) provide 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 that has precise mathematical formulations and consequences; see, for instance, Embrechts, Klüppelberg, and Mikosch (1997, sect. 8.3.3). Recent fraud (hence OpRisk) events have unfortunately shown how much this principle holds in the world of banking. In the case of settlements related to fraudulent marketing of mortgage-backed securities that helped cause the 2007–2009 financial crisis, fines have indeed become huge: Citigroup ($7 billion), JP Morgan Chase ($13 billion) and, at the time of writing these lines, a record $16–17 billion for Bank of America (*Wall Street Journal*, August 6, 2014).

- The kind of OpRisk data we could get access to raises the question whether there is an overall bias in terms of a positive shift of the losses (as large losses are more likely to be reported publicly). However, note that our model effectively deals with such a bias. To see this, assume all exceedances are shifted above by h. Then, the excesses are $Y_{t_i} = (X_{t_i} - h) - u = X_{t_i} - (h + u), i \in \{1, \ldots, n\}$, and thus we are in the same setup as before with the threshold being replaced by (the larger) $h + u$. Since we conducted a threshold analysis, this bias has been taken care of.
- In theory, the model allows for all possible interactions between the covariates; in practice, this is only possible for sufficiently large sample sizes. The simulated example in the "Demonstration of the Dynamic Approaches Based on Simulated Data" section in the Appendix partly considers interactions between the variables "group" and "year." This has been possible because enough data have been drawn for each combination of interaction level. Furthermore, for sufficiently large data sets, 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 the Appendix. Indeed "large enough" is difficult to define. In order to "gain in sample size" for the OpRisk data considered in this article, we found it reasonable not to include the covariate "event type." A larger data set would allow us to fit all possible interactions (combinations) of "business line," "event type," and "year." It would also allow for 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 considered carefully.
- One concern with any regression model is spurious regression. This phenomenon could occur, for example, if there is another hidden (or present) variable, with which the considered losses and covariates are correlated. In principle we cannot rule out such a possibility, but in case other covariates are identified at some point in time, they can be incorporated in our model and do not require to create a new modeling methodology. This feature of our methodology comes from the underlying generalized additive modeling approach. It also allows to consider different

link functions to improve the fitting capabilities, a feature we have only partially made use of in the present article for guaranteeing that λ is positive (with a log-link function).

- In the "Dynamic Block Maxima Method" section, 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, Klüppelberg, and Mikosch (1997, p. 148). See also Theorem 1 (for $\lambda = 0$) and the comment thereafter in Wadsworth, Tawn, and Jonathan (2010). The numerical reduction from $\hat{\xi} \geq 1$ obtained from the original data to $\hat{\xi} < 1$ for the log-transformed data may sometimes be useful; for $\xi \geq 1$ (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 that 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 the language 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 (nontrivial) 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 of our methodology clearly extends to the 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.

APPENDIX A

Derivatives of the Reparameterized Log-Likelihood

Algorithm A.1 requires computing 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

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

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).

Fitting and Bootstrapping Confidence Intervals

Algorithm A.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 (nondynamic) approach described in the "Classical EVT Approach for Modeling Losses" section; see the Fitting and Bootstrapping Confidence Intervals Algorithm below for this step. The formulas `xi.formula` and `nu.formula` refer to the (parametric, nonparametric, or semiparametric) model specification; see, for example, (24)–(29) in the "Application to an OpRisk Loss Database" section. For the derivatives of the reparameterized log-likelihood appearing in the following algorithm, see the "Derivatives of the Reparameterized Log-Likelihood" section.

Algorithm A.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 $z_i = (t_i, x_i, y_{t_i})$, $i \in \{1, \dots, n\}$, 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)

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

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

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

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)

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

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

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

for the specified formula `nu.formula`.

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

Based on Algorithm A.1, the following backfitting algorithm computes the estimators

$$\hat{\boldsymbol{\xi}} = (\hat{\xi}(x_1, t_1), \ldots, \hat{\xi}(x_n, t_n)),$$
$$\hat{\boldsymbol{\beta}} = \left(\frac{\exp(\hat{\nu}(x_1, t_1))}{1 + \hat{\xi}(x_1, t_1)}, \ldots, \frac{\exp(\hat{\nu}(x_n, t_n))}{1 + \hat{\xi}(x_n, t_n)}\right)$$

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

Algorithm A.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}, $i \in \{1, \ldots, n\}$, as described in "The POT Approach" section.

(2.2) Compute $\nu^{\text{MLE}} = \log((1 + \xi^{\text{MLE}})\beta^{\text{MLE}})$.

(2.3) Set $k = 1$, $\boldsymbol{\xi}^{(k)} = (\xi^{\text{MLE}}, \ldots, \xi^{\text{MLE}})$, and $\boldsymbol{\nu}^{(k)} = (\nu^{\text{MLE}}, \ldots, \nu^{\text{MLE}})$.

(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 A.1 (Newton step).

(3.2) Check convergence: If the mean relative differences satisfy

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

stop. In this case, compute

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

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 1.

The following algorithm wraps around Algorithm A.2 to compute a list of estimates (as returned by Algorithm A.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 A.3 (Post-blackend bootstrap; `QRM::gamGPDboot()`)

(1) Compute estimates $\hat{\boldsymbol{\xi}}$, $\hat{\boldsymbol{\nu}}$ (and $\hat{\boldsymbol{\beta}}$) with Algorithm A.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 A.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 $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\beta}}^{(b)}$, $b \in \{1,\dots,B\}$).

Demonstration of the Dynamic Approaches Based on Simulated Data

In this section, we provide an example based on simulated data (sample size of 2,000) 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 `demo(game)` in `QRM`.

The Dynamic POT Approach. We generate a data set 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 (nonstationary) GPD depending on the covariates "year" and "group." We remove the losses of year 2006 in group "A" (to mimic having no losses for this year–group combination) and 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 the "Loss Frequency" section to "The Penalized Maximum Likelihood Estimator and Risk Measure Estimators" section. 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 A1. Overall, as a comparison with the true parameters and values indicates, our methodology can be applied to fit a model to the entire data set and captures the different functional forms driven by the covariates.

The Dynamic Block Maxima Approach. We run a similar simulation study to assess the performance of the dynamic GEV approach introduced in the "Dynamic Block Maxima Method" section. We generate a data set 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 (nonstationary) GEV with location parameter μ depending on the covariates "year" and "group" using nonlinear functions; here we consider all simulated losses (no missing data). We fit the semiparametric 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 the "Dynamic Block Maxima Method" section. We then compute dynamic estimates of VaR_α; see Figure A2.

Figure A1
Simulated Data: Estimates $\widehat{\mathrm{VaR}}_{0.999}$ (on Log Scale) Including Bootstrapped Pointwise Two-Sided 95% Confidence Intervals Depending on Time and Group, Obtained From the Dynamic POT Method

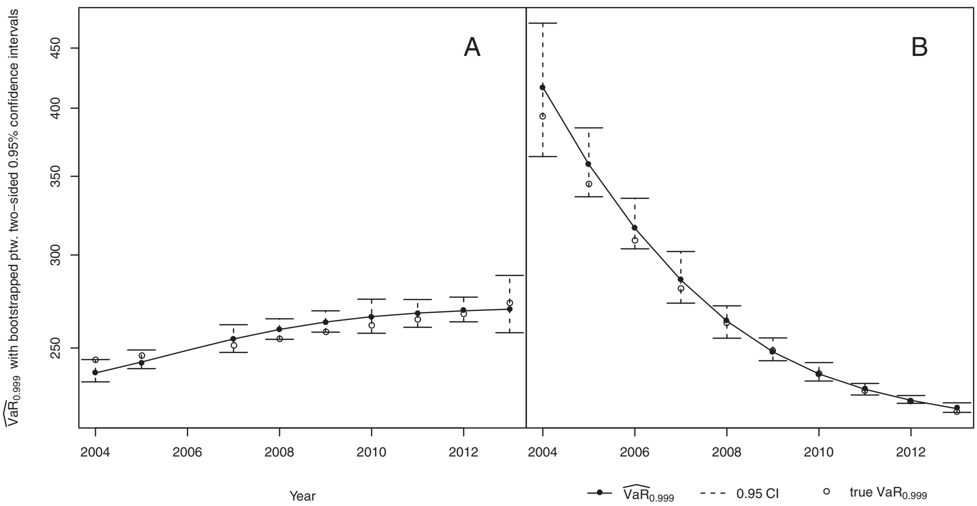
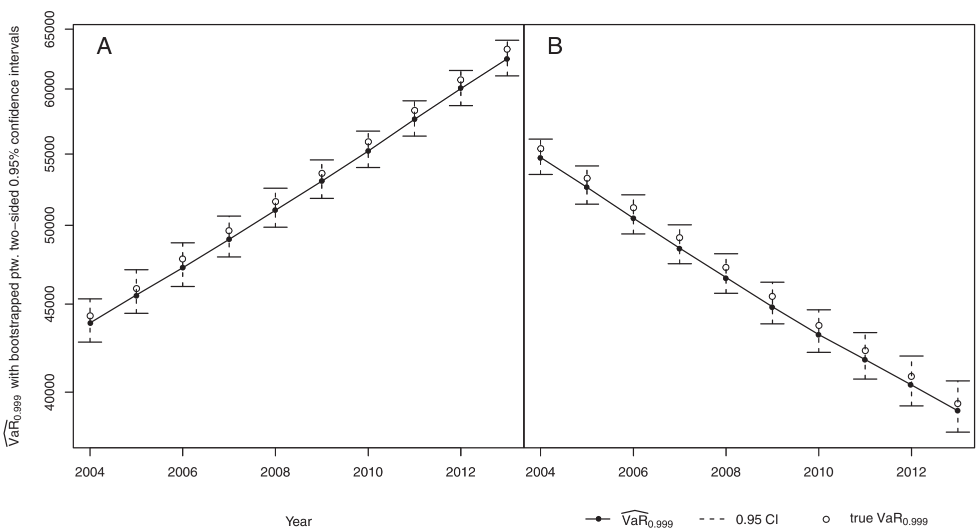


Figure A2
Simulated Data: Estimates $\widehat{\mathrm{VaR}}_{0.999}$ (on Log Scale) Including Bootstrapped Pointwise Two-Sided 95% Confidence Intervals Depending on Time and Group, Obtained From the Dynamic GEV Block Maxima Method



REFERENCES

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

Aue, F., and M. Kalkbrener, 2006, LDA at Work: Deutsche Bank's Approach to Quantifying Operational Risk, *Operational Risk*, 1(4): 49-93.

Bank for International Settlements, 2006, Basel II: International Convergence of Capital Measurement and Capital Standards: A Revised Framework—Comprehensive Version.

Bank for International Settlements, 2012, Fundamental Review of the Trading Book, Consultative Document, May.

Bank for International Settlements, 2013, Fundamental Review of the Trading Book: A Revised Market Risk Framework, Consultative Document, October.

Baud, N., A. Frachot, and T. Roncalli, 2002, Internal Data, External Data and Consortium Data: How to Mix Them for Measuring Operational Risk, Groupe de Recherche Opérationelle Crédit Lyonnais.

Baud, N., A. Frachot, and T. Roncalli, 2003, How to Avoid Over-Estimating Capital Charge for Operational Risk? *risk.net*, Groupe de Recherche Opérationelle Crédit Lyonnais.

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

Böcker, K., and C. Klüppelberg, 2010, Multivariate Models for Operational Risk, *Quantitative Finance*, 10(8): 855-869.

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

Brechmann, E. C., C. Czado, and S. Paterlini, 2014, Flexible Dependence Modeling of Operational Risk Losses and Its Impact on Total Capital Requirements, *Journal of Banking & Finance*, 40, 271-285.

Chavez-Demoulin, V., 1999, Two Problems in Environmental Statistics: Capture-Recapture Analysis and Smooth Extremal Models, PhD Thesis, EPF Lausanne.

Chavez-Demoulin, V., and A. C. Davison, 2005, Generalized Additive Models for Sample Extremes, *Applied Statistics*, 54(1): 207-222.

Chavez-Demoulin, V., and P. Embrechts, 2004, Smooth Extremal Models in Finance and Insurance, *Journal of Risk and Insurance*, 71(2): 183-199.

Chernobai, A., P. Jorion, and F. Yu, 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* (London: Springer-Verlag).

Cope, E., and A. Labbi, 2008, Operational Loss Scaling by Exposure Indicators: Evidence From the ORX Database, *Journal of Operational Risk*, 3(4): 25-46.

Cox, D. R., and N. Reid, 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., C. M. Lewis, and R. Wei, 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 G. Dionnne, 2010, Scaling Models for the Severity and Frequency of External Operational Loss Data, *Journal of Banking and Finance*, 34(7): 1484-1496.
Daníelsson, J., P. Embrechts, C. Goodhart, C. Keating, F. Muennich, O. Renault, and H. S. Shin, 2001, An Academic Response to Basel II, Financial Markets Group, London School of Economics.
Das, B., P. Embrechts, and V. Fasen, 2013, Four Theorems and a Financial Crisis, *International Journal of Approximate Reasoning*, 54, 701-716.
Davison, A. C., and R. L. Smith, 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., J. S. Jordan, V. DeJesus-Ureff, and E. S. Rosengren, 2004, Capital and Risk: New Evidence on Implications of Large Operational Losses, Working Paper 03-5, Federal Reserve Bank of Boston.
de Fontnouvelle, P., E. S. Rosengren, and J. S. Jordan, 2005, Implications of Alternative Operational Risk Modeling Techniques, Working Paper 11103, Federal Reserve Bank of Boston.
Dutta, K., and J. Perry, 2006, A Tale of Tails: An Empirical Analysis of Loss Distribution Models for Estimating Operational Risk Capital, Working Paper 06-13, Federal Reserve Bank of Boston.
El-Gamal, M., H. Inanoglu, and M. Stengel, 2007, Multivariate Estimation for Operational Risk With Judicious Use of Extreme Value Theory, *Journal of Operational Risk*, 2(1): 21-54.
Embrechts, P., and G. Puccetti, 2008, Aggregating Operational Risk Across Matrix Structured Loss Data, *Journal of Operational Risk*, 3(2): 29-44.
Embrechts, P., C. Klüppelberg, and T. Mikosch, 1997, Modelling Extremal Events for Insurance and Finance (Berlin: Springer-Verlag).
Embrechts, P., G. Puccetti, and L. Rüschendorf, 2013, Model Uncertainty and VaR Aggregation, *Journal of Banking and Finance*, 37(8): 2750-2764.
Embrechts, P., G. Puccetti, L. Rüschendorf, R. Wang, and A. Beleraj, 2014, An Academic Response to Basel 3.5, *Risks*, 2(1): 25-48.
Frachot, A., O. Moudoulaud, and T. Roncalli, 2004, Loss Distribution Approach in Practice, in: M. Ong ed., *The Basel Handbook: A Guide for Financial Practitioners* (London: Risk Books).
Ganegoda, A., and J. Evans, 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.
Green, P. J., and B. W. Silverman, 2000, *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach* (London: Chapman & Hall/CRC).
Hastie, T. J., and R. J. Tibshirani, 1990, *Generalized Additive Models* (London: Chapman and Hall).
Ibragimov, R., and J. Walden, 2007, The Limits of Diversification When Losses May Be Large, *Journal of Banking and Finance*, 31, 2551-2569.
Ibragimov, R., and J. Walden, 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., J. Oxman, and Y. Yildirim, 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 edition (New York: McGraw-Hill).
Kagan, Y. Y., and P. Schoenberg, 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., R. Frey, and P. Embrechts, 2005, *Quantitative Risk Management: Concepts, Techniques, Tools* (Princeton, NJ: Princeton University Press).
Mori, A., T. Kimata, and T. Nagafuji, 2007, The Effect of the Choice of the Loss Severity Distribution and the Parameter Estimation Method on Operational Risk Measurement: Analysis Using Sample Data, Reports & Research Papers, 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., J. van den Berg, L. C. Miranda, and M. Leipoldt, 2006, An Econometric Model to Scale Operational Losses, *Journal of Operational Risk*, 1(2): 11-31.
Nelder, J. A., and R. W. M. Wedderburn, 1972, Generalized Linear Models, *Journal of the Royal Statistical Society, Series A*, 135, 370-384.
Nešlehová, J., P. Embrechts, and V. Chavez-Demoulin, 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* (Hoboken: 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., 2011, Modelling Operational Risk Using Bayesian Inference (Heidelberg: Springer).
Shevchenko, P. V., and G. Temnov, 2009, Modeling Operational Risk Data Reported Above a Time-Varying Threshold, *Journal of Operational Risk*, 4(2): 19-42.
Shih, J., A. Khan, and P. Medepa, 2000, Is the Size of an Operational Loss Related to Firm Size? *Operational Risk Magazine*, 2(1): 1-2.
Simonoff, J. S., and C.-L. Tsai, 1999, Semiparametric and Additive Model Selection Using an Improved AIC Criterion, *Journal of Computational and Graphical Statistics*, 8, 1-19.
Soprano, A., B. Crielaard, F. Riacenza, and D. Ruspantini, 2009, Measuring Operational and Reputational Risk: A Practitioner's Approach (Chichester: Wiley).
Wadsworth, J. L., J. A. Tawn, and P. Jonathan, 2010, Accounting for Choice of Measurement Scale in Extreme Value Modeling, *The Annals of Applied Statistics*, 4(3): 1558-1578.
Weitzman, M. L., 2009, On Modeling and Interpreting the Economics of Catastrophic Climate Change, *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).