Modeling the dependence between extreme operational losses and economic factors: a conditional semi-parametric Generalized Pareto approach

J. Hambuckers, C. Heuchenne, O. Lopez

Abstract

In this paper, we model the severity distribution of operational loss data, conditionally to some covariates. Indeed, previous studies [Chernobai et al., 2011, Cope et al., 2012, Chavez-Demoulin et al., 2014a] suggest this distribution might be influenced by macroeconomic and firm-specific factors. We introduce a conditional Generalized Pareto model, where the shape parameter is an unknown function of a linear combination of the covariates. More precisely, we rely on a single-index assumption to perform a dimension reduction that enables to use univariate nonparametric techniques. Hence, we suffer neither from too strong parametric assumption nor from the curse of dimensionality. Then, we develop an iterative approach to estimate this model, based on the maximisation of a semiparametric likelihood function. Finally, we apply this methodology on a novel database provided by the bank UniCredit. We use firm-specific factors to estimate the conditional shape parameter of the severity distribution. Our analysis suggests that the leverage ratio of the company, the proportion of the revenue coming from fees as well as the risk category have an important impact on the tail thickness of this distribution and thus on the probability of suffering from large operational losses.

Key words: Generalized Pareto distribution; Operational risk; Extreme events; Loss distribution approach; Semiparametric regression.

1 Operational losses modeling with conditional Generalized Pareto distributions.

In a series of consultative documents entitled *Operational Risk* and *Operational Risk - Supervisory guidelines for the Advanced Measurement Approach*, the Basel Committee for Banking Supervision (BCBS) discusses the interest for financial institutions to hold capital to cover their operational risks. The BCBS defines these risks as *the risks of direct or indirect loss resulting from inadequate or failed internal processes, people and systems or from external events, excluding therefore legal and reputation risk.* Indeed, for more than 20 years now, events related to this class of risks make regularly the headlines for their huge economical consequences. In 1995, the Barings Bank collapsed due to unauthorized trading positions taken by their senior trader at Singapore, responsible for a \$1.3 billion loss. In 2001, the energy company Enron went bankrupt after the revelation of its accounting malpractices. In a year, the stock price dropped from \$90.75 a share to less than \$1. Citigroup faced charges for helping the accounting fraud and finally provisioned \$2 billion to compensate Enron's investors. In 2003, financial directors of the dairy group Parmalat (Italy) were accused of forgery after revealing that the debt of the company was height times higher than what was stated in the accounting. The company went rapidly bankrupt and the total loss was estimated to \$14.2 billion. More recently, in 2008, rogue trading at Société Générale caused a €4.9 billion loss, whereas Maddof's Ponzi scheme cost \$50 billion to investors. In 2011, the Libor's scandal cost a \$1.5 billion fine to UBS, for its role in fixing rates. All these examples illustrate the fact that operational events (especially employee malpractices) can be the source of sudden and unexpected losses able to put the business continuity in jeopardy. Therefore, it exists a real need for financial institutions to map, model and measure those risks, in order to take proper hedging actions. Following these preoccupations, the econometric literature focused mainly on the issues of modeling and measuring operational risks throughout quantitative techniques.

The BCBS outlines three broad approaches for the quantitative assessment of the operational risks: the basic indicator approach, the standardized approach and the advanced measurement approach (AMA). This last approach quickly gathers the attention of the industry and of the academics. Indeed, whereas the two first approaches rely on simplistic assumptions, the last one acknowledges that banks have specificities to take into account in the computation of their operational risk capital charge. In particular, the Loss Distribution Approach (LDA) offers the possibility to model the total operational loss distribution *via* two specific distributions: one for the number of operational events over a certain horizon of time (the frequency distribution), and one for the severity of these events (the severity distribution). Very intuitively, for a combination business line/event type (the standard classification model of the BCBS), one would supposes that the total (yearly) operational loss amount for the b^{th} business line and the e^{th} event type is given by the sum of all operational events happening during the year:

$$L(b,e) = \sum_{i=1}^{N(b,e)} Z_i, \tag{1}$$

where Z_i is the loss amount of the i^{th} loss and $N(b,e)$ the number of losses over a year for the considered combination (b,e). $N(b,e)$ and Z_i are modeled as random variables, for which we need to assign a density function.

Using this model, a bank can establish a capital charge dedicated to each combination (b,e) to cover potential operational losses. This capital charge is a function of the $99.9th$ order statistic of the distribution of $L(b,e)$ (in the financial literature, this quantity is referred as the Value-at-Risk (VaR) of $L(b,e)$ at the 99.9% level and noted $OpVar(b,e)_{99.9}$). Mathematically, this quantity is defined by

$$\mathbb{P}(L(b,e) \leq OpVar(b,e)_{99.9})) = 0.999 \tag{2}$$

and is simply the 99.9^{th} quantile of $L(b,e)$. In Chapter 3, we focus on the statistical challenges of modeling the tail of the severity distribution (the distribution of Z_i), conditional on some covariates.

With the need to estimate a quantile so far in the (right) tail, most studies [de Fontnouvelle et al., 2004, Moscadelli, 2004, Dutta and Perry, 2006, Aue and Kalkbrener, 2007, Chapelle et al., 2008, Soprano et al., 2009] rely on the Extreme Value Theory (EVT) to model the tail of the severity distribution with a Generalized Pareto distribution (GPD). This distribution has been introduced by Pickands [1975] and Balkema and de Haan [1974]. Indeed, a fundamental result in EVT analysis states that, under certain conditions [see Embrechts et al., 1997, p. 114 and following, for more details], the exceedances z_i of a random variable Z over a suitable large threshold τ approximately follow a GPD. The cumulative distribution function (cdf), with tail parameter γ, location (threshold) parameter τ and scale parameter σ, is given by:

$$GPD(z_i;\gamma,\tau,\sigma) = \begin{cases} 1-\left(1+\gamma\dfrac{(z_i-\tau)}{\sigma}\right)^{-1/\gamma}, & \gamma \neq 0 \\ 1-\mathrm{e}^{-(z_i-\tau)/\sigma}, & \gamma = 0 \end{cases} \tag{3}$$

The probability distribution function (pdf) is given by:

$$gpd(z_i;\gamma,\tau,\sigma) = \begin{cases} \frac{1}{\sigma}\left(1+\gamma\dfrac{(z_i-\tau)}{\sigma}\right)^{-1/\gamma-1}, & \gamma \neq 0 \\ \frac{\mathrm{e}^{-(z_i-\tau)/\sigma}}{\sigma}, & \gamma = 0 \end{cases} \tag{4}$$

The associated p th quantile $q_p(\gamma,\tau,\sigma)$, defined such that $\mathbb{P}(Z-\tau < q_p(\gamma,\tau,\sigma)) = p$, is given by:

$$q_p(\gamma,\tau,\sigma) = \begin{cases} \tau+\sigma((-\log(1-p))^{-\gamma}-1)/\gamma, & \gamma \neq 0 \\ \tau+\sigma(-\log(-\log(1-p))), & \gamma = 0 \end{cases} \tag{5}$$

and is sometimes noted $VaR_p(\gamma,\tau,\sigma)$ (indeed, the pth quantile is the Value-at-Risk with a confidence level equal to p). The distribution of the excess over the threshold (i.e. $Y = Z-\tau$) is $GPD(y_i;\gamma,0,\sigma)$ distributed and then simply noted $GPD(y_i;\gamma,\sigma)$.

This modeling technique has been widely applied in all areas of Finance. See Chavez-Demoulin and Embrechts [2010] for a discussion on the subject. Nevertheless, some authors propose other models for the severity distribution: Degen et al. [2007] and Shevchenko [2010] use the *g-and-h* distribution, whereas Plunus et al. [2012] adapt the CreditRisk+ approach to the operational risk context. However, all these studies assume that the parameters of the frequency and severity distributions are constant conditional on business lines and event types only, while recent studies emphasized that they could be function of other variables: time, macroeconomic indicators, board composition, firm size, regulatory environment, unemployment rate and regulation intensity, among others. Chernobai et al. [2011] investigate the link between the frequency of operational loss events and firm-specific variables (market value of equity, firm age, Tier 1 capital ratio, cash holding ratio, etc.), as well as macroeconomic variables (Moody's Baa-Aaa credit spread, gross domestic product - GDP growth rate, S&P500 returns and volatility, SEC budgets). They find a strong relationship between the frequency and firm-specific variables. They use a conditional Poisson regression model whose process intensity is estimated at a monthly frequency. Wang and Hsu [2013] investigate the role of the board characteristics (i.e. size, age, independence and tenure of the members) as drivers of operational risk events frequency. They use logistic regression models to measure these relationships and find that the heterogeneity of the board size and of the age play a significant role in the likelihood of operational risk events. Cope et al. [2012] study the macroenvironmental determinants of the operational loss severity. They discover that some event-types are sensitive to per-capita GDP, a governance index, shareholder protection laws and supervisory power.

Notwithstanding their empirical contributions, all of these studies suffer from the same drawback: none of them focus on the tail of the severity distribution and none of them provide a way to estimate the GPD parameters conditional on these covariates. A solution would be to perform a maximum likelihood estimation taking into account only subsets of losses (i.e. those that share a same level of a covariate), but the small size of the current database, as well as the continuous nature of some covariates, make this technique very ineffective because a lot of data would be wasted. To the best of our knowledge, only two studies, in the context of financial data modeling, provide methods to estimate the GPD parameters conditional on covariates, taking into account all available data. Beirlant and Goegebeur [2004] propose a local polynomial estimator in the case of a single covariate. Nonetheless, the convergence rate of this estimator in the multivariate case decreases rapidly when the dimension of the number of covariates increases. They illustrate their methodology on a Norwegian fire claim dataset, conditional on the year of the claim. Besides, Chavez-Demoulin et al. [2014a] develop an additive model with spline smoothing to properly use the covariates in the estimation of the GPD and Poisson distribution parameters. Premises of this methodology can be found in Chavez-Demoulin and Embrechts [2004], where a semiparametric smoothing technique is developed. Chavez-Demoulin et al. [2014a] approach supposes an additive structure and uses a cubic spline to estimate the relationship between parameters and covariates. They apply their methodology to a public database of operational losses, using the business lines and the time as covariates. Although their method enables to consider the time in a nonparametric way, other covariates may only linearly influence the GPD parameters.

Despite these recent methodological improvements, covariates investigated in Chernobai et al. [2011], Cope et al. [2012], Wang and Hsu [2013] have never been used as explanatory variables of the conditional severity distribution. Chernobai et al. [2011] and Wang and Hsu [2013] focus on the frequency distribution, whereas Beirlant and Goegebeur [2004] and Chavez-Demoulin et al. [2014a] do not consider these variables. We believe that it is due first to technical considerations: before Beirlant and Goegebeur [2004] for the univariate case and Chavez-Demoulin et al. [2014a] for the multivariate case, no formal method existed to compute GPD parameters conditional on covariates. Second, as noticed by Chavez-Demoulin et al. [2014a], it is well known to be very difficult for academics to get their hands on real operational loss data, furthermore on data associated with covariates. Thus, the issue of developing parameters estimators of the GPD conditional on covariates is challenging but the results are promising. Indeed, it is very likely that the frequency and the severity distribution of the operational losses are driven by factors internal and external to the firm: a poor control environment in the company, a merger, a modification of the volatility of the stock exchange, the acquisition of some buildings in a seismic region, etc., should influence the frequency and the size of operational losses. If, in the estimation process, we use some variables that measure these changes, we could improve the estimation. Moreover, being able to model the severity distribution conditional on to economic factors is of interest for practitioners and policy makers because it would allow to assess the impact of adverse scenarios on banks' capitalizations [see Petrella and Resti, 2013, for more details].

In Section 2, we introduce a new method to estimate the shape parameter ($\gamma(X)$) of the GPD conditional on some covariates, using a semiparametric regression model. This statistical technique combines conveniently para- and nonparametric techniques to form a reasonable compromise between fully parametric and fully nonparametric models. On a methodological point of view, one can effectively assume a parametric link function between the parameters and the covariates, as in the models (21) and (24) to (28) of Chavez-Demoulin et al. [2014a]. Indeed, parametric models usually provide nice convergence rates of the estimators. Nevertheless, they often rely on strong assumptions that can be unrealistic. On the other side, a nonparametric approach has the advantage to let the link function totally unspecified, as in Beirlant and Goegebeur [2004] and the other models in Chavez-Demoulin et al. [2014a]. However, the convergence rates of these estimators decrease rapidly when the number of covariate increases. To solve these issues, our method relies on a single-index model that performs a dimension reduction of the covariates [see, e.g., Härdle et al., 1993, Ichimura, 1993]. Such a model assumes that

$$\gamma(X) = \gamma_{\theta_0}(\theta_0^T X) \tag{6}$$

where $\gamma(X)$ is the conditional shape parameter of the GPD, θ_0 an unknown vector of parameters, γ_{θ_0} an unknown link function and X the vector of covariates. The idea is that, if we knew the true parameter θ_0, we would be back to an univariate case where $\gamma(X)$ can be obtained with a nonparametric estimator, without suffering from the curse of dimensionality. Hence, working with this technique avoids both the curse of dimensionality and inflexible parametric models. Of course, the counterpart is that we need to find sufficiently good estimators of θ_0, to regress correctly $\gamma(X)$ over $\theta^T X$. To achieve this goal, we use a two-step iterative algorithm, that improves at each step the estimation of either θ_0 or $\gamma(X)$. We start first with a preliminary estimator $\hat{\theta}^{(0)}$ of θ_0, using the average derivative technique [Hristache et al., 2001]. Afterwards, we estimate $\gamma(X)$ using nonparametric (e.g. local polynomial technique, as in Beirlant and Goegebeur [2004]). These estimators are later on used in a semiparametric likelihood function that is maximized to obtain a consistent estimator $\hat{\theta}$ of θ_0. The operation is repeated until convergence of a likelihood criterion. This setting is inspired from the one proposed in Hristache et al. [2001]. Notice that our approach is in the same spirit as the one of Chavez-Demoulin and Embrechts [2004] and Chavez-Demoulin et al. [2014a] but differs in the estimation procedure. Indeed, whereas they use a spline smoothing to estimate the link function, we use a kernel smoothing approach over a single-index. Additionally, we discuss the practical estimation of the conditional scale parameter, either with the method of conditional moments or throughout conditional quantile regression techniques. We also discuss briefly the choice of the threshold parameter.

In Section 3, we study the finite sample behaviour of this methodology throughout simulations. We implemented the procedure in MatLab 2013a. We generate GPD data according to a single-index model with two different link functions $\gamma(X)$. To initialize our iterative procedure, we use the average derivative technique based on local polynomial estimators of the conditional moments. We select the bandwidth parameter with a leave-one-out cross-validation procedure. Then, we compare the obtained results with our starting solutions. We see that, with three covariates and the leave-one-out cross-validation procedure, our iterative procedure decreases the mean and median error rates on $\gamma(X)$, and the median error rates on θ. These simulations emphasize the need to select carefully the bandwidth parameter used to construct the semiparametric likelihood function maximized over θ.

Finally, in Section 4, we illustrate our methodology on real data. Our dataset is fairly unique and consists in around 41,000 operational losses from the Italian bank UniCredit. These data have been provided by the operational risk department of UniCredit[1]. We estimate a threshold parameter and assume a conditional GPD for the tail of the severity distribution. Then, we look at the value of the shape parameter conditional on subsets of covariates. We perform both univariate (i.e. we consider the potential effect of the covariates on the shape parameter one by one) and multivariate analyses (i.e. we consider the simultaneous effect of subsets of explanatory variables). To estimate the conditional shape parameter of the severity distribution, we use 3 firm-specific covariates: an efficiency ratio, the percentage of the bank's total revenue coming from fees and the leverage ratio. We combine these continuous covariates with some classification covariates internal to the bank, inspired by the Basel classification. Our analysis suggests that high leverage ratios and low percentages of the revenue coming from fees are associated with high values of $\gamma(X)$. The leverage ratio seems to be the most influential covariate. Also, the event type seems to influence the severity distribution: losses resulting from internal and external frauds have a severity distribution with a higher $\gamma(X)$ than the other event types. Finally, we observe that the effect of the efficiency ratio may depend on the event type. For example, an increase of the efficiency ratio will have a positive effect on $\gamma(X)$ for losses related to the *Employee practices and workplace safety* event type, whereas it has a negative effect on losses related to the *Internal frauds* event type. We observe that a balanced economic situation, i.e. a low leverage ratio and a high % of fee revenue, exhibit lower values of $\gamma(X)$. Overall, we show that taking into account the level of these covariates impacts the shape parameter of the severity distribution. Our analysis reveals that if economic changes happen, it will have undoubtedly some consequences on the total $OpVaR_{99.9}$ and on the capital requirements of a financial institution.

[1] A big thanks to Pr. A. Nassig and Dr. F. Piacenza for their help in obtaining these data.

The rest of the paper is organized as follows: in Section 2, we introduce our semi-parametric approach. In Section 3 we investigate the finite sample behaviour of the estimation procedure. In Section 4 we illustrate the methodology on the UniCredit data whereas we conclude briefly in Section 5.

2 Methodology

We recall first some general results regarding the EVT and the Peak-Over-Threshold (POT) methodology (Subsection 2.1). Then, in Subsection 2.2 we detail the model and our estimation procedure of the shape parameter. We focus on the shape parameter because it controls the tail thickness of the GPD and has a big influence on the traditional risk indicators (VaR, ES,...). Indeed, the moment generating function of the GPD is exponentially linked to the shape parameter, whereas it is only linear in the scale parameter. Eventually, in Subsection 2.3 we present briefly a way to obtain an estimator of the scale parameter in our context, using either the method of conditional moments or kernel quantile regression techniques. We also discuss the choice of a good threshold parameter, this task being a recurrent issue in the EVT approach.

2.1 Peaks-over-Threshold (POT) approach

The POT method is based on a theorem known under the name of *Pickands-Balkema-de Haan theorem* [Pickands, 1975, Balkema and de Haan, 1974]. Let's denote $\{z_i\}$, $i = 1, ..., N(T)$ the exceedances of a r.v. Z, over a sufficiently high threshold τ and along a period of time T (the index i denotes that the realization z of Z exceeds the threshold). Under some mathematical conditions (see below), this theorem states that

1. the number of exceedances $N(T)$ approximately follows a Poisson process with intensity λ, i.e. $N(T) \sim P(\lambda)$,

2. the severity of the excesses $y_i = z_i - \tau$ approximately follows, independently from $N(T)$, a $GPD(y_i; \gamma, \sigma)$ for $y_i \geq 0$ if $\gamma \geq 0$, and for $\sigma > 0$. See equation (4).

Especially, the second result implies that the excess conditional distribution function of i.i.d. realizations of Z (i.e. $\mathbb{P}(Z - \tau \leq z | Y > \tau)$) converges to a GPD distribution with parameters γ and σ, when τ goes to infinity. This result holds under the condition that $F(z)$ (the cumulative distribution function of Z) respects

$$\bar{F}(z) = 1 - F(z) = z^{-1/\gamma} L(z)$$

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

$$\lim_{z \to \infty} \frac{L(\nu z)}{L(z)} = 1$$

for all $\nu > 0$. This condition, known as *the regular variation* property, implies that the tail of the loss distribution decays at a power rate of Z. Hence, relying on this asymptotic result, one can model the distribution of the excesses over τ using a GPD distribution and use maximum likelihood techniques to obtain estimators of γ and σ.

2.2 A single-index assumption for the GPD shape parameter

Starting from the previous subsection, we assume that, given a set of covariates $X \in \mathcal{X} \subset \mathbb{R}^d$, the response variable $Y = Z - \tau$ (i.e. the amount of an operational loss above the threshold τ, conditional on $Z > \tau$) follows exactly a GPD with conditional tail parameter $\gamma(X)$ and scale parameter $\sigma(X)$:

$$Y \sim GPD(\gamma(X), \sigma(X)). \tag{7}$$

The GPD function is given by equation 3. Moreover, we assume that the tail parameter depends from the covariates only through an unknown linear combination of the covariates, that is,

$$\gamma(X) = \gamma_{\theta_0}(\theta_0^T X), \tag{8}$$

where θ_0 is the single-index parameter (parametric part), and γ_{θ_0} is the unknown link function (nonparametric part). The first element of θ_0 is assumed equal to one, to ensure the proper identifiability of the model. Alternatively, this condition can be replaced by the assumption that the norm of θ_0 is equal to one. Using these hypotheses, we perform a dimension reduction, since all the relevant information on the covariates are summarized into a single-index $U = \theta_0^T X$. For a sample of size n, $\{(X_j, Y_j) : j = 1, \cdots, n\}$, once we know θ_0, the problem of finding a good nonparametric estimator $\hat{\gamma}(X)$ of $\gamma(X)$ is reduced to an univariate nonparametric regression problem over $u_j = \theta_0^T x_j$, $j = 1, \cdots, n$ (x_j being a realization of the random vector X_j), where we do not suffer from the curse of dimensionality. One could rely on a semiparametric log-likelihood function, where the conditional shape parameter is expressed as a local polynomial in U:

$$\begin{aligned}(\hat{\gamma}(U), \hat{\gamma}'(U)) = arg \max_{a_j, b_j} \sum_{i=1}^{n} l\left(a_j + b_j(\theta_0^T X_i - U); \frac{Y_i}{\sigma(X_i)}\right)& \\ \times K\left(\frac{\theta_0^T X_i - U}{h}\right),&\end{aligned} \tag{9}$$

where $K(\cdot)$ is a kernel function with integral one, h is a bandwidth parameter and $l(a_j + b_j(\theta_0^T X_i - U); Y_i/\sigma(X_i))$ is the density function of the GPD for observation Y_i with parameters $a_j + b_j(\theta_0^T X_i - U)$ and $\sigma(X_i)$ (see equation 4). In other words, $l(\gamma; y)$ is equivalent to:

$$l(\gamma; y) = -\left(\frac{1}{\gamma} + 1\right) \log(1 + y\gamma).$$

The estimator described in equation 9 has been studied in Beirlant and Goegebeur [2004], and is designed to maximize a localized version of the likelihood function. In practice, the true shape parameter $\gamma(X)$ is never known, this is why we prefer to use a log-likelihood approach instead of minimizing some L2-score functions. If θ_0 were unknown and were estimated by $\hat{\theta}$, we could simply plug $\hat{\theta}$ in equation 9. The drawback with this approach is that it exists no closed-form solution to this estimator, which is the solution of a nonlinear maximization problem. Instead, one may prefer other estimators of γ_{θ_0}. In particular, observe that for $\gamma(X) < 1$,

$$\gamma(U) = 1 - \frac{\sigma(X)}{m_{\theta_0,1}(U)}, \tag{10}$$

where $m_{\theta_0,1}(U)$ is the first moment of Y, conditional on $U = \theta_0^T X$: $m_{\theta_0,1}(U) = \mathbb{E}\left[Y|U = \theta_0^T X\right]$. Using a one-dimensional local polynomial technique relying on a traditional minimization problem of the sum of squared residuals, we can obtain an estimator $\hat{m}_{\theta_0,1}(U)$ of this quantity, as well as an estimator of its first partial derivative $\hat{m}'_{\theta_0,1}(U)$ with respect to U. Then, we can compute

$$\hat{\gamma}(U) = 1 - \frac{\sigma(X)}{\hat{m}_{\theta_0,1}(U)}, \tag{11}$$

and deduce

$$\hat{\gamma}'(U) = \frac{\sigma(X)\hat{m}'_{\theta_0,1}(U)}{\left[\hat{m}_{\theta_0,1}(U)\right]^2}, \tag{12}$$

the corresponding estimated derivative. The main advantage of this technique is that it exists a closed-form solution to this minimization problem [see, e.g., Ruppert and Wand, 1994]. If $\sigma(X)$ is unknown, one can simply use a consistent estimator of this quantity (see Section 3.1.3) and reinstate it in equation 10. In our approach, we chose not to further model $\sigma(X)$ (this choice is motivated by the fact that $\gamma(X)$ is the most important parameter for statistical inference).

On the other side, if γ_{θ_0} were known and θ_0 were unknown, an estimator $\hat{\theta}$ could be obtained using a maximum likelihood approach. This estimator is defined by:

$$\hat{\theta} = \max_{\theta} \frac{1}{n} \sum_{i=1}^{n} l(\gamma_{\theta_0}(\theta^T X_i); Y_i/\sigma(X_i)), \tag{13}$$

where $l(\gamma_{\theta_0}(\theta^T X_i); Y_i/\sigma(X_i))$ is the density function of the GPD for observation Y_i with conditional shape parameter $\gamma_{\theta_0}(\theta^T X_i)$ and scale parameter $\sigma(X_i)$. In practice, this link function is hardly known. Therefore, we could act exactly as in equation 9: replace the unknown quantities by some estimators. For example, we could replace $\gamma_{\theta_0}(\theta^T X_i)$ by our estimator $\hat{\gamma}(X_i)$ defined in equation 9 or in equation 11.

We see that both estimations are intrinsically inter-related: we can compute $\hat{\theta}$ only if we have some estimators of $\gamma(X)$ at hand, and vice-versa. Also, if one of the two quantities is badly estimated, it will have a big impact on the other estimation. A solution could be to perform a joint maximization over θ and $\gamma(X_i)$, but this is a hard task, requiring complex numerical procedures. Instead, we propose to use an iterative procedure, inspired by the set-up of Hristache et al. [2001] to obtain sequences of $(\hat{\gamma}(X), \hat{\theta})$ until our final estimation has converged sufficiently, according to some criteria. This iterative procedure can be summarized in the following way:

1. Set $k = 0$. Find a starting value $\hat{\theta}^{(0)}$ (see below for more details).
2. Chose a bandwidth parameter $h_{(k)}^{opt}$ to compute $\hat{\gamma}$, using e.g. a cross-validation procedure (see below for more details) where $U_i = (\hat{\theta}^{(k)})^T X_i$.
3. Set $k = k+1$ and obtain an updated estimator $\hat{\theta}^{(k)}$, maximizing the following expression over θ:
$$\hat{\theta}^{(k)} = arg\max_{\theta} \frac{1}{n} \sum_{i=1}^{n} l(\hat{\gamma}(X_i); Y_i/\sigma(X_i)). \tag{14}$$
   where $\hat{\gamma}$ is obtained using the bandwidth $h_{(k-1)}^{opt}$
4. Go back to step 2 until convergence of the log-likelihood function .

Hambuckers et al. [2015] demonstrated the consistency and asymptotic normality of the estimator $\hat{\theta}$, computed with equation 14.

In step 3, we suppose that we have a closed-form solution of the semiparametric likelihood function. This is the case when $\gamma(X) < 0$ and if we use Nadaraya-Watson-type or local linear estimators of the conditional first moments (see equation 11). If we use the local-polynomial estimator of Beirlant and Goegebeur [2004] instead, step 3 can be decomposed into an estimation of $\gamma(X)$ and a maximization over θ. Then, the iterative procedure becomes:

1. Set $k = 0$. Find a starting value $\hat{\theta}^{(0)}$.
2. Chose the bandwidth parameter $h_{(k)}^{opt}$, using e.g. a cross-validation procedure for $\hat{\gamma}$ (see below for more details) where $U_i = \hat{\theta}^{(k)} X_i$.
3. Estimate the conditional shape parameter with $\theta = \hat{\theta}^{(k)}$, at points $u_j = \hat{\theta}^{(k)T} x_j$, $j = 1, \cdots, n$, using the local polynomial estimator of Beirlant and Goegebeur [2004] and $h_{(k)}^{opt}$:
$$(\hat{\gamma}(U), \hat{\gamma}'(U)) = arg\max_{a_j, b_j} \sum_{i=1}^{n} l(a_j + b_j(\hat{\theta}^{(k)T} X_i - U); Y_i/\sigma(X_i)) \times K\left(\frac{\hat{\theta}^{(k)T} X_i - U}{h}\right).$$
   This equation is similar to equation 9 with $\hat{\theta}^{(k)}$ replacing θ_0.

4. Set $k = k+1$. Replace a_j and b_j (estimators of $\gamma_{\theta_0}(\theta_0^T x_j)$ and its first derivative) in equation 14 to build a semiparametric likelihood function. Then, obtain an updated estimator $\hat{\theta}^{(k)}$ by maximizing the following equation over θ:

$$\hat{\theta}^{(k)} = arg\max_{\theta} \sum_{j=1}^{n} l(a_j + b_j(\theta^T X_j - U_j); Y_j/\sigma(X_j)). \tag{15}$$

5. Go back to step 2 until convergence of the log-likelihood function.

In step 4, we use the fact that we suppose the tail parameter locally linear in $\theta^T X_j$ in a small window around U_j. We *freeze* the local structure of γ and using this linear approximation, we estimate θ_0.

Regarding the stopping conditions, we assume that the estimation has converged once we observe a decrease or an increase of less than a specified threshold (e.g. $10^{-8}\%$) of the global likelihood function $\frac{1}{n}\sum_{i=1}^{n} l(\gamma(X_i); Y_i/\sigma(X_i))$, or when a predetermined number of steps has been performed (e.g. 10 steps).

Now, two issues remain: selecting an adequate bandwidth parameter and obtaining the starting estimation $\hat{\theta}^{(0)}$. An usual method for bandwidth selection is the leave-one-out cross-validation. In our case, it consists in selecting the bandwidth $h^{opt} \in \mathcal{H}$ that maximizes

$$h^{opt} = arg\max_{s} \sum_{j=1}^{n} l(\hat{g}_{s,-j}(U_j); Y_i/\sigma(X_i)), \tag{16}$$

with

$$(\hat{g}_{s,-j}(U_j), \hat{g}'_{s,-j}(U_j)) = arg\max_{a_j,b_j} \sum_{i\neq j}^{n} l(a_j + b_j(U_i - U_j); Y_i/\sigma(X_i)) \times K\left(\frac{U_i - U_j}{h_s}\right), \tag{17}$$

for $j = 1, \cdots, n$ and for $s = 1, \cdots, S$ (the number of elements in $\mathcal{H}$). Beirlant and Goegebeur [2004] suggest to use the same technique. If we use equations 11 and 12 instead of equation 9, h^{opt} is simply chosen to minimize the cross-validated sum of squared residuals $\sum_{i=1}^{n}(Y_i - \hat{m}_{\theta,1}^{CV})^2$, where $\hat{m}_{\theta,1}^{CV}$ is the leave-one-out estimator of $m_{\theta,1}$.

Regarding the initialization of the iterative procedure, we propose to use the average derivative technique. A detailed explanation of the average derivative technique can be found in Hristache et al. [2001]. If the model is true, we have the following relationships

[Hristache et al., 2001]:

$$\begin{aligned}\nabla_x \gamma(X_i) &= \theta_0 \gamma'_{\theta_0}(\theta_0^T X_i), && (18)\\ \frac{1}{n}\sum_{i=1}^{n} \nabla_x \gamma(X_i) &= \theta_0 \frac{1}{n}\sum_{i=1}^{n} \gamma'_{\theta_0}(\theta_0^T X_i), && (19)\\ \theta_0 &= \lambda^{(0)} \left(\frac{1}{n}\sum_{i=1}^{n} \nabla_x \gamma(X_i)\right), && (20)\end{aligned}$$

where $\lambda^{(0)}$ is a normalizing constant ensuring that the first element of θ_0 is equal to one (remember the identifiability condition stated in equation 7). Then, a preliminary estimator $\hat{\theta}^{(0)}$ of θ_0 is given by:

$$\hat{\theta}^{(0)} = \lambda^{(0)} \left(\frac{1}{n}\sum_{i=1}^{n} \nabla_x \hat{\gamma}^{(0)}(X_i)\right), \quad (21)$$

where $\nabla_x \hat{\gamma}^{(0)}(X_i)$, $i = 1, \cdots, n$ are preliminary estimators of $\nabla_x \gamma(X_i)$. To compute $\nabla_x \hat{\gamma}^{(0)}(X_i)$, we can start from the fact that, for $\gamma(X) < 1$,

$$\gamma(X) = \frac{1}{2} - \frac{m_1(X)^2}{2\left[m_2(X) - m_1(X)^2\right]}, \quad (22)$$

where $m_1(X)$ and $m_2(X)$ are respectively the first and second conditional moments of Y: $m_1(X) = \mathbb{E}\left[Y|X\right]$ and $m_2(X) = \mathbb{E}\left[Y^2|X\right]$ (these moments are not conditioned by any values of θ). Using a local polynomial smoothing technique, we can obtain multivariate estimators $\hat{m}_1(X)$ and $\hat{m}_2(X)$ of these quantities, as well as estimators of their first partial derivatives $\nabla_x \hat{m}_1(X)$ and $\nabla_x \hat{m}_2(X)$, with respect to the d components of X. Then, we can compute

$$\hat{\gamma}^{(0)}(X) = \frac{1}{2} - \frac{\hat{m}_1(X)^2}{2\left[\hat{m}_2(X) - \hat{m}_1(X)^2\right]}, \quad (23)$$

and deduce $\nabla_x \hat{\gamma}^{(0)}(X)$, the corresponding estimated gradient vector. We define a local polynomial estimator of $m_1(X)$, $m_2(X)$ and of their gradients in the following way:

$$(\hat{m}_1(X), \nabla_x \hat{m}_1(X)) = arg \min_{\delta_0, \delta_1, \cdots, \delta_d} \left\{ \sum_{i=1}^{n} Y_i - \delta(X_i - X) \right\}^2 \times \prod_{l=1}^{d} K\left(\frac{X_i(l) - X(l)}{h_1}\right), \quad (24)$$

$$(\hat{m}_2(X), \nabla_x \hat{m}_2(X)) = arg \min_{\delta_0, \delta_1, \cdots, \delta_d} \left\{ \sum_{i=1}^{n} Y_i^2 - \delta(X_i - X) \right\}^2 \times \prod_{l=1}^{d} K\left(\frac{X_i(l) - X(l)}{h_2}\right). \quad (25)$$

where $\delta = \{\delta_0, \delta_1, \cdots, \delta_d\}$ is a $d+1$ dimensional vector of parameters, $X_i(l)$ is the l^{th} element of X_i, h_1 and h_2 are d-dimensional vectors of bandwidths associated to covariates $j = 1, ..., d$ and $K(.)$ a univariate kernel function with integral one. d is the number of covariates (the dimension of X). The multivariate d-dimensional kernel function is obtained by convolution of the univariate kernel densities. This problem is a typical weighted least squares minimization problem, whose solution is given by

$$(\hat{m}_1(X), \nabla_x \hat{m}_1(X))^T = (X_x^T W_x X_x)^{-1} X_x^T W_x Y, \tag{26}$$

where

$$X_x = \begin{bmatrix} 1 & (X_1 - X) \\ \vdots & \vdots \\ 1 & (X_n - X) \end{bmatrix},$$

and

$$W_x = diag\Big\{ \prod_{l=1}^{d} K\left(\frac{X_1(l) - X(l)}{h_2}\right), \cdots, \prod_{l=1}^{d} K\left(\frac{X_n(l) - X(l)}{h_2}\right) \Big\}.$$

More details regarding this expression can be found in Ruppert and Wand [1994]. Notice that equation 10 uses a univariate nonparametric regression, whereas here we rely on a multivariate nonparametric technique that suffers rapidly from the curse of dimensionality. Alternatively, when estimators of the gradient are not needed, we can use simple Nadaraya-Watson estimators, using the following definition of $\hat{m}_1(X)$ and $\hat{m}_2(X)$:

$$\hat{m}_1(X) = \frac{\sum_{i=1}^{n} \prod_{j=1}^{d} K\left(\frac{X_i(l) - X(l)}{h_1}\right) Y_i}{\sum_{i=1}^{n} \prod_{j=1}^{d} K\left(\frac{X_i(l) - X(l)}{h_1}\right)}, \tag{27}$$

$$\hat{m}_2(X) = \frac{\sum_{i=1}^{n} \prod_{j=1}^{d} K\left(\frac{X_i(l) - X(l)}{h_1}\right) Y_i^2}{\sum_{i=1}^{n} \prod_{j=1}^{d} K\left(\frac{X_i(l) - X(l)}{h_2}\right)}. \tag{28}$$

However, when $\gamma(X) > 1$, the relationship stated in equation 10 is not valid and the method of conditional moments does not work. As explained in Chavez-Demoulin et al. [2014a] and de Fontnouvelle et al. [2004], real-life situations with $\gamma(X) > 1$ are not unusual, and it would be interesting to obtain starting values in this case, too.

To solve this issue, we need to find another way to get some estimators of $\nabla_x \gamma(X)$. We propose to rely on the relationship between the quantile $q_p(X)$ of level p of the GPD and its parameters ($\sigma(X)$ and $\gamma(X)$). We define $q_p(X)$ such that, for $Y \sim GPD(\gamma(X), \sigma(X))$, $\mathbb{P}(Y < q_p(X)) = p$. For $\gamma \neq 0$, $p \in [0, 1]$, the following relationship holds:

$$q_p(X) = \frac{\left((1-p)^{-\gamma(X)} - 1\right) \sigma(X)}{\gamma(X)}. \tag{29}$$

Assuming $\sigma(X) = \sigma$, $\forall X \in \mathcal{R}$, and defining $s = 1 - p$, we have also:

$$\nabla_x q_p(X) = (\sigma \nabla_x \gamma(X)) \frac{-\exp^{-\gamma(X)\log(s)} \log(s)\gamma(x) - \left(s^{-\gamma(X)} - 1\right)}{[\gamma(X)]^2}. \tag{30}$$

The idea is to obtain, for a given X, estimations $\hat{q}_{p_1}(X)$ and $\hat{q}_{p_2}(X)$, as well as $\nabla_x \hat{q}_{p_1}(X)$ and $\nabla_x \hat{q}_{p_2}(X)$ (let's say, for $p_1 = 0.25$ and $p_2 = 0.75$). Replacing $q_p(X)$ by $\hat{q}_p(X)$, for $p = \{p_1, p_2\}$ in equation (29), we obtain a system of two equations that can be solved for $\sigma(X)$ and $\gamma(X)$. Afterwards, using the obtained estimations of $\sigma(X)$, we can compute an estimator of the global scale parameter:

$$\hat{\sigma} = \frac{1}{n} \sum_{i=1}^{n} \hat{\sigma}(X_i),$$

for n the number of points in our design and $\hat{\sigma}(X)$ an estimator of $\sigma(X)$. Then, for a given p (let's say $p = .5$), we can simply replace σ, $\gamma(X)$, $q_p(X)$ and $\nabla_x q_p(X)$ in equation 30 by their corresponding estimators $\hat{\sigma}$,$\hat{\gamma}^{(0)}(X)$, $\hat{q}_p(X)$ and $\nabla_x \hat{q}_p(X)$, and solve for $\nabla_x \gamma(X)$. Finally, $\nabla_x \hat{\gamma}(X)$ is plugged in equation (21) to obtain our preliminary estimator $\hat{\theta}^{(0)}$ of θ_0 *via* the average derivative technique.

Practically speaking, $q_p(X)$ and $\nabla_x q_p(X)$ can be estimated using a local polynomial quantile regression procedure, based on the check loss function [see, Yu and Jones, 1998, for a discussion on the subject]. In this approach, $\hat{q}_p(X)$ and $\nabla_x \hat{q}_p(X)$ are the solutions of the following minimization problem:

$$\begin{aligned}(\hat{q}_p(X), \nabla_x \hat{q}_p(X)) = \arg \min_{\kappa_0, \kappa_1, \cdots, \kappa_d} \Bigg\{ &\sum_{i=1}^{n} p\, C_i \mathbb{1}(C_i > 0) \\ &+ (1-p)\, C_i \mathbb{1}(C_i < 0) \Bigg\} \prod_{l=1}^{d} K\left(\frac{X_i(l) - X(l)}{h}\right), \end{aligned} \tag{31}$$

with

$$C_i = |Y_i - \kappa_0 - \sum_{j=1}^{d} \kappa_j (X_i(j) - X(j))|, \tag{32}$$

and where $\kappa = \{\kappa_0, \kappa_1, \cdots, \kappa_d\}$ is a $d+1$ dimensional vector of parameters, $X_i(l)$ is the l^{th} element of X_i, h is d-dimensional vectors of bandwidths associated to covariates $j = 1, \cdots, d$ and $K(.)$ a univariate kernel function with integral one. d is the number of covariates (the dimension of X). The multivariate d-dimensional kernel function is obtained by convolution of the univariate kernel densities.

2.3 Estimating the scale and the threshold parameters

Eventually, we need also estimators of the conditional scale and threshold parameters to fully characterize the conditional severity distribution of operational losses.

For the scale parameter, we propose once again to use the method of conditional moments since, for $\gamma(X) < 1$, we have:

$$\sigma(X) = \frac{m_1(X)}{2}\left(1 + \frac{m_1(X)^2}{m_2(X) - m_1(X)^2}\right), \tag{33}$$

with the same definition of $m_1(X)$ and $m_2(X)$ as in Subsection 3.1.2 (see equations 28 and following). These moments (as well as their gradients) can be estimated and then plugged in equations 23 and33 for the estimation of $\hat{\sigma}(X)$ and $\hat{\theta}^{(0)}$. If we expect $\gamma(X) > 1$, then we can use the conditional quantile regression technique described at the end of Subsection 2.2. If we want a global estimator $\hat{\sigma}$ of the scale parameter, instead of conditional ones, we can simply use:

$$\hat{\sigma} = \sum_{i=1}^{n} \hat{\sigma}(X_i). \tag{34}$$

which is the same estimator as the one used in the quantile regression approach.

Besides, the estimation of the threshold parameter is a delicate operation in the POT approach. Indeed, as explained in Embrechts et al. [1997], Chavez-Demoulin et al. [2014a,b] and Scarrot and McDonald [2012], the choice of the threshold involves to find a balance between bias and variance: the threshold needs to be sufficiently high to ensure that the asymptotic convergence to a GPD approximately holds, in order to have a small bias. On the other hand, a large threshold reduces the sample size, and increases the variance of the other estimators. Usually, one works with a fixed threshold approach, meaning that the threshold is chosen before fitting the other parameters. In our empirical application, we use a combination of graphical diagnostic tools (like the Hill plot) and rules of thumb to make this choice. Chapelle et al. [2008] use an algorithm based on the modified Hill estimator of Huisman et al. [2001] and the empirical cumulative distribution function of the data. Moscadelli [2004] uses a graphical analysis of the mean excess plot. Chavez-Demoulin et al. [2014a] prefer to simply use the median, after some robustness tests with statistics of higher orders. Overall, as noticed by Chavez-Demoulin et al. [2014b] (p.46), the threshold selection has a degree of arbitrariness in practice. Ideally, in our case, the threshold should be also a function of the covariates, but this raises difficult theoretical questions that are beyond the scope of this thesis. Therefore we use the following tools to choose a threshold:

1. A graphical analysis of the mean excess plot. For a set of increasing thresholds $\tau_j > 0$, between 0 and $\max\{Z_1, \cdots, Z_n\}$, for $j = 1, 2, \cdots$, we compute

   $$\mathbb{E}(Z - \tau_j | Z > \tau_j) = \sum_{i=1}^{m} \mathbb{1}(Z_i > \tau_j)(Z_i - \tau_j) / \sum_{i=1}^{m} \mathbb{1}(Z_i > \tau_j),$$

   where $\mathbb{1}$ is an indicator function taking value one if the condition in parentheses is met, zero otherwise. m is the number of observed losses Z_i. If the data follow effectively a GPD, the mean excess function should appear linear. Hence, one will select the lowest threshold so that the mean excess function is linear in τ. See, e.g., Ghosh and Resnick [2010] for a discussion on the subject.

2. A graphical analysis of the modified Hill estimator (obtained by a weighted least square approach) of Huisman et al. [2001]. A good threshold supposes to find a correct balance between bias (if the threshold is too low, the tail parameter is estimated on non-GPD data) and variance (if the threshold is too high, the tail parameter is estimated on too few data and the estimator is very volatile). For a sequence of increasing threshold, the sequence of estimated tail parameters should stabilize at some point (before this point, it increases or decreases, and beyond this point, it becomes very volatile). The Hill estimator being biased for small samples, Huisman et al. [2001] proposed a modified version of this estimator, less biased. We use it in our graphical analysis.

3 Simulations

In this section, we perform a small simulation study to illustrate the potential of the proposed method. Especially, we want to know how our iterative procedure improves the average derivative estimator of θ and the resulting initial estimation of $\gamma(X)$. We generate $B = 250$ samples of size $n = 2000$ from the GPD, following the single-index model described in the previous section. We specify two different settings, with two different link functions γ. The threshold parameter (τ) and the scale parameter (σ) are assumed known, constant and equal to respectively 0 and 1. The covariates X are composed of 3 components uniformly distributed on [0, 1]. These components are correlated between each other throughout a Gaussian copula, with a correlation matrix M (see below). In both settings, $\theta_0 = [1\,0.5\,1.5]$. In the iterative procedure, we use the estimator of $\gamma(X)$ described by equations 11, for which we have a closed-form solution. The conditional first moment is estimated using a local polynomial estimator [Ruppert and Wand, 1994]. At each step of the iterative procedure, we select the bandwidth parameter throughout a leave-one-out procedure. The bandwidth is chosen over [0.15 , 5]. Analytically:

$$Y(X) \sim GPD(\gamma(X), \sigma(X)). \tag{35}$$

$$X^{(d)} \sim U(0,1),\ d = 1, \cdots, 3, \tag{36}$$

with

$$M = \begin{bmatrix} 1 & 0.5 & 0.6 \\ 0.5 & 1 & 0.72 \\ 0.6 & 0.72 & 1 \end{bmatrix}$$

and the two different link functions are:

Model 1: $\gamma(X) = .05 + 0.2\,(\theta_0^T X)$,

Model 2: $\gamma(X) = .3\,\log(1 + 0.4\,\theta_0^T X) + 0.05$,

We compare the quality of the estimations (obtained either with the average derivative or with the proposed iterative procedure) for both θ and $\gamma(X)$. Regarding the estimation of θ, we use the mean squared error (MSE_θ) and the mean absolute error (MAE_θ) criteria. These quantities are estimated in the following way:

$$\widehat{MSE}_\theta = \frac{1}{B}\sum_{b=1}^{B}\|\hat{\theta}_b - \theta_0\|^2, \tag{37}$$

$$\widehat{MAE}_\theta = \frac{1}{B}\sum_{b=1}^{B}\|\hat{\theta}_b - \theta_0\|. \tag{38}$$

To obtain indicators less sensitive to extreme values, we also compute the median squared error and the median absolute error, given by:

$$\widehat{mSE}_\theta = Md(\|\hat{\theta}_b - \theta_0\|^2),\ b = 1, \cdots, 500, \tag{39}$$

$$\widehat{mAE}_\theta = Md(\|\hat{\theta}_b - \theta_0\|),\ b = 1, \cdots, 500. \tag{40}$$

where Md stands for median. Regarding the estimation of $\gamma(X)$, we use the mean integrated squared error ($MISE_\gamma$) and the mean integrated absolute error ($MIAE_\gamma$) criteria. These quantities are estimated in the following way:

$$\widehat{MISE}_\gamma = \frac{1}{B\,n}\sum_{b=1}^{B}\sum_{i=1}^{n}(\hat{\gamma}(X_{b,i}) - \gamma(X_{b,i}))^2, \tag{41}$$

$$\widehat{MIAE}_\gamma = \frac{1}{B\,n}\sum_{b=1}^{B}\sum_{i=1}^{n}|\hat{\gamma}(X_{b,i}) - \gamma(X_{b,i})|. \tag{42}$$

As for θ, we also compute the median integrated squared error and the median integrated absolute error:

$$\widehat{mISE}_\gamma = Md\{\frac{1}{n}\sum_{i=1}^{n}(\hat{\gamma}(X_{b,i}) - \gamma(X_{b,i}))^2\}, \tag{43}$$

$$\widehat{mIAE}_\gamma = Md\{\frac{1}{n}\sum_{i=1}^{n}|\hat{\gamma}(X_{b,i}) - \gamma(X_{b,i})|\}. \tag{44}$$

The results for both models are displayed in Table 1. For each model, we start our iterative procedure with the average derivative solution.

To compare our procedure to the average derivative technique, we compute $\widehat{MSE}_\theta(1)/\widehat{MSE}_\theta(0)$, where $\widehat{MSE}_\theta(1)$ is the estimated MSE obtained with our procedure and $\widehat{MSE}_\theta(0)$ the MSE obtained with the average derivative technique. Similar ratios are computed for $\widehat{MAE}_\theta$, $\widehat{MIAE}_\gamma$, $\widehat{MISE}_\gamma$ and the corresponding median errors. A ratio below 1 indicates that our procedure provides better results.

Finally, some word regarding the stopping conditions of our iterative procedure: when we observe a decrease of the global likelihood function described in the previous section, or an increase lower than 10^{-8}, we stop the procedure. The maximal number of iterations is fixed to 10.

For the first model, we see that our procedure improves between 5% and 8.5% the criteria related to γ. Especially, the $\widehat{MISE}_\gamma$ criterion is 8.5% lower than the error rate obtained with the average derivative method. Regarding the performance over θ, we see that on the $\widehat{MSE}_\theta$ and $\widehat{MAE}_\theta$ criteria, we are equivalent to the average procedure. This is due to some very large errors in our samples, caused presumably by the selection of bandwidths that are too small or too large. Indeed, if we look at $\widehat{mSE}_\theta$ and $\widehat{mAE}_\theta$, our procedure decreases the error rates up to 11% for the $\widehat{mSE}_\theta$ criterion. These error may not have too much influence on the estimation of γ, though. Indeed, the nonparametric estimation of γ can compensate for these errors.

For the second model (model 2), we observe similar results: the iterative procedure performs well regarding the γ criteria, where the error rates are decreased up to 12% for the $\widehat{mSE}_\gamma$ criterion. Overall, the error rates decrease between 8% and 12% with our procedure. Once again, we see that the mean error rates are a bit higher than the median error rates, indicating that some large errors happen in a few samples. Regarding the performance over θ, we don't beat the average derivative technique in term of $\widehat{MSE}_\theta$ but it seems due to some large errors, as for model 1. Indeed, taking the median error ($\widehat{mSE}_\theta$), the error rate decreases by 10%. These results emphasize the need to select carefully the bandwidth parameter and indicate that our procedure improves the estimation of both $\gamma(X)$ and θ. Here, this choice is complicated because we start with a preliminary estimator that can, sometimes, be quite far from the true solution. Therefore, an efficient automatic selection (especially in the context of a simulation) is quite hard to put in place.

	Model 1			Model 2		
Stat.	Av. derivative	Iter. procedure	Ratio	Av. derivative.	Iter. procedure	Ratio
MSE_θ	11.333	11.491	1.014	10.799	11.353	1.051
mSE_θ	1.773	1.581	0.892	3.328	3.053	0.908
MAE_θ	2.757	2.752	0.999	3.205	3.178	0.991
mAE_θ	1.597	1.539	0.964	2.294	2.189	0.954
$MISE_\gamma$	6.361	5.823	0.915	2.193	2.002	0.913
$mISE_\gamma$	4.319	3.951	0.914	1.947	1.712	0.879
$MIAE_\gamma$	73.421	70.421	0.959	36.751	34.453	0.938
$mIAE_\gamma$	55.142	52.224	0.947	36.074	33.223	0.921

Table 1: Values of the various error rates and the ratio statistics, obtained with the average derivative technique (columns *Av. derivative*) and our iterative procedure (columns *Iter. procedure*), for both models. The columns Ratio display the $\widehat{MSE}_\theta(1)/\widehat{MSE}_\theta(0)$ ratio statistics. A ratio below 1 indicates that our model is better. Statistics related to γ are expressed in 10^{-3}.

4 Empirical illustration: Unicredit database

In this section, we use the described methodology to model the severity distribution of an operational loss database, conditional on some explanatory variables.

First, we compute several descriptive statistics using the usual business line - event type categorization. We select a threshold and estimate a global scale parameter. Then, we estimate the conditional shape parameter using different subsets of covariates.

This database consists in 40,871 operational losses registered between 2005 and 2014. The data have been provided by UniCredit's Operational Risk department and are scaled by an unknown factor to preserve the confidentiality. The scaled collection threshold is 2000€ and corresponds to the threshold above which the data quality is certified by the risk department. The losses are dispatched into 7 risk categories (and 12 subcategories), identified by a code (from 10 to 73): *Internal fraud*, *External fraud* (related to payments and others), *Employment practices and workplace safety*, *Clients - products - business practices* (related to derivatives, financial instruments and others), *Damage to physical assets*, *Business disruption and system failure*, *Execution - delivery and process management* (related to financial instruments, payments and others). The codes associated to each category can be found in Appendix. Figure 1 shows the distribution of the losses conditional on these categories. The categories with the largest numbers of losses are *Clients, products and business practices*, and *Execution, delivery and process management.* Table 7 provides descriptive statistics conditional on the risk category (as well as descriptive statistics for all the losses).

Additionally to these basic covariates, we decide to focus on 3 firm-specific covariates, computed at a quarterly frequency: an efficiency ratio, the percentage of the total revenue coming from fees and the leverage ratio. The efficiency ratio is defined as the *ratio of non-interest expenses for the fiscal interim to the total revenue less interest expenses for the same period, and is expressed as a percentage.* It measures the cost to the bank of each unit of revenue (lower values of the ratio are considered better). It may be viewed as a measure of the bank's efficiency to conduct its business. In our opinion, it is very likely that the more the bank is efficient (i.e. the lower the efficiency ratio is), the better it is organized. This variable should act as a (relative) proxy of the quality of internal controls. Besides, the percentage of the total revenue coming from fees (denoted by % fee revenue) can be interpreted as a measure of the economic well-being of the bank: the higher this ratio is, the less dependent the bank is from market interest rates (a quantity hardly under its control). We expect that if the bank is well-managed (i.e. if the % fee revenue is high), the severity distribution of its operational losses should have a thinner tail compared to a situation where it is badly managed. The same reasoning holds for the leverage ratio (defined here as the asset-to-equity ratio): a low leverage ratio indicates a balanced investment policy, that should limit the potential severity of operational losses. At the contrary, a highly leveraged bank exposes itself to potentially huge losses because it relies too much on external debts. In our analysis, we use these variables to obtain estimations of the conditional shape parameter $\gamma(X)$. Figure 2 shows the distributions of these covariates. We decide to focus on these ratios because the bank is supposed to have a certain control over them (e.g. the percentage of the total revenue coming from fees could be increased throughout a more aggressive marketing strategy, the leverage ratio could be increased by issuing new shares, etc.). Thus, knowing the type of relationship between these quantities and $\gamma(X)$ would help risk managers to take informed actions to mitigate these risks.

To determine the threshold, as indicated in Section 2.3, we display the mean excess plot and a plot of the Hill and modified Hill estimators of Huisman et al. [2001] (Figure 3 and 4). The mean excess plot indicates that 2000€ could be a good threshold (the mean of the excesses, for a sequence of increasing thresholds, is linear starting from 2000€). The plot of the Hill and modified Hill estimators suggests a higher threshold (around 420,000€ for the Hill estimator, around 150,000€ for the modified Hill estimator). Such thresholds would give us samples of excesses of respectively 529 and 1196 losses. These thresholds correspond to the 0.988 and 0.971 empirical quantile. We decide to use a threshold of 150,000€ , as a compromise between 2000€ and 420,000€. It gives us a sample of 1196 excesses. The sensitivity of the results to the threshold is discussed later in this section.

We use the conditional quantile regression technique to obtain a set of conditional scale parameters. Then, we simply take the mean of these estimators as our global (unconditional) estimated scale parameter $\hat{\sigma}$. The bandwidth parameters are selected by trials and errors. Once again, this selection seems to not have much impact on the final estimation. For the nonparametric estimation of γ, we use the estimator of Beirlant and Goegebeur [2004], as in the first database. Indeed, we may also encounter $\gamma(X) > 1$ [see de Fontnouvelle et al., 2004]. Because the computing time may be substantial to estimate $\gamma(X)$ (due to the absence of a closed-form solution), the bandwidth is chosen by cross-validation at the first iteration only and kept constant for the rest of the procedure.

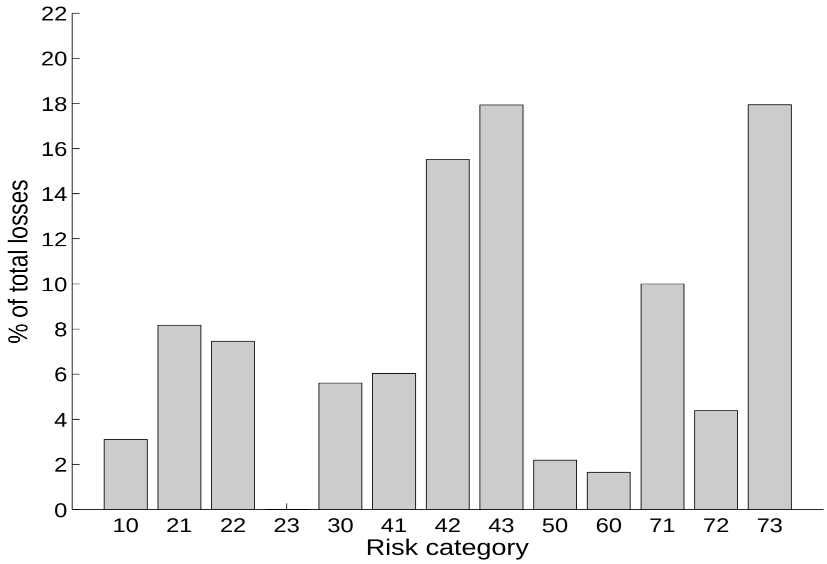


Figure 1: Break-down of the losses between the 12 risk categories (according to internal classifications).

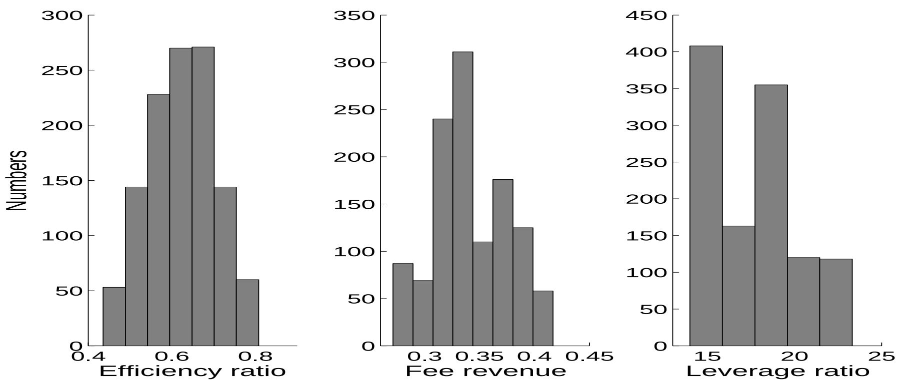


Figure 2: Histogram of the 3 continuous covariates used to estimate the conditional shape parameter $\gamma(X)$. From left to right: efficiency ratio, proportion of fee revenue and leverage ratio.

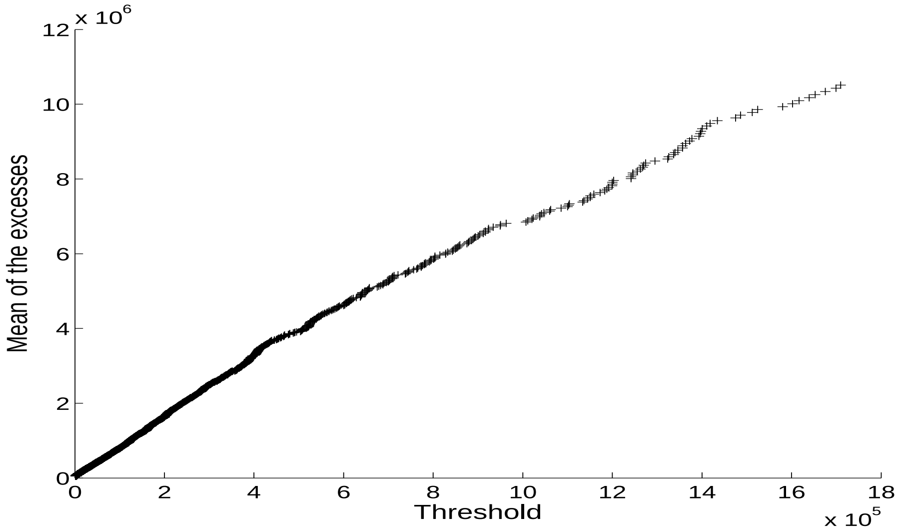


Figure 3: Mean excess plot for a sequence of increasing thresholds. The x axis is expressed in € Million.

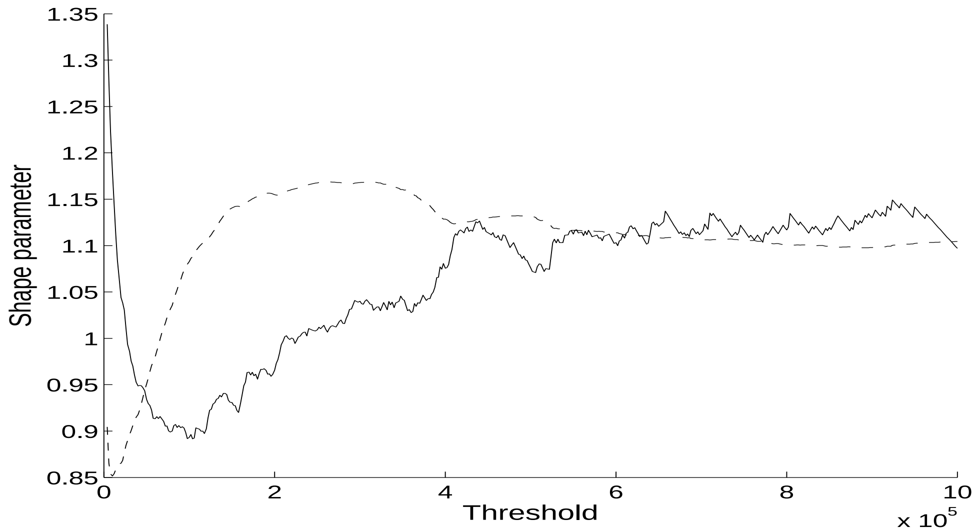


Figure 4: Value of the Hill estimator (solid) and modified Hill estimator (dashed) of Huisman et al. [2001] for a sequence of increasing thresholds. The x axis is expressed in € 100,000.

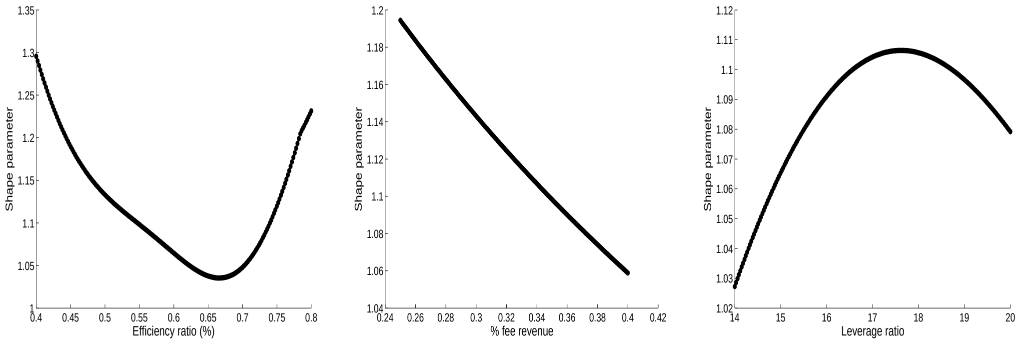
Now, we apply our iterative procedure to obtain conditional estimations of $\gamma(X)$. Table 5 shows the final estimations of θ_0 for all subsets of covariates tested. Tables 2 and 3 show a recap of all the tested combinations of covariates. Figures 5 to 7 show the estimated conditional shape parameter, with respect to the various combinations of firm-specific (continuous) covariates. We see that adding covariates increases the value of the log-likelihood function. However, we don't have a procedure to test if these differences are significant. Considering the covariates one by one, the estimated $\gamma(X)$ parameters vary roughly between 1 and 1.3, suggesting an infinite mean model. The efficiency ratio appears to have a negative impact on the shape parameter up to a value of 0.7. The % fee revenue seems also to have a negative impact on $\gamma(X)$, whereas the leverage ratio has a positive impact up to a value of 17.5, and then a negative impact. When we use the variables two by two and all together, the single-index parameters are all positive. Figures 6(b) and (c) show that the value of the single index seems mainly driven by the leverage ratio: the correlation between the single index value and the leverage ratio is close to -0.99. The relationship with the % fee revenue and the leverage ratio appears quite clear: a high % fee revenue is synonym of a bank in good financial health, whereas a low leverage ratio indicates more independence from creditors. Both situations are associated with lower levels of $\gamma(X)$, thus with lower risks. Regarding the efficiency ratio, we expected it to be positively related to $\gamma(X)$, as low values of this ratio indicate a better economic situation. It is the case only for values of the ratio between 0.65 and 0.8 when we use the efficiency ratio as the only covariate. When we combine this variables with other covariates, we observe a negative effect followed by a positive effect (i.e. an increase of the efficiency ratio decreases $\gamma(X)$ then increases it - see Figures 6(a), (c) and 7). This result may be due to the fact that we do not control for the risk category: some categories of losses may have $\gamma(X)$ positively related to the efficiency ratio, others not. In the next paragraph, we look if this effect can vary according to the categorical variables.

Cov.	1	2	3	4	5	6	7
Efficiency ratio	$\surd$	-	-	$\surd$	-	$\surd$	$\surd$
% fee revenue	-	$\surd$	-	$\surd$	$\surd$	-	$\surd$
Leverage ratio	-	-	$\surd$	-	$\surd$	$\surd$	$\surd$
$\hat{\sigma}$	123890	117001	121911	124043	125172	121432	124821
NLL	16469.75	16470.34	16469.82	16470.7	16467.49	16467.09	16466.96

Table 2: Summary of the tested combinations of covariates (continuous covariates only). $\surd$ indicates that the covariate is taken into account to perform the conditional estimation of γ. The line NLL gives us the final value of the negative log-likelihood function. The line $\hat{\sigma}$ gives us the value of the estimated scale parameter used to perform the estimation of $\gamma(X)$.

Cov.	8	9	10	11	12
Efficiency ratio	$\surd$	$\surd$	-	-	-
% fee revenue	-	-	$\surd$	-	$\surd$
Leverage ratio	-	-	-	$\surd$	$\surd$
EPWS	$\surd$	-	-	-	-
CPBP	-	$\surd$	-	-	-
Fraud	-	-	$\surd$	$\surd$	$\surd$
Services	-	-	$\surd$	$\surd$	$\surd$
$\hat{\sigma}$	141480	163910	116050	117680	118590
NLL	16470.62	16475.64	16468.28	16468.39	16466.2

Table 3: Summary of the tested combinations of covariates (continuous and binary covariates). $\surd$ indicates that the covariate is taken into account to perform the conditional estimation of $\gamma(X)$. The line NLL gives us the final value of the negative log-likelihood function. The line $\hat{\sigma}$ gives us the value of the estimated scale parameter used to perform the estimation of $\gamma(X)$.



(a) (b) (c)

Figure 5: From left to right, estimation of the shape parameter as a function respectively of the efficiency ratio, the % fee revenue and the leverage ratio.

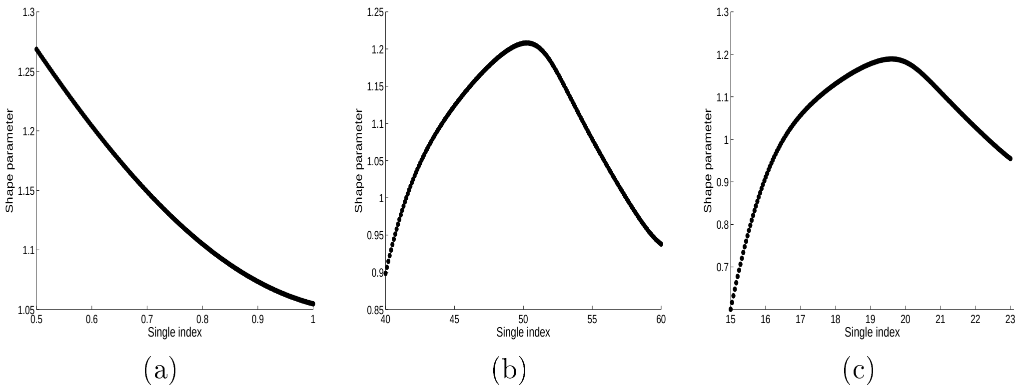


Figure 6: From left to right, estimation of the shape parameter as a function respectively of the efficiency ratio and the % fee revenue, of the % fee revenue and the leverage ratio, and finally of the efficiency and the leverage ratio.

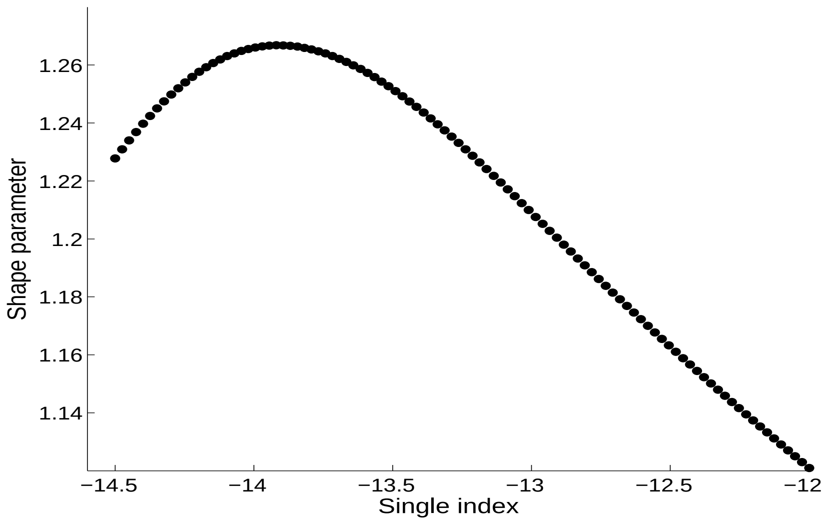


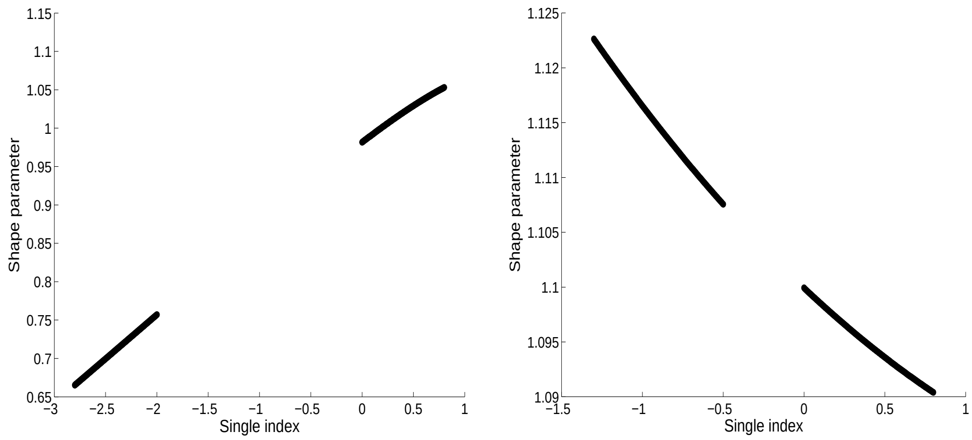
Figure 7: Estimation of the shape parameter as a function of the efficiency ratio, the % fee revenue and the leverage ratio.

What happens if we use the categorical variables? First, we aggregate the category 50 and 60 (*Damage to physical assets* and *Business disruption and system failure*) into a single category (each category has respectively 8 and 9 losses only). Then, we perform a preliminary analysis, where the 6 risk categories are mapped into 6 binary variables, taking value 1 when the loss belongs to the risk class, 0 otherwise (subclasses 21 and 22, then 41 to 43 and 71 to 73 form 3 classes: 20, 40 and 70). These classes are mutually exclusive: a loss belongs only to a single class. Each variable is used independently to estimate $\gamma(X)$ conditional on the risk class. Results are displayed in Table 4. We see that the classes *Internal frauds* and *External frauds* are the riskiest, whereas the aggregate category composed of the events *Damage to physical assets* and *Business disruption and system failures* has the lowest $\gamma(X)$. In the previous paragraph, we suggest that the strange shape of the link function for the efficiency ratio may be due to some heterogeneity related to the risk classification. Clearly, it appears that $\gamma(X)$ varies across classes. This heterogeneity may have an impact on the estimation if we do not take it into account. These results are not surprising: the recent history shows us that both external and internal frauds can cause enormous losses. For example, institutional investors lost hundreds of millions in Madoff's Ponzi scheme; the Société Générale lost several billions in rogue trading and the bank Barings went bankrupt because of trading malpractices.

Risk class	10	20	30	40	50+60	70
$\hat{\gamma}(X)$	1.2248	1.2327	0.7406	1.0946	0.4191	1.087
$\hat{\sigma}$	117030	121900	124890	123630	119910	120960
NLL	164720.5	164732.2	164724.7	164733.5	164785.7	164728.8

Table 4: Summary of the estimated $\gamma(X)$ parameters, conditional on the risk categories. See Appendix 3.5 for details about these classes. $\hat{\sigma}$ is obtained using a local polynomial quantile regression. NLL stands for the value of the negative log-likelihood function whereas $\hat{\sigma}$ gives us the value of the estimated global scale parameter.

What happens if we use categorical and continuous variables together? Figure 8 shows the estimation of $\gamma(X)$ conditional on the efficiency ratio and the risk class 40 (*Customers, products and business process* - CPBP) on one side, and the efficiency ratio combined with the risk class 30 (*Employment practices and workplace safety* - EPWS) on the other side. We see that both figures exhibit different link functions (the link function is increasing with EPWS, and decreasing with CPBP), suggesting that the effect of a covariate can vary across event types (the single index parameters have the same signs). Based on the results of the previous paragraph and to keep the number of binary variables limited, we aggregate some risk classes together. We combine the classes 10 (*Internal fraud*) and 20 (*External fraud*) into a *Fraud* class. Then, we aggregate the classes 40 (CPBP) and 70 (*Execution, delivery and process management* - EDPM) into a *Services* class. Finally, we pool all the other classes (30, 50 and 60) together, into a *Workplace safety* class. It gives us 3 super-categories that could be mapped into 2 new binary variables: the variable *Fraud* takes value 1 if a loss belongs to the classes 10 or 20, 0 otherwise. The variable *Services* takes value 1 if the loss belongs to the classes CPBP (40) or EDPM (70), 0 otherwise. We combine these variables with the leverage ratio and/or the % fee revenue (we set aside the efficiency ratio). Figures 9 and 10 show the effects of the continuous variables when we control for the event type with these binary variables. In Figure 9, we see that the single index parameter and the link function are globally alike, indicating that both categories have the same kind of effect on $\gamma(X)$. However, in Figure 10 we see clearly that the effects of the continuous covariates vary in the third category (*Workplace safety*). Being in the *Workplace safety* category (i.e. when *Fraud*=0 and *Services*=0) *shifts* us to the bottom right of the link function, where we observe that the effect of the leverage ratio and of the % fee revenue is at the opposite of their effect in the other categories. Also, the general level of shape parameter in this part of the link function is way lower than in the two other categories (as already emphasized in our univariate analysis), suggesting that the *Workplace safety* category has a strong negative impact on $\gamma(X)$. It seems to confirm that the drivers of the operational events may differ across event types. We don't have any economical explanation to these observations, and let this subject open for further discussions. Eventually, the top left of Figure 10 confirms the results display in Figure 9 : the effects of the covariates are similar in both the *Fraud* and *Services* categories, but the variable *Fraud* increases $\gamma(X)$ stronger than the variable *Services*. As argued before, it may be due to the nature of these losses: frauds related to rogue trading and associated trading events can be pretty huge (sometimes, several times the total capital of the bank).



(a) (b)

Figure 8: Estimation of $\gamma(X)$ conditional on (a) the efficiency ratio and the EPWS class, (b) the efficiency ratio and the CPBP class.

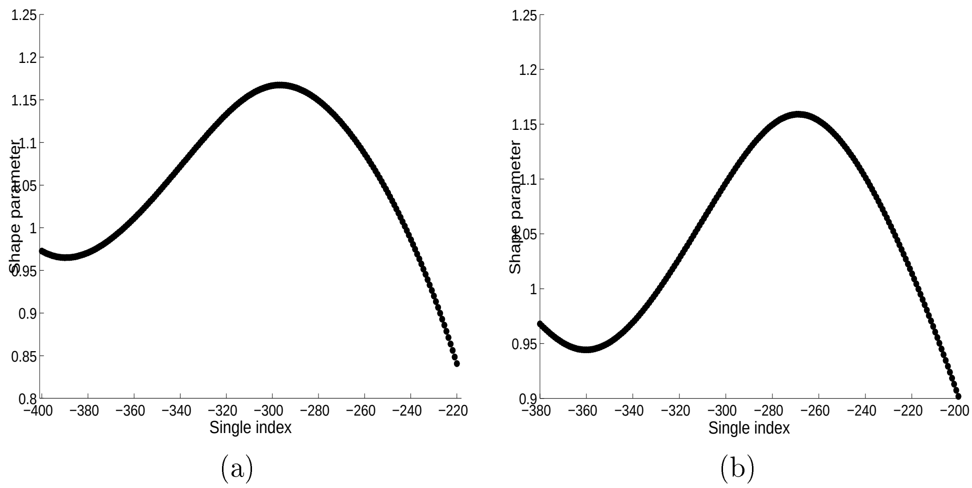


Figure 9: Estimation of $\gamma(X)$ conditional on (a) the % fee revenue, the leverage ratio and the *Fraud* class, (b) the % fee revenue, the leverage ratio and the *Services* class.

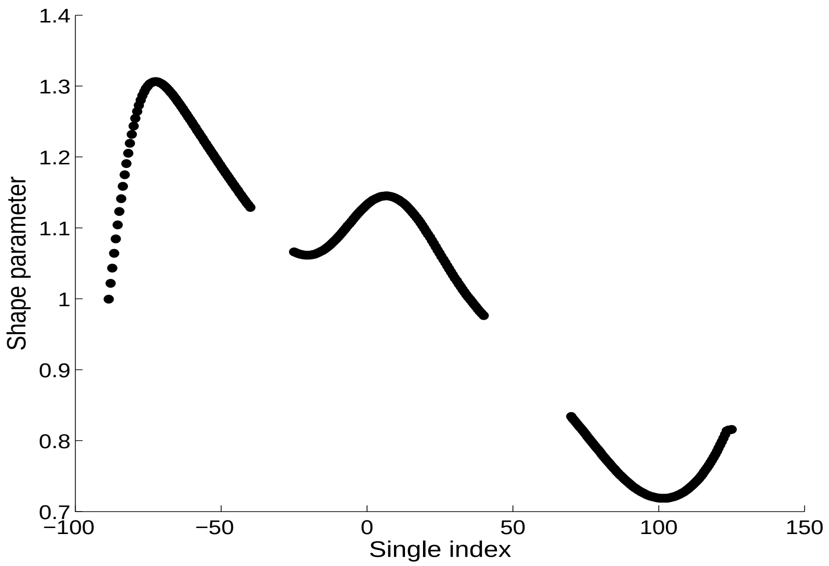


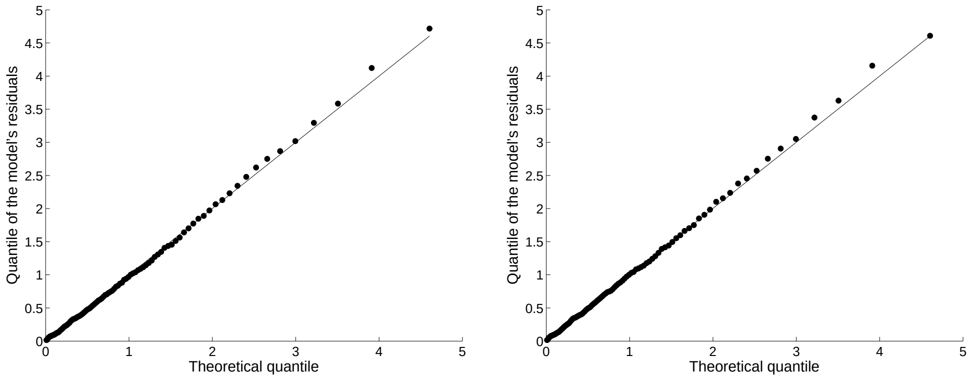
Figure 10: Estimation of $\gamma(X)$ conditional on the % fee revenue, the leverage ratio, the *Fraud* class and the *Services* class.

Cov.	4	5	6	7	8	9	10	11	12
Efficiency ratio	1	1	-	1	1	1	-	-	-
% fee revenu	0.824	-	1	14.3502	-	-	1	1	1
Leverage ratio	-	1.079	2.789	-1.0595	-	-	-17.015	-15.846	6.4611
EPWS	-	-	-	-	-	-3.016	-	-	-
CPBP	-	-	-	-	-1.737	-	-	-	-
Fraud	-	-	-	-	-	-	11.5059	-	-202.163
Services	-	-	-	-	-	-	-	8.553	-110.239

Table 5: Summary of the estimated θ_0 parameters for all tested combinations of covariates.

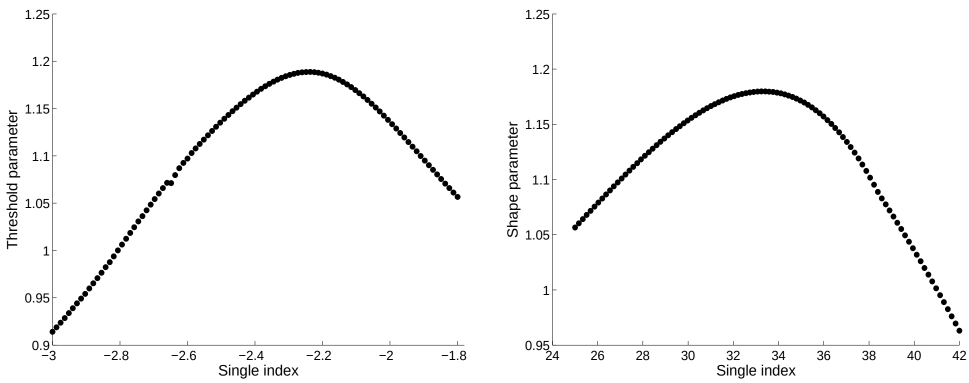
These results emphasize the potential of our methodology. With this technique, we show that the severity distribution of the operational losses may be effectively influenced by some covariates (in particular by the event type and the leverage ratio). We are also able to compute estimations of $\gamma(X)$ conditional on the levels of these covariates *altogether*

Finally, some words regarding the fact that we find several $\gamma(X) > 1$. As said before, these quantities may not make sense, economically speaking, because it implies that the conditional expectation of the loss severity distribution beyond the threshold is infinite. Nevertheless, it is not unusual, even for data from a single bank, to observe such values [see de Fontnouvelle et al., 2004, e.g.]. Neslehova et al. [2006] provide a very interesting discussion on what is implied by this kind of models. Especially, they notice that the results may be dependent from the selected threshold. We do not perform a rigorous threshold sensitivity analysis, because the main purpose of this section was to illustrate the methodology, not to conduct a comprehensive analysis of operational risks' determinants. However, our experience with these data shows that whatever threshold we choose, we never get values of $\gamma(X)$ far from 1. Figure 4 shows nonparametric estimations of an unconditional $\gamma(X)$ with increasing threshold, whereas Figure 12 shows the estimated value of $\gamma(X)$ conditional on the 3 continuous covariates with the 96th and 99.5th quantiles as thresholds (the threshold goes from 150,000€ to 115,000€ and 780,650€). Another explanation to these high values of $\gamma(X)$ could be *data contamination*. As shown in Neslehova et al. [2006], few observations with a very heavy-tailed severity distributions may cause an overestimation of the global shape parameter, and in our case, of the conditional shape parameter. We believe that this may be the case in the public database, where the heterogeneity induced by some very large losses can cause these high $\gamma(X)$ values. In the UniCredit database, we control for a possible heterogeneity using several covariates, but other covariates may exist. Finally, as suggested in Chavez-Demoulin et al. [2014a] and Neslehova et al. [2006], $\gamma(X) > 1$ may also be viewed as a sign that the uncertainty around the quantiles far in the tail is very high. This raises the concern that trying to estimate a quantile (or a conditional expectation) so far in the tail (i.e. the 99.9th quantile, or the conditional expectation beyond this point) may have no economic meaning at all. We let this question pending for future researches. Thus, instead of trying to conform our results to some preconception of the reality, we prefer to let the data speak by themselves. Using the same graphical goodness-of-fit test as Chavez-Demoulin et al. [2014a], we see that despite these concerns, our conditional GPD models fit well the data (see Figure 11 for models 7 and 11 for UniCredit data). Remember Madoff's Ponzi scheme, the rogue trading at Société Générale and Barings, the subprime crisis: before that these events happen, no one would believe that losses or fines of several \$ billions were even possible. Hence, if this is the way things are, it is better to leave it that way.



(a) (b)

Figure 11: Graphical goodness-of-fit tests for the GPD models, using (a) the three continuous covariates or (b) two categorical covariates and two continuous covariates (respectively the sets of covariates 7 and 11). If the models are correct, the models' residuals (given by $e_i = -\log(1 - GPD(Y_i; \sigma(X_i), \gamma(X_i)))$, under the hypothesis that the excesses $Y_i, i = 1, \cdots, n$ are approximately independent, should approximately follow a standard exponential distribution [see Chavez-Demoulin et al., 2014a, , p. 9, equation 12].



(a) (b)

Figure 12: Estimation of the shape parameter, conditional to the 3 continuous covariates, with (a) a threshold equal to the 96th quantile and (b) a threshold equal to the 99.5th quantile. The single index parameters are respectively $\{1, 4.12, -0.23\}$ and $\{1, -77.68, 3.42\}$ (the inverted signs of the parameters explains why the figures appear inverted).

5 Conclusion

In this article, we contribute in several ways to the existing literature on operational losses modeling. First, we provide a full methodology to properly take into account the dependence structure between the tail of the severity distribution and economic factors. Whereas previous studies [mostly Cope et al., 2012, Chernobai et al., 2011] focus on the expected value of the operational losses severity, we show how to model the dependence that may arise in the tail of the distribution. Therefore, this result is of particular interest for regulators and risk managers, because it would allow them to compute proper tail-related risk measures (Value-at-Risk, Expected Shortfall...) conditionally to certain states of the economy. Moreover, this methodology offers novel possibilities to develop some stress-testing procedures related to operational risks. To the best of our knowledge, only Chavez-Demoulin et al. [2014a] provide a way to model this dependency in the multivariate case.

Second, thanks to the single index assumption and the nonparametric specification, our methodology is flexible and may potentially handle high-dimensional covariates. Indeed, with the single index assumption we are able to reduce easily the dimension of the problem and to escape the curse of dimensionality. This assumption allows to extend Beirlant and Goegebeur [2004]approach to the multivariate case more efficiently. Besides, the nonparametric estimation of the link function prevents us from relying on unrealistic dependence structure (especially linear ones) and allows us to model very complicated link functions.

Third, we provide a way to handle the (not so rare) case of infinite mean models, with the use of nonparametric quantile regression techniques. Despite some reluctance from practitioners, it seems that a lot of operational losses data exhibit this property. Therefore, ensuring that our procedure is valid in this context is crucial.

Finally, we apply our methodology on a totally new set of data. We show that the leverage ratio seems to be the biggest driver of the tail thickness of the severity distribution. Especially, an increase of the leverage ratio (defined as the asset-to-equity ratio) increases the probability of suffering from large operation losses. We also show that operational losses related to financial business lines have a conditional severity distribution with thicker tails. These first empirical results suggest that taking into account these type of accounting and financial indicators may help financial institutions to set aside capital reserves that match better their true level of operational risk. Especially, it opens the door to the development of forecasting models related to this area of the risk management.

Acknowledgement

Julien Hambuckers acknowledges the support of the Belgian National Fund for Scientific Research (FNRS), that partially funded this work with a Research Fellow grant.

6 Appendix: descriptive statistics

Code	Risk Categories
10	Internal frauds
21	External frauds: Payments
22	External frauds: Others
30	Employment practices and workplace safety
41	Clients, products and business practices: Derivatives
42	Clients, products and business practices: Financial Instruments (excluding derivatives)
43	Clients, products and business practices: Others
50	Damages to physical assets
60	Business disruption and system failures
71	Execution, delivery and process management: Financial Instruments
72	Execution, delivery and process management: Payments
73	Execution, delivery and process management: Others

Table 6: Risk categories and associated identification codes for the UniCredit database.

Code	Mean	Std	Median	3^{rd} quartile	Skewness	Max	# losses
All	50004	992190	6208.5	170670	65.63	$103.4e^6$	40871
10	139470	$1.06e^6$	12448	48482	20.9098	285600	1271
21	10370	136400	3139	5271	52.4976	76200	3340
22	30320	632600	6866	18850	54.564	348200	3051
30	35260	294600	6533	17949	36.0121	126500	2292
41	127080	$1.55e^6$	26289	68149	32.1107	550600	2466
42	33210	772900	6459	14005	53.0658	432600	6342
43	85250	$1.84e^6$	10438	26214	44.5918	$103.4e^6$	7330
50	9990	60800	3241	5357	22.3868	16400	896
60	14430	40500	4288	10058	7.7892	5200	674
71	35230	559000	5139	11221	43.8784	304900	4085
72	17750	89500	4646	9489	17.0932	22900	1791
73	42990	439900	4951	12670	27.8981	187300	7333

Table 7: Descriptive statistics of the losses, conditional on their risk classification.

References

F. Aue and M. Kalkbrener. LDA at work: Deutsche Bank's approach to quantifying operational risk. *Journal of Operational Risk*, 1(4):49–93, 2007.

A.A. Balkema and L. de Haan. Residual life time at great age. *The Annals of Probability*, 2(5):792–804, 1974.

J. Beirlant and Y. Goegebeur. Local polynomial maximum likelihood estimation for pareto-type distributions. *Journal of Multivariate Analysis*, 89(1):97–118, 2004.

A. Chapelle, Y. Crama, G. Hübner, and J-P. Peters. Practical methods for measuring and managing operational risk in the financial sector: a clinical study. *Journal of Banking & Finance*, 32:1049–1061, 2008.

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

V. Chavez-Demoulin and P. Embrechts. Revisiting the edge, ten years on. *Communications in Statistics: Theory & Methods*, 39(8/9):1674–1688, 2010.

V. Chavez-Demoulin, P. Embrechts, and M. Hofert. An extreme value approach for modeling Operational Risk losses depending on covariates. *Journal of Risk and Insurance*, 2014a.

V. Chavez-Demoulin, P. Embrechts, and S. Sardy. Extreme-quantile tracking for financial time series. *Journal of Econometrics*, 181(1):44–52, 2014b.

A. Chernobai, P. Jorion, and F. Yu. The derminants of Operational Risk in U.S. financial institutions. *Journal of Financial and Quantitative Analysis*, 46(8):1683–1725, 2011.

E.W. Cope, M.T. Piche, and J.S. Walter. Macroenvironmental determinants of operational loss severity. *Journal of Banking & Finance*, 36(5):1362–1380, 2012.

P. de Fontnouvelle, J. Jordan, and E. Rosengren. Implications of alternative operational risk modeling techniques. *Federal Reserve Bank of Boston*, Working paper, 2004.

M. Degen, P. Embrechts, and D. Lambrigger. The quantitative modelling of operational risk: between g-and-h and EVT. *Astin Bulletin*, 38(2), 2007.

K. Dutta and J. Perry. A tale of tails: an empirical analysis of loss distribution models for estimating operational risk capital. *Research Review*, Jul-Dec(6):11–14, 2006.

P. Embrechts, C. Klupperlberg, and T. Mikosch. *Modelling extremal events for insurance and finance*, volume 648. Springer - Verlag, Berlin, 1997.

S. Ghosh and S. Resnick. A discussion on mean excess plots. *Stochastic Processes and their Applications*, 120(8):1492–1517, 2010.

J. Hambuckers, C. Heuchenne, and O. Lopez. A semiparametric model for Generalized Pareto regression based on a dimension reduction assumption. *Work in progress*, 2015.

W. Härdle, P. Hall, and H. Ichimura. Optimal smoothing in single-index models. *The Annals of Statistics*, 21(1):157–178, 1993.

M. Hristache, A. Juditsky, and V. Spokoiny. Direct estimation of the index coefficient in a single-index model. *The Annals of Statistics*, 29(3):595–623, 2001.

R. Huisman, K. Koedijk, C. Kool, and F. Palm. Tail-index estimates in small samples. *Journal of Business & Economic Statistics*, 19(2):208–216, 2001.

H. Ichimura. Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *J. Econometrics*, 58(1-2):71–120, 1993.

M. Moscadelli. The modelling of operational risk: experience with the analysis collected by the Basel Committee. Report, Bank of Italy, 2004.

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

G. Petrella and A. Resti. Supervisors as information producers: Do stress tests reduce bank opaqueness? *Journal of Banking & Finance*, 37(12):5406–5420, 2013.

J. Pickands. Statistical inference using extreme order statistics. *The Annals of Statistics*, 3 (1):119–131, 1975.

S. Plunus, G. Hübner, and J-P. Peters. Measuring operational risk in financial institutions. *Applied Financial Economics*, 22(18):1553–1569, 2012.

D. Ruppert and M. P. Wand. Multivariate locally weighted least squares regression. *The Annals of Statistics*, 22(3):1346–1370, 1994.

C. Scarrot and A. McDonald. A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT*, 10(1):33–60, 2012.

P.V. Shevchenko. Implementing loss distribution approach for operational risk. *Applied Stochastic Models in Business and Industry*, 26:277–307, 2010.

A. Soprano, B. Crielaard, F. Piacenza, and D. Ruspantini. *Measuring Operational and Reputational Risk: A Practitioner's Approach*. Wiley, 2009.

T. Wang and C. Hsu. Board composition and operational risk events of financial institutions. *Journal of Banking & Finance*, 37(6):2042–2051, 2013.

K. Yu and M. C. Jones. Local linear quantile regression. *Journal of the American Statistical Association*, 93(441):228–237, 1998.