What are the determinants of the operational losses severity distribution? A multivariate analysis based on a semi-parametric approach.

J. Hambuckers, C. Heuchenne, O. Lopez

Abstract

In this paper, we analyse a database of around 41,000 operational losses from a single large European bank. We investigate three kinds of covariates: firm-specific, financial and macroeconomic covariates and we study their relationship with the shape parameter of the severity distribution. To do so, we introduce a semi-parametric approach to estimate the shape parameter of the severity distribution, conditionally to large sets of covariates. Relying on a single index assumption to perform a dimension reduction, this approach avoids the curse of dimensionality of pure multivariate non-parametric techniques as well as too restrictive parametric assumptions. We show that taking into account variables measuring the economic well being of the bank could cause the required Operational Value-at-Risk to vary drastically. Especially, high pre-tax ROE, efficiency ratio and stock price are associated with a low shape parameter of the severity distribution, whereas a high market volatility, leverage ratio and unemployment rate are associated with higher tail risks. Finally, we discuss the fact that the considered approach could be an interesting tool to improve the estimation of the parameters in a Loss Distribution Approach and to offer an interesting methodology to study capital requirements variations throughout scenario analyses.

1 Introduction

In a series of consultative documents entitled Operational Risk and Operational Risk - Supervisory guidelines for the Advanced Measurement Approaches, the Basel Committee for Banking Supervision (BCBS) discussed the interest to hold capital to cover the operational risks of a financial institution. It 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[BCBS, 2004]. Indeed, for more than 20 years now, events related to this class of risks made the headlines regularly 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. The stock price dropped from $90.75 a share to less than $1 in a year. 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 estimated to $14.2 billion. More recently, in 2008, rogue trading at Société Générale caused €4.9 billion losses, 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 interest 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. It exists therefore a real need for companies to map, model and measure those risks to take proper hedging actions. Following these preoccupations, the econometric literature focused mainly on the issue of modeling and measuring operational risks throughout quantitative techniques.

Among the technical approaches outlined by the BCBS, the advanced measurement approach (AMA) quickly gathers the attention of the industry and of the academics. Indeed, whereas the other approaches rely on simplistic assumptions, this approach 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, one would suppose 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:

\[
L(b, e) = \sum_{i=1}^{N(b,e)} Y_i,
\]

where \( Y_i \) 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 \( Y_i \) are random variables (as well as \( L(b, e) \)), that need to be modelled by the bank. Using this model, a bank can establish a capital charge (devoted to a specific combination \((b,e)\)) to hold to cover potential operational losses. It is taken as the Value-at-Risk (VaR) of \( L(b, e) \), at the 99.9% level (i.e. the 99.9th order statistic of the distribution of \( L(b, e) \)) and noted \( \text{OpVar}(b,e)_{99.9} \). Mathematically, this quantity is defined throughout

\[
P(L(b, e) \leq \text{OpVar}(b,e)_{99.9}) = 0.999
\]

and is simply the definition of the 99.9th quantile of \( L(b, e) \). 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, Chavez-Demoulin et al., 2014a, 2014b] 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 $y_i$ of a random variable $Y$ over a suitable large threshold $\tau$ approximately follow a GPD. The cumulative distribution function (cdf), with tail parameter $\gamma$, location (threshold) parameter $\tau$ and scale parameter $\sigma$, is given by:

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

The probability distribution function (pdf) is given by:

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

The distribution of the excess over the threshold (i.e. $Z = Y - \tau$) is consequently $GPD(z_i; \gamma, 0, \sigma)$ distributed and then simply noted $GPD(z_i; \gamma, \sigma)$. The associated Value-at-Risk at a confidence level $p$ is given by:

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

This theory has been widely applied in all area of Finance for VaR estimation and quantitative risk management. See Chavez-Demoulin and Embrechts [2010] for a discussion on the subject. Some authors also proposed other models for the severity distribution: [Degen et al., 2007, Shevchenko, 2010] used the g-and-h distribution, whereas [Plunus et al., 2012] adapted the CreditRisk+ approach to the operational risk context. However, all these studies, with the notable exception of Chavez-Demoulin et al. [2014a], assume that the parameters of the frequency and severity distributions are constant conditionally to 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, among others. Chernobai et al. [2011] investigate the link between 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 frequency and firm-specific variables, but only weaker results with respect to the macroeconomic variables. They use conditional Poisson regression models whose process intensities are 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 operational loss severity. They discover that some event-types are sensitive to per-capita GDP, a governance index, shareholder protection law and supervisory power. However, none of these studies provide a way to estimate the GPD parameters conditionally to these covariates. A solution would be 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 three studies, in the context of financial data modeling, provide methods to estimate the GPD parameters conditionally to covariates, taking into account all available data. Beirlant and Goegebeur [2004] propose a local polynomial estimator in the case of a one-dimensional covariate. However, the convergence rate of this estimator is expected to decrease rapidly when the dimension of the covariate increases. A solution would be to increase the size of the dataset but operational losses database are often scarce and not easy to enlarge. They illustrate their methodology on a Norwegian fire claim dataset, conditionally to 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 semi-parametric 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 business lines and years as covariates. Finally, in a more recent work, Heuchenne et al. [2014] introduce a semi-parametric methodology to estimate the shape parameter of a GPD, conditionally to some covariates. Relying on a single-index assumption, they perform a dimension reduction that enables to use the same local-polynomial estimator as Beirlant and Goegebeur [2004], in the univariate case. They illustrate their methodology on the same public database as Chavez-Demoulin et al. [2014a] with up to 3 macroeconomic covariates.

However, until now, covariates investigated in Chernobai et al. [2011], Cope et al. [2012], Wang and Hsu [2013] have never been used as explanatory variable of the conditional severity distribution. Notwithstanding that Chernobai et al. [2011] and Wang and Hsu [2013] focus on the frequency distribution, Beirlant and Goegebeur [2004] and Chavez-Demoulin
et al. [2014a] do not consider these kind of variables. Heuchenne et al. [2014] consider macroeconomic variable, but in an illustrative perspective only. We believe that it is due first to technical considerations: before Beirlant and Goegebeur [2004] for low dimension covariates, Chavez-Demoulin et al. [2014a] and Heuchenne et al. [2014], no formal method existed to compute GPD parameters conditionally to covariates. Second, as noticed by Chavez-Demoulin et al. [2014a], it is well known to be extremely difficult, for academics, to get their hands on real operational loss data, furthermore on data associated with covariates. All these issues together make the parameters estimation of GPD with covariates a challenging - but promising- task. Indeed, having that kind of procedure at hand, as well as knowing what are the variables that influence the severity distribution, is of interest for practitioners and researchers because it would allow to compute a capital charge for specific states of the economy or for specific state of the firm. Intuitively, letting the parameters depending on covariates is very appealing. It seems natural to think that the frequency and the severity 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, acquiring some buildings in a seismic region, etc., offer increased opportunities for more frequent and bigger losses to happen. Using variables that measure these changes could bring useful information in the estimation process.

In this paper, we study a unique database of around 41,000 losses from a single European bank. With the losses, we have access to 26 different covariates: 11 firm-specific covariates, 7 financial covariates and 8 macroeconomic covariates. Using the method introduced in Heuchenne et al. [2014], we perform a comprehensive analysis of the impact of these covariates on the severity distribution of these losses. More precisely, we assume a GPD for the tail of the severity distribution and we look at the value of the shape parameter conditionally to subsets of these covariates. We also solve practical issues not discussed in Heuchenne et al. [2014].

We use this new semi-parametric method both because it offers a convenient middle-path between parametric and nonparametric techniques and because it has not been applied on a large database before. This statistical technique combines conveniently para- and nonparametric techniques to form a reasonable compromise between fully parametric and fully nonparametric modeling. 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] or the other models in Chavez-Demoulin et al. [2014a]. However, the convergence rate of these estimators is expected to decrease rapidly when the dimension of the covariate increases. To solve these issues, our method relies on a single-index model to perform a
dimension reduction of the covariates [see, e.g., Härdle et al., 1993]. Such a model assumes that

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

(6)

where $\gamma(X)$ is the conditional shape parameter of the GPD, $\theta_0$ an unknown vector of parameters, $\gamma_{\theta_0}$ an unknown link function and $X$ the vector of covariates. The idea is that, if we knew the true parameter $\theta_0$, we would be back in a univariate case where $\gamma(X)$ can be obtained with the (nonparametric) local-polynomial approach of Beirlant and Goegebeur [2004], without suffering from the curse of dimensionality. Working with this approach avoids both the curse of dimensionality and inflexible parametric models. Of course, the counterpart is that we need to find sufficiently good estimators of $\theta_0$, to be able to obtain good local-polynomial estimators of $\gamma(X)$. More precisely, we use a two-step iterative algorithm, that improves at each step the estimation of either $\theta_0$ or $\gamma(X)$. We start first with a preliminary estimator $\hat{\theta}^{(0)}$ of $\theta_0$, using the conditional method of moments. Afterwards, we estimate $\gamma(X)$ and its first derivative, using the local-polynomial technique. These estimators are later on used in a semi-parametric likelihood function that is maximized to obtain a new (and better) estimator $\hat{\theta}$ of $\theta_0$. The operation is repeated until convergence. This setting is inspired from the one proposed in Hristache et al. [2001]. A precise description of the estimation procedure and of the model can be found in Section 2. 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 as an estimator of the link function, we use a kernel smoothing approach over a single-index.

In Section 3, we detail our database and apply the estimation procedure with different subsets of covariates. We use firm-specific variables (Tier-I capital ratio, leverage ratio,...) as well as macroeconomic and financial variables to estimate the conditional shape parameters. All variables are computed on a quarterly basis. Overall, we consider 26 covariates, whose values are computed the same quarter than the loss events. We perform both a univariate analysis (i.e. we consider the potential effect of the covariates on the shape parameter one by one) and a multivariate analysis (i.e. we consider subsets of explanatory variables up to 4 different covariates). We show that taking into account the level of these covariates could drastically change the 99.9% quantile of the severity distribution and therefore impact the $OpVaR$ hold by the bank.

Finally, we build a full LDA model and perform a scenario analysis. We show that if we take the impact of the economic environment of the bank when it is in some extreme states, it modifies drastically the total $OpVaR_{99.9}$. We discuss the results and conclude in Section 4.
2 Methodology

In this section we introduce the new semi-parametric statistical approach. Relying on a single-index assumption, we perform a dimension reduction that enables us to use univariate nonparametric techniques. Hence, we suffer neither from too strong parametric assumption nor from the curse of dimensionality.

We recall first some general results regarding the EVT and the Peak-Over-Threshold (POT) methodology in Subsection 2.1. Then, in Subsection 2.2 we detail the model and our estimation procedure of the shape parameter. Related theoretical work can be found in Heuchenne et al. [2014] (see this reference for a detailed version of the underlying hypotheses). We focus mainly on the shape parameter, this one being of prime importance to obtain a good parametric estimation of high order statistics. Indeed, the moment generating function of the GPD is exponentially linked to the tail parameter, whereas it is only linear in the scale parameter. Eventually, in Subsection 2.3 we present briefly a way to obtain an estimators of the scale parameter in our context, using the method of conditional moments. 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 methods 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 \( \{y_i\}, i = 1, ..., N(T) \) the exceedances of a r.v. \( Y \), over a sufficiently high threshold \( \tau \) and along a period of time \( T \) (the index \( i \) denotes that the realization \( y_i \) of \( Y \) exceed the threshold). Under some mathematical conditions, this theorem states that

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

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

Especially, the second result means that the excess conditional distribution function of i.i.d. realizations of \( Y \) (\( P(Y - \tau \leq z | Y > \tau) \)) converges to a GPD distribution with parameters \( \gamma \) and \( \sigma \) when \( \tau \) goes to infinity. This result holds under the condition that \( F(y) \) (the cumulative distribution function of \( Y \)) respects

\[
\bar{F}(y) = 1 - F(y) = y^{-1/\gamma} L(y)
\]

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

\[
\lim_{x \rightarrow \infty} \frac{L(\nu y)}{L(y)} = 1
\]
for all $\nu > 0$. This condition, known as regular variation property, implies that the tail of the loss distribution decays at a power rate of $Y$. Hence, relying on this asymptotic result, one can model the distribution of the excesses over $\tau$ using a GPD distribution and use maximum likelihood techniques to obtain estimators of $\gamma$ and $\sigma$.

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 $Z = Y - \tau$ (i.e. the amount of an operational loss above the threshold $\tau$, conditional to $Y > \tau$) is GPD with a conditional tail parameter $\gamma(X)$ and scale parameter $\sigma(X)$:

$$Z \sim GPD(\gamma(X), \sigma(X)),$$  \hspace{1cm} (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,

$$\gamma(X) = \gamma_{\theta_0}(\theta_0^T X),$$  \hspace{1cm} (8)

where $\theta_0$ is the single-index parameter (parametric part), and $\gamma_{\theta_0}$ the unknown link function (nonparametric part). The first element of $\theta_0$ is assumed equal to one, to ensure the proper identifiability of the model. Using these hypotheses, we perform a dimension reduction, since all the relevant information on the covariates are summarized into a single-index $U = \theta_0^T X$. Hence, for a sample of size $n$, $\{(X_j, Z_j) : j = 1, \cdots, n\}$, once we know $\theta_0$, the problem of finding a good nonparametric estimator $\hat{\gamma}_{\theta_0}$ of $\gamma_{\theta_0}$ is reduced to an univariate nonparametric regression problem of $\gamma(x_j)$ over $u_j = \theta_0^T x_j$, $j = 1, \cdots, n$ ($x_j$ being a realization of the random vector $X_j$). One could use a semi-parametric log-likelihood function, where the conditional shape parameter is expressed as a local polynomial in $U$,

$$(\hat{\gamma}_{\theta_0}(U), \hat{\gamma}'(U)) = \arg \max_{a_j, b_j} \sum_{i=1}^{n} l(Z_i; a_j + b_j(\theta_0^T X_i - U)) K\left(\frac{\theta_0^T X_i - U}{h}\right),$$  \hspace{1cm} (9)

where $K(\cdot)$ is a kernel function with integral one and $h$ is a bandwidth parameter. This estimator has been studied in Beirlant and Goegebeur [2004], among others, 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 a usual L2-score function. If $\theta_0$ is unknown and estimated by $\hat{\theta}$, then we simply plug $\hat{\theta}$ in equation 9

On the other side, if $\gamma_{\theta_0}$ were known and $\theta_0$ unknown, $\hat{\theta}$ could be obtained using a maximum likelihood approach. This estimator is defined by:

$$\hat{\theta} = \max_{\theta} \frac{1}{n} \sum_{i=1}^{n} l(Z_i; \gamma_{\theta_0}(\theta^T X_i), \sigma(X_i)),$$  \hspace{1cm} (10)
where \( l(Z_i; \gamma_0(\theta^T X_i), \sigma(X_i)) \) is the density function of the GPD at point \( Z_i \) with conditional shape parameter \( \gamma_0(\theta^T X_i) \) and scale parameter \( \sigma(X_i) \). However, the link function is unknown but we can act exactly as in equation 9: replace the unknown quantity by some estimators. For example, we can replace \( \gamma_0(\theta^T X_i) \) by our estimator in equation 9.

We see that both estimations are intrinsically inter-related: we can compute \( \hat{\theta} \) only if we have some \( \hat{\gamma} \) at hand, and vice-versa (via the single index \( U \)). Also, if one of the two quantities is badly estimated, it will have a big impact on the other estimation. A solution would be to perform a joint maximization, 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 \((\hat{\gamma}, \hat{\theta})\) until our estimation converges sufficiently. The 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. Using equation (9), estimate the link function by local polynomial regression with \( \theta = \hat{\theta}^{(k)} \), at points \( u_j = (\hat{\theta}^{(k)})^T x_j, j = 1, \ldots, n \).

3. Use the values of \( a_j \) and \( b_j \) (estimators of \( \gamma_0(\theta^T x_j) \) and its first derivative) in equation 10 to build a semi-parametric maximum likelihood function. Set \( k = k + 1 \), then obtain an updated estimator \( \hat{\theta}^{(k)} \) by maximizing the following equation over \( \theta \):

\[
\hat{\theta}^{(k)} = \arg \max_{\theta} \sum_{j=1}^{n} l(Z_j; a_j + b_j(\theta^T X_j - U_j), \sigma(U_j)).
\]

(11)

4. Go back to step 2. Repeat until convergence.

In step 3, we use the fact that the tail parameter is locally linear in \( \theta^T X_j \) in a small window around \( U_j \). Because we have this relationship linking \( \gamma_0 \) and \( \hat{\theta} \), the iterative procedure is able to perform. Instead of a local-polynomial estimation of \( \hat{\gamma}_0 \), another possibility would be to use a Nadaraya-Watson-type estimator of \( \gamma_0 \), but we still need an estimator of the first derivative. Indeed, we cannot assume that \( \gamma(X) \) is a constant on a small local interval, because it would make it independent of \( \theta \), therefore useless in equation (10). Alternatively, one could rely on finite differencing methods. With these methods and our estimator of \( \gamma(X) \), it would be possible to obtain estimators of the first derivatives that converge at a rate faster than \((nh)^{3/2}\). We stop the iterative procedure once we observe a decrease of the likelihood function \( \frac{1}{n} \sum_{i=1}^{n} l(Z_i; a_i; \sigma(X_i)) \), an increase of less than \( 10^{-6} \), or when a predetermined number of steps has been performed (e.g. 10 steps).

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

$$h^{\text{opt}} = \arg \max_k \sum_{j=1}^n l(Z_i; \hat{g}_{k,-j}(U_j), \sigma(X_i)),$$

with

$$(\hat{g}_{k,-j}(U_j), \hat{g}'_{k,-j}(U_j)) = \arg \max_{a_j, b_j} \sum_{i \neq j} l(Z_i; a_j + b_j(U_i - U_j))K \left( \frac{U_i - U_j}{h_k} \right),$$

for $j = 1, \ldots, n$ and for $k = 1, \ldots, K$ (the number of elements in $\mathcal{H}$). Beirlant and Goegebeur [2004] proposed an identical bandwidth selection procedure.

Regarding the initialization of the iterative procedure, Heuchenne et al. [2014] propose to use a combination of the conditional moments method and of the average derivative technique. A detailed explanation of this method can be found in Hristache et al. [2001]. It starts from the fact that, for $\gamma(X) < 0.5$:

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

where $m_1(X)$ and $m_2(X)$ are respectively the first and second conditional moments of $Z$: $m_1(X) = \mathbb{E}[Z|X]$ and $m_1(X) = \mathbb{E}[Z^2|X]$. Using a local polynomial smoothing techniques, we can get 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 $X$ (see the next subsection for more details on these estimators). Then, we can compute

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

and deduce $\nabla_x \hat{\gamma}(X)$, the corresponding gradient vector. Now, using the average derivative technique, we define

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

where $\lambda^{(0)}$ is a normalizing constant ensuring that the first element of $\hat{\theta}^{(0)}$ is equal to one (recall the identifiability condition stated in equation (7)). This result stems from the fact that, if the model is true, we have the two following relationships [Hristache et al., 2001]:

$$\nabla_x \gamma(X_i) = \theta_0 \gamma'_{\theta_0}(\theta_0^T X_i)$$

$$\frac{1}{n} \sum_{i=1}^n \nabla_x \gamma(X_i) = \theta_0 \frac{1}{n} \sum_{i=1}^n \gamma'_{\theta_0}(\theta_0^T X_i)$$

Using the identifiability condition previously stated, we finally obtain equation (16).
2.3 Estimating the scale and the threshold parameters

Eventually, on a practical point of view, we need also estimators of the conditional scale and threshold parameters to use our method in a LDA. With these quantities at hand, we would be able to fully characterize the conditional severity distribution.

For the scale parameter, we propose to use the method of conditional moments since, for $\gamma(X) < 0.5$, we have:

$$\sigma(X) = \frac{m_1(X)}{2} \left( 1 + \frac{m_1(X)^2}{m_2(X) - m_1(X)^2} \right),$$  \hspace{1cm} (19)

with the same definition of $m_1(X)$ and $m_2(X)$ as in Subsection 3.1.2. These moments (as well as their gradients) can be estimated and then plugged in equations 19 and 15 for the estimation of $\hat{\sigma}$ and $\hat{\theta}^{(0)}$. We define first a local polynomial estimator of $m_1(X)$, $m_2(X)$ and their gradients in the following way:

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

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

where $\delta = \{\delta_0, \delta_1, \ldots, \delta_d\}$ 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 $j = 1, \ldots, d$ 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

$$(\hat{m}_1(X), \nabla_x \hat{m}_1(X))^T = (X_x^T W_x X_x)^{-1} X_x^T W_x Z,$$  \hspace{1cm} (20)

where

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

and

$$W_x = \text{diag} \left\{ \prod_{l=1}^{d} K \left( \frac{X_1(l) - X(l)}{h_1} \right), \ldots, \prod_{l=1}^{d} K \left( \frac{X_n(l) - X(l)}{h_2} \right) \right\}.$$
estimators, using the following definition of $\hat{m}_1(X)$ and $\hat{m}_2(X)$:

$$\hat{m}_1(X) = \frac{\sum_{i=1}^{n} \prod_{j=1}^{d} K \left( \frac{X_i(l) - X(l)}{h_1} \right) Z_i}{\sum_{i=1}^{n} \prod_{j=1}^{d} K \left( \frac{X_i(l) - X(l)}{h_1} \right)},$$ (21)

$$\hat{m}_2(X) = \frac{\sum_{i=1}^{n} \prod_{j=1}^{d} K \left( \frac{X_i(l) - X(l)}{h_2} \right) Z_i^2}{\sum_{i=1}^{n} \prod_{j=1}^{d} K \left( \frac{X_i(l) - X(l)}{h_2} \right)},$$ (22)

Besides, the estimation of the threshold parameter is a delicate operation. Indeed, as explained in Embrechts et al. [1997], Chavez-Demoulin et al. [2014a,b] and Scarrot and MacDonald [2012], threshold choice involves 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 estimates. Usually, one work with a fixed threshold approach, meaning that the threshold is chosen before fitting the other parameters. We use a combination of graphical diagnostic tools (like the Hill plot or the threshold stability plot [Scarrot and MacDonald, 2012]) and rules of thumb (like simple fixed quantile rule, e.g. at the 90th percentile), 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 a higher order. Overall, as noticed by Chavez-Demoulin et al. [2014b](p.46), the threshold selection has a degree of arbitrariness in practice. In our empirical application, we use a combination of these tools to try to find an adequate threshold. 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 paper. Therefore we stick to a classical approach and use the following tools to determine some potential thresholds:

1. A graphical analysis of the mean excess plot. For a set of various thresholds $\tau_j > 0$, between 0 and $\max\{Y_1, \cdots, Y_n\}$, we compute

$$\mathbb{E}(Y - \tau_j | Y > \tau_j) = \sum_{i=1}^{n} \mathbb{1}(Y_i > \tau_j)(Y_i - \tau_j)/\sum_{i=1}^{n} \mathbb{1}(Y_i > \tau_j)$$

where $\mathbb{1}$ is an indicator function taken value one if the condition in parentheses is met, zero otherwise. If the data are effectively GPD distributed, the mean excess
function should appear linear. Hence, one will select the lowest threshold so that the mean excess function is linear in \( \tau \). 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 inferred on too few data and 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’s estimator being biased for small samples, Huisman et al. [2001] proposed a convenient modified version of this estimator, less biased. We use it in our graphical analysis.

3. A simple computation of the empirical quantile \( \hat{q}(p) \) with \( p = \{0.1, 0.2, 0.3, 0.4, 0.5\} \).

3 Empirical study

3.1 Description of the data

This database consists in 40,871 operational losses taking place between 2005 and 2014, from a single large European bank. The data have been provided by its 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. We count 71 losses with a value equal to this threshold. The losses are dispatched into 12 risk categories, 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 code associated to each category can be found in Appendix 3.4 (Table 4). Figure 1 shows their distribution among these categories. The categories with the largest number of losses are clients - products - business practices related to financial instruments and others (respectively 6342 and 7330 losses), and execution - delivery and process management related to payments and others. Table 1 provides descriptive statistics conditional to the risk category (as well as descriptive statistics for all the losses). We observe some dissimilarities, especially between the second and eighth classes on one side, and the other classes on the other side.
We have also access to the exact date of the event, therefore we can break down the losses month by months. Table 2 presents descriptive statistics conditionally to the month of the event. We observe that December (the traditional month for accounting closure) seems to suffer bigger losses than the other months (the mean and the standard deviation are particularly large).

Additionally to these basic covariates (i.e. the date and the risk classification), we have access to a large number of firm-specific data, computed at a quarterly frequency: tier-I capital ratio, % of fee revenue, deposit growth, ratio of external operating profit (EOP) from loans versus EOP from deposits, pretax return on assets, pretax return on equity, leverage ratio (assets/equity). We also have access to financial covariates (stock price of the bank, Thomson Reuters European index price, S&P500 price, MIB and VFTSE index price, VIX index price, long term bond rate) and macroeconomic covariates (both at the country level and European level: unemployment rate, wage inflation, real inflation and a consumer confidence index). A detailed description of all the covariates can be found in Appendix 3.4. It gives us a total of 24 variables to investigate. This is a very large set of covariates, therefore we will use subgroups of covariates rather than all of them simultaneously.

![Figure 1: Break-down of the losses between the 12 risk categories (according to internal classifications).](image)
<table>
<thead>
<tr>
<th>Code</th>
<th>Mean</th>
<th>Std</th>
<th>Median</th>
<th>3\textsuperscript{rd} quartile</th>
<th>Skewness</th>
<th>Max</th>
<th># losses</th>
<th>$\gamma^H$</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>50004</td>
<td>992190</td>
<td>6208.5</td>
<td>170670</td>
<td>65.63</td>
<td>103.4e6</td>
<td>40871</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>139470</td>
<td>1.06e6</td>
<td>12448</td>
<td>48482</td>
<td>20.9098</td>
<td>285600</td>
<td>1271</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>10370</td>
<td>136400</td>
<td>3139</td>
<td>5271</td>
<td>52.4976</td>
<td>76200</td>
<td>3340</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>30320</td>
<td>632600</td>
<td>6866</td>
<td>18850</td>
<td>54.564</td>
<td>348200</td>
<td>3051</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>35260</td>
<td>294600</td>
<td>6533</td>
<td>17949</td>
<td>36.0121</td>
<td>126500</td>
<td>2292</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>127080</td>
<td>1.55e6</td>
<td>26289</td>
<td>68149</td>
<td>32.1107</td>
<td>550600</td>
<td>2466</td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>33210</td>
<td>772900</td>
<td>6459</td>
<td>14005</td>
<td>53.0658</td>
<td>432600</td>
<td>6342</td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>85250</td>
<td>1.84e6</td>
<td>10438</td>
<td>26214</td>
<td>44.5918</td>
<td>103.4e6</td>
<td>7330</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>9990</td>
<td>60800</td>
<td>3241</td>
<td>5357</td>
<td>22.3868</td>
<td>16400</td>
<td>896</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>14430</td>
<td>40500</td>
<td>4288</td>
<td>10058</td>
<td>7.7892</td>
<td>5200</td>
<td>674</td>
<td></td>
</tr>
<tr>
<td>71</td>
<td>35230</td>
<td>559000</td>
<td>5139</td>
<td>11221</td>
<td>43.8784</td>
<td>304900</td>
<td>4085</td>
<td></td>
</tr>
<tr>
<td>72</td>
<td>17750</td>
<td>89500</td>
<td>4646</td>
<td>9489</td>
<td>17.0932</td>
<td>22900</td>
<td>1791</td>
<td></td>
</tr>
<tr>
<td>73</td>
<td>42990</td>
<td>439900</td>
<td>4951</td>
<td>12670</td>
<td>27.8981</td>
<td>187300</td>
<td>7333</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Descriptive statistics of the losses, conditionally to their risk classification. A description of the risk category associated to the identification codes can be found in Appendix.
3.2 Threshold selection

To determine the threshold, as indicated in Section 3.1.3, we display a mean excess plot and a plot of the Hill and modified Hill estimators of Huisman et al. [2001] (Figure 2 and 3). 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 420000€ for the Hill estimator, around 150000€ for the modified Hill estimator). Such thresholds would give us samples of excesses of respectively 529 and 1196 losses. The thresholds correspond to the 0.988 and 0.971 empirical quantile.

Table 2: Descriptive statistics of the losses, conditionally to the month of the event.

<table>
<thead>
<tr>
<th>Code</th>
<th>Mean</th>
<th>Std</th>
<th>Median</th>
<th>3\textsuperscript{rd} quartile</th>
<th>Skewness</th>
<th>Max</th>
<th># losses</th>
<th>(\hat{\gamma}^H)</th>
</tr>
</thead>
<tbody>
<tr>
<td>J</td>
<td>40711</td>
<td>572200</td>
<td>5097.9</td>
<td>12802</td>
<td>29.48</td>
<td>18.72e\textsuperscript{6}</td>
<td>2650</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>76085</td>
<td>1.76e\textsuperscript{6}</td>
<td>5251.5</td>
<td>12755</td>
<td>38.78</td>
<td>79.24e\textsuperscript{6}</td>
<td>2741</td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>40092</td>
<td>569500</td>
<td>6555.4</td>
<td>17980</td>
<td>40.99</td>
<td>28.56e\textsuperscript{6}</td>
<td>4468</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>43046</td>
<td>746200</td>
<td>5182.4</td>
<td>12603</td>
<td>39.57</td>
<td>34.82e\textsuperscript{6}</td>
<td>2733</td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>21019</td>
<td>88600</td>
<td>6052.8</td>
<td>14924</td>
<td>20.78</td>
<td>2.78e\textsuperscript{6}</td>
<td>2946</td>
<td></td>
</tr>
<tr>
<td>J</td>
<td>45253</td>
<td>407700</td>
<td>7527.2</td>
<td>22320</td>
<td>36.07</td>
<td>20.28e\textsuperscript{6}</td>
<td>4575</td>
<td></td>
</tr>
<tr>
<td>J</td>
<td>19329</td>
<td>87900</td>
<td>5353.3</td>
<td>12820</td>
<td>26.22</td>
<td>3.40e\textsuperscript{6}</td>
<td>2565</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>23871</td>
<td>173300</td>
<td>5349.8</td>
<td>13062</td>
<td>31.72</td>
<td>6.86e\textsuperscript{6}</td>
<td>2124</td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>53892</td>
<td>1.24e\textsuperscript{6}</td>
<td>6733</td>
<td>20000</td>
<td>56.84</td>
<td>72.16e\textsuperscript{6}</td>
<td>3468</td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>28339</td>
<td>370900</td>
<td>5851.3</td>
<td>15114</td>
<td>50.07</td>
<td>19.57e\textsuperscript{6}</td>
<td>2936</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>61060</td>
<td>1.92e\textsuperscript{6}</td>
<td>5499.2</td>
<td>14025</td>
<td>53.77</td>
<td>103.40e\textsuperscript{6}</td>
<td>2941</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>90955</td>
<td>1.26e\textsuperscript{6}</td>
<td>8363.4</td>
<td>25306</td>
<td>33.96</td>
<td>55.06e\textsuperscript{6}</td>
<td>6724</td>
<td></td>
</tr>
</tbody>
</table>
Figure 2: Mean excess plot for a sequence of increasing threshold. The X-axis is expressed in € Million.

Figure 3: Value of the Hill estimator (solid) and modified Hill estimator (dashed) of Huisman et al. [2001] for a sequence of increasing thresholds.

3.3 Results
We start with a preliminary analysis where we look at the dependence between the shape parameter and the 24 covariates, one by one. As a threshold, we use 150000€, as a com-
promise between no threshold and 420000€. It gives us a sample of 1196 excesses. For the scale parameter, we apply the conditional method of moments. For each explanatory variable, we display the results for contemporaneous and the lagged covariate. Figures 4 to 6 show the estimated conditional shape parameter, with respect to the firm-specific, financial and macroeconomic covariates. The estimated parameters vary roughly between 0.4 and 1. Results related to the estimation of the conditional scale parameter and the \( VaR_{0.999} \) can be found in Appendix 3.4.2.

Figure 4: Estimation of the conditional shape parameter, with respect to a single covariate. From top left to bottom right: efficiency ratio, % fee revenue, deposit growth, Tier-I risk adjusted capital ratio, ratio of external operating profit (EOP) from loans versus EOP from deposits, securities average earning, pre-tax return on assets (ROA), leverage ratio (assets/equity) and pretax return on equity (ROE). X axis: value of the covariate. Y axis: shape parameter. Black: contemporaneous covariates. Red: lagged covariates.
Figure 5: Estimation of the conditional shape parameter, with respect to a single covariate. From top left to bottom right: stock price of the bank, Thomson Reuters European index price, S&P500 price, MIB and VFTSE index price, VIX index price, long term bond rate. Y axis: shape parameter. Black: contemporaneous covariates. Red: lagged covariates.
Figure 6: Estimation of the conditional shape parameter, with respect to a single covariate. From top left to bottom right: unemployment rate, European unemployment rate, wage inflation, European wage inflation, real inflation, European real inflation, a consumer confidence index, a European consumer confidence index. X axis: value of the covariate. Y axis: shape parameter. Black: contemporaneous covariates. Red: lagged covariates.

Now we apply our iterative procedure to obtain multivariate conditional parameter estimations. We build subsets of covariates to obtain multivariate conditional estimators of $\gamma(X)$. 
We consider 8 groups of covariates, whose associated results are displayed in Table 3. We start by considering pairs of covariates. Figures ?? to 10 show the response surface of the shape parameter and of the \( VaR_{0.999} \). Estimated scale parameters and link functions can be found in Appendix 3.4.2. Figures 11 to 14 show the results for the four other sets of covariates. When we have more than 2 covariates, we cannot draw a d-dimensional hyper-surface. Therefore, we set up some covariates to their sample median and look at the variations induced by the other ones. First, let us remark that the shape parameter varies between 0.5 and 0.8, whereas the \( VaR_{0.999} \) is approximately of order \( 10^8 \).

For the first set of covariates, we see that an increase in the efficiency ratio is linked with a decrease of the shape parameter. Based on the sign of the single index parameter and the link function, an increase of the Tier-I capital ratio is linked with an increase of the shape parameter. Notice that for the efficiency ratio covariates, we had to modify several outliers: the third semester 2011 registered an efficiency of 213.4%. This is replaced by the sample median. For the second set of covariate, we observe that a decrease of the pre-tax ROE past a given point (approximately a pre-tax ROE of 0.05) is linked to an increase of the shape parameter. Similarly, an increase of the leverage ratio is associated with an increase of shape parameter. For the third set of covariate, we observe opposite relationships: an increase of VFTSE index is linked with an increase of the shape parameter, whereas an increase of the bank’s stock price is associated to a decrease of this parameter. For the last pair of covariates, we observe that an increase of the unemployment rate is linked with an increase of the shape parameter, whereas an increase of the efficiency ratio is associated with a decrease of this parameter. Looking at the value of the negative log-likelihood function (Table 3), we see that the third pair of covariates as the lowest value, suggesting that these variables could possess the best explanatory power.
<table>
<thead>
<tr>
<th>Covariate</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Efficiency ratio</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Tier-I capital ratio</td>
<td>-1.75</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-0.1039</td>
<td>-</td>
<td>-5.6811</td>
<td>-</td>
</tr>
<tr>
<td>Leverage ratio</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-1.443</td>
<td>-</td>
<td>-1.6808</td>
<td>0.1504</td>
</tr>
<tr>
<td>pre-tax ROE</td>
<td>-</td>
<td>11.6537</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-3.0281</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Stock price</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>0.0219</td>
</tr>
<tr>
<td>VFTSE index</td>
<td>-</td>
<td>-</td>
<td>-1.228</td>
<td>-</td>
<td>-</td>
<td>-0.621</td>
<td>-</td>
<td>0.0093</td>
</tr>
<tr>
<td>ITA unemployment</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-0.0083</td>
<td>-</td>
<td>-0.1063</td>
<td>-</td>
<td>-0.0867</td>
</tr>
</tbody>
</table>

# covariates 2 2 2 2 3 3 4 5
Neg. LL (10^4) 1.687 1.6871 1.6826 1.685 1.6862 1.6815 1.6873 1.6807

Table 3: Subset of covariates used to estimate the conditional shape parameters, as well as the final value of the estimated $\theta_0$ parameter related to each covariate. Value of the covariates are computed the same semesters as the one when the loss takes place.
Figure 7: Estimation of the shape parameter and the $VaR_{0.999}$, conditionally to the first pair of covariates investigated. On the X axis: efficiency ratio. On the Y axis: Tier-I capital ratio. Left: shape parameter. Right: $VaR_{0.999}$.

Figure 8: Estimation of the shape parameter and the $VaR_{0.999}$, conditionally to the second pair of covariates investigated. On the X axis: leverage ratio. On the Y axis: pre-tax ROE. Left: shape parameter. Right: $VaR_{0.999}$. 
Figure 9: Estimation of the shape parameter and the $VaR_{0.999}$, conditionally to the third pair of covariates investigated. On the X axis: stock price. On the Y axis: VFTSE index (i.e. volatility indicator). Left: shape parameter. Right: $VaR_{0.999}$.

Figure 10: Estimation of the shape parameter and the $VaR_{0.999}$, conditionally to the third pair of covariates investigated. On the X axis: efficiency ratio. On the Y axis: Italian unemployment rate. Left: shape parameter. Right: $VaR_{0.999}$.

Regarding the fifth and sixth sets of covariates, we observe the same kind of effects: an increase of the efficiency ratio is associated with a decrease of the shape parameter, whereas both an increase in the Tier-I capital ratio and the leverage ratio are associated with an increase of the shape parameter (and thus with an increase of the $VaR_{0.999}$). Moreover,
we observe that when the variations of the shape parameter associated with variations of the efficiency ratio are more important than variations with the other variables (especially than the Tier-I capital ratio). We also observe a negative relationship between the shape parameter and the stock price, as well as a positive relationship between the shape parameter and the level of the VFTSE index. Italy unemployment appears to have a very limited effect in this configuration. Regarding the value of the negative log-likelihood function, the lowest value is provided by the set of financial and macroeconomic covariates (sixth set, composed of the stock price, the VFTSE index and the Italian unemployment rate).

For the seventh set of covariates, the results are more difficult to interpret. The marginal variations, as well as the link function, are far from being monotonic.

Eventually, we observe that the economic well-being of the bank appears to have an impact on the shape parameter: high pre-tax ROE, efficiency ratio or stock price are associated with low shape parameter, and therefore with lower probabilities of very large losses. At the contrary, indicators of an unbalanced economic situation or of an unfavourable macroeconomic context are linked with an increase of the shape parameter: high market volatility (measured by the VFTSE index), high unemployment rate and high leverage ratio are associated with a high level of the shape parameter. Some of these results are in apparent contradiction with Cope et al. [2012]. Indeed, they found out that the GDP per capita is positively linked to the severity of the losses (in certain event types), postulating that good economic conditions are associated with higher losses. However, using firm-level indicators, it appears that a financially healthy bank bears lesser risks to suffer from large operational losses. This difference may be due to the fact that GDP per capita explain more the severity of the losses associated to legal settlements, whereas our measures are maybe more related to other types of losses (e.g. internal fraud), more dependent to the risk management inside the bank.

Interactions between more than three covariates are difficult to study, and we may suffer from boundary issues. Nevertheless, our methodology seems to provide an interesting way to take several covariates into account.
Figure 11: Top left: estimation of the shape parameter as a function of the single index with the parameters display in Table 3 for subset 5. From left to right and top to bottom: effect, on the shape parameter, of the variation of a single covariate. The other covariates are set up to their sample median.

Figure 12: Top left: estimation of the shape parameter as a function of the single index with the parameters display in Table 3 for subset 6. From left to right and top to bottom: effect, on the shape parameter, of the variation of a single covariate. The other covariates are set up to their sample median.
Figure 13: Top left: estimation of the shape parameter as a function of the single index with the parameters display in Table 3 for subset 7. From left to right and top to bottom: effect, on the shape parameter, of the variation of a single covariate. The other covariates are set up to their sample median.

Figure 14: Top left: estimation of the shape parameter as a function of the single index with the parameters display in Table 3 for subset 8. From left to right and top to bottom: effect, on the shape parameter, of the variation of a single covariate. The other covariates are set up to their sample median.
3.4 Scenarios Analysis

To be completed.

4 Conclusion

In this paper, in the context of modeling the operational loss severity distribution, we study the relationship between the shape parameter ($\gamma$) of a GPD distribution and explanatory variables (macro-economic, firm specific and financial covariates). At this end, we use a database of a large European bank, consisting in around 41,000 operational events. To study these relationships, we estimate $\gamma$ conditionally to these covariates, using a method that combines conveniently a single index hypothesis with semiparametric regression tools.

We show that the economic well being of the considered bank appears to have an impact over the severity distribution: when the bank is in a good situation (as indicated by a high pre-tax ROE, a high efficiency ratio or a high stock price), the shape parameter decreases, indicating that the probability of very large losses is decreased, too. At the contrary, we observe that an unbalanced economic situation could increase this probability: a high leverage ratio, a high financial market volatility (as measured by the VFTSE index) and a high unemployment rate are all associated with an increase of the shape parameter.

These results indicates that the general environment of a bank as an impact on its operational risks. Hence, in the purpose of modeling the distribution of these risks to set a capital level or to monitor its exposure, a bank should calibrate its model conditionally to covariates. We provide here several econometric tools that could be of interest. Finally, if a bank want to perform a scenario analysis, they could apply our approach very easily.
# Appendix

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

Table 4: Risk categories and associated identification code for the European bank database.
Figure 15: Estimation of the conditional scale parameter, with respect to a single covariate. From top left to bottom right: efficiency ratio, % fee revenue, deposit growth, Tier-I risk adjusted capital ratio, ratio of external operating profit (EOP) from loans versus EOP from deposits, securities average earning, pre-tax return on assets (ROA), leverage ratio (assets/equity) and pretax return on equity (ROE). X axis: value of the covariate. Y axis: shape parameter. Dotted: contemporaneous covariates. Black: contemporaneous covariates. Red: lagged covariates.
Figure 16: Estimation of the conditional scale parameter, with respect to a single covariate. From top left to bottom right: stock price of the bank, Thomson Reuters European index price, S&P500 price, MIB and VFTSE index price, VIX index price, long term bond rate. Y axis: shape parameter. Black: contemporaneous covariates. Red: lagged covariates.
Figure 17: Estimation of the conditional scale parameter, with respect to a single covariate. From top left to bottom right: unemployment rate, European unemployment rate, wage inflation, European wage inflation, real inflation, European real inflation, a consumer confidence index, a European consumer confidence index. X axis: value of the covariate. Y axis: shape parameter. Black: contemporaneous covariates. Red: lagged covariates.
Figure 18: Estimation of the link function and the scale parameter, conditionally to the first pair of covariates. On the X axis: efficiency ratio. On the Y axis: Tier-I capital ratio. Left: shape parameter as a function of the single index. Right: response surface of the scale parameter. For the scale parameter, the bandwidth has been chosen by trial-and-error, and is close to half of the sample range of the data.

Figure 19: Estimation of the link function and the scale parameter, conditionally to the first pair of covariates. On the X axis: leverage ratio. On the Y axis: pre-tax ROE. Left: shape parameter as a function of the single index. Right: response surface of the scale parameter. For the scale parameter, the bandwidth has been chosen by trial-and-error, and is close to half of the sample range of the data.
Figure 20: Estimation of the link function and the scale parameter, conditionally to the first pair of covariates. On the X axis: stock price. On the Y axis: VFTSE index. Left: shape parameter as a function of the single index. Right: response surface of the scale parameter. For the scale parameter, the bandwidth has been chosen by trial-and-error, and is close to half of the sample range of the data.

Figure 21: Estimation of the link function and the scale parameter, conditionally to the first pair of covariates. On the X axis: efficiency ratio. On the Y axis: Italian unemployment rate. Left: shape parameter as a function of the single index. Right: response surface of the scale parameter. For the scale parameter, the bandwidth has been chosen by trial-and-error, and is close to half of the sample range of the data.
References


