What are the economic determinants of operational losses severity ? A regularized Generalized Pareto regression approach.

J. Hambuckers[1,†], A. Groll[1,†] and T. Kneib[1,†]

Résumé

We investigate a database of 40,871 extreme operational losses from the bank UniCredit. These data cover a period of 10 years and 7 different event types. We study the dependence between a set of macroeconomic, financial and firm-specific factors with the severity distribution of these losses, assumed to be Generalized Pareto. Answering to this question is of particular interest for banks and regulators to define a risk capital charge in line with the economic situation. To perform covariate selection and identify the relevant explanatory variables, we use a penalized-likelihood approach based on a local quadratic approximation of L_1 penalty terms. Because this method has not been applied yet to Generalized Pareto regression, we study the finite-sample properties of this estimation technique in a simulation study. Then, we conduct the regression analysis with the proposed approach. Our results suggest that only a small subset of the covariates are deemed relevant. Among them, the unemployment rate, the VIX index and the leverage ratio are found to be good explanatory variables of the severity distribution. Last, we illustrate the impact of several economic scenarios on the requested capital if the total loss distribution is conditioned on these scenarios.

Keywords : Operational loss ; Generalized Pareto ; penalized likelihood.

[1] Georg-August-Universität Göttingen, Chair of Statistics, Humboldtallee 3, 37073 Göttingen, Germany.
† corresponding author : julien.hambuckers@mathematik.uni-goettingen.de.

Résumé

Dans cet article, nous étudions une base de données constituée de 40,871 pertes opérationnelles extrêmes, provenant de la banque UniCredit. Ces données ont été collectées sur une période de 10 ans et recouvrent 7 différents types de pertes. Nous étudions la dépendance entre un ensemble de variables économiques et financières, et la distribution de sévérité de ces pertes, supposées suivant une distribution Generalized Pareto. Cette problématique est particulièrement intéressante pour les banques et leurs régulateurs, pour définir un niveau de capital réglementaire qui soit en adéquation avec la situation économique. Afin de sélectionner les variables pertinentes dans un tel modèle, nous employons une approche par fonction de vraisemblance pénalisée, où nous utilisons une approximation quadratique locale pour le terme de pénalisation, de type L_1. Etant donné qu'une telle approche n'a pas encore été utilisée dans le cas de la régression Generalized Pareto, nous étudions brièvement ses propriétés aux moyens de simulations. Appliquant la méthode proposées à nos données, nos résultats indiquent que seule une petite proportion des variables étudiées ont un impact significatif. Parmi celles-ci, le taux de chômage, l'indice VIX et le ratio d'endettement semble avoir un impact sur la distribution de sévérité. En dernier lieu, nous complétons cette analyse en illustrant l'impact de certains scénarios économiques sur le capital requis si la distribution de la perte opérationnelle totale est conditionnée à ces scénarios.

Mots-clés : Perte opérationnelle ; distribution Generalized Pareto ; vraisemblance pénalisée.

1 Summary

Understanding the relationship between the severity distribution of operational losses and economic variables is particularly important for the banking industry, especially for risk management and regulatory purposes. Indeed, if we know some variables that can explain variations in the severity distributions of these losses, we could improve the estimation of the associated risk measures and the adequateness of the requested operational loss capital charge.

Operational losses are defined by the Basel Committee for Banking Supervision (BCBS) as *"direct or indirect losses resulting from inadequate or failed internal processes, people and systems or from external events"* [Basel Committee on Banking Supervision (BCBS), 2004]. The total loss L_t over the time period $[t-1,t]$ is given by

$$L_t = \sum_{i=1}^{N_t} Z_{t,i} \tag{1}$$

where N_t is the number of losses over the considered period and $Z_{t,i}$ the size of the i^{th} loss during the t^{th} period. $Z_{t,i}$ is here our quantity of interest. For $Z_{t,i}$, an usual approach is to rely on the Peak-over-Threshold (POT) technique and to only consider losses above a high threshold (since those losses are the main driver of L_t). Thanks to the Extreme Value Theory (EVT), we know that the Generalized Pareto distribution (GPD) is the limiting distribution of these exceedances [under suitable conditions, see Pickands, 1975]. Therefore, we can use a GPD to model the severity distribution of these large losses.

Starting from this representation of the operational loss phenomenon, we investigate a set of firm-specific, macroeconomic and financial variables that may impact the distribution of large values of $Z_{t,i}$, larger than some threshold τ in the POT idea. Indeed, whereas previous studies [Moscadelli, 2004, Chapelle et al., 2008] focused on modelling the operational losses *independently from the economic conditions* with a model of type (1), the attention of researchers shifted recently towards the *conditional distribution* (i.e. conditional on the economic situation) of operational losses [Chernobai et al., 2011, Cope et al., 2012, Wang and Hsu, 2013, Chavez-Demoulin et al., 2016].

To do so, we study into detail a database of around 42,000 private operational losses from a single bank (UniCredit), over a ten-year period (01/2005-06/2014) (Figure 1, left panel). Relying on the POT technique, the number of losses effectively under study is reduced to 10,217. This is a unique feature of the present paper, as such a huge amount of private data is unusual and extremely difficult to obtain for academics [Chavez-Demoulin et al., 2016].To each loss, we associate macroeconomic as well as financial time series. We were also provided with internal ratios directly related to UniCredit. To the best of our knowledge, the lack and the poor quality of data have always been obstacles in the other empirical studies. Confidentiality, heterogeneity and small sizes of public databases make it difficult to associate covariates to losses and to analyse properly the data. The present study overcomes these difficulties.

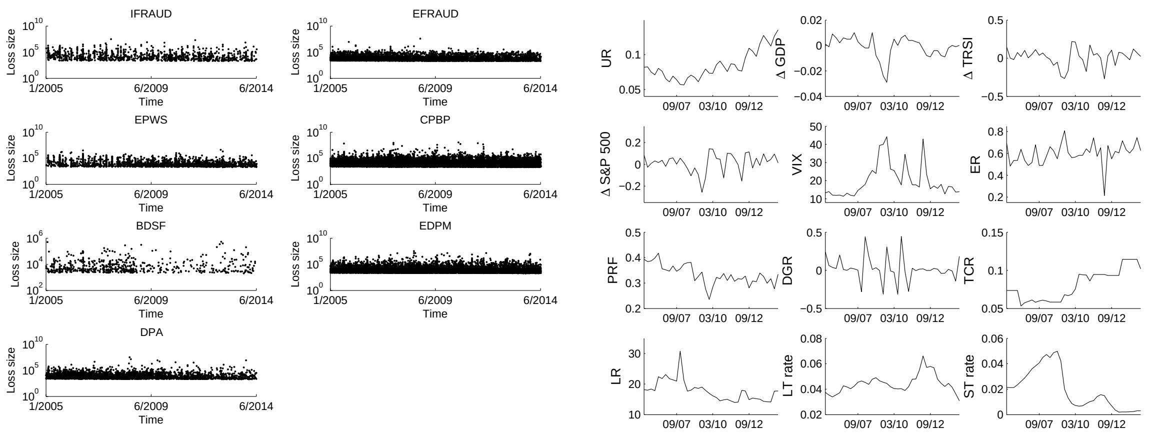

Figure 1 – Left: Operational losses recorded at UniCredit between 2005 and 2014 (y-axis is in log-scale), split between event types. Right: Time series of the economic covariates.

Following Chavez-Demoulin et al. [2016], we apply the POT method in a non-stationary context. We assume that the severity distribution for losses over τ is GPD, but with γ and σ taking values that depend on economic covariates. We make the assumption that for the i^{th} loss larger than τ taking place during the time period $[t-1,t]$, it holds that

$$Y_{t,i} \sim GPD(y_{t,i}; \gamma(X^{\gamma}_{t,i}), \sigma(X^{\sigma}_{t,i})), \tag{2}$$

with $Y_{t,i} = Z_{t,i} - \tau$ given $Z_{t,i} > \tau$, $\gamma(X^{\gamma}_{t,i}) > 0$, $\sigma(X^{\sigma}_{t,i}) > 0$, and where these parameters can be characterized by a structures of the type

$$\log(\gamma(X^{\gamma}_{t,i})) = \alpha^{\gamma}_0 + \sum_{j=1}^{p_\gamma} \alpha^{\gamma}_j X^{\gamma}_{t,i}(j), \tag{3}$$

$$\log(\sigma(X^{\sigma}_{t,i})) = \alpha^{\sigma}_0 + \sum_{j=1}^{p_\sigma} \alpha^{\sigma}_j X^{\sigma}_{t,i}(j). \tag{4}$$

where $X^{\theta}_{t,i}(j)$ denotes the j^{th} component of the vector of covariates $X^{\theta}_{t,i}$ associated to $Y_{t,i}$ for $\theta \in \{\gamma, \sigma\}$. We use a log link function to ensure the positivity of the parameters. Such a model is a parametric GAMLSS under the particular case of a GPD response function [Rigby and Stasinopoulos, 2005].

A recurrent problem in such models is to properly select meaningful covariates. Automatic variable selection procedures, relying on penalized-likelihood techniques, have become more and more popular. The difficulty of using these techniques lies in finding the maximum of an objective function suffering from non-differentiability at certain points. To overcome this issue, Oelker and Tutz [2015] recently developed a general framework that allows approximating different types of penalty terms, ensuring continuity and differentiability, as well as sparsity of the final solution. However, they only considered the case of exponential response function. Therefore, we adapt the procedure proposed in Oelker and Tutz [2015] to the GP regression case. More precisely, we consider the following penalized log-likelihood function:

$$\mathcal{L}_{pen.}(Y, \Theta) = \mathcal{L}(Y; \Theta) - \mathcal{P}_\lambda(\Theta), \tag{5}$$

where $\mathcal{L}(Y;\Theta)$ is the log-likelihood and $\mathcal{P}_\lambda(\theta)$ is the penalty with vector of smoothing parameters $\lambda = \{\lambda_\sigma, \lambda_\gamma\}$. We consider the following penalty:

$$\mathcal{P}_\lambda(\Theta) = \lambda_\sigma \sum_{i=1}^{p_\sigma} a^{\sigma}_i |\Theta^{\sigma}_i| + \lambda_\gamma \sum_{j=1}^{p_\gamma} a^{\gamma}_j |\Theta^{\gamma}_j| \tag{6}$$

where θ^{σ}_i (resp. θ^{γ}_j), $i = 1, \ldots, p_\sigma$ (resp.$j = 1, \ldots, p_\gamma$) consists in the j^{th} (resp. i^{th}) parameter associated to the equation of σ (resp. γ). Oelker and Tutz [2015] suggest that non-differentiability can be overcome by maximizing repeatedly a linearised version of (5), making the following approximation for the penalty and its first derivative:

$$
\begin{aligned}
|\xi| &\approx \sqrt{\xi^2 + c}, && (7)\\
\partial|\xi|/\xi &\approx (\xi^2 + c)^{-1/2}\xi. && (8)
\end{aligned}
$$

where c is a small constant.

Our simulation results indicate that this procedure, combined with BIC for the selection of the smoothing parameter and an additional re-estimation step to reduce the bias, provides excellent results in the GP regression case. Following these observations, we apply the proposed technique to our data. Using the penalized approach, we find that only a small fraction of the considered covariates are included in the final model. As shown on the QQ-plot (Figure 2), this model provides a very good fit.

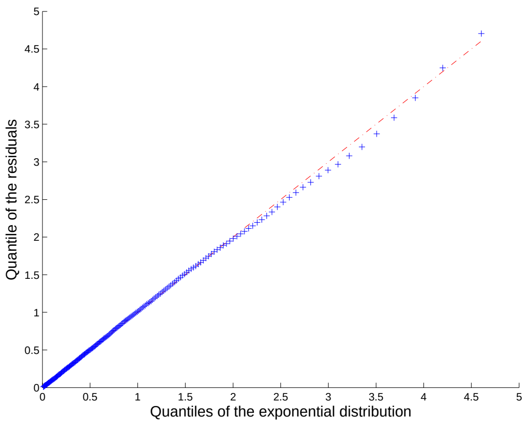

Figure 2 – QQ-plot of the residuals $e_i = (1/\hat{\gamma}_i)\log(1 + \hat{\gamma}_i(z_i - \tau)/\hat{\sigma}_i)$, $i = 1, \ldots, n$

Last, we consider several extensions of this framework by including category-specific interactions, as well as two-way interactions between covariates. We also illustrate the effect of several macroeconomic scenarios on the requested capital: we fit an additional inhomogeneous Poisson process on the frequency distribution, so that we can simulate convolutions between the frequency and the severity processes, and compute estimates of the 99.9% quantile. Economic implications are discussed.

References

Basel Committee on Banking Supervision (BCBS). Basel II: international convergence of capital measurement and capital standards. A revised framework. Technical report, Basel, Switzerland, 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(6):1049–1061, 2008.

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*, 83(3):735–776, 2016.

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

M. Moscadelli. The modelling of operational risk: experience with the analysis collected by the Basel Committee. Technical report, 2004.

M.R. Oelker and G. Tutz. A uniform framework for the combination of penalties in generalized structured models. *Advances in Data Analysis and Classification*, to appear, 2015.

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

R.A. Rigby and D.M. Stasinopoulos. Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society. Series C: Applied Statistics*, 54(3):507–554, 2005.

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