Adjusting the Generalized Pareto Distribution with Evolution Strategies. An application to a Spanish motor liability insurance database[1]

María J. Pérez-Fructuoso [1], Almudena García[3], Antonio Berlanga[2], José M. Molina[2]

[1] Departamento de Economía de la Empresa
[2] Departamento de Informática
Universidad Carlos III de Madrid, Campus de Colmenarejo
Avenida de la Universidad Carlos III, 22. Colmenarejo 28270. Spain
[3] Departamento de Economía de la Empresa
Universidad de Alcalá de Henares
Plaza de la Victoria, 2. Alcalá de Henares 28802. Spain

Abstract. Management of extreme events is required of a special consideration, as well as a sufficiently wide time horizon for solvency evaluation. Whereas their classical adjustment is usually carried out with Extreme Value Theory (EVT)-based distributions (namely, the Generalized Pareto Distribution), Evolutionary Techniques have been tried herein to fit the GPD parameters as an optimisation problem. The comparison between classical and evolutionary techniques highlights the accuracy of the evolutionary process. Data adjusted in this paper come from a Spanish motor liability insurance portfolio.

1. Introduction

Extreme events are low-frequency, high-severity occurrences that classical risk theory cannot entirely capture, as they take place at the tail of the loss probability distribution [1]. Extremes may give rise to higher fluctuations than classical volatility and uncertainty risks, turning notably complex, thus, to specify both loss amounts and capital sums devoted to their coverage.

An accurate estimation of extreme claims is fundamental to assess solvency capital requirements (SCR). Extreme Value Theory (EVT) (classical parametric estimation) is generally used to fit a Generalized Pareto Distribution (GPD) to excesses over a certain threshold depending on the available data. In this work, data are related with the motor liability insurance historical datasets of one representative company operating within the Spanish market. In this work, an Evolution Strategy (ES) is used to obtain those parameters that better fit the GPD to the experimental data. ES lies in the general field of natural metaphor algorithms (simulated annealing, genetic

[1] Funded by CICYT TSI2005-07344, MADRINET S-0505/TIC/0255 and AUTOPIA IMSERSO

programming and so on), and could be named as Darwinian approach or evolutionary programming after [2]. In order to allow an evaluation of the characteristics of the new method, some comparisons have been carried out with classical estimation. This method has been applied before to optimize the parameter values of different possible models in order to adjust them to a real data distribution using simple ES [3] and multiobjetive ES [4]. In this case, models define different distributions representing the behavior of reported claims after a catastrophe occurs. The optimization was carried out using real data recorded from several Spanish catastrophes[2].

This new method provides insurers a useful tool to manage risks, allowing them to infer on a statistical basis the extreme values of either a population or a stochastic process, and hence estimate the probability of yet more extreme events than the historical ones. By modeling extremes aside the global sample data, EVT enables insurers to predict both high values at the tail (outliers) and situations exceeding the historical records, without need to turn to the global distribution of the data observed altogether. Consequently, the study of extreme risk preserves insurers' stability and solvency when faced to the occurrence of extreme losses, by application of statistical models to more precisely measure risk and optimally decide on capital requirements, level of provisions, pricing and cession to reinsurance.

In section 2, the definition of GPD and the classical estimation methods are presented. A brief description of ES [5] is included in section 3. Section 4 is devoted to the results of the classical procedures. ES results appear in section 5 and, finally, a comparison with classical techniques and some conclusions are summarized in section 6.

2. Estimation of Generalized Pareto Distribution

The Pickands-Balkema-De Haan theorem demonstrates that a GPD can be adjusted over a certain threshold. For further information on EVT, among others, we refer the reader to [6], [7], [8], [9] and [10]

The distribution of excesses over u is defined as:

$$F^u(x) = F^u(y+u) = P(X-u \le y / X > u) = \frac{F(x)-F(u)}{1-F(u)} \quad for \quad 0 \le y \le x_0 - u \qquad \textbf{(1)}$$

where x is the total claim amount, x_0 represents the finite or infinite supreme value of the distribution function, and y stands for the excess over the threshold u, with $y = x - u$.

From (1), once the value of the threshold has been optimized, it is possible to fit $F^u(x)$ to a GPD. The GPD is a two parameter distribution with distribution function:

[2] Data from the reinsurance department of "Consorcio de Compensación de Seguros"

$$W_{\gamma,\beta}(x)=\begin{cases}1-\left(1+\xi\dfrac{y}{\beta}\right)^{-1/\xi} & \text{si } \xi\neq 0\\ 1-\exp\left(\dfrac{-y}{\beta}\right) & \text{si } \xi=0\end{cases} \tag{2}$$

where $x\geq 0$ if $\xi\geq 0$, and $0\leq x\leq -\dfrac{\beta}{\xi}$ if $\xi<0$, with ξ and β being the shape and scale parameters. When $\xi>0$, we have the usual Pareto distribution with shape $\alpha=1/\xi$, and the GPD is heavy-tailed. If $\xi<0$, we have a type II Pareto distribution, whereas $\xi=0$ gives the exponential distribution.

The family of GPD's may be extended by adding a location parameter, μ. Then the GPD $G_{\xi,\mu,\beta}(x)$ is defined as $G_{\xi,\beta}(x-\mu)$.

The steps to perform the classical theoretical adjustment of the GPD to a extreme data series are the following:

1) Choose the optimum threshold over which the GPD may be fitted to the corresponding values over the threshold, using to that aim the empirical mean excess function.
2) Estimate the parameters of the model according to the heavy-tailedness of the distribution, by means of those estimators that minimize the Mean Squared Error (MSE):

- MLE (GP): Maximum likelihood estimator for parameters ξ and β based on data sets governed by a GPD.
- Moment (GP): Moment estimate for parameter ξ of GPD's (a better name would be Log-Moment estimate). The estimate is related to the Hill estimate and computed with the *k* largest values of the active data set.
- L-Moment (GP): The L-Moment estimator in the full GP model. Take care that the true shape parameter ξ is smaller than 1.
- Drees-Pickands (GP): Drees-Pickands estimate of the parameter ξ of GPD. The estimate is a mixture of Pickands estimates based on the *k* largest values of the underlying sample.
- Slope(GP): The slope β of the least squares line, fitted to the mean excess function right of the *k* largest observations, is close to $\xi/(1-\xi)$, and, therefore, $\xi(k)=\mu/(1+\mu)$ is a plausible estimate of ξ.

3) Check the goodness-of-fit of the underlying distribution through the Quantile-Quantile plot (QQ plot).

4) Perform the inference on the basis of our estimated conditional model, and calculate the marginal probabilities leading to determine the unconditional data distribution.

Our analysis will focus on one representative Spanish insurer' motor liability portfolio along a ten year-period. It exhibits a recent history (from ten years on), although significantly improved over the last years of the interval. We have introduced a distortion in order to keep the company identity undisclosed.

Two different concepts are assumed as forming the loss amount:

- The cost of settled claims, summing all the net recapture payments already made out
- The cost of non settled claims, comprising all the net recapture payments already made out, and/or the reserves for estimated future payments still pending.

Data have been updated to 2006 values. Table 1 lists below, on an annual basis, the number of claims of the company, with both their total and average individual costs expressed in nominal currency units. Data in Table 1 indicates that the insurer lacks of a stable behavior as regards the average cost evolution. This may be attributed to three main explanatory factors: diverse coverages integrating the final cost of the claims, different claim settlement periods, and the occurrence of extreme events.

The history of the firm (Table 1) shows a weighted loss frequency as high as 45.13 percent during the last four years of the interval, although decreasing until stability with the growth of policies in portfolio.

Table 1. History of the firm

Year	Total cost	Annual variation (%)	Average cost	Annual variation (%)	Number of claims	Annual variation	Number of policies	Annual variation	claims/policies ratio
1	65.487		31,20		2.099		0		
2	182.745	179%	32,08	2,84%	5.697	171,34%	7.913		72,00%
3	284.724	56%	30,29	-5,58%	9.400	65,01%	14.207	79,54%	66,16%
4	375.414	32%	32,45	7,13%	11.569	23,08%	18.454	29,90%	62,69%
5	444.348	18%	32,39	-0,17%	13.717	18,56%	22.300	20,84%	61,51%
6	585.843	32%	35,73	10,29%	16.397	19,54%	28.564	28,09%	57,40%
7	831.326	42%	39,73	11,21%	20.923	27,60%	41.027	43,63%	51,00%
8	1.177.518	42%	42,00	5,69%	28.039	34,01%	59.467	44,95%	47,15%
9	1.408.682	20%	40,93	-2,53%	34.415	22,74%	78.297	31,67%	43,95%
10	1.774.348	26%	44,92	9,74%	39.501	14,78%	93.444	19,35%	42,27%

3. Evolution Strategies

Evolution strategies (ES) developed by Rechenberg and Schwefel, have been traditionally used for optimization problems with real-valued vector representations. As Genetic Algorithms (GA), ES are heuristics search techniques based on the building block hypothesis. Unlike GA, however, the search is basically focused on the gene mutation. This is an adapting mutation based on the likely the individual represents the problem solution. The recombination plays also an important role in the search, mainly in adapting mutation.

Evolutionary Algorithms combine characteristics of both classifications of classical optimization techniques: volume-oriented and path-oriented methods. Volume-oriented methods (Monte-Carlo strategies, clusters algorithms) apply the searching process scanning the feasible region while path-oriented methods (pattern search, gradient descent algorithms) follow a path in the feasible region. A definition of a restricted search space of finite volume and the starting point is required to volume-oriented and path-oriented methods respectively.

Evolutionary Algorithms characteristics change during the evolutionary process and both exploitation and exploration search takes places. ES are techniques widely used (and more appropriated than Genetic Algorithm) in real-values optimization problems. Evolutionary computation algorithms offer practical advantages facing difficult optimization problems [11]. These advantages are: conceptual simplicity, broad applicability, potentiality to use knowledge and hybridize with other methods, implicit parallelism, robustness to dynamic changes, capability for self-optimization and capability to solve problems that have no known solutions. A general ES is defined as an 8-tuple [12]:

$$ES= (I, \Phi, \Omega, \Psi, s, \iota, \mu, \lambda) \tag{3}$$

where $I = (\vec{x}, \vec{\sigma}, \vec{\alpha}) = \Re^n \times \Re_+^{n_\sigma} \times [-\pi, \pi]^{n_\alpha}$ is the space of individuals, $n_\sigma \in \{1,\ldots, n\}$ and $n_\alpha \in \{0, (2n-n_\sigma)(n_\sigma-1)/2\}$, $\Phi : I \rightarrow \Re = f$, is the fitness function, $\Omega=\{m_{\{\tau, \tau', \beta\}}: I^\lambda \rightarrow I^\lambda\} \cup \{r_{\{rx, r\sigma, r\alpha\}}: I^\mu \rightarrow I^\lambda\}$ are the genetic operators, mutation a recombination operators. $\Psi(P)= s(P \cup m_{\{\tau, \tau', \beta\}}(r_{\{rx, r\sigma, r\alpha\}}(P)))$ is the process to generate a new set of individuals, s is the selection operator and ι is the termination criterion.

In this work, the definition of the individual has been simplified: the rotation angles n_α have not been taken into account, n_α=0.

The mutation operator generates new individuals as follows:

$$\sigma_i' = \sigma_i \cdot \exp(\tau' \cdot N(0,1) + \tau \cdot N_i(0,1)) \tag{4}$$

$$\vec{x}' = \vec{x} + \sigma_i' \cdot \vec{N}(\vec{0},1) \tag{5}$$

ES has several formulations, but the most common form is (μ, λ)-ES, where $\lambda > \mu \geq 1$, (μ, λ) means that μ-parents generate λ-offspring through recombination and mutation in each generation. The best μ offspring are selected deterministically from the offspring and replace the current parents. Elitism and stochastic selection are not used. ES considers that strategy parameters, which roughly define the size of mutations, are controlled by a “self-adaptive” property of their own. An extension of the selection scheme is the use of elitism; this formulation is called $(\mu+\lambda)$-ES. In each generation, the best μ-offspring of the set μ-parents and λ-offspring replace current parents. Thus, the best solutions are maintained through generation. The computational cost of (μ, λ)-ES and $(\mu+ \lambda)$-ES formulation is the same.

4. Problem Resolution with Classical Techniques

The classical GPD adjustment to a data sample involves two fundamental issues:

- The choice of the optimal threshold *u*, since not all the priorities render a good adjustment of the parametrical function.
- The estimation of the distribution shape (μ) and scale (β) parameters

The mean excess plot for the 1000 highest claims, with pairs $\left(X_{k+1};E_{k,n}\right)\ for\ \ k=1,...,n-1$, is useful to select the threshold (see Figure 1).

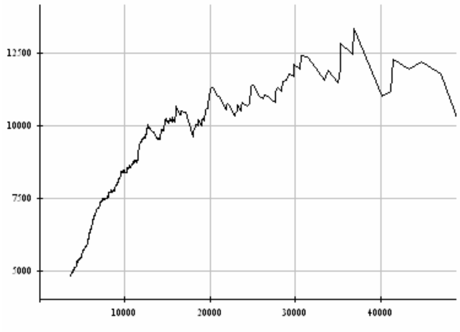

Fig. 1 Mean excess function for the 1000 most severe claims

The plot is growing linear from a quite low priority through roughly 30000, where the function becomes plain or even decreasing. Then, one may wonder whether the extreme losses are heavy-tailed or not, and therefore, if they may be adjusted to a GPD.

The QQ plot versus the exponential distribution, by means of the pairs $\left(F_n^{-1}(p);X_{r,n}\right)$, where $F_n^{-1}(p)=-\ln(1-p)=X_p$ and $p=\frac{r}{(n+1)}$, or $p=\frac{r-1/3}{n+1/3}$ provides the answer to both questions. A straight line would mean the goodness-of-fit of the distribution (namely, the exponential distribution), whereas a curve would indicate a more heavy-tailed distribution (Figure 2).

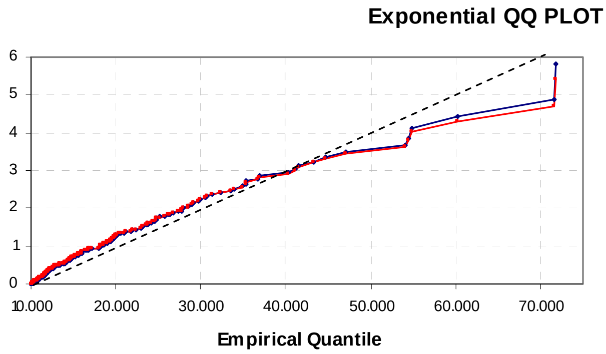

Fig. 2 Exponential QQ Plot

The slight curvature of the empirical quantiles reveals that the tail cannot be modelled with the exponential distribution. But as the plot is very close to linearity, peak losses are heavy-tailed, although not quite. That's to say, the GPD adjusting parameter is positive, but very close to zero.

Being aware that the distribution is heavy-tailed and ignoring at the same time the optimal threshold, we opt by the XTREMES algorithm to both optimize the choice of the threshold to fit the GPD and estimate its parameters. Results are shown in Table 2.

Table 2. Classical Estimation Results

Estimation	Threshold	Tail index ξ	Scale parameter β
MLE	11908	0,137872	8454,29
ME	11908	0,222948	8072,61
L-Moment	11908	0,159826	8220,06
Drees-Pickands	11908	0,154199	8297,47
Slope	11908	0,0433713	8664,96

5. Problem Resolution with Evolution Strategies

The global optimization procedure must adjust the GPD parameters for each value of the threshold minimizing the MSE. The general optimization problem is summarized in the following definition [13]:

Given a function $f: M \subseteq \Re^n \rightarrow \Re$, $M \neq \emptyset$, for $x \in M$ the value $f^* := f(x^*) > -\infty$ is called a global minimum, iff: $\forall x \in M$: $f(x^*) \leq f(x)$

Then x^* is a global minimum point, f is called objective function, and the set M is called the feasible region. In this case, the global optimization problem has a unique restriction: $\xi > 0$. This restriction is included in the codification and all individuals are processed to become feasible ones. Then, in spite of this restriction, the solutions space does not have infeasible regions.

The type of recombination used in this work is the discrete recombination and the strategy (μ+λ)-ES was used to select the individual to the next generation. The parameters of the ES are summarized in Table 3. Besides, different runs where achieved changing the random seed.

Table 3. Setting of exogenous parameters of the ES.

Parameter	**Value**
Initial standard deviations $\sigma_i(0)$	Randomly generated in range [0.01,10.0]
Number of rotation angles n_α	0
Parent population size μ	20
Offspring population size λ	40
Termination criterion	Number of generation step = 500

Results are shown in Figures 3, 4 and 5. In Figure 3, the value of MSE is depicted as function of the threshold value. Each threshold value represents an optimization process to obtain the optimal parameters of GPD for this value. The best value is shown in Figure 3, settled the threshold to 11908. Figure 4 and 5 show the values of parameters for each threshold.

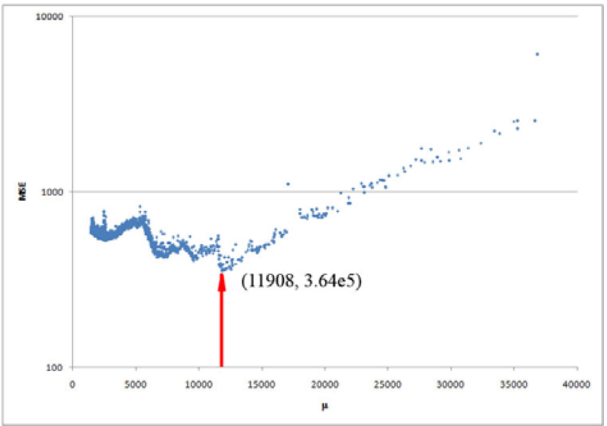

Fig. 3 Optimized MSE value for each threshold

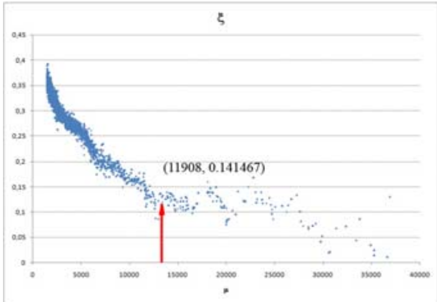

Fig. 4 Optimized ξ for each threshold

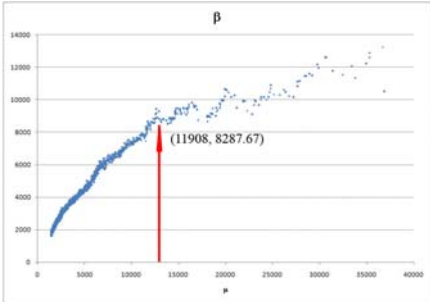

Fig. 5 Optimized β for each threshold

The minimum MSE (3.64E+05) is obtained for a value threshold of 11908. This point is the optimum threshold with values of $\xi = 0.141467$ and $\beta = 8287.67$.

6. Conclusion and discussion

Reinsurers, as well as cedent insurers, share a deep concern in being able to accurately estimate the probability of claims over a certain threshold. Their expertise in handling extreme risks is decisive to determine how much financial capacity is required to assume the cost of extreme losses, or, instead, cede them to reinsurance. Fitting a GPD to claims above a high threshold is a powerful tool to model the tail of severe losses. We have proposed in this work an ES-based optimization method to estimate the GPD parameters. Table 4 shows a comparative analysis of classical techniques and the proposed method.

Table 4. Setting of exogenous parameters of the ES.

Estimation	Threshold μ	Tail index ξ	Scale parameter β	MSE
Evol. Strategies	11908	8287.67	0.141467	3.64E+05
MLE	11908	8454.29	0.137872	3.71E+05
Drees-Pickands	11908	8297.47	0.154199	4.59E+05
L-Moment	11908	8220.06	0.159826	4.88E+05
Slope	11908	8664.96	0.0433713	2.98E+06
Moment	11908	8072.61	0.222948	3.72E+06

Results show that evolution strategies fit the distribution better than EVT classical techniques. We can conclude that ES are an efficient technique to solve the global minimization problem (minimizing the MSE) without any domain knowledge. Then, we probe that ES could be considered in actuarial, financial and economic areas as a solver mechanism able to approximate any mathematical model to a set of real data, as we show in previous works [3],[4] by adjusting the claims rate from a single and multi objective perspective.

7. References

[1] Këllezi, E., Gilli, M., "Extreme Value Theory for Tail-Related Risk Measures". Kluwer Academic Publishers, 2000.
[2] Holland J.H, Adaptation in natural an artificial Systems, MIT Press, Bradford Books edition, Michigan, MI, 1975.
[3] García, J., Pérez, M.J., Berlanga, A., Molina, J. M., "Adjustment of Claims Rate with Evolution Strategies in a Mathematical Model for Insurance Loss Ratio", Internet, e-com and Artifcial Intelligence. III International Workshop on Practical Applications of Agents and Multiagent Systems, pp 173-182, Burgos, Spain, 2004.
[4] Pérez, M. J., García, J., Martí, L., Molina, J. M., "Multiobjective Optimization Evolutionary Algorithms in Insurance-Linked Derivatives". Handbook of Research on Nature Inspired Computing for Economy and Management, Ed. Jean-F. Bernard, vol II, pp 885-908. Idea Group Inc., 2006.
[5] Schewefel, H.-P. "Evolutionary learning optimum-seeking on parallel computer architectures". In A. Sydow, S. G. Tzafestas, and R. Vichnevetsky, editors, Proceedings of the International Symposium on Systems Analysis and Simulation 1988, I: Theory and Foundations, pp. 217-225. Akademie-Verlag, Berlin. 1988
[6] Beirlant, J., Teugels J.L., Vynckier P., "Practical Analysis of Extreme Values". Leuven University Press, Leuven, 1996
[7] De Haan, L., Ferreira A., "Extreme value theory: An introduction". Springer Series in Operations Research and Financial Engineering, New York, 2006
[8] Embrechts, P., Klüppelberg C., Mikosch T., "Modelling extremal events for Insurance and Finance". Applications of Mathematics, Springer-Verlag Berlin Heidelberg, 1997
[9] Kotz, S, Nadarajah S., "Extreme value distributions. Theory and Applications". Imperial College Press, London, 2000.
[10] Reiss, R. D., Thomas M., "Statistical Analysis of Extreme Values (with applications to insurance, finance, hydrology and other fields)", Birkhäuser Verlarg, 2nd edition, 2001
[11] Fogel, D.B., "The Advantages of Evolutionary Computation," Proc. of BCEC97: BioComputing and Emergent Computation, D. Lundh, B. Olsson, and A. Narayanan (eds.), World Scientific, Singapore, pp. 1-11, 1997
[12] Bäck, T.,"Evolutionary Algorithms in Theory and Practice". Oxford University Press, Inc. 1996
[13] Törn, A., Zilinskas, A., "Global Optimization", vol. 350 of Lecture Notes in Computer Science. Springer, Berlin. 1991