Operations Value-at-Risk to Estimate Maximum Claims of Potential House Fire Risk using the Portfolio Approach as an Information Supply Chain in Decision Making

Endang Soeryana Hasbullah[1], Sukono[1], Riaman[1], S. Supian[1], Moch Panji Agung[1], Asep Saepulrohman[2]

[1]*Department of Mathematics, Faculty of Mathematics and Natural Sciences, Padjadjaran University, Indonesia*
[2]*Department of Computer Sciences, Faculty of Mathematics and Natural Sciences, Pakuan University, Indonesia*
*Corresponding author: endang.soeryana@unpad.ac.id

***Abstract*— The occurrence of house fires in densely populated areas has a high-risk level. One of city in Indonesia that has high-risk level of this incident is Bandung City. That high-risk incident cause anxiety community, and also cause many house fire insurance product arise. Insurance product is made to protect consumers from risk and guarantee by a premium. Insurance company formulate premium based on analysis calculation from expected claim, cost, commission, and margin. This paper aims to estimate maximum expected claim using portfolio approach. There are several steps in this research. The first step is resampling the data used Maximum Entropy Bootstrapping (ME Boot). Next, determine threshold value to get extreme data value. Then, conduct Kolmogorov-Smirnov test to fit the data with Generalized Pareto Distribution (GPD). Afterwards, estimate Generalized Pareto Distribution (GPD) parameter. Then, calculate Operational Value-at-Risk (OpVaR Portfolio) as maximum expected claim measurement. The results from this research are the expected maximum claim of IDR. 18.690.352.676,615 for next one year with 95% confidence level. The results of this maximum claim estimate are very useful as a supply chain of information in the consideration of determining the house fire insurance premium for Bandung City residents.**

***Keywords*—** *Portfolio approach, expected claim maximum, Operational Value-at-Risk (OpVaR), house fire risk.*

1. Introduction

Extreme events are events that have a rare frequency of occurrence but have a large or serious impact on the object that experiences them [1]. One of the extreme problems is the house fire. The risk of house fires increase in densely populated areas these extreme risks become more worrying for the community. One of the big cities that have this risk is the City of Bandung. The risk that causes anxiety is finally led to home fire insurance products. A property insurance such as home is one of the solutions offered by insurance companies guaranteed by insurance premiums. Insurance companies calculate a premium based on an analysis of potential claims calculation, fees, commissions, and margins.

Potential claims are an important part in forming an insurance product but in the insurance industry itself there are many obstacles in estimating potential claims. To get an estimate of a potential claim it is necessary to identify the factors that influence the size and probability of the claim [2]. One factor that indicates the size of a potential claim is the value of the risk guaranteed in the claim. In house fire insurance product, the risk form is the loss value of the house in fire incident and the risk value is one of the factors to measure the potential claims.

Several previous studies have discussed the potential claims of a risk and the Extreme Value Theory method. Gourier et. al has modeled data that has big data tails using Extreme Value Theory and introduced Copula theory that showed the Value-at-Risk is a measure of risk that occurs. The results of this study indicate that the possibility of diversification is not appropriate when the distribution is mean-infinite [3]. Baran and Witzany conducted a study comparing Extreme Value Theory with standardized estimation methods (variance, covariance, historical simulation) to produce Value-at-Risk [4]. This different Value-at-Risk search method was compared with back testing procedures and give rise to volatility returns that vary in the period.

Gilli and Kellezi applied Extreme Value Theory to measure risk. Extreme Value Theory was considered to provide the basis for extreme statistical modeling, which many fields of modern science and engineering must deal with rare events that have significant consequences or can be called extreme events. The research aimed to explain the basics of Extreme Value Theory and tactical aspects in estimating and assessing statistical models for measuring risk of an extreme event [5]. Based on this description, the problem in this study was how to use the Extreme Value Theory method in estimating the maximum potential claim of the house fire risk in the city of Bandung for the next one year. Estimates of the potential claims are expected to be a supply chain of information for consideration in the manufacture of house fire insurance products.

2. Material and Method

2.1 Object of research

The object in this study used data on losses from home fire incidents that occurred in the city of Bandung within 12 years period, from 2007 to 2018. The loss data were obtained from the Bandung City Fire and Disaster Management Agency. The loss data taken was the value of losses and the causes of home fire incidents that occurred in Bandung.

2.2 Maximum Entropy Bootstrapping (ME Boot)

Bootstrap was first introduced as a resample data method by [6]. Then in [7], a journal entitled Maximum Entropy Bootstrap for Time Series in 2009 has developed a bootstrapping based on the principle of maximum entropy and commonly referred to as (ME Boot). ME Boot is essentially a method for deriving strong estimates of standard errors and confidence intervals to estimate proportions, averages, medians, odds ratios, correlation coefficients or regression coefficients.

ME Boot conducted in this study was assisted with R software, Package ME boot to make it easier to manage the data needed. The command functions used were me boot (x, reps, trim = list (trim = 0.10,), reachbnd = TRUE, expand.sd = FALSE, force.clt = TRUE, scl.adjustment = TRUE).

2.3 Selection of threshold

The threshold value is the initial value on the tail distribution that meets the extreme value distribution. Choosing a threshold value basically seeks an optimal balance in order to obtain model errors and parameter errors to a minimum. One method for determining threshold values is the percentage method. Determining of the threshold value using the percentage method is more practical and easier to apply.

The threshold value selection method in this study was the percentage method, due to the practical reasons mentioned above. Based on extensive simulation studies, Chavez-Demoul in recommended choosing thresholds such that the data above the threshold is approximately 10% of the total data [8].

Total of extreme data obtained through equation

$$m = 10\% \times n \quad (1)$$

m is the number of data above threshold or extreme data and n is total observation data. The threshold value is u using the following equation.

$$u = m + 1 \quad (2)$$

2.4 Extreme value identification

Identification of extreme values from data of loss can be conducted by two methods. The first is the block maxima method, which is the traditional method used to analyse seasonal data. Each block of the period was determined the maximum loss. Second, the method of Peak's Over Threshold (POT) used data more efficiently by identifying extreme values which are above a value or the maximum loss or certain threshold value [1]-[9]. This study used the Peak’s Over Threshold (POT) method in determining extreme values.

2.4.1 Peaks over threshold (POT)

A peak over threshold (POT) identifies extreme values by setting certain threshold values and ignoring the time of occurrence. Extreme values are data that are above the threshold value. Later this extreme value will be modelled the distribution. The POT method is applied the Pickland Dalkema-DeHann theorem which states that the higher the threshold, the distribution for data above the threshold will follow the generalized pareto distribution (GPD) [10]-[11]. Assuming the data was above the threshold will follow the GPD, it was obtained by looking at the tail distribution of the data away from the close line. Large tail data distribution or heavy tail discovered by making QQ-Plot against the data above the threshold.

2.4.2 Generalized Pareto Distribution (GPD)

Generalized Pareto Distribution (GPD) is defined as the distribution limit of scaled excesses above the threshold value. For example, X is a random variable from daily loss with 2 GPD parameters, the GPD distribution function of X is as follows [9]-[10].

$$g_{\xi,\beta}(x) = \begin{cases} \frac{1}{\beta}\left(1+\frac{\xi}{\beta}x\right)^{-1-\frac{1}{\xi}}, \xi \neq 0 \\ \\ \frac{1}{\beta}\exp\left(-\frac{x}{\xi}\right) \quad , \xi \neq 0 \end{cases} \tag{3}$$

which $\beta > 0$; $x \geq 0$ if $\xi > 0$; $0 \leq x \leq -\frac{\beta}{\xi}$ if $\xi < 0$ with ξ: shape parameter) and β: scale parameter.

There are three types of distribution in Generalized Pareto Distribution (GPD). Distribution of GPD that can be differentiated into three types based on the value of the shape parameter (ξ) which is the exponential distribution when the value ξ = 0, the distribution of Pareto type I when the value of ξ > 0 and the pareto distribution type II if the value of ξ < 0 [9].

2.4.3 GPD distribution suitability test

Distribution testing can be done using the Kolmogorov-Smirnov test. This test was conducted by adjusting the sample distribution function (empirical) with certain theoretical distributions. According to Frank [12] to get the conclusions then comparing D_{count} with $D_{1-\alpha}$ on the Kolmogorov-Smirnov table with significance level (α). Reject H0 if $D_{count} > D_{1-\alpha}$.

In this study the process of testing the GPD distribution suitability on extreme data taken above the threshold value was carried out with Easy Fit software. The package or command in used was the Goodness of Fit Tool which took the results of the distribution suitability test with the Kolmogorov-Smirnov test.

2.4.4 Estimating GPD Parameters

Davidson and Smith have discussed the Maximum Likelihood Estimation for estimating GPD parameters [13]-[14]. The parameter estimation formula is obtained by the Maximum Likelihood Estimation (MLE) method as follows:
Shape parameter:

$$\hat{\xi} = \frac{n^2 s - \sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i - n\sum_{i=1}^{n} x_i} \tag{4}$$

With ξ: shape parameter, n: the number of extreme data, s: standard deviation of extreme data, and x_i: extreme data –i.
Scale parameter:

$$\hat{\beta} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{5}$$

with β: scale parameter, n: the number of extreme data, and X_i: extreme data –i

2.5 Application of Portfolio Approach in Operational Value-at-Risk (OpVaR)

Extreme data based on the EVT method is the basis for the application of the maximum potential claim of an operational risk of house fire. Operational Value-at-Risk (OpVaR) is a method to measure the loss caused by operational risk with a certain confidence level [1] - [15] The OpVaR used in this study illustrates the operational risk of claims that may occur. The magnitude of the operational risk sought is VaR with p% quantile from the distribution of the total loss value. OpVaR in this study uses a 95% confidence level.

OpVaR for each risk can be searched based on the threshold value of extreme data and the estimated value of extreme data distribution parameters so that the value of OpVaRs can be found using the formula [1]:

$$\text{OpVaR} = u + \frac{\beta}{\xi}\left\{\left[\frac{n}{m}(1-p)\right]^{-\xi} - 1\right\} \tag{6}$$

which u : Threshold, β : scale parameter, ξ : Shape parameter, n : total number of observational data, m : number of data above the threshold, p : Level of confidence.

In the world of insurance, the portfolio tangent theory of the risk in an insurance product is a combination of the risks borne into becoming a product with certain amount of premium. Insurance Portfolio also can be used to minimize the risk of claims that occurred. OpVaR of the portfolio is expected to be much smaller than the sum OpVaR of each risk. Therefore, making it possible to reduce the value of premiums of the insurance products. The equation models in used are almost similar with the Markowitz portfolio. However, the rate of return that is expected (expected return) is replaced with the level of claims that are expected (expected claim). Portfolio calculation in this study was a portfolio calculation of the risks that occurs. While the portfolio calculation stages were:

a. Calculate the weight of each risk

$$\text{Frequency} = \frac{\text{OpVaR}}{\text{Claim Value}} \tag{7}$$

$$W_i = \frac{f_i}{\sum_{i=1}^{n} f_i} \tag{8}$$

that W_i: : weight of risk - i , f_i: : OpVaR frequency– i

b. OpVaR Portfolio

After getting the weight W_i then OpVaR portfolio obtained using the calculation as follow:

$$\text{OpVaR Portofolio} = \sum \text{OpVa}\,R_i \times W_i \tag{9}$$

2.6 Research Stages

This research was carried out in several steps, as follows: 1) Resample data with ME Boot assisted by R software in accordance with the available packages; 2) Perform extreme data collection with equation (1); 3) Determine the threshold value with equation (2); 4) Testing extreme data with the Kolmogorov-Smirnov test against GPD assisted by Easyfit software; 5) Calculate the estimated GPD parameters with equation (4) for the shape parameters and equation (5) for the scale parameters; 6) Calculate the value of the OpVaR Portfolio as a maximum potential claim size using equation (6).

3. Results and Discussion

3.1 Data characteristics

The data in this study were the value of losses from house fires in the city of Bandung in 2007-2018 based on the risk factors; Stove and gas, electric current, and other combustible objects. The data was briefly presented in Figure 1, Figure 2, and Figure 3.

Data on house fire losses for each risk showed characteristics that were not in accordance with the assumptions needed to identify extreme values with Peaks Over Threshold. Therefore, the next step was to process the resample data with maximum entropy bootstrapping (ME Boot).

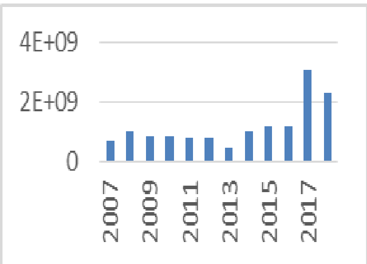

Figure 1. Data of house fire losses due to stove and gas risks.

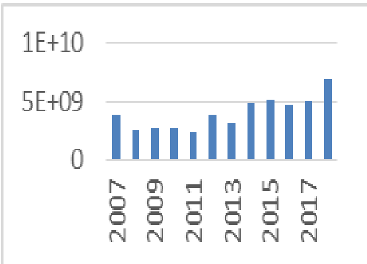

Figure 2. Data of house fire losses due to electricity risks

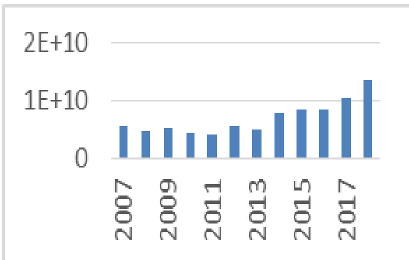

Figure 3. Data of house fire losses due to other combustible objects

3.2 Processing Maximum Entropy bootstrapping (ME Boot) from the Data Losses

Loss data was processed by ME Boot assisted by Software R. Then, the threshold was taken with a 10 percentage and a lot of extreme data above the threshold value. The summary results of the processed data are in Table 1. The ME Boot summary results in Table 1 showed the total fire loss data was 1200 data. The ME Boot data was taken because it was in accordance with the assumption of a large data tail distribution. The distribution of large data tails was based on the results of the QQ-Plot against the results of ME Boot. This QQ-plot was used to see the suitability of the ME Boot data with the nature of its extreme data which had a large data tail as an indication of extreme data with Generalized Pareto Distribution (GPD) distribution. QQ-Plot was done with the help of R software, QQ-Plot Package software. The following QQ-Plot results from ME Boot data could be seen in Figure 4, Figure 5, and Figure 6.

Figure 4, Figure 5, and Figure 6 showed the results that correspond to the desired assumption which was the large tail data or away from near the normal line. This assumption resulted in the interpretation that extreme data will be in accordance with the GPD distribution.

3.3 Kolmogorov-Smirnov Extreme Data Test on GPD

Extreme data that assumed GPD was tested for compatibility with the Kolmogorov-Smirnov test assisted by Easyfit software. The results of the suitability test could be seen in Figure 4.

From the Kolmogorov-Smirnov test results in Figure 7, Figure 8, Figure 9, it could be concluded that the extreme data were in accordance with the GPD distribution because there was no rejection or hypothesis assumption that the GPD data distribution is accepted. Therefore, it could be continued to estimate the parameters.

Table 1. ME Boot Summary Data on House Fire Losses in Bandung City

Event Type	Repeat	Data Losses	Extreme Data	Threshold (IDR.)
Gas stove	90	1080	108	IDR 1.611.991.700
Electricity	100	1200	120	IDR 5.944.737.000
Other Combustible Objects	90	1080	108	IDR 3.391.266.200

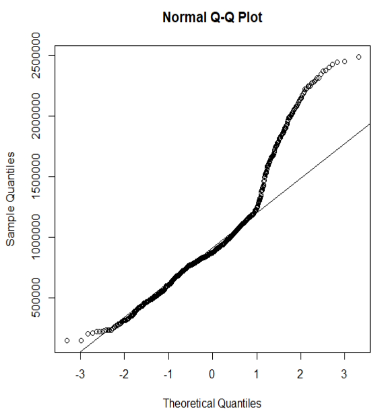

Figure 4. QQ-Plot Data of ME Boot with Stove and Gas Risk Type

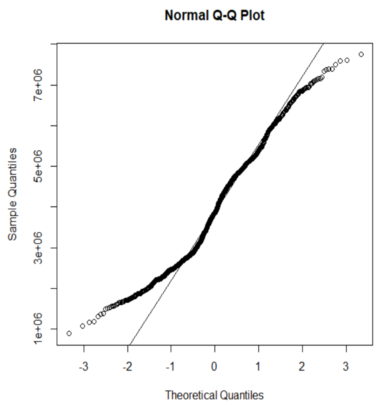

Figure 5. QQ-Plot Data of ME Boot with Electricity risk Type

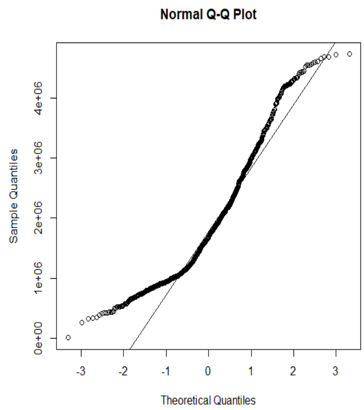

Figure 6. QQ-Plot Data of ME Boot with other type of combustible objects risk

Goodness of Fit - Details [hide]

Gen. Pareto [#24]

Kolmogorov-Smirnov

Sample Size Statistic P-Value Rank	108 0,03452 0,99904 2		
α	0,2	0,1	0,05
Critical Value	0,10325	0,11768	0,13067
Reject?	No	No	No

Figure 7. Kolmogorov-Smirnov Test Results of Extreme Data from stove and Gas risks with GPD in EasyFit Software.

Goodness of Fit - Details [hide]

Gen. Pareto [#24]

Kolmogorov-Smirnov

Sample Size Statistic P-Value Rank	120 0,04239 0,97611 2		
α	0,2	0,1	0,05
Critical Value	0,09795	0,11164	0,12397
Reject?	No	No	No

Figure 8. Kolmogorov-Smirnov Test Results of Extreme Data from the electricity risks with GPD in EasyFit Software.

Goodness of Fit - Details [hide]

Gen. Pareto [#24]

Kolmogorov-Smirnov

Sample Size Statistic P-Value Rank	108 0,05526 0,87809 2		
α	0,2	0,1	0,05
Critical Value	0,10325	0,11768	0,13067
Reject?	No	No	No

Figure 9. Kolmogorov-Smirnov Test Results of Extreme Data from the Other Combustible Objects Risk with GPD in Easyfit Software.

Table 2. Descriptive Statistics of Extreme Data for Every Risk

Descriptive Statistics			
	Gas stove	Electric current	Other objects
Data	108	120	108
Mean	1946237286	6516909073	3983901107
Standard Deviation	236687314,8	437337470,7	385690501
Sample Variance	5,60209E+16	1,91264E+17	1,48757E+17
Kurtosis	-0,814683294	-0,283730295	-1,182345992
Skewness	0,486812917	0,709237341	0,137464936
Minimum	1618630100	5949430000	3396458800
Maximum	2484881600	7748068000	4726594000
Sum	210193626900	782029088800	430261319604

3.4 GPD Parameter Estimation

Calculation of shape and scale parameters required standard deviations (s), the number of extreme data (n), and the total of extreme data value $(\sum_{i=1}^{n} x_i)$ obtained from the descriptive statistics of extreme data as follows.

Based on Table 2, *s*; *n* and $\sum_{i=1}^{n} x_i$ for every extreme data. Furthermore, the shape and scale parameters were found by equations (4) and (5). The results of the two GPD parameters showed the distribution function if the value $\xi < 0$ then x that satisfied the distribution was $0 \leq x \leq -\frac{\beta}{\xi}$ which β is scale parameter For example for Stove and Gas risks, the upper limit x was

$$\frac{\beta}{\xi} = -\frac{1946237286}{-0{,}113403547} = 17162049436{,}6654.$$

The value was in accordance with the extreme data in Table 2 where the largest (maximum) data did not exceed the upper limit. Other parameters of risk were also in accordance with the extreme data limit. After the two GPD parameters have been obtained and in accordance with the extreme data used in GPD. The OpVaR could be calculated for each risk and followed by weighting to get OpVaR Portfolios.

Table 3. Parameter Estimation Results for Extreme Data for Each Risk

	Gas stove	Electricity	Other combustible objects
$\hat{\xi}$	-0,113403547	-0,059268685	-0,088371261
$\hat{\beta}$	1946237286	6516909073	3983901107

Table 4. OpVaR Results for Extreme Data for Each Risk

	Gas and Stove	Electricity	Other Combustible Objects
$p = 95\%$	2909362498,78036	10370384671,5821	6069822039,22743

Table 5. Weighting results and OpVaR Portfolios with a 95% confidence level

Risk Type	OpVaR (p = 95%)	Frequency	Wi	OpVaR portfolio
Stove	5556045588	740,8060783	0,172407828	957905750,36527
Electricity	19971581252	1997,158125	0,464798689	9282764781,04042
Combustible Objects	11691447885	1558,859718	0,362793483	4241581104,18332
	Total	4296,823922	1	**14482251635,58901**

3.5 Estimating Potential Claim with Operational Value-at-Risk (OpVaR) Portfolios

After the two GPD parameters have been obtained, then the Operational Value-at-Risk (OpVar) calculation was performed as an estimated value of the potential claims of the risk of house fires. This OpVaR calculation was obtained using a confidence level of 95%. The OpVaR calculation used equation (6), so the following results were obtained.

The OpVaR results for each risk in Table 4 illustrated the expected claim value of each risk at each level of confidence. The OpVaR electricity risk with a 95% confidence level of IDR.10,370,384,671.58 was that we believed 95% that the expected amount of claims originating from the risk of electric current in the next year was IDR.10,370,384,671.58. Likewise, for the understanding of OpVaR with other risks which described the expected claims of each risk in the next year with a confidence level of 95%.

The OpVaR value of each risk in Table 4 must been given weight to carry out the risk pool. Giving weights will produce a portfolio OpVaR which will be the basis for calculating premiums as the maximum potential claim from the portfolio approach. The results of the frequency calculation of each risk with equation 7, then the weight of the OpVaR for each risk with equation 8, and continued with the calculation of the portfolio OpVaR with equation 9, were summarized in Table 5.

OpVaR portfolio is a potential fire disaster claim from a combination of three risks (stove and gas, electricity, and other combustible objects) with a confidence level of 95% resulting in a value of IDR. 7,517,216,152.51. which meant we believed in 95% that the magnitude of the expected claim originating from the three risks for the next year was Rp7,517,216,152.51.

The maximum claim potential value can be used as an information supply chain to calculate premium, taking into account the maximum claim coverage value that is likely to occur in the next year. Therefore, the premium price is adjusted to be sufficient to cover potential claims. Many events can occur with various probabilities. The impact of these events on supply chain performance is measured to determine supply chain risk. Supply chain management manages the flow of materials and services from upstream to downstream. This definition of material includes not only tangible material, but also intangible material. Therefore, the scope of the meaning of material in supply chain management is very broad. It includes information on the risk of insurance claims, insurance premiums, premium reserves, types of insurance products, insurance policies (benefits), probability values, interest rates, information, and so on, which are very dynamic in the supply chain process in insurance companies.

4. Conclusion

Based on the data processing results, the value of losses from house fires in the city of Bandung using a portfolio approach to produce an estimate of the maximum potential claim price based on the Operational Value-at-Risk (OpVaR) portfolio was Rp7,517,216,152.51 with a 95% confidence level. The maximum claim price estimation could be used for the purposes of establishing a home fire insurance premium so the premium was sufficient to cope with the potential of the claim. Therefore, the insurance product that is made will be suitable for long-term needs by considering the supply chain of claim risk information.

Acknowledgements

Acknowledgments are conveyed to the Director of General of Higher Education of the Republic of Indonesia (Dikti), and Chancellor, Director of the Directorate of Research, Community Engagement, and Innovation (DRPMI), and the Dean of the Faculty of Mathematics and Natural Sciences (FMIPA), Universitas Padjadjaran (Unpad), who have provided the Basic Research Scheme Grant with Contract Number**: 1827/UN6.3.1/LT/2020.**

References

[1] Hubbert, S. *Essential mathematics for market risk management*, New York: John Wiley & Sons, 2011.

[2] Bortoluzzo, A. B., Claro, D. P., Caetano, M. A. L., and Artes, R. “Estimating total claim size in the auto insurance industry: a comparison between tweedie and zero-adjusted inverse gaussian distribution.” *BAR-Brazilian Administration Review*, Vol. 8, No. 1, pp. 37-47, 2011.

[3] Abbate, D., Gourier, E., and Farkas, W. “Operational risk quantification using extreme value theory and copulas: from theory to practice.” *Journal of Operational Risk*, Vol. 3, 2009.

[4] Baran, J., and Witzany, J. “A comparison of EVT and standard VaR estimations.” *Available at SSRN 1768011*, pp. 1-27, 2011.

[5] Gilli, M. “An application of extreme value theory for measuring financial risk.” *Computational Economics*, Vol. 27, No. 2-3, pp. 207-228, 2006.

[6] Efron, B. *The jackknife, the bootstrap, and other resampling plans*, New York: SIAM, 1982.

[7] Vinod, H. D., and López-de-Lacalle, J. “Maximum entropy bootstrap for time series: the meboot R package.” *Journal of Statistical Software*, Vol.29, No. 5, pp. 1-19, 2009.

[8] Chavez-Demoulin, V. *Two problems in environmental statistics: Capture-recapture analysis and smooth extremal models*, PhD thesis, EPFL, Lausanne, Switzerland, 1999.

[9] V. Chavez-Demoulin, A. Davison, and A. McNeil, “A point process approach to value-at-risk estimation.” *National Centre of Competence in Research Financial Valuation and Risk Management*, vol. 134, pp. 1-30, 2003.

[10] Kang, S., and Song, J. “Parameter and quantile estimation for the generalized Pareto distribution in peaks over threshold framework.” *Journal of the Korean Statistical Society*, Vol. 46, No. 4, pp. 487-501, 2017.

[11] Degen, M., Embrechts, P., and Lambrigger, D. D. “The quantitative modeling of operational risk: between g-and-h and EVT.” *ASTIN Bulletin: The Journal of the IAA*, Vol. 37, No. 2, pp. 265-291, 2007.

[12] Massey Jr, F. J. “The Kolmogorov-Smirnov test for goodness of fit.” *Journal of the American statistical Association*, Vol. 46, No. 253, pp. 68-78, 1951.

[13] Davison, A. C. “Modelling excesses over high thresholds, with an application.” In *Statistical extremes and applications,* pp. 461-482, 1984. Springer, Dordrecht.

[14] Smith, R. L. “Maximum likelihood estimation in a class of nonregular cases.” *Biometrika*, Vol. 72, No. 1, pp. 67-90, 1985.

[15] Esterhuysen, P. Styger, and G. V. Vuuren, “Calculationg Operational Value-at-Risk in a Retail Bank.” *South Adrican Journal of Economic and Management Sciences*, Vol. 11, No. 1, pp. 1-16, 2008.