Modeling Correlations in Operational Risk

M. Karwański* and U. Grzybowska

Department of Informatics, Faculty of Applied Informatics and Mathematics, Warsaw University of Life Science,
Nowoursynowska 159, PL-02776 Warszawa, Poland

The key demand for banks' economic capital methodology is to ensure that the model covers all relevant sources of risk in the right way. Operational risk models treat the arising losses as stochastic variables. One of the problems encountered in modeling is the need of taking into account correlations between events. It is possible to build models for correlated events based on copula functions. But the problem is that the losses are related to isolated events and simple applications of copulas are not allowed. The authors present a new algorithm that shows a modified application of copulas to calculating operational risk. The calculations were done on real data that allows for examining the correlation impact on risk measurement. As an additional evaluation of the algorithm a reference model based on the Pareto–Lévy copulas was used.

1. Introduction

Operational risk (OR) models are used to examine losses that arise as a result of operational events that can be treated as stochastic variables. Modeling operational risk is a difficult task and a very important one due to a high magnitude of losses it concerns [1]. According to the Basel II/III definition, as it appears in the final document (Basel Committee on Banking Supervision 2016 [2]), operational risk is defined as "the risk of loss resulting from inadequate or failed internal processes, people and systems or from external events". Banks are allowed to develop their own empirical strategies called advanced measurement approach (AMA) to measure this risk. A common industry approach in AMA is the loss distribution approach (LDA) [3–6]. Under this approach, bank estimates two probability distributions: one of the severity (single event impact) and the other of one period event frequency, using its internal data. With these two distributions, the probability distribution of the aggregate operational loss is computed.

Operational risk modeling involves various problems. One of them is the issue of scarce of data as events included in this risk are not frequent. Therefore estimation of OR in a bank is supported by external data. In our research we have used both data sourced from various segments (cells) of a bank and external data. As the segments are often related the losses are dependent. Therefore we face another problem how to deal with dependences between various segments or cells of the same bank. Dependences between the loss events occur in frequency and size. It has been strongly recommended in Regulation (EU) (Article 322(2)(d)) [7] to introduce dependences into the OR models. In the paper we show a method of introducing correlations into OR model in such a way that they would capture all the above mentioned problems.

Operational risk modeling is under permanent development. Various, previously recognized methods are no longer permitted or regarded proper. It is also reflected by continuously changing external regulations. OR modeling is subject to restrictive directives. According to the recent approach some types of copulas can be applied but for a few years now banks have not been permitted neither to use normal copulas in dependences modeling nor to apply correlations between frequencies in various segments of the bank. Correlations can be considered exclusively for the severity of events [8]. A thorough discussion of different approaches in OR modeling, starting from the easiest, not permitted any more, and ending with advanced analytical approaches can be found in [9–11].

In our paper we have investigated the problem of introducing correlations into the aggregate loss model and we propose a new algorithm to solve that problem. We also demonstrate its performance on real data. In our approach we follow the pattern presented in [12]. The method presented there is however based on identifying risks that should be paired to calculate correlations. In our algorithm we omit that difficulty. The algorithm is a two-stage based on scenario aggregation algorithm. We compare its performance with approach based on the Pareto–Lévy copulas [3]. This model is in our belief the only one analytically tackleable approach that can be used as a reference or a benchmark.

The paper is organized as follows. Section 2 discusses briefly the problem of aggregate loss in operational risk. A good literature overview can be found in [9, 10], therefore we refer only to selected papers which we regard vital for our research. We also provide overview of necessary regulations and some basic facts for copula approach in dependences modeling. The steps of our algorithm are presented in Sect. 3. Short description of data used in the analysis and methodology application are presented

---

*corresponding author

in Sect. 4 while the results of our research are shown in Sect. 5. Concluding remarks are presented in Sect. 6.

2. Operational risk and methodology of its estimation

Basel II distinguishes eight business lines and seven risk types which makes a total of 56 operational risk cells. In practice, however, that division does not have to be strictly applied [13]. In OR modeling framework, the total aggregate loss, L, is computed as

$$L = Loss_{total} = \sum_{j \in J} Loss_j = \sum_{j \in J} \sum_{i \in N_j} X_{ij}, \qquad (1)$$

where J denotes the number of bank segments distinguished according to business activities, e.g., type of business line or event type and N_j refers to the loss associated with the cell j, which is simply given by the sum of the severities of the individual losses X_{ij} with $i = 1, ..., N_j$. Aggregate losses result from two distinct sources of randomness (i.e., frequency and severity) which both have to be modeled. In essence the LDA model as used in operational risk or in actuarial sciences assumes the following three properties within each segment of risk:

- N_j and $(X_{j1}; X_{j2}; ...)$ are independent random variables,

- $X_{j1}; X_{j2}, ...;$ is a set of independent random variables,

- $X_{j1}; X_{j2}; ...$ follow the same marginal distribution.

There is a number of ways how to introduce dependences in (1). Dependences among frequencies and severities can exist within or between cells. The latter is an important focus of research on operational risk [1]. The reason is that the standard LDA assumes that for a given cell j, quantities $X_{j1}, X_{j2}, ..., X_{jN}$, and Nj are independent and that the X_{ji} are independent and identically distributed random variables.

2.1. Directives of Basel Committee on Bank Supervision

European Union and European Banking Authority provide precise directives on operational risk modeling (see regulation (EU) No. 575/2013, Article 321, 322(2)d) [7], or [14]. According to the definition of Basel Committee on Bank Supervision the regulatory capital requirement (or capital-at-risk) can be computed within the AMA methodology as an economic capital equal to the sum of expected loss (EL) and unexpected loss (UL) for a one year holding period and a 99.9% confidence interval. This definition implies that frequency distribution must be understood on a yearly basis. In the spirit of a value-at-risk-like measure, the economic capital requirement K is the 99.9% percentile of distribution F of the aggregate loss

$$K = \text{value at risk} = F^{-1}(99.9\%)$$

meaning that one expects to incur a loss higher than K ("in average") once every 1000 years.

According to Basel II/III it is first suggested to sum the J capital charges separately in each cell, i.e.

$$L = Loss_{total} = \sum_{j \in J} Loss_j. \qquad (2)$$

As it is now well known this assumes implicitly that aggregate losses L_j are perfectly correlated.

Assuming perfect correlation involves adding together the stand alone risk capital amounts. It ignores potential diversification benefits and produces an upper bound for the economic capital. Mathematically this is equivalent to assuming a perfect dependence between risks, e.g., 100% correlation. According to [14] Article 222 p. 37: "Risk measurements for different operational risk estimates must be added for purposes of calculating the regulatory minimum capital requirement. However, bank may be permitted to use internally determined correlations in operational risk losses across individual operational risk estimates, provided it can demonstrate to the satisfaction of the national supervisor that its systems for determining correlations are sound, implemented with integrity, and take into account the uncertainty surrounding any such correlation estimates (particularly in periods of stress). Bank must validate its correlation assumptions using appropriate quantitative and qualitative techniques". This involves modeling dependences e.g., by copulas and gives a possible deviation from a 100% correlation principle.

Let two random variables L_1 and L_2 be given with distributions F_1 and F_2, respectively. Economic capital requirements associated with these variables at the confidence level α are $Cap_i(\alpha) = F_i^{-1}(\alpha)$. The result of capital aggregation can be computed as $L = L_1 + L_2$. Let F be the distribution of L. Then

$$F(x) = \int_{x_1 + x_2 < x} \mathrm{d}F(x_1, x_2) =$$
$$\int_{x_1 + x_2 < x} \mathrm{d}Copula(F_1(x_1), F_2(x_2)). \qquad (3)$$

EU regulation [8] grants significant flexibility in building the operational risk measurement system for calculating the economic capital. The methodology should adequately take into consideration the correlations structure and source of randomness of risk drivers. The aggregation principle (2) which consists in summing up all the individual capital charges, corresponds to the case where the dependence between the random losses is perfect (or in other words the random losses are perfectly correlated). If the dependences are not perfect one can split the problem into two parts:

- the first part which describes the individual behavior of each risk in isolation, i.e. the stand-alone marginal risk distribution and

- the second part (which itself is a distribution function) describing the dependence structure between the risks. The second part is where the copula comes in.

2.2. Copulas and their application

Copulas are the basic tool for modeling dependences in various multivariate risk estimations. They allow for description of more complex dependences between variables than the Pearson correlation and can be implemented in the Monte Carlo framework used in LDA [15]. Therefore copulas have been frequently used to model dependences in operational risk [3, 5, 16, 17].

An n-dimensional copula is a multivariate join distribution on $[0,1]^n$ such that each marginal distribution is uniform $U(0,1)$. We have the following:
$C : [0,1]^n \to [0,1]$ is a copula if:

- $C(u) = 0$ when u has at least one 0 component;

- $C(u) = u_i$ when $u = (1, \ldots, u_i, 1, 1)$;

- $C(u)$ is n-increasing.

The theorem of Sklar [18] is a fundamental tool in application of copulas. We propose a new model to include correlations into aggregate loss distribution and we will use T-copulas to model dependences. At the moment, T-copula is the most commonly used copula in operational risk modeling [1, 16]. T-copula is a d-dimensional Student distribution with linear correlation matrix R. Its feature is a tail dependence meaning that one extreme event or series of events triggers risks which are normally assumed to have low correlation. T-copula is symmetrical and its left and right dependences are equal. Therefore T-copula is not a perfect solution but it is quite well suited to be a multidimensional risk measure. It is relatively easy to understand, mathematically tractable and can be programmed easily to generate simulated output within an economic capital model. The other Archimedean copulas used in practice require hierarchical approach and are difficult to extend to multivariate case. On the other hand, Gaussian copulas cannot be applied because of zero tail dependences. Therefore we have chosen T-copulas to calculate VaR and we compare this results with results obtained with the Pareto-Lévy copulas [3, 5].

3. Algorithm

EU regulation [8] states that institutions should recognize correlations in operational risk losses across severity of individual OR estimates. The correlations in frequency domain lead to incorrect results. Risk drivers cause tail events in operational risk generally in form of isolated, single incidents. Therefore it is not clear how to measure correlations between them. Copula models used in VaR calculations need some modifications. We propose below a new algorithm for constructing dependences based on copulas adapted to the operational risk environment. We have based OR calculations on our algorithm of VaR estimation. Due to different frequencies of losses in considered segments one cannot apply copu-

las directly to simulate severity of a single loss. We have used a two-stage algorithm in which copulas are applied after the aggregates are estimated. The steps of algorithm are presented for seven bank segments $i = 1, ..., 7$, not for separate cells as suggested in [13]. In the first step we build a system for simulating scenarios and their frequencies and in the second step we simulate sets of events that constitute a basis for calculating aggregates. The steps of our algorithm are:

Step 1: Generating independent uniformly distributed random variables for $i = 1, \ldots, 7$

$$Generator_i U(0,1) \rightsquigarrow \{x_i\}_i \in (X_i),$$

$$Generator_i U(0,1) \rightsquigarrow \{n_i\}_i \in (N_i),$$

where x_i, $i = 1, ..., 7$ are uncorrelated.

Step 2: Calculating scenarios of annual aggregates independently for each segment $i = 1, ..., 7$ as convolutions of severity and frequency distributions

$$F_i^{-1}(x_i) * N_i^{-1}(n_i) \rightsquigarrow \{g_i\}_{i \in K}$$

where

- $F_i()$ is a distribution of severity in segment i;

- $N_i()$ is a distribution of frequency in segment i;

- g_i — a sample from the aggregate for segment i for the years $1, \ldots, k$.

Step 3: Generating independent uniformly distributed random variables $i = 1, 2, \ldots, k$

$$Copula_{T(Corr,100)} \rightsquigarrow \tilde{x}_1, \ldots, \tilde{x}_7,$$

where

$\tilde{x}_i \rightsquigarrow U(0,1)$ are "correlated" through T-copula.

Step 4: Calculating scenarios and annual aggregates for each segment for $i = 1, ..., 7$:

$$\hat{G}_{emp,i}^{-1}(\tilde{x}_i) \rightsquigarrow h_i,$$

where

- $\hat{G}_{emp,i}$ is a distribution of aggregates for segment i, calculated based on empirical sample distribution g_i;

- h_i are correlated annual loss samples.

4. Data and methodology application

Calculations were done based on data sourced from one of the biggest Polish banks. Due to insufficient historical data an additional external source was used. Both the external and internal data was considered in model building. The external data was sourced from the consolidated data base ZORO [†]. The bank was divided into

several segments according to business lines (the problem of that division is beyond the scope of our research and will not be discussed here).

Some risk types were very rare in terms of the number of losses. As a consequence these event types and business lines were aggregated into the so-called segments. The division was approved by supervisory authorities. Finally, 7 segments were used for risk calculation. The internal data sourced from the segments covered the years 2013–2016. In order to obtain the data not sensitive to changes in business environment of the bank the data was transformed through a suitable compensation method using several key risk indicators. The external data has been transformed according to the category of banks from which they were taken and scaled to match the internal data. In order to assure the confidentiality, the calculations presented in the paper were done on modified data. The modifications preserved necessary relationship between estimated parameters. Exemplary loss severity data for some segments (5,7) gathered from two sources is shown in Fig. 1.

For modeling the loss distribution, parametric distribution functions were used. We used weighted hybrids as models of loss distributions. Hybrids consisted of two base distributions: one for the so-called body and other distribution for the so-called tail. As a cut-point dividing the distributions included in the hybrids, the lower threshold of losses registered in the ZORO database was chosen. This choice guaranteed that the tail of the distribution was not estimated only on the basis of external data. A separate problem was the lower threshold for internal data registration. Hybrid models are necessary because losses registered in ZORO database have lower limit (threshold), different from the lower limit of internal data and the sticky point of the data leads to an additional requirement imposed on the smoothness of the distribution function [19]. As base models for severity, the following distributions were taken into account: log-normal, logistic, log–log, Weibull, exponential, Pareto, generalized Pareto, Burr, inverse Burr, inverse Gauss, Frechet, Gumpel, gamma. For frequency modeling Poisson and negative binomial distributions were taken into account. As allowed by [8] we have used the Akaike information criterion for evaluating the relative performance of loss distributions and selecting the optimal LDA model. Some distributions used in models for segments (5,7) are shown in Fig. 2.

---

†The ZORO (Event System from Operational Risk Area) system was designed as an auxiliary tool and an external database to assist banks in estimating operational risk. It was the answer to Recommendation M [21] and the requirements of Resolution No. 76/2010 of the Polish Financial Supervision Authority No. 76/2010 dated 10 March 2010 (as amended). In the database there was collected information about events resulting from operational risk or on the border of operational and credit risks, whereby the bank has suffered a financial loss or has been exposed to potential losses.
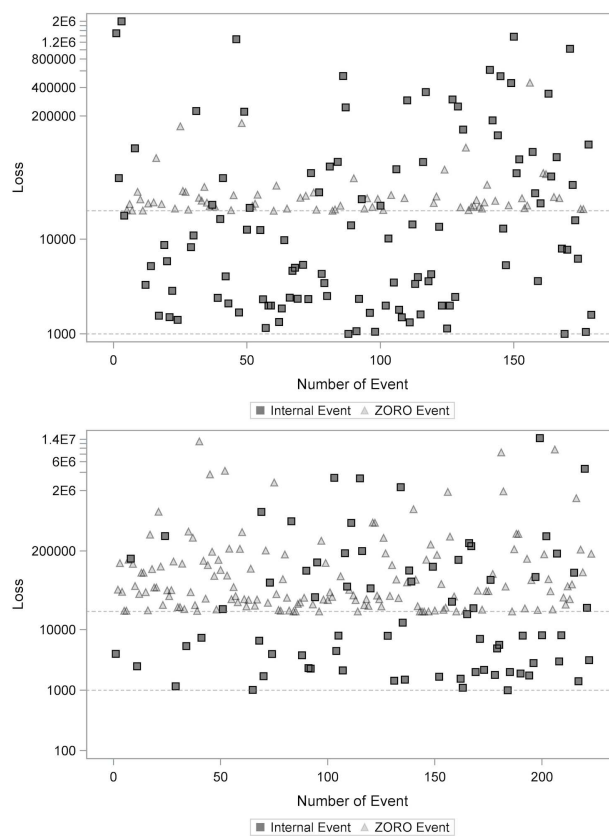

Fig. 1. Raw internal data and appropriate ZORO data for segments 5 and 7.

In each segment parametric distributions for both frequency and severity were estimated. They were the components of the total loss distribution $Loss_{total}$. Next, the value of risk, VaR, was calculated for the following cases:

- Perfectly correlated;

- Correlation modeled by T-copulas;

- The Pareto–Lévy copulas model, which was used as a benchmark [3, 5].

As a benchmark method we have used first-order approximations of OR based on the Pareto–Lévy copulas [3].

Monte Carlo approach was used to estimate the split of two distributions (body and tail). However the MC pseudo simulations are in fact random and the exact results are not known. As recommended by authority the error due to this randomness should be at most 1%. It caused the need to process about 2 million pseudo simulations. According to [19], tail dependence among operational risk categories, thus including the possibility of simultaneous occurrence of large losses of different types, is a critical condition in the calculation process. To estimate T-copula parameters we have used the method of moments. The Kendall tau rank correlation was used to derive estimates of the elements of correlation matrix.
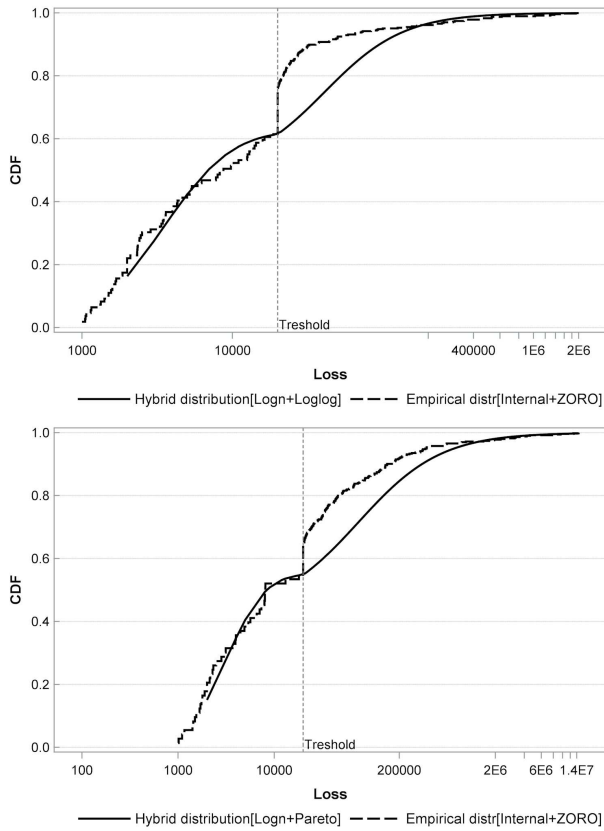
Fig. 2. Empirical and parametric distributions used for segment 5 and 7.

5. Results

We compare VaR calculated for the case of perfect correlation (simple sum), correlation obtained with T-copula and correlation modeled by the Pareto–Lévy copulas.

Numerical results showed that the proposed method involving T-copulas (Table I) gives for $p = 0.999$ reserves about 19.1% less than naive simulation for computing tail loss probabilities and conditional expectations based on simple sum of random risks (Table II). For $p = 0.999$ results obtained by our proposal and benchmark (Table III) differ by about 8.5%. This means that the model presented in the paper gives different but acceptable results. It should be noted that not always lower risk values are desirable from the point of view of the bank. The model based on the Pareto–Lévy copulas (Table III) was suited to compute very high percentiles of loss distribution. Therefore the difference between methods for smaller percentiles shown in Tables I, II, and III is significant.

TABLE I

VaR based on T-copulas.

Average	VaR 50.0%	VaR 75.0%	VaR 99.9%
3,289,140	22,112,365	30,117,185	510,354,897

TABLE II

VaR calculated with perfect correlation.

Average	VaR 50.0%	VaR 75.0%	VaR 99.9%
3,776,454	18,875,581	28,571,145	631,440,840

TABLE III

VaR calculated based on Pareto–Lévy copulas.

Average	VaR 50.0%	VaR 75.0%	VaR 99.9%
7,449,366	204,532	272,709	466,846,501

It has to be added that in [20] information about influence of incorporating correlations on the amount of economic capital across different regions was collected. The summary indicates that the value of 9% is significant. Also practical experience of authors based on cooperation with three Polish banks having regulatory capital of 500 mln PLN, confirms that application of AMA reduced the capital by 20%.

6. Conclusions

In this article we discussed how to take into account the correlations (the diversification effect) for operational risk. The traditional LDA methodology is based on MC techniques applied separately for each segment or cell of the bank. The corrections due to diversification effect are very difficult to include even through the copula approach. In the paper we proposed a statistical methodology that can be applied in practice. It is based on traditional techniques dealing with operational risk analysis using Monte Carlo simulations. It allows to combine Monte Carlo simulation models with copula functions. It is based on a two-stage algorithm that allows to introduce dependences between correlated events, without imposing restrictions on the boundary distributions used to model them. The proposed methodology allows to use copula functions to model severity of isolated events which depend on external factors operating at the aggregate level. This is an important problem in practice. Our numerical results showed that the proposed method is efficient in comparison to a simple summation simulations for computing tail loss probabilities and conditional expectations. We also compared the proposed method with another reference model based on the Pareto–Lévy copulas [3, 5]. On this basis it can be stated that the proposed algorithm gives promising results.

Unlike market or credit risk, operational risk methodology has not worked out effective backtesting tools. Directives concerning this risk type require that experts' opinions should be taken into account in calculating economic capital. The practice confirms that the influence of expert knowledge is similar in value to LDA estimation. The model presented here allows to introduce correlations also in this case.

Banks are interested in incorporating correlations in risk estimation because it makes the calculations more

realistic. As noticed in [21] and [1] considering correlations does not have to lead to a decrease in capital requirement. On the contrary, as shown in [20] incorporating correlations within AMA decreases significantly capital (over 22%) relative to 100% correlation and increases slightly (8.3%) relative to full independence (0% correlation). Our calculations confirm that fact. Application of our algorithm gave results differing by 8.5% from results obtained by the Pareto–Lévy copulas, which indicates results consistency. The results of our further research (not presented here) confirm however that our model is more stable than the approach based on the Pareto–Lévy copulas.

References

[1] S. Mittnik, S. Paterlini, T. Yener, *J. Oper. Risk* **8**, 83 (2013).

[2] Basel Committee on Banking Supervision (2016).

[3] K. Böcker, C. Klüppelberg, *J. Oper. Risk* **24**, 3 (2008).

[4] N. Wagner, T. Wenger, "Integrating Op Risk into Total VaR", in: *Operational Risk toward Basel III. Best Practice and Issues in Modeling, Management and Regulation*, Wiley, New Jersey 2009, p. 131.

[5] K. Böcker, C. Klüppelberg, "First-Order Approximations to operational risk: Dependence and Consequences", in: *Operational Risk toward Basel III. Best Practice and Issues in Modeling, Management and Regulation*, Wiley, New Jersey 2009, p. 219.

[6] A. Frachot, T. Roncalli, E. Salomon, *Operational Risk – Risk's Newsletter* (2004).

[7] Regulation (EU) No 575/2013 EUR-Lex, 2013.

[8] Final Draft Regulatory Technical Standards on the specification of the assessment methodology under which competent authorities permit institutions to use Advanced Measurement Approaches (AMA) for operational risk in accordance with Article 312 of Regulation (EU) No 575/2013, EBA/RTS/2015/02 (2015).

[9] P. Shevchenko, *Operational Risk using Bayesian Inference* Springer-Verlag, Berlin 2011.

[10] G. Cruz, G.W. Peters, P.V. Shevchenko, *Fundamental Aspects of Operational Risk and Insurance Analytics: A Handbook of Operational Risk*, Wiley, 2015.

[11] R. Giacometti, S. Rachev, A. Chernobai, M. Bertocchi, *J. Oper. Risk* **3**, 3 (2008).

[12] V. Chavez-Demoulin, P. Embrechts, J. Neslehova, *J. Bank. Financ.* **30**, 2635 (2006).

[13] F. Aue, M. Kalkbrener, *J. Oper. Risk* **1**, 49 (2006).

[14] Basel Committee on Banking Supervision Consultative Document, operational risk — Supervisory Guidelines for the Advanced Measurement Approaches.

[15] U. Cherubini, E. Luciano, W. Vecchinto, *Copula Methods in Finance*, Wiley, New York 2001.

[16] O. Rachedi, D. Fantazzini, "Multivariate Models for operational risk: A Copula Approach Using Extreme Value Theory and Poisson Shock Model", in: *Operational Risk toward Basel III. Best Practice and Issues in Modeling, Management and Regulation*, Wiley, New Jersey 2009, p. 197.

[17] D. Fantazzini, L.D. Valle, P. Giudici, *IJRAM* **9**, 238 (2008).

[18] A. Sklar, *Publ. Inst. Statist. Univ. Parist* **8**, 229 (1959).

[19] M. Karwański, U. Grzybowska, *The Strategies of Combining Loss Data from Different Sources in Operational Risk*, 2017, unpublished, in review (in Polish).

[20] Basel Committee on Banking Supervision Results from the 2008 Loss Data Collection Exercise for Operational Risk (2009).

[21] *Recommendation M on the management of operational risk at banks*, KNF, 2013 (in Polish).