Data Scaling for Operational Risk Modelling

H.S. Na, L. Couto Miranda, J. van den Berg and M. Leipoldt

ERIM REPORT SERIES *RESEARCH IN MANAGEMENT*	
ERIM Report Series reference number	ERS-2005-092-LIS
Publication	December 2005
Number of pages	19
Persistent paper URL	
Email address corresponding author	jvandenberg@few.eur.nl
Address	Erasmus Research Institute of Management (ERIM) RSM Erasmus University / Erasmus School of Economics Erasmus Universiteit Rotterdam P.O.Box 1738 3000 DR Rotterdam, The Netherlands Phone: + 31 10 408 1182 Fax: + 31 10 408 9640 Email: info@erim.eur.nl Internet: www.erim.eur.nl

Bibliographic data and classifications of all the ERIM reports are also available on the ERIM website:
www.erim.eur.nl

ERASMUS RESEARCH INSTITUTE OF MANAGEMENT

REPORT SERIES
RESEARCH IN MANAGEMENT

ABSTRACT AND KEYWORDS	
Abstract	In 2004, the Basel Committee on Banking Supervision defined Operational Risk (OR) as the risk of loss resulting from inadequate or failed internal processes, people and systems or from external events. After publication of the new capital accord containing this dfinition, statistical properties of OR losses have attracted considerable attention in the financial industry since financial institutions have to quantify their exposures towards OR events. One of the major topics related to loss data is the non-availability of a suficient amount of data within the Financial Institutions. This paper describes a way to circumvent the problem of data availability by proposing a scaling mechanism that enables an organization to put together data originating from several business units, each one having its specific characteristics like size and exposure towards operational risk. The same scaling mechanism can also be used to enable an institution to include external data originating from other institutions into their own exposure calculations. Using both internal data from different business units and publicly available data from other (anonymous) institutions, we show that there is a strong relationship between losses incurred in one business unit respectively institution, and a specific size driver, in this case gross revenue. We study an appropriate scaling power law as a mechanism that explains this relationship. Having properly scaled the data from different business units, we also show how the resulting aggregated data set can be used to calculate the Value-at-OR for each business unit and present the principles of calculating the value of the OR capital charge according the minimal capital requirements of the Basel committee.
Free Keywords	Operational Risk, Power Law Scaling, Loss Distribution, Value at Operational Risk, Minimal Capital Requirements
Availability	The ERIM Report Series is distributed through the following platforms: Academic Repository at Erasmus University (DEAR), DEAR ERIM Series Portal Social Science Research Network (SSRN), SSRN ERIM Series Webpage Research Papers in Economics (REPEC), REPEC ERIM Series Webpage
Classifications	The electronic versions of the papers in the ERIM report Series contain bibliographic metadata by the following classification systems: Library of Congress Classification, (LCC) LCC Webpage Journal of Economic Literature, (JEL), JEL Webpage ACM Computing Classification System CCS Webpage Inspec Classification scheme (ICS), ICS Webpage

Data Scaling for Operational Risk Modelling

H.S. Na, L. Couto Miranda, J. van den Berg, and M. Leipoldt

Erasmus University Rotterdam, Faculty of Economics,

Dept. of Computer Science, room H10-19, P.O. Box 1738, 3000 DR, Rotterdam, the Netherlands

ABN AMRO N.V., Group Risk Management,

Gustav Mahlerlaan 10 (HQ1073), 1082PP, Amsterdam, The Netherlands

ABN AMRO REAL S.A., Brazilian Risk Management,

Av. Paulista 1374, 13 Floor, Sao Paulo, SP, Brazil

e-mail: heryoes@yahoo.com, lourencomiranda@uol.com.br, jvandenberg@few.eur.nl, marc.leipoldt@nl.abnamro.com

Abstract

In 2004, the Basel Committee on Banking Supervision defined Operational Risk (OR) as the risk of loss resulting from inadequate or failed internal processes, people and systems or from external events. After publication of the new capital accord containing this definition, statistical properties of OR losses have attracted considerable attention in the financial industry since financial institutions have to quantify their exposures towards OR events. One of the major topics related to loss data is the non-availability of a sufficient amount of data within the Financial Institutions. This paper describes a way to circumvent the problem of data availability by proposing a scaling mechanism that enables an organization to put together data originating from several business units, each one having its specific characteristics like size and exposure towards operational risk. The same scaling mechanism can also be used to enable an institution to include external data originating from other institutions into their own exposure calculations. Using both internal data from different business units and publicly available data from other (anonymous) institutions, we show that there is a strong relationship between losses incurred in one business unit respectively institution, and a specific size driver, in this case gross revenue. We study an appropriate scaling power law as a mechanism that explains this relationship. Having properly scaled the data from different business units, we also show how the resulting aggregated data set can be used to calculate the Value-at-OR for each business unit and present the principles of calculating the value of the OR capital charge according the minimal capital requirements of the Basel committee[1].

Keywords. Operational Risk, Power Law Scaling, Loss Distribution, Value at Operational Risk, Minimal Capital Requirements.

[1]The content of this report is largely based on the research executed by H.S. Na, the results of which are extensively presented and discussed in Master thesis [1].

1 Introduction

1.1 Operational Risk Modelling and the lack of loss data

In probabilistic terms, OR Modelling is about modelling loss distributions. The loss distribution portrays loss events which are defined as incidents where damages are suffered. The corresponding losses are the currency amounts of the suffered damages. Having fitted the loss distributions, the goal is to use them in order to make inferences about future behavior of losses according to the risk profile of the Bank. For example, a type of Value at Risk calculation can be applied in order to estimate the capital charge for OR. A future risk profile is usually estimated using internally experienced loss information of a group of events (Event Categories). This information may be readily found in the accounts of the Institution or in a specific repository of losses built for that purpose. The latter is the result of the gathering effort made by the various Business Lines (BLs) of the Bank.

Generally speaking however, Financial Institutions do not have enough information regarding the OR losses that occurred under their premises in order to make meaningful forecasting and valid estimations of the minimum acceptable capital for their own operational risk profile. Therefore, a major issue is to find out how external losses collected from a group of other Banks can be incorporated into the internal loss history so that one figure can be set up representing the whole capital due to operational risk for that particular institution. The problem now resides in the fact that not all institutions have the same risk profile. For modelling purposes, it is usually (implicitly) assumed that the group of events is homogeneous enough that one can perform meaningful and reliable statistics with it. In such cases one could, for example, calculate, from its own loss data, a credible and unbiased sample mean or a standard deviation, or some other parameter of a hypothesized loss distribution, and one could hypothesize that this distribution would be a good representative of the behavior of an entire population of losses. In cases we deal with data pooled from several institutes, the above-mentioned assumption generally does not hold and consequently, it is not advisable to incorporate external information into the internal risk profile indiscriminately.

Following the same reasoning used for external data, internal data collected from different BLs might not be homogeneous either. Each individual internal data comes from a specific BL having its internal characteristics like size and control profile by the time the event occurred. Therefore, one may e.g. expect that BLs that are different in size show a greater or lesser aggregated loss per period.

1.2 Scaling of Operational Risk Data

For the above-given reasons, it is not advisable to simply put together data originating from different Business Lines or different Financial Institutions: it may result in statistical flaws, especially for high severity-low frequency events [2]. Instead, certain rigorous statistical treatments are required like correction for 'truncation above a specific threshold' [2] or the application of 'scaling' [3] of relevant variables in order to enable comparison of OR data from different Banks and different BLs, and to achieve comparable standards.

For already quite a long time, physicists have been fascinated by power laws (see, e.g., the references in [3]). The main reason for this is related to the property of 'universality' which means that for many physical phenomena close to the so-called 'critical point', the scaling laws found hold independently of the microscopic details of the phenomenon. Related to this, it has been found that different materials may have the same 'universal' values of critical-point exponents. Power laws are also observed in economic and financial data [4] by, among many others, scientists like Pareto and Mandelbrot [5, 6]. In many cases, power laws in economics are formulated as probability distributions like the famous Pareto probability distribution of wealths (W):

$$\Pr(W) = \frac{W_0^{\mu}}{W^{1+\mu}}, \quad W >> W_0. \tag{1}$$

Here, μ characterizes the decay of the distribution for increasing wealth values W.

In this paper, we are interested in the application of *direct scaling* of variables according to a power law in order to enable comparison of OR data from different banks and different BLs. As suggested by certain researchers [7, 8], suitable scaling variables may be related to bank (BL)-specific factors such as the size (measured by exposure indicators like gross revenue, transaction volumes, number of employees, etc.) and the risk control environment, all taken per Line of Business. Their analysis however was limited to scaling properties of the severity of OR losses. We take these suggestions as a starting point and extend the analysis: we will analyze the distributions of the *aggregate loss* and the *frequency* of OR losses per unit of time, and of the *severity* of OR losses per loss event. The analysis starts by presenting a few theoretical results which can be considered as hypotheses that are assumed to hold. Next we illuminate the way the simulations are set up and finally, we show and discuss the outcomes of certain simulations and statistical tests.

The rest of this paper is structured as follows. In section 2, we present (the assumptions underlying) our modelling approach, discuss some mathematical properties, and introduce a method for calculating the Value-at-Risk of OR. In section 3, we present the way we organized our experiments and in section 4, we present the outcomes of the simulations. In section 5, we include a short discussion. We finish by giving some conclusions and presenting an outlook.

2 Scaling Operational Risk Loss Data

2.1 Modelling Approach

Using the so-called 'Loss Distribution Approach' [9], frequency and severity distributions of OR loss data can be combined in order to produce a loss distribution describing the distribution of the aggregate loss per period [10]. Here, the frequency distribution describes the number of OR loss events per unit of time and the severity distribution yields the distribution of loss amounts per single OR loss event. By running a statistical simulation [10, 11], the distribution of the sum of OR losses per unit of time is found yielding a typical distribution of OR losses as shown in Figure 1. This figure can be used to calculate the Value-at-Risk for OR capital can be calculated. In case of using an Advanced Measurement Approach without having taken specific provisional measures against OR losses [12], this capital charge equals the sum of expected and 'unexpected' loss, the latter being the difference between the VAR at a confidence

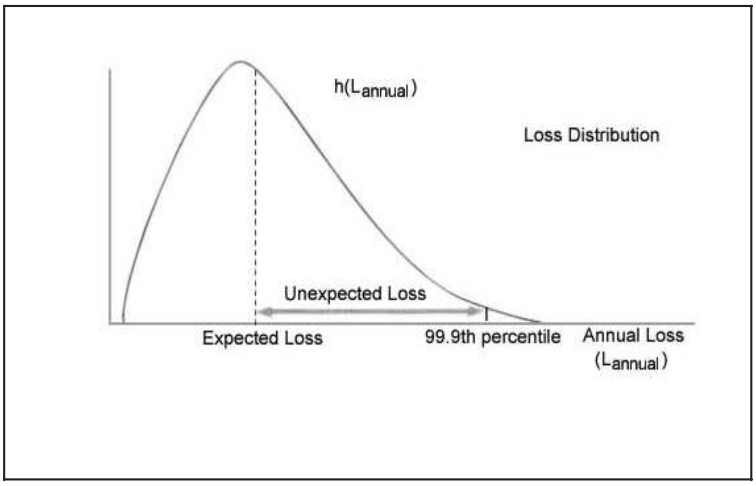

Figure 1: A typical distribution of the aggregate loss [10].

level of 99.9% and the expected loss, everything calculated per year: see again Figure 1.

We now continue by discussing the statistical properties of the aggregate loss (stochastic) variable. A bank is usually divided into various Business Lines (BLs) $b, (b = 1, \ldots, B)$, each one of which can, for modelling purposes, be viewed as an independent financial institution. The stochastic variable L_b describing the aggregate loss of BL b per time unit, can be thought of as being caused by two components, namely, the 'common component' R^{com} and the 'idiosyncratic component' r_b^{idio} [8]:

$$L_b = u(r_b^{idio}, R^{com}). \tag{2}$$

In our approach, the common component R^{com} is assumed to be stochastic and refers to the statistical influence on OR losses caused by general factors such as the macroeconomic, the geopolitical and the culture environment, the general human nature, and more. The idiosyncratic component r_b^{idio} is assumed to be deterministic and refers to OR arising from more specific factors such as size and exposure towards operational risk of BL b. As a consequence of these assumptions, the effect of R^{com} on the probability distribution of L_b is thought to be common to all Lines of Business, while the effect of r_b^{idio} is specific for that BL. This latter effect we shall try to compensate for.

In order to be able to implement a suitable compensation method, we need to make assumptions about the precise effect of the idiosyncratic component on the distribution of the stochastic variable L_b. The first assumption made here is that the total effect of R^{com} and r_b^{idio} can be decomposed according to

$$L_b = u(r_b^{idio}, R^{com}) = g(r_b^{idio}).h(R^{com}), \tag{3}$$

where r_b^{idio} is an indicator of *s*ize of BL b. Below, this indicator will be denoted as s_b and will be represented by the gross income of the particular BL. Next to this, we assume that L_b scales with the gross income per period for a particular BL according to the following power law:

$$L_b = g(r_b^{idio}).h(R^{com}) = (s_b)^{\lambda}.h(R^{com}), \tag{4}$$

where $\lambda > 0$ is a universal exponent, i.e., a number that is equal for all BLs b. Equation (4) expresses that the larger the BL, the larger the loss aggregate suffered. The proportion between the losses of different BLs is given by the scaling factor $(s_b)^\lambda$. Considering several BL's $B = 1, 2, \ldots$, we can rewrite equation (4) as

$$\frac{L_1}{(s_1)^\lambda} = \frac{L_2}{(s_2)^\lambda} = \ldots = h(R_{com}) = L_{st}, \tag{5}$$

showing that $h(R_{com})$ represents the aggregate loss L_{st} per period of the 'standard' BL with size $s_{st} = 1$. We observe that equation (4) can compactly be written as

$$L_b = (s_b)^\lambda . L_{st}, \tag{6}$$

or, equivalently,

$$L_b . (s_b)^{-\lambda} = L_{st}. \tag{7}$$

So, having available a set of OR loss sample data from several BLs and re-scaling them using (7), i.e., multiplying the sample values originating from BL b by $(s_b)^{-\lambda}$, $(b = 1, 2, \ldots)$, we may assume that the complete set of re-scaled data originates from just *one* distribution, namely, that of stochastic variable L_{st}. Vice versa, according to (6) data corresponding to the distribution of L_b are found by multiplying all samples (thought to be) emanated from L_{st} with re-scaling factor $(s_b)^\lambda$.

2.2 Mathematical properties

In this subsection, we discuss several mathematical properties following from the above-given modelling assumptions.

1. First we observe that equations (6) and (7) describe a transformation from one stochastic variable to another. Then, if we know the probability density function (pdf) of one, it is not difficult to calculate the pdf of the other using the 'change of variable technique' [13]: let $f(l_{st})$ be the pdf of L_{st}, then the pdf $f'(l_b)$ of L_b is given by

$$f'(l_b) = f((s_b)^{-\lambda} . l_b) \, |\frac{dl_{st}}{dl_b}| = f((s_b)^{-\lambda} . l_b) * (s_b)^{-\lambda}. \tag{8}$$

   A simple example based on a uniform distribution may clarify the meaning of equation (8). Let $f(l_{st})$ be uniformly distributed according to

$$f(l_{st}) = \begin{cases} 1 & \text{if } l_{st} \in [0, 1] \\ 0 & \text{elsewhere,} \end{cases} \tag{9}$$

   then it follows from equation (8) that $f'(l_b)$ is (also uniformly) distributed conform

$$f'(l_b) = \begin{cases} (s_b)^{-\lambda} & \text{if } l_b \in [0, (s_b)^\lambda] \\ 0 & \text{elsewhere.} \end{cases} \tag{10}$$

   So if, for example, $(s_b)^\lambda = 2$, then after multiplication of all sample values of L_{st} by 2, the density of the uniform distribution is halved and the domain is doubled.

2. Using standard transformation properties of stochastic variables [13], the next scaling formulas are a direct consequence of the assumptions made in subsection 2.1:

$$\mu_{L_b} = (s_b)^\lambda . \mu_{L_{st}}, \tag{11}$$

and

$$\sigma_{L_b} = (s_b)^\lambda . \sigma_{L_{st}}, \tag{12}$$

so, the scaling of the variable values also holds for their means and standard deviations.

3. Equation (11) can be interpreted as a regression line of data pairs (s_b, μ_{L_b}) where for each s_b, the value μ_{L_b} equals the mathematical expectation of L_b with pdf $f'(l_b)$. A linearized equation can be obtained by taking the logarithm yielding

$$\ln(\mu_{L_b}) = \lambda . \ln(s_b) + \ln(\mu_{L_{st}}). \tag{13}$$

Note that in equation (13), the variable $\ln(\mu_{L_b})$ is a function of $\ln(s_b)$ and that both λ and $\ln(\mu_{L_{st}})$ are constant. Equation (13) describes the hypothesis that all data pairs $(\ln(s_b), \ln(\mu_{L_b}))$ lie on one straight line. If for each BL b, a data pair $(\ln(s_b), \ln(\mu_{L_b}))$ is available, this hypothesis can be tested using linear regression.

4. By defining $l_\mu = \ln(\mu_{L_b})$, $s = \ln(s_b)$, and $\imath_\mu = \ln(\mu_{L_{st}})$, equation (13) can shortly be written as

$$l_\mu = \lambda . s + \imath_\mu. \tag{14}$$

Analyzing equation (12) in a similar way as equation (11), we obtain a linearized equation which can shortly be written as

$$l_\sigma = \lambda . s + \imath_\sigma. \tag{15}$$

Having data pairs $(\ln(s_b), \ln(\sigma_{L_b}) = (s, l_\sigma))$ available, we can test whether the linear equation (15) indeed holds by again applying regression. Note that we expect to find equal values for λ in equations (14) and (15).

5. Having found λ, we can re-scale the given data set according to equation (6) or (7). After having performed an appropriate re-scaling, the data can be considered as samples from the distribution of the aggregate loss per time unit of one specific BL and, from this, a corresponding OR loss *capital charge* can be calculated according the minimal capital requirements as fixed by the Basel committee. The OR capital charge calculation is based on a standard Value-at-Risk (VAR) estimation as mentioned in section 2.1.

2.3 Extending the analysis

Sofar, we focussed our presentation on the aggregate loss per unit of time (L_b) but it should be clear that a similar discussion can be set up on with respect to the frequency and severity of OR losses. In addition, like has been explained in the introduction, the same scaling techniques can be used in order to combine internal data with data originating from external sources, i.e., other Banks.

3 Experimental Set-up

In this section we sketch the experiments performed and describe the data sets available.

3.1 Experiments Performed

In order to verify the above-given considerations concerning the scaling of OR losses, a straightforward approach would be (1) to collect OR loss data from all BLs/Banks available, (2) to estimate the pdfs $f'(l_b)$, and (3) to analyze the (statistical) properties of these pdfs in order to test the validity of the given theoretical framework. In practice, a limited set of OR loss data per BL appeared to be available. So, due to the small amounts of data available per BL, move 2 of the above-given scheme, that of direct estimating the full pdfs per BL, could not be made. Therefore, we were forced to limit the statistical analysis during step 3 to the verification of just some basic properties of the pdfs that are supposed to hold.

We decided to focus first on the validity of equations (11) and (12). The reason is simple: an important issue is whether the estimated value of λ in both equations appears to be the same. If so, we would have found evidence that the two most important parameters of the pdfs scale in the same way. This is a *necessary condition* for the validity of the theoretical framework introduced in section 2. The two estimations of the gradient λ (denoted as λ_μ and λ_σ respectively) as well as as the estimations of the intercept values $\imath_\mu$ and $\imath_\sigma$ (denoted as i_μ and i_σ respectively) are found using linear regression. The resulting regression lines can be written as

$$l_\mu = \lambda_\mu . s + i_\mu . \tag{16}$$

and

$$l_\sigma = \lambda_\sigma . s + i_\sigma . \tag{17}$$

In addition, we verified whether the linear relationships as described by equations (16) and (17) do indeed hold by applying statistical testing. The regression simulation is performed using external data, internal data as well as a combination of internal and external data. Next, we performed a Value-at-Risk calculation for the aggregate loss. Having concluded all these investigations, we further performed similar experiments and analysis related to the frequency and severity of OR loss data.

3.2 Available Data Sets

For the variable of size s_b of each Business Line (BL) or Business Unit (BU) b, we use the daily gross income data of the year 2003 which are summed up to the gross income per week by division with the number of weeks in a year. The data concerning the aggregate loss, frequency, and severity of OR losses have been made available by ABN-AMRO.

The internal loss data of the bank are given per Business Unit (like, for example, the BU Brazil and the BU North America) while the external loss data are given per Business Line (like, for example, the BL Corporate Finance and the BL Retail Banking), the latter according to the Basel Lines of Business categorization [11]. For the reason of simplicity we have chosen to use these data directly instead of trying to transfer the internal bank data per BU into data per BL. For each BL/BU, the OR losses per week have been calculated. Next, these losses have been put together in a histogram (see figure 2) showing an approximation of the pdf of the aggregate losses per week for each BL/BU. We observe that there

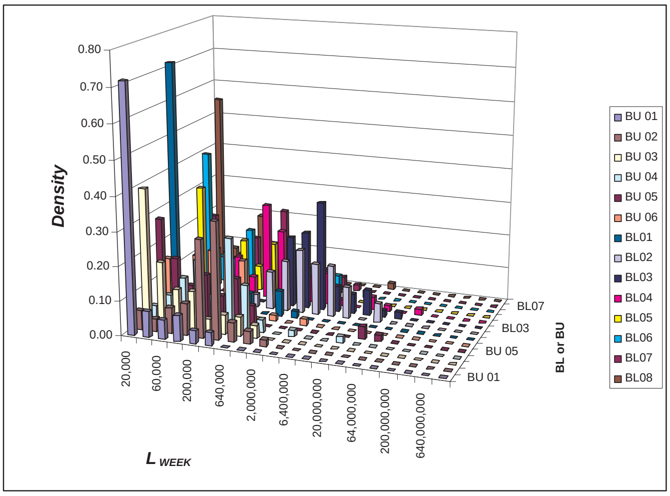

Figure 2: Histograms of aggregate losses per week, for each BL and BU.

exist large differences between these histograms caused by different scales and, more importantly, lack of substantial amounts of data. It was further an easy exercise to calculate, for each BL/BU, the sample mean and standard deviation of (*i*) the sum of OR losses per time unit, (*ii*) the frequency of OR losses per time unit, and (*iii*) the severity per OR loss event.

4 Simulation Results

In this section, we present and analyze the outcomes of simulations concerning the relationship between the aggregate losses across different BLs/BUs and the size of these BLs/BUs according the experimental setup described in section 3. We also present the results of similar studies for the frequency and the severity of OR losses.

4.1 Analyzing the aggregate losses per week

4.1.1 Regression lines based on the means of OR losses

We start by showing the linear regression results based on equation (16) using both the internal data per BU and the external data per BL: see figure 3. The data concern OR losses per week. We observe that the values of the gradient parameter λ_μ of the two regression lines are quite close. The overall impression from this figure is that the regression lines are almost in line with each other supporting the above-mentioned choices made, including those concerning the data sources used. We also combined the internal and external data sets into one set and then found the regression line as visualized in Figure 4. As might be expected, the value of the gradient parameter λ_μ is between the two values found above and turns out to be close to one.

We found several numerical results concerning the three linear regressions mentioned above. The results are presented in Table 1. The first column describes the data source, the second column the value of the squared correlation coefficient, the third the type of regression parameter (the intercept i_μ

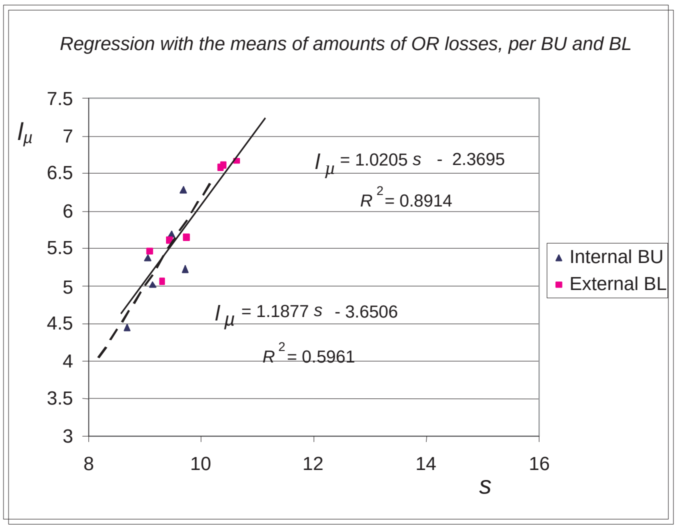

Figure 3: Regression lines according to equation(16), based on internal BU data and external BL data.

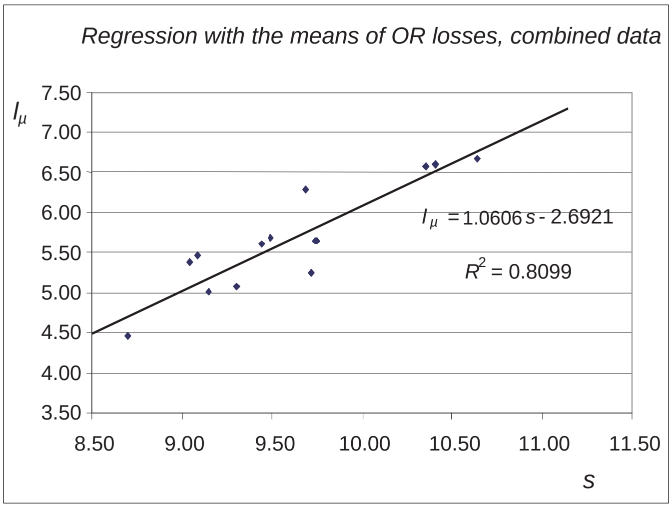

Figure 4: Regression lines according to equation(16), based on a combination of the internal BU data and external BL data.

or gradient λ_μ), the value of which is given in the fourth column. The standard error of the regression parameters (expressing how far they are likely to be from their expected value [14]) is given in the fifth, column. The value of the t-statistic mentioned in the sixth column results from the t-test which is used to examine whether the regression parameter values are significantly different from zero. The corresponding P-value mentioned in the seventh column is related to the level of significance α.

For the external data, the interpretation is as follows: since the P-values in the sixth column are here both less than $\alpha = 0.10$ means that the value of both regression parameters is significantly different from zero at a significance level of 99%. The gradient parameter is even significant at a level of 99.95% because the P-value is (slightly) less than $\alpha = 0.0005$. These results suggest that there exist a power-law between the aggregate losses per week and the size & exposure towards OR per week for the external Bls. We also observe that the idiosyncratic component can explain a big proportion of the variability in

Data set	R^2	Parameter	Value	Standard Error	t-statistic	P-value
External	0.8914	i_μ	-2.3695	1.1828	-2.0033	0.0920
		λ_μ	1.0205	0.1455	7.0163	0.0004
Internal	0.5961	i_μ	-3.6506	3.7060	-0.9851	0.3804
		λ_μ	1.1877	0.4889	2.4294	0.0720
Combined	0.8099	i_μ	-2.6921	1.1718	-2.2975	0.0404
		λ_μ	1.0606	0.1483	7.1511	0.0000

Table 1: Parameter and statistical values of the regression with the means of the aggregate losses (Figures 3 and 4).

the aggregate loss per week as we can see from the high value of R^2.

For the internal data we observe that the value of $R^2 = 0.5961$ is less convincing. However, we can still reject the hypothesis that $\lambda_\mu = 0$ at a level of 90% since the P-value is smaller than $\alpha = 0.10$. These results suggest that there exist a power-law between the aggregate loss per week and the size & exposure towards OR per week for the internal BUs, although the evidence is not as strong as in the case of external data. For the combination of internal and external data, we again observe a quite high value of R^2 (0.8099). The P-value of the intercept i_μ is smaller than 0.05 and indicates a 95% level of significance. The gradient $\lambda_\mu = 1.0606$ is almost equal to 1 and is significant at a level of significance of 99.99% because its P-value is slightly less than 0.0001.

4.1.2 Regression lines based on the standard deviations of OR losses

Next we show the linear regression results of the standard deviations of OR losses based on equation (17) using both the internal data per BU and external data per BL: see Figure 5.

We also combined the data again into one set and then found the regression line as visualized in figure 6. In addition, we found the following statistical results concerning the three linear regressions mentioned above: see Table 2. We observe that the values of R^2 are of similar size but, on average, somewhat smaller than the values of R^2 found in the previous section. For the external data, the P-value of λ_σ (0.0154) is less that 0.05 showing a significance level of 95% for the gradient parameter. For the internal data, the significance level equals 90% (P-value (0.0663) < 0.10). Combining the internal and external data yields the highest value of R^2 (0.6643) and the P-value of the gradient (0.0004) corresponds to a confidence level of 99.95%. We also do the important observation that, in case of using the combination of internal and external data, the values of the gradients $\lambda_\mu = 1.0606$ and $\lambda_\sigma = 1.0167$ are both significant and almost equal. These results again suggest the existence of a power law as hypothesized in section 2.

4.1.3 Constructing the 'standard distribution' of the aggregate losses

Assuming the existence of a power law according to equation (4), the data from all different BUs/BLs can thought having been drawn from the standard distribution of the aggregate losses per week provided

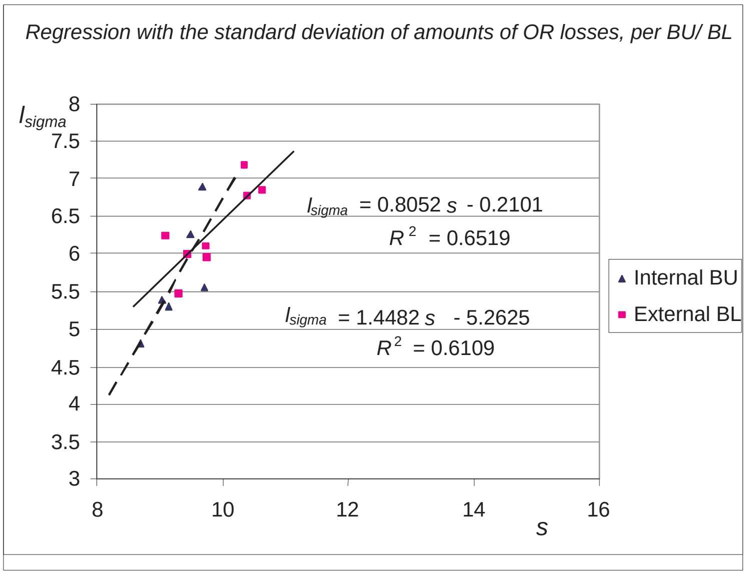

Figure 5: Regression lines according to equation(17), based on internal BU data and external BL data.

Data set	R^2	Parameter	Value	Standard Error	t-Statistic	P-value
External	0.6519	i_σ	-2.101	1.9534	-0.1075	0.9179
		λ_σ	0.8052	0.2402	3.3521	0.0154
Internal	0.6109	i_σ	-5.2625	4.3808	-1.2013	0.2959
		λ_σ	1.4482	0.5779	2.5060	0.0663
Combined	0.6643	i_σ	-1.9553	1.6482	-1.1864	0.2584
		λ_σ	1.0167	0.2086	4.8732	0.0004

Table 2: Parameter and statistical values of the regression with the standard deviation of the aggregate losses (Figures 5 and 6).

we apply re-scaling according to equation (7). After having performed these calculations and having aggregated the data from all BUs and BLs, we found the histogram as shown in Figure 7 which represents an approximation of the true, continuous distribution of stochastic variable L_{st} of OR losses taken per week. We observe that the shape of this histogram resembles the ones described in the literature, an example of which was shown in Figure 1.

4.1.4 Calculating the Value-at-Operational-Risk of a BU/BL

As mentioned in section 2.1, a straightforward procedure can be followed to calculate the Value-at-Risk (VAR) of the aggregate losses per week. Using the aggregated data set as described in the previous section, the aggregate losses per week for the standard BU/BL are first sorted, a visualization of which is shown in Figure 8. Having all together N sample values and choosing a confidence level α, the VAR of the aggregate losses of the standard BU/BL is given by the j-th sorted loss where $j = \alpha.N$. In our case, this VAR has been calculated for three confidence levels. The resulting values are summarized in

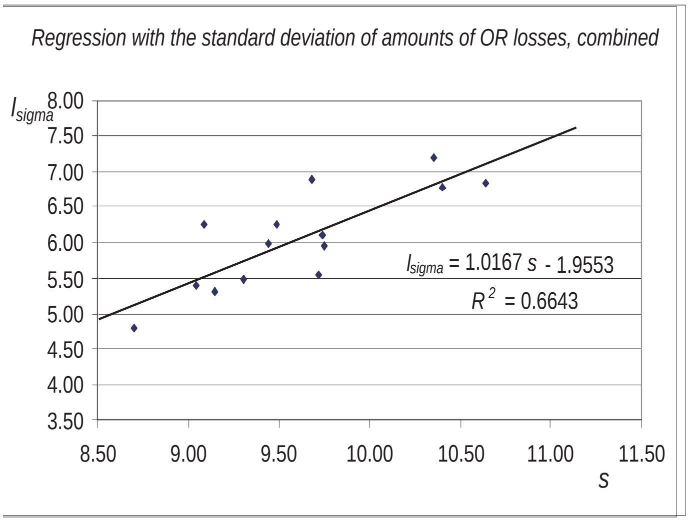

Figure 6: Regression lines according to equation(17), based on a combination of the internal BU data and external BL data.

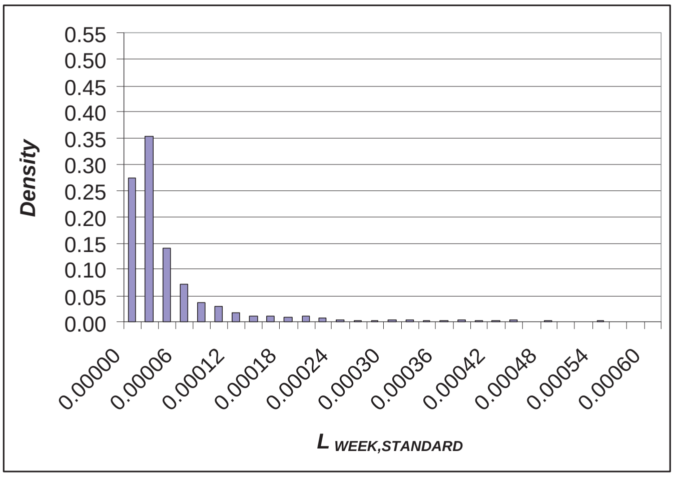

Figure 7: Approximation of the pdf of the aggregate losses per week of the standard BU/BL.

Table 3. So for a time horizon of one week, the VAR of the aggregate losses for the standard BU/BL at

Confidence level (α, in %)	j	VAR (per week, standard BU/BL)
95	705	0.00019
99	735	0.00067
99.9	741	0.00363

Table 3: The Value-at-Operational Risk for the standard BL/BU, for several confidence levels.

confidence level 99.9% equals 0.00363 (for reasons of confidentiality, the precise meaning of this number is not illuminated here). To find the VAR of the aggregate losses for another BU/BL, we can apply scaling

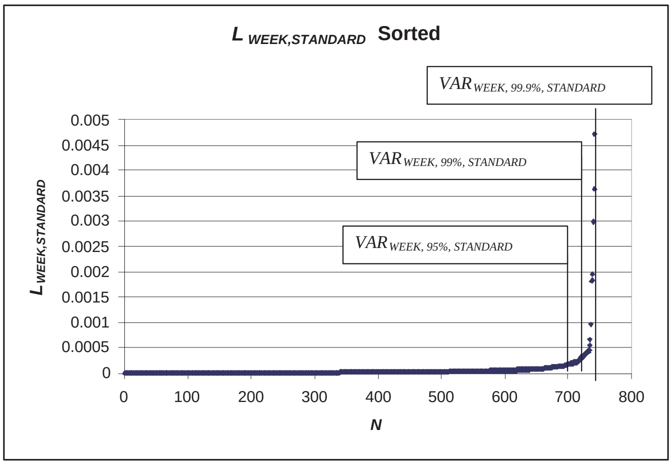

Figure 8: The sorted aggregate losses per week for the standard BU/BL, with VAR-values at 95%, 99%, and 99.9% confidence level.

in a similar way as has been applied in equation(11) and (12):

$$VAR_{L_b} = (s_b)^{\lambda}.VAR_{L_{st}}. \tag{18}$$

So, it is an easy exercise to assess the VAR of the aggregate losses of a given BU/BL for any other time horizon. Knowing the *VAR* of a BU/BL per year, we can also easily calculate the so-called OR capital charge as dictated by the Basel Committee the general procedure of which was sketched in item 5 of section 2.2. We do not elaborate the details here. Interested readers can found these in Master thesis [1].

4.2 Analyzing the frequency of OR losses

As explained in section 2, an operational risk loss distribution is usually estimated by a compound of the frequency and severity distributions. It is therefore interesting to investigate whether the power-law relationship comes from the frequency element, the severity element, or even from both elements. For this reason, a similar experiment is conducted to the frequency and the severity distributions. We will start with the frequency element in this section.

4.2.1 Regression lines based on the means of OR losses

The experiments related to the frequency distributions have been executed in precisely the same way as those related to the aggregate losses described in the previous section (precise details are given in [1]). We confine ourselves here to presenting the relevant tables containing the numerical values related to the various regression lines, both for the external data and the internal data, and the combination of the two:

The coefficient i_μ of the intercept as well as the coefficient λ_μ of the gradient are significant at a confidence level of 99.9%, because their P-values are less than 0.001. These result suggest that there is a power-law relationship between the frequency of operational loss per week and the size & exposure towards operational risk per week of the external Lines of Business. The idiosyncratic component can explain a

Data set	R^2	Parameter	Value	Standard Error	t-statistic	P-value
External	0.8923	i_μ	-7.5987	1.1529	-6.5907	0.0006
		λ_μ	0.9996	0.1418	7.0507	0.0004
Internal	0.2142	i_μ	-3.0449	3.1297	-0.9729	0.3857
		λ_μ	0.4311	0.4129	1.0442	0.3553
Combined	0.6912	i_μ	-5.7064	1.1784	-4.8425	0.0004
		λ_μ	0.7731	0.1492	5.1831	0.0002

Table 4: Parameter and statistical values of the regression with the means of the frequency of OR losses.

big proportion of the variability in the frequency of operational loss per week, as we can conclude from the high value of R^2 (0.8923).

The latter value for the internal data is less impressive (0.2142). In addition, the P-values are too high for rejecting the null hypotheses that the coefficients are equal to zero. This result suggests that a power-law relationship between the frequency of operational loss per week and the size & exposure towards operational risk per week of the internal Business Units is not present. If we combine the internal and external data, the picture changes completely: $R^2 = 0.6912$ expresses that 69.12% of the frequency of OR losses is attributable to explaining variable (being the logarithm of indicator of size s_b: compare linearized equation (13)). The values of the coefficients are significant at the 99.95% significance level because their P-values are less than 0.0005.

Table 4 shows the results of the regression related to the standard deviation of the frequency of OR losses. The interpretation is straightforward. Roughly spoken, the pattern of results is equal to that of

Data set	R^2	Parameter	Value	Standard Error	t-statistic	P-value
External	0.7770	i_σ	-4.9689	1.2082	-4.1127	0.0063
		λ_σ	0.6793	0.1486	4.5722	0.0038
Internal	0.1354	i_σ	-1.0113	1.5330	-0.6597	0.5455
		λ_σ	0.1601	0.2022	0.7916	0.4729
Combined	0.6848	i_σ	-4.0514	0.9732	-4.6397	0.0006
		λ_σ	0.5643	0.1105	5.1058	0.0003

Table 5: Parameter and statistical values of the regression with the standard deviation of the frequency of OR losses.

the regression with the mean of the frequency of OR losses. Again, the confidence interval is increasing when we perform the regression on the combination of internal Business Units and external Lines of Business. This result suggests that there exists a universal power-law relationship between the frequency of operational loss per week and the size & exposure towards operational risk per week of the combination of external Lines of Business & internal Business Units. We further observe that the values of λ_μ and λ_σ are more varying in the tables shown of this section than in those presented in section 4.1.

4.3 Analyzing the severity of OR losses

Finally we take a short look at the severity distributions of OR losses. Please note the fact that the severity of operational losses is the financial loss amount of individual events. This means that severity is measured per individual event and not related to a time interval. Therefore, in this section we study just the relationship between the severity of operational loss and the size & exposure towards operational risk of different Lines of Business and Business Units. For the rest, the experiments have (again) been executed in the same way (once again, much more details can be found in [1]). The two tables summarizing the quantitative results of the regressions performed are as follows. These results are not convincing at all.

Data set	R^2	Parameter	Value	Standard Error	t-statistic	P-value
External	0.0028	i_μ	5.1931	1.5766	3.2938	0.0165
		λ_μ	0.0209	0.1600	0.1306	0.9004
Internal	0.3532	i_μ	-1.9103	4.7626	-0.4011	0.7088
		λ_μ	0.7566	0.5119	1.4779	0.2135
Combined	01757	i_μ	2.5185	1.7300	1.4558	0.1711
		λ_μ	0.2875	0.1798	1.5991	0.1358

Table 6: Parameter and statistical values of the regression with the means of the severity of OR losses.

Data set	R^2	Parameter	Value	Standard Error	t-statistic	P-value
External	0.1095	i_σ	3.5082	2.8459	1.2327	0.2638
		λ_σ	0.2480	0.2888	0.8587	0.4235
Internal	0.5898	i_σ	-8.1924	5.7157	-1.4335	0.2250
		λ_σ	1.4733	0.6143	2.3983	0.0745
Combined	0.3315	i_σ	-0.3200	2.4954	-0.1282	0.9001
		λ_σ	0.6327	0.2594	2.4395	0.0312

Table 7: Parameter and statistical values of the regression with the standard deviation of the severity of OR losses.

E.g., the values of R^2 are generally very low, the values of λ_μ and λ_σ are strongly different, and null hypotheses can not be rejected with a sufficient level of significance. In short, no power-law relationship can be concluded between the severity of operational loss and the size & exposure towards operational risk of the external Lines of Business & internal Business Units, not from the internal data, not from the external data, and not from the combination of the two data sets.

5 Discussion

5.1 Considerations of technical nature

In the previous section we examined the relationship between the aggregate loss, frequency and severity of OR losses on the one hand and the size & exposure towards operational risk of the external Lines of Business, of the internal Business Units, and of the combination of external Lines of Business & internal Business Units on the other hand. The results with respect to the aggregate losses are convincing since the value of the correlation coefficient appeared to be quite high in all cases, the values of the regression coefficients appeared to be significant, and the estimated values λ_μ and λ_σ are generally quite close. This result suggests that the mean of the aggregate losses scale in the same way as their standard deviation and that the average of the estimated values λ_μ and λ_σ can be used to represent the value of the universal scaling parameter λ. It is further remarkable that the estimates of λ_μ and λ_σ are very close to 1. This suggests that the universal power-law relationship can be regarded as a linear one. In other words, the aggregate loss per week relates almost linearly to the size & exposure towards operational risk per week, in particular for the combination of external Lines of Business & internal Business Units.

The above-given findings support the believe that there exists a power-law relationship between the aggregate loss per week and the size & exposure towards operational risk per week. However, although the mean and standard deviation values of OR losses scale in the same way, it can not be guaranteed that the full probability density function of operational losses scales in the same way. To conclude this more general result, additional research is required where regression is applied with all data of the aggregate losses per week against the idiosyncratic component of every Line of Business and Business Unit.

Since the distribution of the aggregate losses has two underlying components namely the frequency of OR losses and the severity of OR losses, we also investigated whether a power-law relationship exist for each of these two components. Our findings with respect to the frequency of OR losses are that both for the external BLs and for the combination of internal BUs and external BLs a power-law relationship exists for the scaling of the mean and the scaling of the standard deviation of OR losses, although their scaling parameters λ are not equal. In other words, no universal scaling parameter has been found. Our findings with respect to the severity of OR losses show that no power-law relationship according to equation (4) exists.

5.2 Considerations of OR Management nature

Even allowing for the limited data that was available for use in these experiments, we think that the results found provide interesting discussion points for the OR Management (ORM) practice. Especially from a Capital Adequacy perspective, the outcome is encouraging and opens up avenues for further research. In this section, we shall review the ORM implications of the main findings and suggest areas for refinement.

The main finding can be summarized as follows:

A clear relation exists between Gross Income of a BU or BLs and the aggregated loss amount per period.

In a way, this preliminary finding is both helpful and unhelpful for those banks that are now considering adopting an Advanced Measurement Approach (AMA) for their Capital Adequacy regime under Basel II. It is helpful, since, if this relation can be further established and hardened, it will give a useful proxy for the determination or the validation of capital amounts that are computed using the Loss Distribution Approach. If the relation can be further substantiated, it may go some way to further refine the alpha for the Basic Indicator Approach and the beta's for the Standardised Approach [12].

The finding that the outlined relation hold true for the Aggregate loss amounts per period, but not for Severity per event and to a limited extent for the number of losses per period points to a clear direction in ORM terms. The lack of a relation between size and severity is a generally accepted rule. In fact, extreme losses have been incurred by small outfits within banks, while being large by no means exempts a bank from incurring extreme events. This part of the finding is therefore well within expectation. From a theoretical point of view, one would expect that the variability of the aggregate amount then, is accounted for by the frequency information. The fact, however, that the equation holds much stronger for aggregate amounts than it does for frequency seems to belie this. We believe that the explanation should be sought in the looseness with which loss recording is done. In technical analysis, we are assuming the losses to have been accurately recorded. In practice, however, it is well known that in the ORM discipline, debate over the way in which losses are recorded (is this one event of 10 million, or is it 10 events of 1 million) can lead to political and fierce debates. In that sense, we trust the aggregate loss amounts per period to be a far more accurate representation and far more comparable between BUs, BLs and firms than either frequency or severity data by itself.

This lack of uniform or accurate reporting may also be a stumbling block when analyzing loss data by event category. The categories are neither mutually exclusive nor are they comprehensive exhaustive. In fact losses have been known to be re-classified as more information surrounding the loss becomes apparent. As an example, an external fraud should be re-classified as an Internal fraud when it is discovered that an insider was involved. This situation also applies to those who would perform causal analysis along these lines.

In a more complicated way, however, the main finding is also unhelpful, since it may suggest that ORM needs to concern itself only with Gross Income. This is clearly too simple a message. It is unlikely that exposure to Operational Risk in Business Units is not related to the quality of their processes, internal control, complexity, growth rates, market pressure, staff training and the particular culture in the business. These items are not irrelevant, they are simply harder to measure. Our preliminary conclusion should therefore more properly read as:

A clear relation exists between Gross Income of a BU or BL and the aggregated loss amount per period, ceteris paribus, the specific business and environmental control environment.

6 Conclusions and Outlook

In this paper we have shown the results of an investigation of the relationship between the aggregate loss L_b (incurred in a financial institution/BU b within a certain time period) and an indicator s_b of size &

exposure towards the OR of that financial institution. We have found evidence that the power-law form described by

$$L_b = (s_b)^\lambda . L_{st}, \tag{19}$$

(where λ is a universal constant and L_{st} represents the aggregate loss of the standard financial institution/BU) can be used to explain this relationship.

Based on the existence of the power-law relationship, we were able to apply the scaling mechanism to remove financial institutions' specific characteristics, so that the external data can be considered to have the same characteristics as the internal data. Instead of investigating at the aggregate level (view each bank as a single entity), we have chosen to investigate at the Line of Business level. The choice of examining on the BLs' level is particularly based on the information available from the external data. We can only tell from which Line of Business, but not from which bank, an operational loss comes from. This information is not given away in the external data. In our data set, the internal loss data of the bank is given per Business Unit. For the reason of simplicity, we directly used the Business Units of the bank instead of mapping them into the Basel Lines of Business categorization. We used Gross Income as the indicator for the size & exposure to operational risk of a Line of Business.

We have also shown how the resulting aggregated data set can be used to (i) scale the data appropriately and (ii) calculate the Value-at-OR for each business unit. In addition, we presented the principles of calculating the value of the OR capital charge of each BL and BU, all according to the minimal capital requirements of the Basel committee.

Finally, we have tried to observe whether the power-law relationship - between the aggregate loss per week and the size & exposure towards operational risk per week of the external BLs, of the internal BUs, and of the combination of external BLs & internal BUs - comes from the frequency element, the severity element, or even from both elements. The obtained results suggest that the scaling mechanism always holds for the aggregate loss data, sometimes holds for the frequency of OR loss data, and never holds for the severity of OR loss data. In a discussion, we presented some ideas that may explain these observations.

Extensions to the current study are widely open. Related to the main theme of this paper, we propose to try the following studies related to the scaling of OR loss data:

- Mapping the BUs of the bank first into the Basel Lines of Business (BLs) and use these BLs as the internal data of the bank.

- Estimating the value of the universal scaling parameter λ by running the regressions on individual data (both aggregate loss data, frequency of OR loss data, and severity of OR loss data) instead of on their mean and standard deviation per BU/BL.

- Studying the development of the universal scaling parameter λ across time.

- Using other time horizons, e.g., monthly instead of weekly.

- Testing the existence of universal power-law relationship by using other loss data than the external data and the bank's internal data.

References

[1] Heru Sataputera Na, *Operational Risk, Analysing and Scaling Operational Risk Loss Data*, Master thesis, informatics & economics, Erasmus University Rotterdam, 2004, Available from http://www.few.eur.nl/few/people/jvandenberg/masters.htm, #50.

[2] N. Baud, A. Frachot, and T. Roncalli, "Internal data, external data and consortium data for operational risk measurement: How to pool data properly?," Tech. Rep., 2002, Groupe de Recherche, Crédit Lyonnais, France.

[3] H.E. Stanley, "Scaling, universality, and renormalization: Three pillars of modern critical phenomena," *Reviews of Modern Physics*, vol. 71, pp. S358–S366, 1999.

[4] J.-P. Bouchaud, "Power-laws in economy and finance: some ideas from physics," *Quantitative Finance*, vol. 1, pp. 105–112, 2001.

[5] V. Pareto, *Cours d'économie politique*, Droz, Geneva, Reprinted as a volume of Oevres Complètes.

[6] B.B. Mandelbrot, *Fractals and Scaling in Finance*, Springer, 1997.

[7] J. Shih, AH Samad-Khan, and P. Medapa, "Is the size of an operational loss related to firm size," *Operational Risk*, January 2000.

[8] J. Shih, "On the use of external data for operational risk quantification," *Risk Architecture*, Citigroup.

[9] A. Chapelle, Y. Crama, G. Hübner, and J.-P. Peters, "Measuring and managing operational risk in the financial sector, an integrated framework," Tech. Rep., Social Science Reserach Network, February 2005, http://ssrn.com/abstract=675186.

[10] C. Alexander, *Operational Risk Regulation, Analysis and Management*, Prentice Hall, 2003.

[11] R. Cole et al., "Working paper on the regulatory treatment of operational risk," Tech. Rep., Basel Committee on Banking Supervision, Basel, Switzerland, September 2001.

[12] Basel Committee on Banking Supervision, "International convergence of capital measurement and capital standards, a revised framework," Tech. Rep., Bank for International Settlements, Basel, Switzerland, June 2004.

[13] R.V. Hogg and A.T. Craig, *Introduction to Mathematical sStatistics*, Macmillan, New York, 1970.

[14] R.D. Hoffman (The Animated Software Company), "Internet glossary of statistical terms," http://www.animatedsoftware.com/statglos/sgsderro.htm.

Publications in the Report Series Research* in Management

ERIM Research Program: "Business Processes, Logistics and Information Systems"

2005

On The Design Of Artificial Stock Markets
Katalin Boer, Arie De Bruin and Uzay Kaymak
ERS-2005-001-LIS
http://hdl.handle.net/1765/1882

Knowledge sharing in an Emerging Network of Practice: The Role of a Knowledge Portal
Peter van Baalen, Jacqueline Bloemhof-Ruwaard, Eric van Heck
ERS-2005-003-LIS
http://hdl.handle.net/1765/1906

A note on the paper Fractional Programming with convex quadratic forms and functions by H.P.Benson
J.B.G.Frenk
ERS-2005-004-LIS
http://hdl.handle.net/1765/1928

A note on the dual of an unconstrained (generalized) geometric programming problem
J.B.G.Frenk and G.J.Still
ERS-2005-006-LIS
http://hdl.handle.net/1765/1927

Privacy Metrics And Boundaries
L-F Pau
ERS-2005-013-LIS
http://hdl.handle.net/1765/1935

Privacy Management Contracts And Economics, Using Service Level Agreements (Sla)
L-F Pau
ERS-2005-014-LIS
http://hdl.handle.net/1765/1938

A Modular Agent-Based Environment for Studying Stock Markets
Katalin Boer, Uzay Kaymak and Arie de Bruin
ERS-2005-017-LIS
http://hdl.handle.net/1765/1929

Lagrangian duality, cone convexlike functions
J.B.G. Frenk and G. Kassay
ERS-2005-019-LIS
http://hdl.handle.net/1765/1931

Operations Research in Passenger Railway Transportation
Dennis Huisman, Leo G. Kroon, Ramon M. Lentink and Michiel J.C.M. Vromans
ERS-2005-023-LIS
http://hdl.handle.net/1765/2012

Agent Technology Supports Inter-Organizational Planning in the Port
Hans Moonen, Bastiaan van de Rakt, Ian Miller, Jo van Nunen and Jos van Hillegersberg
ERS-2005-027-LIS
http://hdl.handle.net/1765/6636

Faculty Retention factors at European Business Schools
Lars Moratis, Peter van Baalen, Linda Teunter and Paul Verhaegen
ERS-2005-028-LIS
http://hdl.handle.net/1765/6559

Determining Number of Zones in a Pick-and-pack Orderpicking System
Tho Le-Duc and Rene de Koster
ERS-2005-029-LIS
http://hdl.handle.net/1765/6555

Integration of Environmental Management and SCM
Jacqueline Bloemhof and Jo van Nunen
ERS-2005-030-LIS
http://hdl.handle.net/1765/6556

On Noncooperative Games and Minimax Theory
J.B.G. Frenk and G.Kassay
ERS-2005-036-LIS
http://hdl.handle.net/1765/6558

Optimal Storage Rack Design for a 3-dimensional Compact AS/RS
Tho Le-Duc and René B.M. de Koster
ERS-2005-041-LIS
http://hdl.handle.net/1765/6730

Strategies for Dealing with Drift During Implementation of ERP Systems
P.C. van Fenema and P.J. van Baalen
ERS-2005-043-LIS
http://hdl.handle.net/1765/6769

Modeling Industrial Lot Sizing Problems: A Review
Raf Jans and Zeger Degraeve
ERS-2005-049-LIS
http://hdl.handle.net/1765/6912

Cyclic Railway Timetabling: a Stochastic Optimization Approach
Leo G. Kroon, Rommert Dekker and Michiel J.C.M. Vromans
ERS-2005-051-LIS
http://hdl.handle.net/1765/6957

Linear Parametric Sensitivity Analysis of the Constraint Coefficient Matrix in Linear Programs
Rob A. Zuidwijk
ERS-2005-055-LIS
http://hdl.handle.net/1765/6991

Diffusion of Mobile Phones in China
Sunanda Sangwan and Louis-Francois Pau
ERS-2005-056-LIS
http://hdl.handle.net/1765/6989

An Elementary Proof of the Fritz-John and Karush-Kuhn-Tucker Conditions in Nonlinear Programming
S.I. Birbil, J. B. G. Frenk and G. J. Still
ERS-2005-057-LIS
http://hdl.handle.net/1765/6992

General model for automated diagnosis of business performance
Emiel Caron and Hennie Daniels
ERS-2005-058-LIS
http://hdl.handle.net/1765/6987

Exploring retailers' sensitivity to local sustainability policies
H.J. Quak and M.B.M. de Koster
ERS-2005-066-LIS
http://hdl.handle.net/1765/7057

Setting the holding cost rates in a multi-product system with remanufacturing
Umut Corbacıoğlu and Erwin A. van der Laan
ERS-2005-072-LIS
http://hdl.handle.net/1765/7095

A Dynamic Pricing Model for Coordinated Sales and Operations
Moritz Fleischmann, Joseph M. Hall and David F. Pyke
ERS-2005-074-LIS
http://hdl.handle.net/1765/7124

Optimal Redesign of the Dutch Road Network
M. Snelder, A.P.M. Wagelmans, J.M. Schrijver, H.J. van Zuylen and L.H. Immers
ERS-2005-091-LIS

Data Scaling for Operational Risk Modelling
H.S. Na, L. Couto Miranda, J. van den Berg and M. Leipoldt
ERS-2005-092-LIS

---

* A complete overview of the ERIM Report Series Research in Management:
https://ep.eur.nl/handle/1765/1

ERIM Research Programs:
LIS Business Processes, Logistics and Information Systems
ORG Organizing for Performance
MKT Marketing
F&A Finance and Accounting
STR Strategy and Entrepreneurship