Article:

New results on the correlation problem in operational risk

The Journal of Financial Perspectives

EY Global Financial Services Institute

July 2014 | Volume 2 - Issue 2

Building a better working world

The EY Global Financial Services Institute brings together world-renowned thought leaders and practitioners from top-tier academic institutions, global financial services firms, public policy organizations and regulators to develop solutions to the most pertinent issues facing the financial services industry.

The Journal of Financial Perspectives aims to become the medium of choice for senior financial services executives from banking and capital markets, wealth and asset management and insurance, as well as academics and policymakers who wish to keep abreast of the latest ideas from some of the world's foremost thought leaders in financial services. To achieve this objective, a board comprising leading academic scholars and respected financial executives has been established to solicit articles that not only make genuine contributions to the most important topics, but are also practical in their focus. *The Journal* will be published three times a year.

gfsi.ey.com

Executive summary

New results on the correlation problem in operational risk

by Vivien Brunel, Head of Risk and Capital Modeling, Société Générale

Internal models of operational risk are all built based on the same guidelines provided by the regulators. However, we observe a broad range of practices among banks concerning modeling choices and calibration methods. This paper discusses the relative importance of the main drivers and modeling choices of the operational risk capital charge. Many studies in the literature have focused on the modeling of the tails in the severity distributions. Here, we use a class of analytical models for operational risk in order to assess the relative importance of all parameters of the model. In particular, we show that the bank's capital charge is not very sensitive to the dispersion in correlations, the average level of correlations being a much more critical parameter of the operational risk capital charge. We show that the assumption of uniform correlations is robust, contrary to what is often advised by internal auditors or regulators.

New results on the correlation problem in operational risk

Vivien Brunel
Head of Risk and Capital Modeling, Société Générale[1]

Abstract
Internal models of operational risk are all built based on the same guidelines provided by the regulators. However, we observe a broad range of practices among banks concerning modeling choices and calibration methods. It is thus relevant to discuss the relative importance of the main drivers and modeling choices of the operational risk capital charge. Many studies in the literature have focused on the modeling of the tails in the severity distributions. In this paper, we use a class of analytical models for operational risk in order to assess the relative importance of all parameters of the model. In particular, we show that the bank's capital charge is not very sensitive to the dispersion in correlations, the average level of correlations being a much more critical parameter of the operational risk capital charge. We show that the assumption of uniform correlations is robust, contrary to what is often advised by internal auditors or regulators.

1 I would like to thank Pavel Shevchenko for the very stimulating correspondence we had together. This article reflects the author's opinions and not necessarily those of his employers.

1. Introduction

The current regulatory framework allows banks to compute their capital charges for operational risk under an internal model, which is often based on the loss distribution approach (LDA). In this approach, loss distributions are calibrated at the cell level (a cell is the elementary risk unit per business line and type of risk) and the bank's capital charge is estimated by aggregating cell loss distributions under some dependence assumption [Chernobai et al. (2007)].

The Basel Committee [Bank for international Settlements (2011)] provides some guidelines about how banks should appropriately reflect the risk profile in their internal models. However, banks benefit from some flexibility in their modeling choices that may lead to some discrepancies in capital charges for similar risk profiles. The broad range of practices observed among banks results, in particular, from different distributional or dependence assumptions in the models.

Many studies have focused on the modeling of the tails in the severity distributions [Dutta and Perry (2007), Moscadelli (2004)], but the bulk of the correlation problem is still unresolved and controversial. There is a strong debate about the choice of the copula function for losses across cells, since the scarcity of the data makes it quite difficult to solve for this issue. The regulators advise banks to determine sound correlations and to retain conservative assumptions. Some institutions have selected the simplest option and use equal correlations between cell losses. This assumption is, of course, questionable and may embed some model risk, but regulators, as well as practitioners, have great difficulties in asserting arguments about realistic and conservative correlation levels. Some authors believe that correlations between cell losses are as low as 4% [Frachot et al. (2004)].

Most of the knowledge we have about operational risk quantification comes from complex models and heavy Monte Carlo simulations, and, as far as we know, there is no analytical model that takes into account risk and correlation dispersion among cells. This article fills this gap. Under the asymptotic single risk factor (ASRF) assumption, we obtain new results about the bank's capital charge sensitivity to the critical parameters of the model. In particular, we show that the capital charge is not that sensitive to correlation dispersion, and the constant correlation assumption is robust.

This new result is obtained with few specifications, and we conjecture that it remains valid, at least qualitatively, for real bank portfolios that have a finite number of cells. We believe that our approach is also relevant for pioneering a new way to compute capital charges and challenge internal model assumptions as exemplified in this paper.

This paper is organized as followed. In section 2, we provide some real data evidence about cell loss distributions and correlations. We will also solve the ASRF model with lognormal losses at the cell level, even when individual cells have various risk profiles. In section 3, we solve the case of non-equal correlations between cells and provide some key results about the capital charge sensitivity to the main critical parameters of the model. Section 4 concludes the paper.

2. Some empirical facts about cell loss distributions and correlations

In the LDA framework, the aggregate operational loss for cell number i is equal to the sum of individual losses:

$$L_i = \sum_{n=1}^{N_i} X_n^i \quad (1)$$

where L_i is the aggregate loss of cell number i, N_i is the number of events over 1 year, and $(X_n^i)_{1 \le n \le M}$ is the sequence of the individual loss severities for cell number i. The aggregate loss process is a compound Poisson process, and accordingly, the model is based on the following assumptions:

- The number of events and severities are independent
- Severities are independent and identically distributed random variables

2.1 Cell loss distribution parameters

Concerning loss distributions, there exists a number of studies that look at individual loss distributions [see for instance Dutta and Perry (2004) and Moscadelli (2004)], but there are very few empirical studies about aggregate cell losses. We aim to fill this gap here.

We have conducted our study based on the SAS OpRisk Global database, which as of November 2013 included 6,402 events that have occurred in financial firms since 2002, the date from which financial institutions started collecting and reporting

Confidence level	Average	St.dev
95%	98%	41%
97.5%	99%	39%
99%	107%	44%
99.5%	112%	46%
99.9%	124%	48%
All	107%	42%

Table 1: Parameter σ implied value from real data

their operational losses systematically. We have calibrated the frequency of events and lognormal severity distributions for each of the 21 cells that have more than 30 losses. Direct calibration of the aggregate loss distribution from real data is, of course, impossible because there is only one observation per year. However, it is possible to assess the compliance with the lognormal distribution of the aggregate loss distribution obtained through the LDA.

Let us consider that the loss distribution for cell number i is lognormal with parameters μ_i and σ_i; the ratio between the expected value and any quantile depends only on the parameter σ_i:

$$\frac{\text{Expected value}(i)}{\text{VaR}_q(i)} = e^{\frac{\sigma_i^2}{2} + \sigma_i F_q} \qquad (2.1)$$

$$\sigma_i = -F_q - \sqrt{F_q^2 + 2\ln \frac{\text{Expected value}(i)}{\text{VaR}_q(i)}} \qquad (2.2)$$

where VaR_q is the q-percentile of the lognormal distribution and $F_q = N^{-1}(1-q)$. Inverting equation (2.1) leads to two different solutions; we have chosen the one with a minus sign in front of the square root in equation (2.2) because we require the parameters σ_i to decrease with the ratio of expected value to quantile for all cells. We observe that broader distribution assumptions for cell losses in the model can naturally be taken into account by choosing the plus sign solution in equation (2.2).

The LDA leads to the following ratios for each cell in the tail of the loss distribution ($q > 95\%$). For several values of the confidence level, Table 1 provides the observed average value and standard deviation of parameters σ_i over all cells, implied from equation (2.2).

The range of values of parameters σ_i is rather stable when the confidence level changes: the average value over all cells and confidence intervals is equal to 107%, and the observed standard deviation is equal to 42%. To assess the robustness of these estimates, we have computed the median of observed values for the parameters σ_i, which is equal to 108.5%, and is very close to the average value, and the med-med estimator (median value of the spread with the median) that is equal to 31%, which is lower than the measured standard deviation.

2.2 Cell loss correlations

For most of the studies [see for instance Aue and Kalkbrener (2007), Frachot et al. (2004)], cell loss correlations are generated by the dependence of the number of events between cells rather than the dependence of severities. Under the assumption of lognormal severity distributions $X^i \sim LN(m_i, s_i)$, Frachot et al. (2004) show that the loss correlation between cell 1 and cell 2 is equal to:

$$\text{corr}(L_1, L_2) = \text{corr}(N_1, N_2).e^{-\frac{1}{2}s_1^2 - \frac{1}{2}s_2^2} \qquad (3)$$

The correlation of the number of events is linked to the loss frequencies of cells 1 and 2. Bivariate Poisson variables are obtained by considering three independent Poisson variables Z, Y_1, and Y_2 with parameters r, $\lambda_1 - r$ and $\lambda_2 - r$, respectively; the variables $N_i = Z + Y_i$ are also Poisson with intensities λ_i, and their correlation is equal to:

$$\text{corr}(N_1, N_2) = \frac{r}{\sqrt{\lambda_1 \lambda_2}} \leq R = \sqrt{\frac{\min(\lambda_1, \lambda_2)}{\max(\lambda_1, \lambda_2)}} \qquad (4)$$

The upper bound R for the correlation comes from the inequalities $\lambda_1 \geq r$ and $\lambda_2 \geq r$. Aue and Kalkbrener (2007) observed that frequency correlations were about 10%, and higher correlations were specific to some couples of cells only. Frachot et al. (2004) claimed that loss correlations were as low as 4%: we recover this result when we take $\text{corr}(N_1, N_2) = 38.5\%$ [which is the expected value estimated by Brunel (2014)] and $s_1 = s_2 = 1.5$ [which is the lowest value observed by Frachot et al. (2004) for these parameters] in equation (3). From SAS OpRisk data, we observe that the parameters si have an average value equal to 2.03, a standard deviation equal to 0.42 and are ranged between

1.34 and 2.90. Correlation upper bounds can be computed with equations (3) and (4) from this data. We find an expected value equal to 1.33% and a standard deviation equal to 1.61%. Correlation upper bounds all range between 0% and 4%, except a few of them; the highest correlation has an upper bound at 11.27% and is found between "Execution, delivery and process management" and "Internal fraud" cells of the retail brokerage business line. All these studies confirm that we expect low levels of correlation between cells.

2.3 Correlation parameters in the Gaussian copula model

In the Gaussian copula framework with lognormal marginal cell losses, the correlation parameter $\rho_{i,j}$ between two cells is related to the cell loss correlation:

$$\mathrm{corr}(L_i, L_j) = \frac{e^{\rho_{ij}\sigma_i\sigma_j} - 1}{\sqrt{(e^{\sigma_i^2} - 1)(e^{\sigma_j^2} - 1)}} \qquad (5)$$

This formula with parameters $\sigma_i = \sigma_j = 107\%$ and a conservative assumption for loss correlations corr(L_i, L_j) = 4%, leads to a correlation parameter of the Gaussian copula equal to $\rho_{ij} = 7.2\%$. External data support the assumption of very low correlation parameters in the copula framework, much lower than 10%.

3. Critical parameter analysis of the capital charge

3.1 A simplified LDA model

In the rest of this article, we build a simple portfolio model for operational risk. We assess that the bank's operational risks is a portfolio of N operational risks at cell level. We make the following four assumptions:

- **Lognormal distributions:** the loss for cell number i is a lognormal random variable L_i with parameters μ_i and σ_i. As shown in section 2, we assume that the σ_i parameters have an expected value equal to σ = 107% and a variance v = 18% (except in section 3.2).
- **Gaussian copula:** pair-wise correlations ρ_{ij} may be different from each other. For numerical estimations, we assume that the average correlation is equal to 10% (this is a conservative assumption as seen in section 2.3).
- **One factor model:** cell losses are sensitive to the same systemic factor called F. This factor is assumed to be a standard normal random variable. The specific part of the risk is embedded in another independent normal random variable called $\in_i$ (i = 1, ... , N). Systemic and specific factors are all assumed to be independent from each other.
- We assume that the parameters are not dependent on the number of cells N.

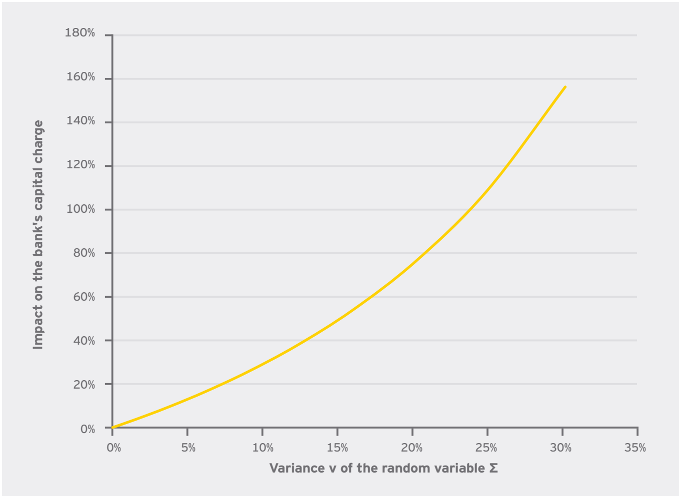

Figure 1: Sensitivity of the bank's capital charge as a function of the parameter v(σ = 107% and ρ = 10%)

In this framework, the annual loss for a cell can be written as the exponential function of a normal random variable, which is a linear combination of the systemic and specific factors. We get for cell number i (i = 1, ... , N):

$$L_i = e^{\mu_i - \sigma_i(\beta_i F + \sqrt{1-\beta_i^2} \in_i)} \qquad (6)$$

The parameters β_i are linked to the pair-wise correlations of the Gaussian copula: $\rho_{ij} = \beta_i\beta_j$. Because cells may have very different risk characteristics, and because correlations may be very different for different pairs of cells, we assume that the parameters μ_i, σ_i and β_i are the observations of i.i.d. random variables called M, Σ and B, respectively. In the limit N→∞, the bank's loss is equal to N.L(F) and is a function of the common factor F, as in Vasicek's model for granular homogeneous loan loss distributions [Vasicek (2002)]:

$$L(F) = \lim_{N\to\infty} \frac{1}{N}\Sigma_{i=1}^{N} L_i = E\left[e^{M-\Sigma(BF+\sqrt{1-B^2}\in_i)} \mid F\right] = E\left[e^{-\Sigma BF + \Sigma^2(1-B^2)/2} \mid F\right].E\left[e^{M}\right]$$

(7)

We are now going to study the following cases for the random

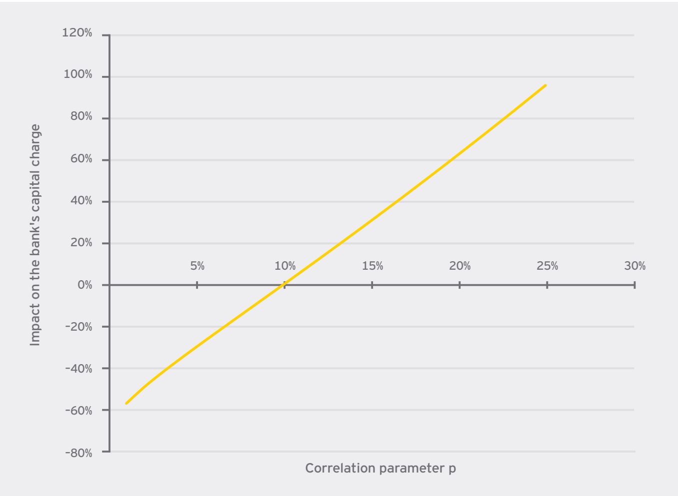

Figure 2: Sensitivity of the bank's capital charge as a function of the correlation parameter ρ(σ = 107%)

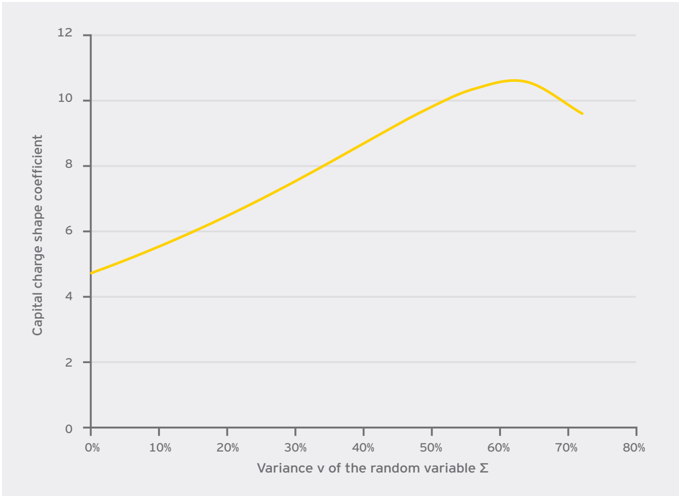

Figure 3: Function S(v)

variables Σ and B:

- Case 1: the parameters σ_i are normally distributed, i.e., $\Sigma = N(\sigma, v)$, and correlations are constant, i.e., $B = \sqrt{\rho}$
- Case 2: the parameters σ_i are constant, i.e., $\Sigma = \sigma$, and correlations are normally distributed, i.e., $B \equiv N(\sqrt{\rho}, w)$.
- Case 3: the parameters σ_i are constant, i.e., $\Sigma = \sigma$, and the variable B is uniformly distributed between $\sqrt{\rho} - \sqrt{3w}$ and $\sqrt{\rho} + \sqrt{3w}$ (bounds are chosen so that the expected value and the variance are equal to $\sqrt{\rho}$ and w, respectively).

3.2 Heterogeneous risk and identical correlations

The parameter σ drives the average risk of each cell and is, of course, a critical parameter of the model. We will show here that the bank's capital charge is also very sensitive to the dispersion of the risks at cell level (measured by the parameter v) and to the correlation parameter ρ.

We assume that we are in case 1 and, for numerical computations, that σ = 107%. In the limit N→∞, the bank's loss can be computed exactly from equation (7) as a Gaussian integral [Brunel (2014)]. In Figure 1, we plot the bank's capital charge as a function of the cell risk dispersion parameter v (the value of the correlation parameter ρ is equal to 10%). We see that the parameter v is a critical parameter of the model because the bank's capital charge is very sensitive to cell risk dispersion.

In Figure 2, we plot the capital charge impact as a function of the correlation parameter ρ compared with the reference capital charge, which is computed with v = 42% and ρ = 10%.

Surprisingly, we observe that the capital charge is almost an affine function of the correlation parameter, which could not be inferred a priori from the closed-form formulas (see Brunel, 2014). The bank's capital charge is a function $K(\sigma, v, \rho)$ of the parameters of the model, and we have the approximate formula:

$$\text{Capital impact} = \frac{K(107\%, v, \rho) - K(107\%, v, 10\%)}{K(107\%, v, 10\%)} \sim S(v)(\rho - 10\%)$$

For v = 42%, the shape coefficient S(v) is equal to 6.25, and this linear formula is very accurate (R^2 = 99.97%). We can observe that linearity is maintained in a wide range of values for the parameter v and remains very accurate (R^2 . 99.5%) in the range of values we have studied for the parameter v; we plot the function S(v) in Figure 3.

The shape function S(v) is increasing, meaning that when cell risk dispersion is getting higher, the correlation parameter is getting more and more critical.

3.3 Uncertain correlations

Correlations are not identical to each other, as illustrated in section 2, but estimating them from real data is a challenge from a statistical viewpoint. Data is scarce and limited only to one observation per year for the aggregate loss. Estimation of the number of events correlation is no longer robust for the same reason and severity correlations are only observable for cells that exhibit a sufficient number of events per year. Assuming identical correlations among cells is a current practice even if, in reality, correlations are unknown parameters. In what follows, we remain in the limit $N \to \infty$ and correlation uncertainty is included in the model by assuming that the random variable B has an expected value equal to $\beta = \sqrt{\rho}$ and a variance equal to w. For the sake of clarity we assume that individual risks are all equal among cells, i.e., Σ is a constant equal to σ.

As pair-wise correlations are equal to $\rho_{ij} = \beta_i \beta_j$, there is a direct link between the variance of ρ_{ij} and the variance of the sensitivity parameters β_i. Because of the independence of the β_i, we can write:

$$\mathrm{var}\,(\rho_{ij}) = E\,(\beta_i^2 \beta_j^2) - E\,(\beta_i)^2 E\,(\beta_j)^2 = w\,(w + 2\beta^2)$$

This leads, by solving the second order equation in w, to:

$$w = \sqrt{\beta^4 + \mathrm{var}\,(\rho_{ij})} - \beta^2 \qquad (8)$$

For $\beta^2 = \rho = 10\%$ and $\sqrt{\mathrm{var}\,(\rho_{ij})} = 3\%$ (which is a conservative value compared to what is measured from observed data; see section 2.2), we have w = 0.44%, i.e., the standard deviation of the parameter β is equal to 6.6%.

The ratio of the capital charge including model risk (w > 0) to the capital charge without model risk (w = 0) measures the increase in capital due to dispersion or uncertainty on correlations. We plot this quantity in Figure 4 as a function of the mean deviation of the correlation parameter (√w) in cases 2 and 3.

The curves corresponding to cases 2 and 3 are close together: the shape of the distribution function for the random variable B is not a critical choice. Moreover, we show that the impact of the mean deviation of the correlation parameter is lower than 2% for √w = 6.6% in both cases.

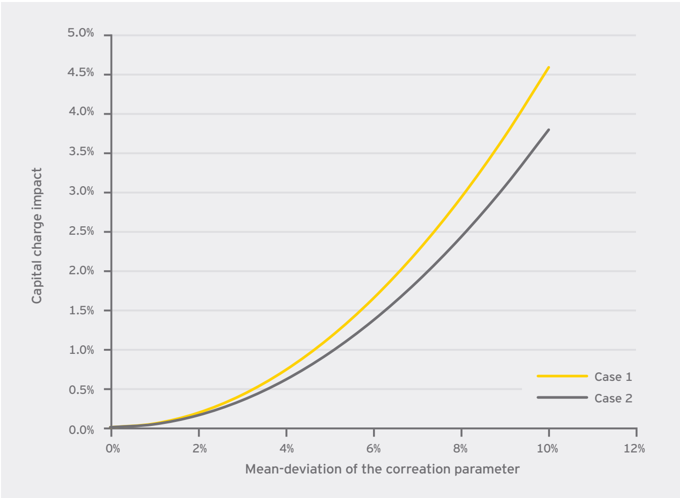

Figure 4: Capital charge impact as a function of the correlation dispersion for normal and uniform laws (β^2 = 10% and σ = 107%)

Even with a much more conservative choice for the individual cell risk parameter σ = 200%, the impact of correlation dispersion on the bank's capital charge would be about +5%. Our conclusion is that correlation dispersion (measured by parameter w) is, by far, not as critical as the other parameters of operational risk models (average cell risk σ, cell risk dispersion v and average correlation parameter √ρ).

4. Discussion and conclusion

This article explores the relative importance of parameters and assumptions in operational risk capital models. In particular, we obtain some new results on the criticality of the risk parameters that are driving the capital charge and on the correlations.

Our approach is based on a class of simplified analytical models that incorporate dispersion in individual cell risks and correlation levels. The main finding of this paper is that uniform correlation is a robust assumption for capital charge modeling. This result is important because it means that model risk associated with the value of correlations is not a major issue for capital measurement: the assumptions related to the calibration of individual cell risk or to the average level of correlations is much more critical. Moreover, regulators often challenge internal models on their accuracy and require refined and complex models. This requirement is relevant when modeling individual cells loss distributions. This paper shows that this is not the case for correlations. Differentiated pair-wise correlations have a less significant impact on the capital charge than uniform correlations, and generate more model risk in the calibration.

Concerning operational risk, there are still other unanswered questions that have as yet not been considered in the academic literature and for which no theoretical basis has been established. Among these issues, we can mention the number of cells (business lines and types of risks) problem. Because the calibration of the loss distribution is done independently across cells and cells risks are aggregated in a second step, the global calibration of the model is not based on a portfolio approach. As a result, the capital charge depends on the number of cells retained in the model whereas the global risk of the bank is independent of any risk classification. This is an open field of research that could have concrete practical implications on the design of internal models for operational risk.

These kinds of theoretical studies are necessary for banks when they are negotiating with the regulators to obtain approval for their internal models. Indeed, in many circumstances, data is too scarce to provide by themselves a formal proof of the model assumptions or methodological choices. Theoretical arguments associated with observed data (external or internal) are a powerful way to assess the robustness of the models.

References

Aue, F., and M. Kalkbrener, 2007, "LDA at work: Deutsche Bank's approach to quantifying operational risk," Journal of Operational Risk 1, 49–93

Bank for International Settlements, 2011, "Operational risk – supervisory guidelines for the advanced measurement approaches," BCBS 196

Brunel, V., 2014, "Analytical models of operational risk and new results on the correlation problem," Risk Magazine, July

Chernobai, A., S. T. Rachev, and F. Fabozzi, 2007, Operational risk: a guide to Basel II capital requirements, models and analysis, John Wiley & Sons

Dutta, K., and J. Perry, 2007, "A tale of tails: an empirical analysis of loss distribution models for estimating operational risk capital," The Federal Reserve Bank of Boston, working paper

Frachot, A., T. Roncalli, and E. Salomon, 2004, "The correlation problem in operational risk," Groupe de Recherche opérationnelle, Crédit Agricole, France

Moscadelli, M., 2004, "The modeling of operational risk: experience with the analysis of the data collected by the Basel committee," Technical Report 517, Banca d'Italia

Neslehova, J., P. Embrechts, and V. Chavez-Demoulin, 2006, "Infinite mean models and the LDA for operational risk," Journal of Operational Risk 1(1), 3-25

Vasicek, O, 2002, "Loan portfolio value," Risk, December, 160–162

Accredited by the American Economic Association

ISSN 2049-8640