# Operational Value-at-Risk in Case of Zero-inflated Frequency

Younès Mouatassim[1], El Hadj Ezzahid[2] & Yassine Belasri[3]

[1] Zurich Insurance, Casablanca, Morocco

[2] Mohammed V-Agdal University, Faculty of Law and Economics, Rabat, Morocco

[3] Turaz Global, Casablanca, Morocco

Correspondence: David Wolf, School of Management, Northern Canada University, Toronto, ON., M3A 2K7, Canada. Tel: 1-613-947-3592. E-mail: davidwolf@gc.ca

**Abstract**

In this paper we analyze operational risk in case of zero-inflated frequency data. We show that standard Poisson distribution does not suit correctly excess zero counts data. Alternatively, Zero-inflated Poisson (ZIP) distribution fits better such data. To assess the benefits of the use of ZIP distribution on operational risk management, we develop two separate aggregate distributions. The first one is based on standard Poisson distribution and the second on ZIP distribution. Note that the severity model is the same for both aggregations. Results show that operational capital charge based on standard Poisson distribution is underestimated by 5% at a very high level of confidence (99.99%).

**Keywords:** operational risk, lda, poisson distribution, excess zeros, over-dispersion, zero-inflated poisson distribution, capital charge, basel ii accord

## 1. Introduction

In 1995, Barings bank (233 years) collapsed because of unauthorized speculations led by the trader Nick Leeson (Fay, 1996). This event was a starting point for intensified works related to operational risk initiated by Basel committee on banking supervision (BCBS). As a result, several definitions of operational risk had emerged (Eladlouni, Ezzahid and Mouatassim, 2011). The British Bankers Association, for example, defines operational risk as the "risk associated with human error, inadequate procedures and control, fraudulent criminal activities; the risks caused by technological shortcomings, system breakdowns; all risks which are not "banking" and arise from business decisions as competitive actions, pricing, etc.; legal risk and risk to business relationships, failure to meet regulatory requirements or an adverse impact on the bank's reputation; "external factors" including natural disasters, terrorist attacks and fraudulent activity, etc." (Britich Bankers Association, 1997).

The definition provided by BCBS (2001b) is the most popular. It defines operational risk as "the risk of loss resulting from inadequate or failed internal processes, people, systems, or from external events". It includes legal risk but excludes strategic, reputational and systematic risks. These definitions reveal the complexity of operational risk because of its various types and causes. To face this heterogeneity, BCBS (in its document Basel II) identified seven types of operational risk: "**i.** Internal fraud: an act intended to defraud, misappropriate property or avoid regulations, law or company policy, excluding diversity/discrimination events, which involve at least one internal party, **ii.** External fraud: an act of a type intended to defraud, misappropriate property or circumvent the law, by an external party, **iii.** Employment practices and workplace safety: an act inconsistent with employment, health or safety laws or agreements, from payment of personal grievance claims, or from diversity/discrimination events, **iv.** Clients, products, and business practices: an unintentional or negligent breakdown to meet a professional obligation to specific clients (including fiduciary and suitability requirements), or from the nature or design of a product, **v.** Damage to physical assets: the loss or damage to physical assets from natural disaster or other events, **vi.** Business disruption and system failures: disruption of business or system failures, and **vii.** Execution, delivery, and process management: failed transaction processing or process management, from relations with trade counterparties and vendors…" (Eladlouni, Ezzahid and Mouatassim, 2011, p. 100).

In Basel II framework, BCBS recommended to banks to develop more adequate techniques to identify, assess, control and mitigate the consequences of operational risk. In pillar I of Basel II, BCBS recommended to banks to determine the minimum capital requirements commensurate with their exposition to risks. For this reason, BCBS proposed three approaches: **1.** The Basic Indicator Approach (BIA), **2.** The Standardized Approach (SA), **3.** The

Advanced Measurement Approach (AMA). The basic indicator approach is based on flat percentage (15%) of positive gross income over the past three years. Under the standardized approach, banks are required to hold capital charge for each business line. Banks' activities are divided into eight business lines: **i.** Corporate finance, **ii.** Trading and sales, **iii.** Retail banking, **iv.** Commercial banking, **v.** Payment and settlements, **vi.** Agency services, **vii.** Assets' management, **viii.** Retail brokerage. A flat percentage, ranging from 12% to 18%, is applied to the three-years average positive gross income for each business line (Panjer, 2006). The aggregate minimum capital is the sum over all business lines.

The Advanced Measurement Approach (AMA) is based on the collected internal and external loss data. For a bank intending to use the AMA, it is required to prove, for each Basel II 56 risk cells (eight business lines × seven risk types), its ability to estimate unexpected losses based on the combined utilization of four sources of data: **i.** Internal loss data, **ii.** External data, **iii.** Scenario analysis, **iv.** Business environment and internal control factors (BEICFs). The AMA can be implemented via one of the four subsequent techniques: **i.** Scorecard Approach (SCA), **ii.** Scenario-Based AMA (sbAMA), **iii.** Internal Measurement Approach (IMA), **iv.** Loss Distribution Approach (LDA) (Karam and Planchet, 2011). Among these techniques, the LDA is the most used. It belongs to the family of probabilistic methods.

In this paper, we will focus on the LDA. To simplify the notation, we shall present LDA for one class of risk which is determined by one Basel II risk cell. Mathematical formulation of LDA necessitates the definition of a random sum of individual losses:

$$S = \sum_{i=1}^{N(t)} X_i \tag{1}$$

Where $S$ is the aggregate loss, $N(t)$ the number of losses occurring during the period $t$, and $X_i$ the severity of the ith individual loss. Note that $N(t)$ and $X_i$ are assumed to be random. The formula (1) shows that the aggregate loss can be decomposed into two variables the frequency and the severity of a loss.

In LDA, it is assumed that: i. $N(t)$ and $(X_1, X_2, ....)$ are independent random variables, ii. $(X_1, X_2, ....)$ is a set of independent and identically distributed random variables. Under these assumptions, the cumulative distribution' function of $S$ is given by:

$$F_S = P(S \le s) = \begin{cases} \sum_{n=0}^{\infty} P(N(t) = n) F_X^{n*}(s) & if \quad s > 0 \\ P(N(t) = 0) & if \quad s = 0 \end{cases}$$

Where $F_X^{n*}$ is $n - times$ the convolution of $F_X$ with itself, defined by:

$$F_X^{1*} = F_X$$
$$F_X^{n*} = F_X^{(n-1)*} * F_X$$

The mean and variance of $S$ are defined as follows:

$$E(S) = E(N(t)) * E(X)$$

$$Var(S) = E(N(t)) * Var(X) + Var(N(t)) * [E(X)]^2$$

Four methods can be used to obtain the convolution $F_S$ : **i.** the Fast Fourier Transforms (Rioux and Klugman, 2004), **ii.** the Recursive algorithm of Panjer (Panjer, 1981), **iii.** the Monte Carlo Simulation Approach (Rioux and Klugman, 2004), and **iv.** the Single-loss Approximation (Böcker and Klüppelberg, 2005). In practice, the required capital charge is measured as $(1 - \alpha) * 100$ quantile of the convolution $F_S$ over one year period. This quantile is also called value-at-risk (VaR) defined as follows:

$$VaR_{t+1/t} = F_S^{-1}(\alpha) \tag{2}$$

The total capital charge is obtained by aggregating the one year Value-at-Risk measures across all combinations, based on the compounded losses. To avoid the over estimation of the total capital charge, the correlations between Basel II cells defined within (8 business line × 7 type of risk) should be taken into account (Frachot and al., 2004).

The remaining of this paper is organized as following. The theoretical framework of Zero-inflated Poisson distribution is reviewed in section 2. In section 3, we model operational risk using LDA technique. The type of risk

concerned is Damage to Physical Assets (DPA). The frequency is modeled by Poisson and Zero-inflated Poisson distributions. The severity is modeled by Weibull distribution. Aggregate distributions and capital charge are developed in section 4.

## 2. Zero-inflated Poisson Distribution Framework

Poisson distribution is usually used to model operational risk events. However, the equality of mean and variance, assumed by Poisson distribution, is rarely confirmed in practice. Over-dispersion, which is an excess of variance from the mean, is due to two causes: heterogeneity of population and excess of zeros. The heterogeneity can be detected when population can be divided into sub-populations which are equi-dispersed. The excess of zeros phenomenon is shown when observed zeros exceed largely what is expected by a statistical distribution. In operational risk, the excess of zeros can be due to two main factors: scarcity of operational losses or the existence of a threshold above from which institutions start to collect losses (Chernobai and al., 2010).

Alternatively, Negative binomial distribution is usually used when over-dispersion is due to heterogeneity of data. However, this distribution remains unable to reproduce the number of observed zeros when data are zero inflated. An alternative way for modeling count events when zeros are preponderant is zero-inflated Poisson distribution (Johnson and Kotz, 1968). This distribution assumes that outcomes are generated by two processes. The first process attempts to model zero inflation outcomes by introducing a proportion of $1-\phi$ for extra zeros and a proportion $\phi e^{-\lambda}$ for zeros from Poisson distribution. The second process models the nonzero outcomes by zero-truncated Poisson distribution.

Let $N(t)$ be the number of operational losses occurred during the period t. Johnson and Kotz (1968) defined Zero-inflated Poisson distribution for the sequence $(N(t))_{0 \le t \le T}$ as follows:

$$P(N(t) = n) = \begin{cases} (1-\phi) + \phi e^{-\lambda} & if \quad n = 0 \\ \phi \lambda^n \dfrac{e^{-\lambda}}{n!} & if \quad n > 0 \end{cases} \tag{3}$$

The mean and the variance of the Zero-inflated poisson (ZIP) distribution are given by: $E[N(t)] = (1-\phi)\lambda$ and $Var[N(t)] = (1-\phi)(\lambda + \phi \lambda^2)$. Johnson and Kotz (1968) have proposed different approaches using the method of moments for estimating the parameter $\phi$. Martin and Katti (1965) have used the maximum likelihood method to estimate the parameters of ZIP distribution. In this paper, maximum likelihood method is used for obataining the estimated values of $\lambda$ and $\phi$. The likelihood function is:

$$L(\lambda, \phi) = \prod_{k=0} \left( \phi + (1-\phi)e^{-\lambda} \right) \prod_{k>0} (1-\phi)\lambda^k \frac{e^{-\lambda}}{k!} \tag{4}$$

Suppose that $N$ is the number of outcomes and $N_0$ is the number of zeros in the data. The likelihood becomes:

$$L(\lambda, \phi) = \left( \phi + (1-\phi)e^{-\lambda} \right)^{N_0} (1-\phi)^{(N-N_0)} \prod_{n>0} \lambda^n \frac{e^{-\lambda}}{n!} \tag{5}$$

The log likelihood is:

$$\ln L(\lambda, \phi) = N_0 \ln\left( \phi + (1-\phi)e^{-\lambda} \right) + (N - N_0)\left[ \ln(1-\phi) - \lambda \right] + \ln(\lambda)\sum_{n>0} n - \sum_{n>0} \ln(n!) \tag{6}$$

The equilibrium conditions are:

$$\frac{\partial}{\partial \lambda} \ln L(\lambda, \phi) = -N_0 \frac{(1-\phi)e^{-\lambda}}{\phi + (1-\phi)e^{-\lambda}} - (N - N_0) + \frac{\sum\limits_{n>0} n}{\lambda} = 0 \tag{7}$$

$$\frac{\partial}{\partial \phi} \ln L(\lambda, \phi) = N_0 \frac{1 - e^{-\lambda}}{\phi + (1-\phi)e^{-\lambda}} - (N - N_0)\frac{1}{(1-\phi)} = 0 \tag{8}$$

Dividing the equation (7) by (8):

$$\frac{\lambda}{1-e^{-\lambda}} = \frac{\sum_{k>0} n}{N - N_0} = \frac{N\bar{n}}{N - N_0} \tag{9}$$

It is clear that equation $(9)$ is only function of $\lambda$. Hence, numerical algorithms can be used to find the estimated value of $\lambda$, denoted by $\hat{\lambda}$. The estimated value of $\phi$, denoted by $\hat{\phi}$, is then determined easily by replacing $\lambda$, in (7) or (8), by its estimated value $\hat{\lambda}$.

Singh (1963) found the following formulas:

$$Var\left(\hat{\lambda}\right) \cong (1-\phi)^{-1} \lambda \left(1 - e^{-\lambda}\right)\left(1 - e^{-\lambda} - \lambda e^{-\lambda}\right)^{-1} \tag{10}$$

$$Var\left(\hat{\phi}\right) \cong (1-\phi)\left[\phi\left(1 - e^{-\lambda}\right) + (1-\phi)e^{-\lambda}\right]\left(1 - e^{-\lambda} - \lambda e^{-\lambda}\right)^{-1} \tag{11}$$

Yoneda (1962) and Johnson and Kotz (1968) have extended the zero-inflated Poisson distribution. They developed a general modified Poisson distribution. Under this distribution, any kind of excess in frequency data is allowed, that is the values $n = 0, 1, 2, ..., K$ are inflated counts while the remaining values $n = K+1, K+2, ..., N$ follow a Poisson process (Johnson and al., 1992).

Lambert (1992) has introduced covariates into Zero-inflated Poisson distribution and has fitted it to defects in manufacturing data. Recently, Lambert's model, called also Zero-inflated Poisson regression, becomes largely used in many fields such as public health, epidemiology, psychology and others.

## 3. Operational Risk Modeling

In this section, we develop LDA model for Damage to Physical Assets (DPA). The data set contains daily internal events from March 21, 2007 through February 01, 2009. We ignore the notion of business line because of scarcity of data. Table 1 shows some descriptive statistics of our data.

Table 1. Descriptive statistics

| Statistics | Freq. | Severity |
|---|---|---|
| Minimum | 0.00 | 1.22 E+003 |
| Maximum | 5.00 | 3.32 E+005 |
| Mean | 0.65 | 4.80 E+004 |
| Standard deviation | 1.09 | 4.06 E+004 |
| Median | 0.00 | 4.41 E+004 |
| Coefficient of variation | 1.68 | 0.84 |
| Skewness coefficient (Cs) | 2.02 | 2.56 |
| Kurtosis coefficient (Ck) | 4.04 | 12.19 |

To compute capital charge using LDA technique, we have to model separately frequency and severity of operational losses. For frequency modeling, we fit Poisson distribution and Zero-inflated Poisson distribution. The severity, in its turn, is modeled using Weibull distribution. The convolution is determined using the Monte Carlo Simulation Approach.

### 3.1 Modeling Operational Risk Frequency

The mean of DPA frequency is 0.654 and the variance is 1.20. We note that the variance is greater than the mean. This fact is called over-dispersion in counts data modeling. Since the histogram (see figure1) is highly picked at zero values, we conclude that our DPA frequency is zero inflated. The histogram shows, also, that the ZIP distribution expects the number of zeros value better than the standard Poisson distribution. To determine an appropriate distribution for data, we use two Goodness-of-Fit Tests: QQ plot test and Kolmogorov-Smirnov (KS) test. The first one is graphical and the second is analytical. These tests allow testing the equality of the distributions of two

samples. The hypotheses are defined as follows:

$H_0$:  $N(t)$    follows theoretical distribution fitted to data.

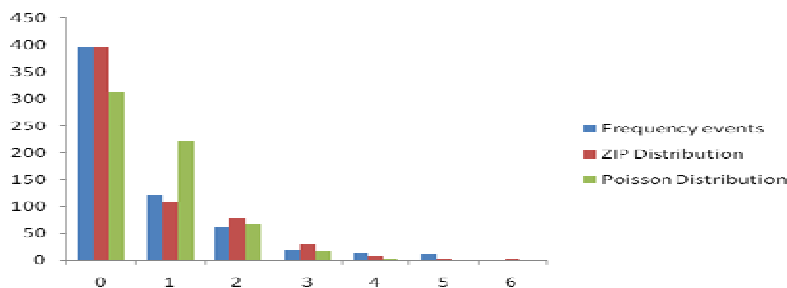$H_1$ :  $N(t)$  does not come from theoretical distribution fitted to data.



Figure 1. Histogram of DPA frequency

Figure 2 shows the results of QQ plot test. Based on this test, we can state that ZIP distribution fits frequency data better than the standard Poisson distribution.
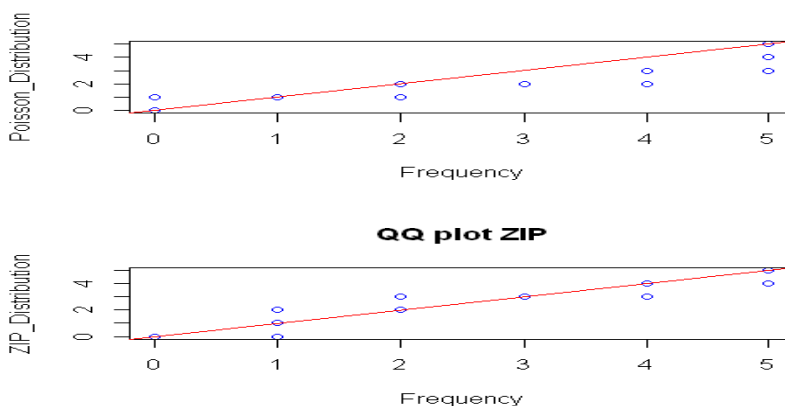


Figure 2. QQ plot tests

To support or deny analytically the results carried out by QQ plot test, we use KS test which tests the equality of empirical distribution and a candidate theoretical distribution. If the P-value associated is lower than significance level (here 5%),  $H_0$  is rejected. If P-value is greater than significance level test (here 5%), we can state that sample data comes from the candidate distribution.

Table 2 shows that the occurances of Damage to capital assets does not come from standard Poisson distribution since the P-value associated to KS test (P-value=0.006073) is lower than 5%. However, the KS test (P-value=0.9333) leads to the equality of the empirical distribution of frequency data and ZIP distribution (with parameters mu = 1.3244026 and sigma = 0.5064397).

Table 2. Kolmogorov Smirnov test

| Distribution | statistic | P-value |
|---|---|---|
| Poisson | D = 0.0966 | 0.006073 |
| ZIP | D = 0.0306 | 0.9333 |

Based on both tests (QQ plot and KS test) and referring to the histogram, we can state that frequency data with excess of zeros are bad fitted by standard Poisson distribution. In such case, standard Poisson distribution is unable

to reproduce the observed number of zeros in the sample. It underestimates the observed dispersion. In contrary, the number of expected zeros generated by Zero-inflated Poisson distribution is close to the number of observed zeros. This power of ZIP distribution emanates from the fact that it assumes a model for zero outcomes and another model for nonzero outcomes.

*3.2 Modeling Operational Severity*

Lognormal, Gamma, Weibull, Pareto and other continuous parametric distributions can be candidates for fitting the severity of operational risk losses (Frachot, 2001 and Nigel, 2004). If losses are extremes, picks-over-thresholds are used (Gourier and al., 2009). For our data, the QQ plot test (See Figure 3) and KS test (D = 0.0734, P-value = 0.5999) prove that the Weibull distribution is an appropriate distribution for operational risk losses subject of our analysis.
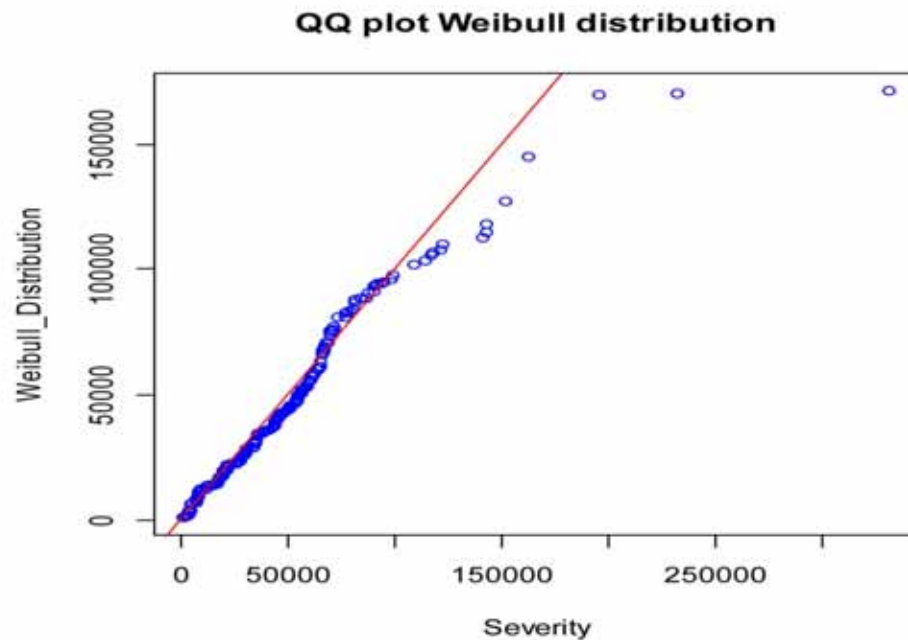


Figure 3. QQ plot test for severity

## 4. Determination of Capital Charge and Discussion

In this section, we estimate the capital charge a bank needs for facing the expected losses due to Damage to Capital Assets (DPA). The Monte Carlo Simulation Approach is used to calculate the aggregate distribution which combined frequency and severity. The steps of the utilized algorithm are:

Step 1. Generate $n$ number of losses per day using ZIP distribution or Poisson distribution fitted to internal data.

Step 2. Generate $n$ losses severity $X_i(1,2,...,n)$ per day using Weibull distribution fitted to internal data.

Step 3. Repeat Step 1 and Step 2 - 365 times and do the sum of all generated $X_i$, noted $S$.

Step 4. The total annual loss based on internal data is S.

Step 5. Repeat *Step1* through *Step 4* a large number of times (*1.000.000*) for obtaining the aggregate distribution of total loss.

Step 6. The VaR is then calculated as $(1-\alpha)*100$ quantile of the empirical distribution of total losses.

To assess the impact of ZIP distribution on the operational capital charge, we constitute two aggregate distributions. The difference between them relies in the distribution of frequency data. The first aggregate distribution is determined by standard Poisson and Weibull distributions whereas the second is a combinaison of ZIP and Weibull distributions.

Table 3. VaR for different significance levels

| 1-α | VaR1: VaR based Poisson | VaR2: VaR based ZIP | Ratio : VaR1/VaR2 |
|---|---|---|---|
| 99.99% | 15 137 520 | 15 858 978 | 95% |
| 99.90% | 14 577 165 | 15 143 577 | 96% |
| 99.50% | 14 032 680 | 14 504 826 | 97% |
| 99.00% | 13 789 075 | 14 210 949 | 97% |

Table 3 presents the capital charge for DPA risk for a given significance level (Column 1) estimated by aggregate distribution based on standard Poisson distribution (Column 2) and aggregate distribution based on ZIP distribution (Column 3). The fourth column gives the ratio (VaR1/VaR2) in order to show the difference between the two models. The results in table 3 shows that the capital charge issued from aggregate distribution based on standard Poisson distribution is lower than that emanated from model based on ZIP distribution. In addition, the differences increase with tolerance level. At 99.99%, the capital charge based on standard Poisson distribution is 5%-lower than that based on ZIP distribution. Since both aggregate distributions are based on the same model of severity, these differences are due to the distribution used for modeling frequency. As shown above, Standard Poisson distribution underestimates the observed dispersion when data are zero-inflated. Thus, the capital charge calculated using aggregate model based on it is underestimated and cannot reflect the reality of bank's losses.

Other zero-inflated discrete distributions can be utilized for modeling operational risk frequency in the case of the preponderance of observed zeros. We cite for example zero-inflated binomial, zero-inflated negative binomial etc…. Zero-inflated negative binomial distribution, for example, is preferred when data are over-dispersed due to heterogeneity and excess zeros simultaneously.

## References

Basel Committee on Banking Supervision. (2001b). *Sound Practices for the Management and Supervision of Operational Risk.*

Basel Committee on Banking Supervision. (2006). *International Convergence of Capital Measurement and Capital Standards - A Revised Framework*. Basel: Bank for International Settlements.

Böcker K., & Klüppelberg C. (2005). Operational VaR: a closed-form solution. *RISK Magazine*, December, 90–93.

Böcker K., & Sprittulla J. (2006). Operational VaR: meaningful means. *Risk Magazine*, December, 96–98.

Britich Bankers Association. (1997). *Operational risk management survey*.

Chernobai C., Menn C., Rachev ST., & Stefan Trück S. (2010). *Estimation of Operational Value-at-Risk in the Presence of Minimum Collection Thresholds*. Risk Books, London.

El Adlouni S., Ezzahid E., & Mouatassim Y. (2011). Mixed Distributions for Loss Severity Modelling with zeros in the Operational Risk losses. *International Journal of Applied Mathematics & Statistics, 21*, 96- 109.

Fay, S. (1996). *The Collapse of Barings*. New York: W.W. Norton.

Frachot A., Georges P., & Roncalli T. (2001). Loss Distribution Approach for Operational Risk. Working Paper, Crédit Lyonnais, Groupe de Recherche Opérationnelle.

Frachot A., Roncalli T., & Salomon E. (2004). The Correlation Problem in Operational Risk. Working Paper, Crédit Lyonnais, Groupe de Recherche Opérationnelle.

Gourier E., Farkas W., & Abbate D. (2009). Operational risk quantification using extreme value theory and copulas: from theory to practice. *The Journal of Operational Risk, 4*, 1-24.

Johnson N. L., & Kotz S. (1968). *Discrete Distributions*. New York: Wiley. Mood, A. M.

Johnson N. L., Kotz S., & Kemp A. W. (1992). *Univariate Discrete Distributions.* John Wiley & Sons, Wiley-interscience publication.

Karam E., & Planchet F. (2011). Operational Risks in Financial Sectors. Cahiers de recherche de l'ISFA, WP21X.

Lambert D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, *34*, 1-14.

Martin D., & Katti S. (1965). Fitting of some contagious distributions to some available data by the maximum likelihood method. *Biometrics, 21*, 34–48. http://dx.doi.org/10.2307/2528350

Nigel Da Costa L. (2004). *Operational Risk with Excel and VBA, Applied Statistical Methods for Risk Management.* John Wiley & Sons.

Panjer H. (1981). Recursive evaluation of a family of compound distributions. *ASTIN Bulletin (International Actuarial Association), 12*(1), 22–26.

Panjer H. (2006). *Operational risk modeling analytics.* John Wiley & Sons.

Rioux J., & Klugman S. (2004). Toward a unified approach to fitting loss models. http://www.iowaactuariesclub.org/library/lossmodels.pdf

Singh S. N. (1963). A note on zero inflated Poisson distribution. *Journal of the Indian Statistical Association, 1*, pp. 140-144.

Yoneda K. (1962). Estimations in Some Modified Poisson Distributions. *Yokohama Mathematical Journal, 10*, 73-96.