The Cascade Bayesian Approach for a controlled integration of internal data, external data and scenarios

Bertrand Hassani, Alexis Renaudin

The Cascade Bayesian Approach for a controlled integration of internal data, external data and scenarios

Bertrand K. HASSANI, Alexis RENAUDIN

The Cascade Bayesian Approach for a controlled integration of internal data, external data and scenarios

February 12, 2013

Authors

- Bertrand K. Hassani: Santander UK and Université Paris 1 Panthéon-Sorbonne CES UMR 8174, 2 Triton Square NW1 3AN, London, United Kingdom, phone: +44 (0)2077565147, e-mail: bertrand.hassani@santander.co.uk

- Alexis Renaudin: Aon Global Risk Consulting, 8 Devonshire Square, EC2M 4PL London, United Kingdom, phone: +44 (0)2070860359, e-mail: alexis.renaudin@aon.co.uk

Abstract

According to the last proposals of the Basel Committee on Banking Supervision, banks under the Advanced Measurement Approach (AMA) must use four different sources of information to assess their Operational Risk capital requirement. The fourth including "business environment and internal control factors", i.e. qualitative criteria, the three main quantitative sources available to banks to build the Loss Distribution are Internal Loss Data, External Loss Data, and Scenario Analysis. This paper proposes an innovative methodology to bring together these three different sources in the Loss Distribution Approach (LDA) framework through a Bayesian strategy. The integration of the different elements is performed in two different steps to ensure an internal data driven model is obtained. In a first step, scenarios are used to inform the prior distributions and external data informs the likelihood component of the posterior function. In the second step, the initial posterior function is used as the prior distribution and the internal loss data inform the likelihood component of the second posterior. This latter posterior function enables the estimation of the parameters of the severity distribution selected to represent the Operational Risk event types.

Key words: Operational Risk, Loss Distribution Approach, Bayesian Inference, Markov Chain Monte Carlo, Extreme Value Theory, Non-parametric statistics, Risk Measures.

1   Introduction

Basel II requires bank to access the capital charge associated to operational risks (BCBS (2001; 2010)). There are three different approaches to measure operational risks, the basic, the standard and the advanced measurement approach (AMA), representing increasing levels of control and difficulty of implementation. The AMA requires a better understanding of the exposure to implement an internal model.

Basel II capital accord defines the capital charge as a risk measure obtained on an annual basis, at a given confidence level on a loss distribution that integrates four sources of information: internal data, external data, scenario analysis, business environment and internal control factors. The regulatory capital is given by the 99.9% percentile of the Loss Distribution and the economical capital by a higher percentile related to the rating of the financial institution which is usually between 99.95% and 99.98%.

The purpose of using multiple sources of information is to build an internal model based on the largest set of data possible, in order to increase the robustness, the stability and the conservatism of the final capital evaluation. However, the different sources of information have different characteristics which taken in isolation can be misleading. Internal loss data represent the entity risk profile, external loss data characterize the industry risk profile and scenarios offer a forward looking perspective and the unexpected loss from an internal point of view (Guégan and Hassani (2012a)). Figure 1 illustrates this point by assuming that the internal loss data tends to represent the body of the severity distribution, the scenarios the extreme tail and the external data the section in between[1].

This paper expands on the Frequency $\times$ Severity framework (Loss Distribution Approach), where the Loss Distribution Function (Frachot et al. (2001), Cruz (2004) and Chernobai et al. (2007)) G is a weighted sum of γ-fold convoluted severity F distribution where γ represents the order of convolution. The weight is provided by the frequency distribution p. Mathematically this

---

[1]External loss data usually overlap both internal data and scenario analysis and is therefore represented as a link between the two previous components.
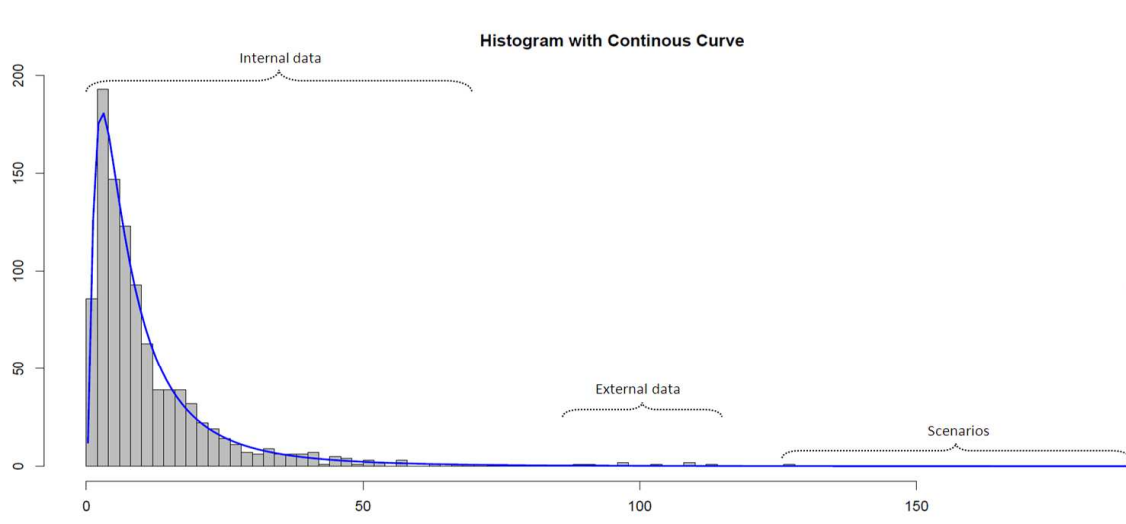
Figure 1: Combination of internal loss data, external loss data and scenario analysis: depending on the risk profile or the quantity of data available, the representation may be slightly different.

4

function corresponds to:

$$G = \sum_{\gamma=1}^{\infty} p(k; \bullet) F^{\otimes \gamma}(x; \bullet), \, x > 0, \tag{1.1}$$

with

$$G = 0, \, x = 0.$$

$\otimes$ denotes the convolution operator. Denoting g the density of G, we have

$$g = \sum_{k=\gamma}^{\infty} p(k; \bullet) f^{\otimes \gamma}(x; \bullet), \, x > 0. \tag{1.2}$$

Assuming a Poisson distribution to model the frequencies, the parameter is estimated by maximum likelihood on the internal loss data. However, as collection thresholds are being set up, the frequency distribution parameter is adjusted using the parameterised severity distribution. Consequently, the frequency distribution will indirectly be informed by the three different components.

The focal point of this paper is to construct severity distributions by combining the three data sources presented above. As the level of granularity of the risk taxonomy increases, the quantity of data available per risk category tends to decrease. As a result, the traditional estimation methods such as the maximum likelihood or the method of moments tend to be less reliable. Consequently, we opted to bring together the three components within the Bayesian Inference theoretical framework (BI) (Box and Tiao (1992), Shevchenko (2011)). Despite the numerous hypotheses surrounding the BI, it is known to be efficient in situations where there are only a few data points available.

In statistics, the Bayesian inference is a statistical method of inference in which Bayes' theorem (Bayes and Prince (1763)) is used to update the probability estimate of a proposition as additional information becomes available. The initial degree of confidence is called the prior and the updated degree of confidence, the posterior.

Consider a random vector of loss data $\boldsymbol{X} = (X_1, ..., X_n)$ whose joint density for a given vector of parameters $\boldsymbol{\phi} = (\phi_1, ..., \phi_K)$, is $h(\boldsymbol{x}|\boldsymbol{\phi})$. In the Bayesian approach, both observations and parameters are considered to be random. Then the joint density is

$$h(\boldsymbol{x}, \boldsymbol{\phi}) = h(\boldsymbol{x}|\boldsymbol{\phi})\pi(\boldsymbol{\phi}) = \pi(\boldsymbol{\phi}|\boldsymbol{x})h(\boldsymbol{x}), \tag{1.3}$$

5

where, $\pi(\boldsymbol{\phi})$ is the probability density of the parameters, known as the prior density function. Typically, $\pi(\boldsymbol{\phi})$ depends on a set of further parameters that are usually called "hyper" parameters [2]. $\pi(\boldsymbol{\phi}|\boldsymbol{x})$ is the density of parameters given data $\boldsymbol{X}$, known as the posterior density, $h(\boldsymbol{x}, \boldsymbol{\phi})$ is the joint density of observed data and parameters and $h(\boldsymbol{x}|\boldsymbol{\phi})$ is the density of observations for given parameters. This is the same as a likelihood function if considered as a function of $\boldsymbol{\phi}$, i.e. $l_{\boldsymbol{X}(\boldsymbol{\phi})} = h(\boldsymbol{x}|\boldsymbol{\phi})$. $h(\boldsymbol{x})$ is the marginal density of $\boldsymbol{X}$ that can be written as $h(\boldsymbol{x}) = \int h(\boldsymbol{x}|\boldsymbol{\phi})\pi(\boldsymbol{\phi})d|\boldsymbol{\phi}$.

The Bayesian inference approach permits a reliable estimation of distributions' parameters even if the quantity of data denoted n is limited. And as n becomes larger, the weight of the likelihood component increases such that if $n \to \infty$, the posterior distribution tends to the likelihood function, and consequently parameters obtained from both approaches converge. As a result, the data selected to inform the likelihood component may lead the model and, as a consequence the capital charge. Therefore, in our two-step approach, there is an opportunity for operational risk managers and modelers to integrate all the aforementioned components in a way they do not have to justify a capital charge increase due to an extreme loss an another entity would have suffered.

This paper proposes applying two Bayesian inference approaches sequentially in order to obtain the parameter of the statistical distribution used to characterise the severity. Scenarios are used to build the prior distributions of the parameters denoted $\pi(\phi)$, which is refined using external data to inform the likelihood component. This results in an initial posterior function $(\pi(\phi|Y))$ which is then used as a prior distribution and the likelihood part is informed by the internal data. This leads to a second posterior distribution which allows the estimation of the parameters of the severity distribution used to build the loss distribution function in the LDA approach.

The next section presents the Cascade strategy and the underlying assumption allowing a data driven model. The third section deals with the implementation in practice using real data sets. A fourth section presents the results and a fifth section concludes.

---

[2] These parameters are omitted here for simplicity of notation, these are the parameters of the densities used as a prior distribution (e.g. a Gamma or Beta distribution).

2 A Bayesian Inference in Two steps for Severity Estimation

This section outlines the two step approach to parametrize the severity distribution. The cascade implementation of the Bayesian inference approach is justified by the following property. The Bayesian posterior distribution implies that the larger the quantity of data used, the larger the weight of the likelihood component. Consequently,

$$\pi(\phi; x_1, ..., x_k) \propto \pi(\phi) \prod_{i=1}^{k} f_i(x_i|\phi) \stackrel{k \to \infty}{\Rightarrow} \prod_{i=1}^{k} f_i(x_i|\phi), \tag{2.1}$$

where, f is the density characterizing the severity distribution, x_i a data point, and ϕ the set of parameters to be estimated. As a result, the order of the Bayesian integration of the components is significant.

This strategy is based on the construction of two successive posterior distribution functions. The first uses the scenario values to inform the prior and the external loss data to inform the likelihood function. Using a Markov Chain Monte Carlo algorithm (Gilks et al. (1996)), a first posterior empirical distribution is obtained. This is used as prior distribution in a second Bayesian approach for which the likelihood component is informed by internal loss data. Due to the Bayesian Approach property presented above, in the worst case, the final posterior distribution is entirely driven by internal data. The method may be formalised as follows. f is the density function of the severity distribution, y_i represent the internal data, x_i the external data and ϕ the set of parameters to be estimated.

1. Prior π_0 is informed by the scenarios and the likelihood component by external data:

$$\pi_1(\phi; x_1, ..., x_k) \propto \pi_0(\phi) \prod_{i=1}^{k} f_i(y_i|\phi). \tag{2.2}$$

2. The aforementioned posterior π_1 used as prior and the likelihood component is informed by internal data:

$$\pi_2(\phi; x_1, ..., x_k) \propto \pi_1(\phi) \prod_{i=1}^{k} f_i(x_i|\phi). \tag{2.3}$$

As a result, the first posterior function becomes a prior, and only a prior, limiting the impact of the first two components on the parameters.

Level 1	Level 2	Nb used	min	median	mean	max	sd	skewness	kurtosis
Internal Fraud	Global	665	4 137	28 165	261 475	46 779 130	1,9E+06	21,4	407,1
External Fraud	Payments	1 567	4 091	12 358	36 133	1 925 000	9,2E+04	11,7	185,2
Execution, Delivery & Process Management	Other	3 602	4 084	10 789	96 620	30 435 400	9,5E+05	24,8	653,8

Table 1: The table presents the statistical moments of the internal loss data used in this paper, as well as other statistics such as the minimum value, the maximum value and the number of data point available. NB: Level 2 "Other" gathers "Execution, Delivery & Process Management" losses other than "Financial Instruments" and "Payments".

Level 1	Level 2	Nb used	min	median	mean	max	sd	skewness	kurtosis
Internal Fraud	Global	2 956	20 001	88 691	697 005	130 715 800	4,5E+06	17,6	387,6
External Fraud	Payments	1 085	20 006	36 464	326 127	106 772 200	4,3E+06	20,8	461,2
Execution, Delivery & Process Management	Other	31 126	20 004	47 428	271 974	585 000 000	4,1E+06	107,2	14068,6

Table 2: The table presents the statistical moments of the external loss data used in this paper, as well as other statistics such as the minimum value, the maximum value and the number of data available. NB: Level 2 "Other" gathers "Execution, Delivery & Process Management" losses other than "Financial Instruments" and "Payments".

3   Carrying out the Cascade Approach In Practice

This section presents how the methodology may be carried out in practice. Firstly we will introduce the data, secondly we will detail the approach and the estimation, and finally we will present the resulting capital charges.

3.1   The Data Sets

The results presented were obtained using OpBase, the external data base developed by Aon Limited. It captures the losses corresponding to the Operational risk claims and the losses in the public domain(PKM). The scenarios correspond to those we obtained through a scenario workshop process. There are many different scenario approaches that are compatible with our approach, therefore this will not be discussed any further (Guégan and Hassani (2012a)) in this paper. the internal loss data were provided by a first tier European bank.

8

Level 1	Level 2	1 in 10	1 in 40
Internal Fraud	Global	6,0E+06	5,2E+07
External Fraud	Payments	1,5E+06	2,5E+07
Execution, Delivery & Process Management	Other	2,5E+07	5,0E+07

Table 3: The table presents the scenario values used in this paper. 1 in 10 and 1 in 40 respectively denote the biggest loss that may occur in the next 10 and 40 years. NB: Level 2 "Other" gathers "Execution, Delivery & Process Management" losses other than "Financial Instruments" and "Payments".

3.2 The Priors

This approach is applicable to any type of the traditional statistical distribution. In this paper, Internal Fraud has been modeled using a lognormal distribution, External Fraud/Payment via by a Weibull distribution and Execution, Delivery & Process Management / Other than Payments is represented by a mixture model using a lognormal distribution in the body and a Generalized Pareto Distribution at the tail (Guégan et al. (2011)).

Two criteria drive the selection of the prior distributions. They are either chosen so that their supports are consistent with the acceptable values for the parameters, e.g. the shape parameter of a Generalized Pareto distribution cannot be greater than 1, otherwise the infinite mean model obtained leads to unrealistic capital values. Therefore, a beta distribution defined on a finite support has been selected. Alternatively, the priors are selected such that the joint prior is a conjugate distribution i.e. the posterior distribution belongs to the same family as the prior distribution, with a different set of parameters.

In the general case, the prior π on the severity parameters can be written as:

$$\pi(\phi) = \pi(\phi_1, \phi_2) = \pi(\phi_1|\phi_2)\pi(\phi_2) = \pi(\phi_2|\phi_1)\pi(\phi_1)$$

If the priors are independent then the previous function becomes

$$\pi(\phi) = \pi(\phi_1, \phi_2) = \pi(\phi_1)\pi(\phi_2)$$

9

Label	Density	Parameters
Beta	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$	$\alpha = shape,\ \beta = scale$
Gamma	$\frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}$	$\alpha = shape,\ \beta = scale$
Gaussian	$\frac{1}{b\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2b^2}}$	$a = location,\ b = standard\,Deviation$

Table 4: This table presents the priors used to parametrize: 1 - the Lognormal distribution, i.e. the Gaussian distribution for μ and the Gamma distribution for σ, 2 - the Weibull distribution, i.e. two Gamma distributions for the shape and the scale, 3 - the GPD, i.e. a Beta distribution for the shape and a Gamma for the scale. These are informed by the scenarios.

Thereafter, priors are assumed to be independent.

For each distributions parameters, a set of prior distributions have been selected (Table 4 lists the corresponding prior density functions):

- Lognormal - a Gaussian and a Gamma distribution

- Weibull - Gamma distributions for both the scale and the shape

- Generalized Pareto Distribution - A Beta distribution for the shape and a Gamma distribution on the scale

3.3 Estimation

To obtain the global severity parameters[3] of the selected distributions, three different approaches may be carried out. These are listed below in order of complexity.

1. One can choose conjugate priors for the parameters in the first step of the Bayesian Inference estimation. In this case, the (joint) distribution of the posterior parameters is directly known and it is possible to sample directly from this distribution to recreate the marginal posterior distributions of each severity parameter. These may then be used to generate the posterior empirical densities required to compute a Bayesian point estimator of the parameters. Admissible estimators are the median, mean, and mode of the posterior distribution.

---

[3]These take into account the 3 different data sources

The mode (also called MAP, Maximum A Posteriori) can be seen as the 'most probable estimator' and ultimately coincides with the Maximum Likelihood Estimator (Lehmann and Casella (1998)). Despite having good asymptotic properties, finding the mode of an empirical distribution is not a trivial matter and often requires some additional techniques and hypothesis (e.g. smoothing). This paper therefore uses posterior means is used as point estimators.

To our knowledge, the only conjugate priors for continuous distributions were studied by Shevchenko (2011), for the lognormal severity case . Conjugate approach requires some assumptions that may not be sustainable in practice, particularly for priors that are modelled with 'uncommon' distributions (e.g. Inverse Chi Square). This might lead to difficulties in the step known as "elicitation", i.e. calibrating the prior hyper-parameters from the chosen scenario values.

2. Another solution is to release the conjugate prior assumption, and to use a Markov Chain Monte Carlo approach in the first step to sample from the first posterior. One can then use a parametric or non-parametric method to compute the corresponding densities. This enables the posterior function π_2 to be evaluated (see equation 2.3). Maximizing this function directly gives the MAP estimators of the severity parameters (see above). Even if this method is sufficient to compute values for the global parameters, it misses the purpose of the Bayesian Inference, which is to provide a distribution as a final result instead of a single value. Additionally, it may also suffer from all the drawbacks that an optimization algorithm may suffer (e.g. non-convex functions, sensitivity to starting values, etc.)

3. The final alternative is to use a MCMC approach at each step of the aforementioned Cascade Bayesian Inference. This method is more challenging to implement but is the most powerful as it generates the entire distribution of the final severity parameters, from which any credibility intervals and/or other statistics may be evaluated.

In this paper, the third alternative has been implemented using the Metropolis-Hasting algorithm. This allows sequential sampling from the two posterior distributions. This algorithm then enables us to build the distributions for the parameters contained in the ϕ vector. However, these distributions are used as priors in the next step of the cascade and as mentioned in

11

the second point above, their densities are required from the sample obtained from the MCMC. Two solutions may be carried out, a parametric and a non parametric one. The parametric solution requires fitting statistical to empirical distributions, which may bias the construction. In order to remain as close as possible to the empirical distributions generated, and therefore to the data, a non parametric approach based on a Kernel density estimation has been chosen. Using an Epanechnikov kernel and a cross validation method to calibrate the bandwidth value, we obtain the non parametric density representing the parameter's distributions (Appendix A and B).

As a result, each of the different elements of the new posterior distribution have been built. The new prior densities in the second step result from the first posterior. The new likelihood function is informed by another data set, for instance the Internal Loss data to guarantee an internal loss data driven model.

Remark 3.1. *The innovation of this paper lies in the Cascade Bayesian Approach and its use to combine different sources of information; consequently little emphasis has been placed on our implementation of the MCMC with the Metropolis-Hasting algorithm, which is a known and widespread topic in literature. The interested reader could for instance refer to Gilks et al. (1996).*

4   Results

The Cascade Bayesian approach enables the updating of the parameters of the distributions considered to model particular risk events. Table 7 presents the parameters estimated considering the different pieces of information used to build the LDF, i.e. the parameters of the selected distributions - for instance, the lognormal, the Weibull and the Generalized Pareto distribution - are estimated on each of these pieces of information without considering the information brought by the others. It results in a large variance in the parameters, which once translated into financial values may be inconsistent with the observations. For example, considering the lognormal distribution for Internal Fraud, the median is equal to $40,783.42$ when considering only the scenarios, $6,047.53$ using the external data and $11,472.33$ considering internal data. The parameter implied variance of the losses is also quite different across the different elements.

12

The variation is similar for the Weibull distribution that models external fraud. It may be even worse in the case of the GPD, as the risk measures are extremely sensitive to the shape parameter.

Table 8 shows the evolution of the parameters as they are updated with respect to the different pieces of information.

- Scenarios Severity is derived from the calibration of the priors distributions to the scenario values. The theoretical means obtained from the calibrated priors provide scenario severity estimates.

- Intermediate Severity refers to the estimation of the severity on the first obtained posterior, i.e. the mean of the posterior distribution obtained from scenario values updated with external loss data.

- Similarly, final severity represents the severity estimation on the second and last posterior distribution, which includes scenarios, external and internal loss data. A 95% Bayesian Confidence Interval (also known as a Credible Interval) derived from this final posterior distribution is also provided. It is worth noticing that this interval – formally defined as containing 95% of the distribution mass – is not unique and is chosen here as the narrowest possible interval. It is therefore not necessarily symmetric around the posterior mean estimator.

From implementing the Cascade approach, it results in final parameters located within the range of values obtained from the different elements taken independently. As the dispersion of the information increases, the variance of the theoretical losses tends to increase, as does the theoretical kurtosis. For example, the evolution of the lognormal parameters is significant, as the introduction of data belonging to the body (Table 7, first line) tends to decrease the mean (the μ parameter is the mean of the log-transformed of the data), but naturally the dispersion increases, and therefore, the variance tends to increase (the σ parameter is a representation of the standard deviation of the log-transform of the data) (Table 8, first line). The impact of the severity on frequency is given in Table 5. The conditional frequencies are naturally increasing with respect to the severity distributions.
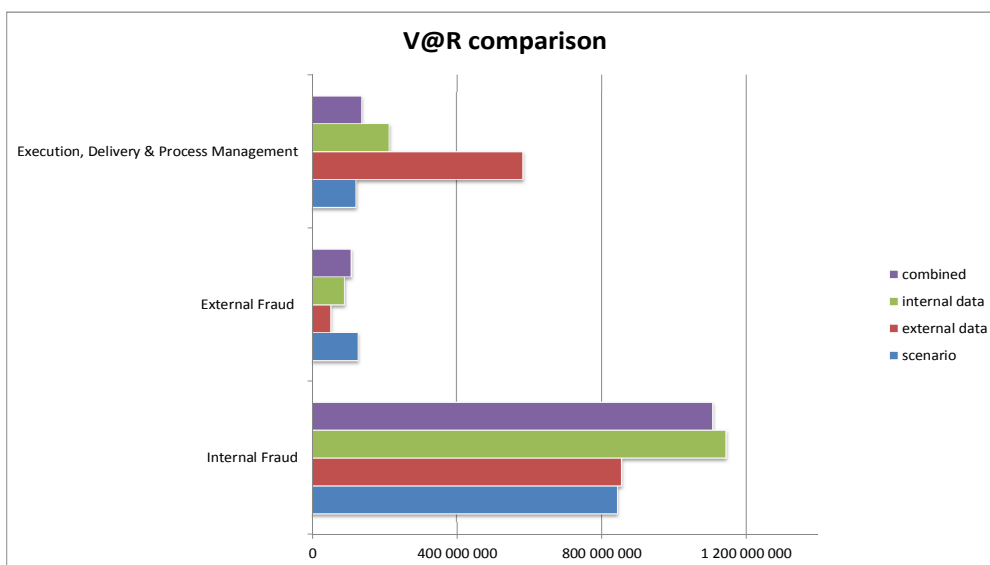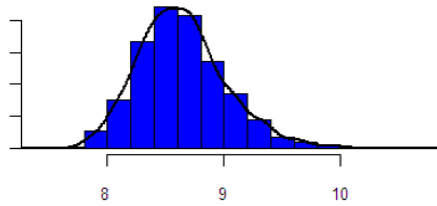
13

Figure 2: Comparison of Capital charge value obtained from the three components on a stand-alone basis and in combination.
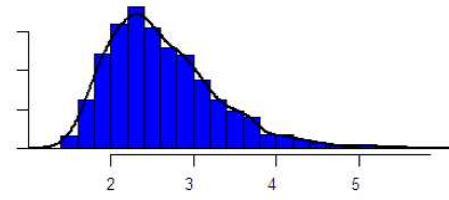
It also results in conservative capital charges considering the Value-at-Risk (Appendix D) obtained from the different elements taken independently, as it tends to be a weighted average of the VaR obtained from each part. The weights are automatically evaluated during the cascade and directly related to the quantity and the quality of information integrated.

14

As a final step, these results may be illustrated for the whole Bayesian Cascade estimation in the simple lognormal case. Figure 3 shows the final posterior distribution obtained for μ and σ. This is used to derive the 95% Bayesian Credible Interval given in Table 8. We also show the results of the convergence of the posterior mean as a function of the number of MCMC iterations for μ and σ. One can see that the values stabilize after 100 to 500 iterations in this example. In the general case we chose to sample 3000 values and discard the first 1000 MCMC generated values (*burn-in period*) to ensure the stability of our estimates.

15

(a) Obtained Posterior Distribution for each severity parameter (μ and σ)

(b) Convergence of the Severity Estimation as the Posterior Mean in the MCMC Sampling

Figure 3: Posterior Distributions and Convergence of the estimations obtained on Internal Fraud (lognormal case)

5 Conclusion

This paper presents an intuitive approach to building the loss distribution function using the Bayesian Inference Framework and combining the different regulatory components.

This approach enables a controlled integration of the different elements through the Bayesian Inference. The prior functions have the same role as a penalization function on the parameters, and therefore behave as constraint during the estimation procedure. This results in capital charges driven by internal data (as shown in Figure 2), that are not dramatically influenced by external data or extreme scenarios.

16

Hence, with our approach, the capital requirement calculation is inherently related to the risk profile of the target financial institution and therefore provides senior management greater assurance of the validity of the results.

The next step to evaluate the financial institution global operational risk capital requirement would be the construction of the multivariate distribution function linking the different loss distributions which characterize the various event types through a copula. In the case, the approach developed by Guégan and Hassani (2012b) may be of interest.

References

Bayes, T. and Prince, R. (1763), 'An essay towards solving a problem in the doctrine of chance. by the late rev. mr. bayes, communicated by mr. price, in a letter to john canton, m.a. and f.r.s.', *Philosophical Transactions of the Royal Society of London* **53(0)**, 370–418.

BCBS (2001), 'Working paper on the regulatory treatment of operational risk', *Bank for International Settlements, Basel* .

BCBS (2010), 'Basel iii: A global regulatory framework for more resilient banks and banking systems', *Bank for International Settlements, Basel* .

Bowman, A. (1984), 'An alternative method of cross-validation for the smoothing of density estimates.', *Biometrika* **71**, 353–360.

Box, G. and Tiao, G. (1992), *Bayesian Inference in Statistical Analysis.*, Wiley Classics Library, JohnWiley & Sons, New York-Chichester-Brisbane.

Chernobai, A., Rachev, S. T. and Fabozzi, F. J. (2007), *Operational Risk: A Guide to Basel II Capital Requirements, Models, and Analysis*, John Wiley & Sons, New York.

Cruz, M. (2004), *Operational Risk Modelling and Analysis*, Risk Books, London.

Frachot, A., Georges, P. and Roncalli, T. (2001), 'Loss distribution approach for operational risk', *Working Paper, GRO, Crédit Lyonnais, Paris* .

Gilks, W., Richardson, S. and Spiegelhalter, D. (1996), *Markov Chain Monte Carlo in Practice.*, Chapman & Hall/CRC, London.

Guégan, D. and Hassani, B. (2012*a*), 'A mathematical resurgence of risk management : an extreme modeling of expert opinions.', *Documents de travail du Centre d'Economie de la Sorbonne 2011.57 - ISSN : 1955-611X* .

Guégan, D. and Hassani, B. (2012*b*), 'Multivariate vars for operational risk capital computation : a vine structure approach.', *Working Paper, University Paris 1, HAL : halshs-00587706, version 3* .

Guégan, D., Hassani, B. and Naud, C. (2011), 'An efficient threshold choice for the computation of operational risk capital.', *The Journal of Operational Risk* **6(4)**, 3–19.

Lehmann, E. and Casella, G. (1998), *Theory of Point Estimation, 2nd edition*, Springer, Berlin.

Rudemo, M. (1982), 'Empirical choice of histograms and kernel density estimators.', *Scand. J. Statist.* **9**, 65–78.

Shevchenko, P. (2011), *Modelling OPerational Risk Using Bayesian Inference*, Springer-Verlag, Berlin.

A   Appendix: Epanechnikov kernel

$$K(t) = \begin{cases} \frac{3}{4}\frac{(1-\frac{1}{5}t^2)}{\sqrt{5}}, \text{ for } |t| < \sqrt{5} \\ 0, \text{ otherwise} \end{cases} \tag{A.1}$$

The efficiency is equal to 1, and the canonical bandwidth is equal to $15^{\frac{1}{5}} \approx 1.3510$

B   Appendix: Least Square Cross Validation

One of the most famous method to estimate the Kernel bandwidth is known as the least square cross validation (Rudemo (1982) and Bowman (1984)). The structural idea is that the MISE might be written has:

$$\zeta(f(\hat{x}, h) - f(x)) = \zeta(\hat{f}(x; h)) - 2\int_{-\infty}^{+\infty} \hat{f}(x; h)f(x)dx + \zeta(f(x)) \tag{B.1}$$

Obviously, the last term of the equation does not depend on f therefore, we shall have the same h minimizing the full MISE or only the first part,

$$\zeta(\hat{f}(x; h)) - 2\int_{-\infty}^{+\infty} \hat{f}(x; h)f(x). \tag{B.2}$$

An unbiased estimator for (B.1) is given by,

$$h_{LSCV} = \zeta(\hat{f}(x, h)) - 2\frac{1}{n}\sum_{i=1}^{n} \hat{f}_{-i}(X_i; h), \tag{B.3}$$

where $\hat{f}_{-1}(x)$ is the density estimate constructed from all the data points except X_i:

$$\hat{f}_{-1}(x) = \frac{1}{h(n-1)}\sum_{j\neq i}^{n} K\left(\frac{(x-X_j)}{h}\right). \tag{B.4}$$

C   Appendix: The Metropolis-Hastings algorithm

The Metropolis-Hastings algorithm is almost a universal algorithm used to generate a Markov chain with a stationary distribution $\pi(\phi|x)$. It has been developed by Metropolis *et al.* in mechanical physics and generalised by Hastings in a statistical setting. It can be applied to a variety of problems since it requires the knowledge of the distribution of interest up to a constant only. Given a density $\pi(\phi|x)$, known up to a normalization constant, and a conditional density $q(\phi^*|\phi)$, the method generates the chain $\phi^{(1)}, \phi^{(2)}, ...$ using the following algorithm:

1. Initialise $\phi^{l=0}$ with any value within a support of $\pi(\phi|x)$;

2. For $l = 1, ..., L$

   (a) Set $\phi^l = \phi^{l-1}$

   (b) Generate a proposal ϕ^* from $q(\phi^*|\phi^{(l)})$

   (c) Accept proposal with the acceptance probability:

   $$p(\phi^{(l)}, \phi^{(*)}) = \min \left\{ 1, \frac{\pi(\phi^{(l)}|x)q(\phi^{(*)}|\phi^{(l)})}{\pi(\phi^{(*)}|x)q(\phi^{(l)}|\phi^{(*)})} \right\}. \tag{C.1}$$

   i.e. simulate U from the uniform distribution function $\mathcal{U}(0, 1)$ and set $\phi^{(l)} = \phi^{(*)}$ if $U < p(\phi^{(l)}, \phi^{(*)})$. Note that the normalization constant of the posterior does not contribute here;

3. Next l (i.e. do an increment, $l = l + 1$, and return to step 2).

D   Appendix: Risk Measure evaluation

For financial institutions, the capital requirement pertaining to operational risks is related to a VaR at 99.9%. It may be defined as follows:

Given a confidence level $\alpha \in [0, 1]$, the VaR associated with a random variable X is given by the smallest number x such that the probability that X exceeds x is not larger than $(1 - \alpha)$

$$VaR_{(1-\alpha)\%} = \inf(x \in \mathbb{R} : P(X > x) \leq (1 - \alpha)). \tag{D.1}$$

And we can compare these results to those obtained based on the Expected Shortfall (ES) defined as follows:

For a given α in $[0, 1]$, η the $VaR_{(1-\alpha)\%}$, and X a random variable which represents losses during a prespecified period (such as a day, a week, or some other chosen time period) then,

$$ES_{(1-\alpha)\%} = E(X|X > \eta). \tag{D.2}$$

21

Event Type	Initial λ	Corrected λ		
Internal Fraud (Global) - lognormal	133	261.5885		
External Fraud (Payments) - Weibull	313.4	485.0353		
Event Type	Initial λ body	Initial λ tail	Initial Global λ	Corrected Global λ
Execution, Delivery, and Product Management (Financial Instruments) - GPD	144.08	9.2	153.28	1882.327

Table 5: Frequency distribution parameters used in the Capital charge calculations. The standard deviations are given in brackets.

Event Type	μ	σ
Execution, Delivery, and Product Management (Financial Instruments) - lognormal body	8.092849	1.882122

Table 6: Body distribution (lognormal) parameters for Execution, Delivery, and Product Management / Financial Instruments. The standard deviations are provided in brackets.

22

Label	Scenarios		External Data			Internal Data		
	ϕ_1	ϕ_2	ϕ_1	ϕ_2	Size	ϕ_1	ϕ_2	Size
Internal Fraud (Global) - lognormal	10.616031	2.024592	8.707405	2.693001	2956	9.347693	2.485755	665
External Fraud (Payments) - Weibull	0.37642	4.6951E+04	1.644924	5.4293E+04	1085	0.292442	976.922586	1567
Execution, Delivery, and Product Management (Financial Instruments) - GPD	0.4705893	1.5309E+06	0.732526	1.1532E+06	31126	0.82231	7.0388E+05	3602

Table 7: This table presents the standalone parameters estimated for each components. Internal Fraud severities are modelled using a lognormal distribution, External fraud / Payments using a Weibull and Execution, Delivery, and Product Management considering a mixture model combining a lognormal distribution in the body and a Generalized Pareto distribution in the tail. The first column presents the initial parameters estimated for the scenarios. ϕ_1 and ϕ_2 represent the severity parameters of the chosen distribution, i.e. (resp.) μ and σ for the lognormal and shape and scale for the Weibull or GPD distribution. The second column shows the values obtained for the external data. The third column shows the parameters obtained for the internal data.

Label	Scenarios severity from priors		Intermediate severity (scenarios + External data)		Final severity (scenarios + External data + Internal Data)	
	ϕ_1	ϕ_2	ϕ_1	ϕ_2	ϕ_1	ϕ_2
Internal Fraud (Global) - Lognormal	10.616031	2.024592	9.123964	2.539844	8.727855	2.684253
(95% Bayesian Credibility Interval)	-	-	-	-	*[7.98176; 9.95824]*	*[1.23225; 4.42314]*
External Fraud (Payments) - Weibull	0.3761	46942	0.38873	49942	0.39910	46622
(95% Bayesian Credibility Interval)	-	-	-	-	*[0.23127; 0.50673]*	*[37632; 58981]*
Execution, Delivery, and Product Management (Financial Instruments) - GPD	0.4705893	1.5309E+06	0.4452626	1.5109E+06	0.6021	6.05E+05
(95% Bayesian Credibility Interval)	-	-	-	-	*[0.3313; 0.9089]*	*[4.52E+05; 8.30E+05]*

Table 8: The table presents the evolution of the parameters obtained carrying out the Cascade Bayesian Approach. Internal Fraud severities are modelled using a lognormal distribution, External fraud / Payments using a Weibull and Execution, Delivery, and Product Management considering a mixture model combining a lognormal distribution in the body and a Generalized Pareto distribution in the tail. The first column presents the initial parameters estimated from the scenarios. ϕ_1 and ϕ_2 represent the severity parameters of the chosen distribution, i.e. (resp.) μ and σ for the Lognormal and shape and scale for the Weibull or GPD distribution. The second column shows the values obtained after the first refinement, i.e. after the incorporation of the external data. The third column shows the final parameters following the second refinement, i.e. after the integration of the internal data. The figures in brackets represent a 95% Bayesian Credible Interval obtained from the final posterior distribution.

24

Label	Scenarios	External Data	Internal Data	Combination	
	VaR	VaR	VaR	VaR	ES
Internal Fraud (Global)	843 445 037	855 464 158	1 143 579 396	1 106 692 211	2 470 332 812
External Fraud Payments (Payments)	128 243 367	13 664 057	3 945 116	119 637 592	126 935 364
Execution, Delivery, and Product Management (Financial Instruments)	122 544 708	583 519 888	137 456 938	213 982 300	281 157 384

Table 9: This table presents the stand alone Value-at-Risk for each of the three components, as well as the VaR and the Expected Shortfall computed by combining the three elements by Cascade Bayesian Integration for each of the three different event types. Internal Fraud severities are modelled using a lognormal distribution, External fraud / Payments using a Weibull and Execution, Delivery, and Product Management considering a mixture model combining a lognormal distribution in the body and a Generalized Pareto distribution in the tail. The parameters of the distributions used to compute these values are shown in tables 7 and 8.