Operational Risk: Combining Internal Data, External Data and Expert Opinions

Pavel V. Shevchenko[1,*] Mario V. Wüthrich[2]

Draft, version from 16 February 2010

Book chapter in *Rethinking Risk Management and Reporting: Uncertainty, Bayesian Analysis and Expert Judgement* (2010), edited by Klaus Böcker, Risk Books.

[1] *CSIRO Mathematical and Information Sciences, Sydney, Locked Bag 17, North Ryde, NSW, 1670, Australia; e-mail: Pavel.Shevchenko@csiro.au*

[2] *ETH Zurich, Department of Mathematics, CH-8092 Zurich, Switzerland*

[*] *Corresponding author*

Abstract

Many banks adopt the Loss Distribution Approach (LDA) to quantify the operational risk capital charge under the current regulatory framework for banking supervision, referred to as Basel II. Under the LDA, banks quantify the frequency and severity distributions of operational risk losses to estimate the capital charge as the 0.999 quantile of the annual compound loss distribution. One of the main challenges for the LDA is to combine bank's internal data, external data from other banks and opinions from experts to estimate the model parameters. In this chapter, we consider methods that allow to accomplish this task. It is not only one of the Basel II requirements but also a critical step to obtain reliable estimates. This is because the internal data for the low frequency high impact risks are very limited and parameter uncertainty implied from the use of internal data only is too large to make satisfactory inferences. In particular, we advocate and demonstrate the use of the Bayesian inference approach to combine different data sources and quantify the impact of the parameter uncertainty. The latter is typically significant for the capital requested at the 0.999 quantile level. Ignoring this uncertainty or using Gaussian approximations might lead to a significant underestimation of the necessary underlying risk.

Keywords: Operational Risk, Loss Distribution Approach, Bayesian Inference.

1 Introduction

Under the Loss Distribution Approach (LDA) for the Basel II Advanced Measurement Approaches (AMA), banks should quantify the distribution of operational risk losses for each risk cell (business line/event type) over a one year time horizon. The commonly used LDA model for the annual loss is a compound loss process for loss frequencies and severities. For a recent review of quantitative methods suggested for implementation of the LDA, see Shevchenko (2009). In this chapter, we consider a single risk cell (business line/event type), where the annual loss is the sum of individual losses and is given by

$$Z = \sum_{i=1}^{N} X_i. \tag{1}$$

Here, N is the annual number of events modelled as a random variable from some discrete distribution (typically Poisson) and X_i are the severities of the events modelled as independent random variables from a continuous distribution. Frequency N and severities X_i are assumed to be independent. Note that the independence assumed here will be conditional on parameters of the distributions.

Estimation of the frequency and severity distributions of operational risk is a challenging task for low frequency high impact losses due to limited data. To improve the estimation, the actual data are supplemented with expert opinions and external data. In fact, it is mandatory to include these data sources into the model estimation to meet the regulatory requirements. In particular, Basel II AMA requires, see BCBS (2006, p.152), that:

> Any operational risk measurement system must have certain key features to meet the supervisory soundness standard set out in this section. These elements must include the use of internal data, relevant external data, scenario analysis and factors reflecting the business environment and internal control systems.

Combining these different data sources for model estimation is certainly one of the main challenges in operational risk. Bayesian inference is a statistical technique well suited to incorporate expert opinions into data analysis. There is a broad literature covering Bayesian inference and its applications to the insurance industry as well as to other areas. For a brief introduction to Bayesian statistics, see chapter by Robert and Rousseau of this volume. Good textbook overviews on Bayesian inference are Berger (1985) and Robert (2001). The methods allow for structural modelling where external data and expert opinions are incorporated into the analysis via specifying distributions (so-called prior distributions) for model parameters. These are updated by the internal data as they become available.

Bayesian methods, in the context of operational risk, have been briefly mentioned in the early literature, for example, in King (2001, Chapter 12), Cruz (2002, Chapter 10), and Panjer

(2006, Section 10.5). However, Bayesian methods have not really merged into operational risk developments until recently. One of the first detailed and illustrative publication of the Bayesian inference methodology for the estimation of operational risk was Shevchenko and Wüthrich (2006). Then, an example of a "toy" model for operational risk, based on the closely related credibility theory, was presented in Bühlmann et al. (2007). The Bayesian methodology was extended to combine three data sources (expert opinion, internal and external data) in Lambrigger et al. (2007); and developed further in Peters et al. (2009) for a multivariate case with dependence modelling between risks. Moreover, it is the main topic of the forthcoming book Shevchenko (2010). Currently, Bayesian methods for operational risk is an active research area as can be seen over the last few years in The Journal of Operational Risk available on *www.journalofoperationalrisk.com*. In this chapter, we describe Bayesian techniques within the context of operational risk and provide several examples of its application for operational risk quantification. We demonstrate how three data sources (internal data, external data and expert opinions) can be combined simultaneously.

The organisation of this chapter is as follows. Section 2 lists several ad-hoc approaches used to combine internal data, external data and expert opinions. Section 3 describes the Bayesian inference framework in the operational risk context. Bayesian model estimation is reviewed in Section 4. Section 5 considers an example of modelling loss frequencies using a standard Bayesian framework combining two data sources. The framework is extended to combine three data sources (internal data, external data and expert opinions) in Section 6. Estimation of the bank capital against operational risk is described in Section 7. Discussions and conclusions are presented in Section 8.

2 Combining Data Using Heuristic Approaches

Often in practice, accounting for factors reflecting the business environment and internal control systems is achieved via scaling of data. Then ad-hoc procedures are used to combine internal data, external data and expert opinions. For example:

- Fit the severity distribution to the combined samples of internal and external data and fit the frequency distribution using internal data only;
- Estimate the Poisson annual intensity for the frequency distribution as $w\lambda_{int} + (1-w)\lambda_{ext}$, where the intensities λ_{ext} and λ_{int} are implied by the external and internal data respectively, using expert specified weight w;
- Estimate the severity distribution as a mixture $w_1 F_{SA}(x) + w_2 F_I(x) + w_3 F_E(x)$, where $F_{SA}(x)$, $F_I(x)$ and $F_E(x)$ are the distributions identified by scenario analysis, internal data and external data respectively, using expert specified weights w_1, w_2, $w_3 = 1 - w_1 - w_2 \in [0, 1]$;

- Minimum variance principle – the combined estimator is a linear combination of the individual estimators obtained from internal data, external data and expert opinion separately with the weights chosen to minimise the variance of the combined estimator.

An easy procedure is the minimum variance principle. The rationale behind this principle is as follows. Consider unbiased independent estimators $\widehat{\Theta}^{(k)}$ for the parameter θ, i.e. $\mathrm{E}[\widehat{\Theta}^{(k)}] = \theta$ and $\mathrm{Var}[\widehat{\Theta}^{(k)}] = \sigma_k^2$, $k = 1, \ldots, K$. Then, the combined unbiased linear minimum variance estimator is

$$\widehat{\Theta}_{tot} = w_1 \widehat{\Theta}^{(1)} + \cdots + w_K \widehat{\Theta}^{(K)}, \quad w_1 + \cdots + w_K = 1, \tag{2}$$

and the weights are found by minimizing $\mathrm{Var}[\widehat{\Theta}_{tot}]$. The explicit expressions for the weights is given by the following theorem, see for instance Wüthrich and Merz (2008, Lemma 3.4).

Theorem 2.1 (Minimum variance estimator)
Assume that we have $\widehat{\Theta}^{(1)}, \ldots, \widehat{\Theta}^{(K)}$ *unbiased and independent estimators of* θ *with variances* $\sigma_k^2 = \mathrm{Var}[\Theta^{(k)}]$. *Then the unbiased linear estimator (2) minimizing the variance has weights*

$$w_k = \frac{1/\sigma_k^2}{\sum_{i=1}^{K} (1/\sigma_i^2)}. \tag{3}$$

Moreover, it has variance

$$\mathrm{Var}[\widehat{\Theta}_{tot}] = \left(\sum_{k=1}^{K} \frac{1}{\sigma_k^2} \right)^{-1}.$$

The weights behave as it is expected in practice. In particular, $w_k \to 1$ for $\sigma_k^2 \to 0$. Heuristically, the minimum variance principle can be applied to almost any quantity e.g. distribution parameters or distribution characteristics such as mean, variance, etc. The assumption that the estimators are unbiased for θ is probably reasonable when combining estimators from different experts (or from experts and internal data). However, it is certainly questionable if applied to combine estimators from external and internal data. Below, we focus on Bayesian inference methods that can be used to combine these data sources in a consistent statistical framework.

3 Combining Data Using Bayesian Approach

In this section we describe the Bayesian inference framework and introduce some notations used throughout this chapter. Consider a random vector of data $\mathbf{X} = (X_1, X_2, \ldots, X_K)'$, whose joint density for a given vector of parameters $\boldsymbol{\Theta}$ is $f(\mathbf{x}|\boldsymbol{\theta})$. In the context of operational risk, $\mathbf{X}$ may represent the loss frequencies or severities. We shall use upper case symbols to represent random variables, lower case symbols for their realizations and bold symbols for vectors; for example,

θ is realization of a random variable Θ. In the Bayesian approach, both observations $\mathbf{X}$ and parameters $\boldsymbol{\Theta}$ are considered to be random. Then, Bayes' theorem can be formulated as follows

$$f(\mathbf{x}, \boldsymbol{\theta}) = f(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) = \pi(\boldsymbol{\theta}|\mathbf{x})f(\mathbf{x}), \tag{4}$$

where

- $f(\mathbf{x}, \boldsymbol{\theta})$ is the joint density of the data $\mathbf{X}$ and parameters $\boldsymbol{\Theta}$;
- $f(\mathbf{x}|\boldsymbol{\theta})$ is the density of the observations for given parameters $\boldsymbol{\Theta}$;
- $\pi(\boldsymbol{\theta})$ is the probability density of the parameters $\boldsymbol{\Theta}$, the so-called prior density function. Typically, $\pi(\boldsymbol{\theta})$ depends on a set of further parameters that are called hyper-parameters, omitted here for simplicity of notation;
- $\pi(\boldsymbol{\theta}|\mathbf{x})$ is the density of parameters given the data $\mathbf{X}$, the so-called posterior distribution;
- $f(\mathbf{x}) = \int f(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$ is the marginal density of $\mathbf{X}$.

Predictive distribution. The objective is to estimate the predictive distribution of a future observation X_{K+1}, conditional on all available information $\mathbf{X} = (X_1, X_2, \ldots, X_K)$. Assume that conditionally, given $\boldsymbol{\Theta} = \boldsymbol{\theta}$, X_{K+1} and $\mathbf{X}$ are independent and X_{K+1} has conditional density $f(x_{K+1}|\boldsymbol{\theta})$. Then, the conditional density of X_{K+1}, given $\mathbf{X} = \mathbf{x}$, is

$$f(x_{K+1}|\mathbf{x}) = \int f(x_{K+1}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{x})d\boldsymbol{\theta}. \tag{5}$$

Posterior distribution. Using (4), the posterior density of the parameters $\boldsymbol{\Theta}$ can be written as

$$\pi(\boldsymbol{\theta}|\mathbf{x}) = f(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})/f(\mathbf{x}) \propto \pi(\boldsymbol{\theta})f(\mathbf{x}|\boldsymbol{\theta}). \tag{6}$$

Here, the density $f(\mathbf{x}|\boldsymbol{\theta})$ is a likelihood function of observations, and $f(\mathbf{x})$ plays the role of the normalizing constant. Thus, the posterior distribution can be viewed as a product of a prior knowledge $\pi(\boldsymbol{\theta})$ with a likelihood function of observed data $f(\mathbf{x}|\boldsymbol{\theta})$.

In the context of operational risk, one can go through the following three logical steps:

- The prior distribution $\pi(\boldsymbol{\theta})$ should be estimated subjectively by expert opinions (pure Bayesian approach, see Section 4.1) or using external data (empirical Bayesian approach, see Section 4.2);
- Then, the prior distribution should be updated with the observed data $\mathbf{X}$ using formula (6) to get a posterior density $\pi(\boldsymbol{\theta}|\mathbf{x})$;

- Formula (5) is used to calculate the predictive distribution of X_{K+1}, given the data $\mathbf{X}$.

The iterative update procedure for priors. If the data $\mathbf{X} = (X_1, X_2, \ldots, X_K)$ are conditionally, given $\boldsymbol{\Theta} = \boldsymbol{\theta}$, independent and X_k is distributed with density $f_k(\cdot|\boldsymbol{\theta})$, then the likelihood function of $\mathbf{X}$ can be written as $f(\mathbf{x}|\boldsymbol{\theta}) = \prod_{i=1}^{K} f_i(x_i|\boldsymbol{\theta})$. Denote the posterior density calculated after k observations as $\pi_k(\boldsymbol{\theta}|x_1, \ldots, x_k)$, then using (6),

$$\pi_k(\boldsymbol{\theta}|x_1, \ldots, x_k) \propto \pi(\boldsymbol{\theta}) \prod_{i=1}^{k} f_i(x_i|\boldsymbol{\theta}) \propto \pi_{k-1}(\boldsymbol{\theta}|x_1, \ldots, x_{k-1}) f_k(x_k|\boldsymbol{\theta}). \tag{7}$$

Thus, we obtain an iterative update procedure. Only the posterior distribution calculated after $k-1$ observations and the k-th observation are needed to calculate the posterior distribution after k observations. Thus the entire loss history over many years is not required, making the model easier to understand and manage, and allowing experts to adjust the priors after a new loss has occurred. Formally, the posterior distribution calculated after $k-1$ observations can be treated as a prior distribution for the k-th observation. In practice, initially, we start with the prior distribution $\pi(\boldsymbol{\theta})$ identified by expert opinions and external data only. Then, the posterior distribution $\pi(\boldsymbol{\theta}|\mathbf{x})$ is calculated, using (6), when actual data arrive. If there is a reason (for example, a new control policy introduced in a bank), then this posterior distribution can be adjusted by an expert and treated as the prior distribution for subsequent observations.

Conjugate prior distributions. Sometimes the posterior density (6) can be calculated in closed form which is useful in practice when Bayesian inference is applied. This is the case for so-called conjugate prior distributions where the prior and posterior distributions are of the same type.

Definition 3.1 (Conjugate prior) *Let F denote the class of density functions $f(\mathbf{x}|\boldsymbol{\theta})$, indexed by $\boldsymbol{\theta}$. A class U of prior densities $\pi(\boldsymbol{\theta})$ is said to be a conjugate family for F if the posterior density $\pi(\boldsymbol{\theta}|\mathbf{x})$ is in the class U for all $f \in F$ and $\pi \in U$.*

Formally, if the family U contains all distribution functions then it is conjugate to any family F. However, to make a model useful in practice it is important that U should be as small as possible while containing realistic distributions. Below we present the Poisson-gamma pair. Several other pairs (binomial-beta, gamma-gamma, exponential-gamma) can be found in e.g. Bühlmann and Gisler (2005).

4 Estimating Prior Parameters

In general, the parameters of the prior distributions (hyper-parameters) can be estimated subjectively using expert opinions (pure Bayesian approach) or using data (empirical Bayesian approach).

4.1 Pure Bayesian Approach

In a pure Bayesian approach, the prior distribution is specified subjectively (that is, in the context of operational risk, using expert opinions only). Section 2 of this volume is solely devoted to expert judgement and elicitation of subjective probabilities. In particular, Berger (1985) lists several methods:

- **Histogram approach**: split the parameter space of $\mathbf{\Theta}$ into cubes and specify the subjective probability for each cube. From this, a smooth density of the prior distribution can be determined.
- **Relative likelihood approach**: compare the intuitive likelihoods of the different values of $\mathbf{\Theta}$. Again, the smooth density of prior distribution can be determined. It is difficult to apply this method in the case of unbounded parameters.
- **CDF determinations**: subjectively construct the distribution function for the prior and sketch a smooth curve.
- **Matching a given functional form**: assume some functional form for the prior distribution to match prior beliefs (on the moments, quantiles, etc) as close as possible.

Below, using the method of matching a given function form, we consider the estimation of the prior distribution parameters for Poisson-gamma pair. The use of a particular method is determined by a specific problem and expert experience. Usually, if the expected values for the quantiles (or mean) and their uncertainties are estimated by the expert then it is possible to fit the priors.

4.2 Empirical Bayesian Approach

Instead of the subjective approach above, the prior distribution can also be estimated empirically from industry data, collective data in the bank, etc. For example, consider a specific risk cell (event type/business line) in J banks with frequency or severity data $\mathbf{X} = \{X_{j,k},\ k = 1,\ldots,K_j,\ j = 1,\ldots,J\}$. Here, K_j is the number of observations in bank j. Assume that $X_{j,k}, k = 1,\ldots,K_j$, are conditionally independent and identically distributed from $f(\cdot|\boldsymbol{\theta}_j)$, for given $\mathbf{\Theta}_j = \boldsymbol{\theta}_j$. That is, data of different banks are modelled by the same parametric distribution, however, with different bank-specific parameter $\boldsymbol{\theta}_j$. Assume now that $\boldsymbol{\Theta}_j$, $j = 1, \ldots, J$, are independent and identically distributed from $\pi(\cdot)$, that is we assumed that the risk cells of different banks are the same a priori (before we have any observations). Then, the likelihood of all observations can be written as

$$\ell(\mathbf{x}) = \prod_{j=1}^{J} \int \left[\prod_{k=1}^{K_j} f(x_{j,k}|\boldsymbol{\theta}_j) \right] \pi(\boldsymbol{\theta}_j) d\boldsymbol{\theta}_j. \tag{8}$$

The hyper-parameters of $\pi(\cdot)$ can now be estimated by maximizing the above likelihood.

5 Example: Poisson-Gamma Case

For illustrative purposes we now consider the case of Poisson distributed loss frequencies combined with a gamma prior distribution. Operational risk severities can be modelled in a similar fashion. For the lognormal-normal-inverse chi squared model we refer to Shevchenko and Wüthrich (2006, Appendix A), for Pareto-gamma model to Shevchenko and Wüthrich (2006, Section 3.3).

5.1 Model

Model Assumptions 5.1

- *Suppose that, given $\Lambda = \lambda > 0$, frequencies $N_1, \ldots, N_{K+1}$ are independent random variables from a Poisson distribution Poisson(λ), i.e. for $k = 1, \ldots, K+1$,*

$$f(n|\lambda) = \Pr[N_k = n|\Lambda = \lambda] = e^{-\lambda} \frac{\lambda^n}{n!}, \quad n \in \mathbb{N}. \tag{9}$$

- *The prior distribution for Λ is a gamma distribution $\Gamma(\alpha, \beta)$ with density*

$$\pi(\lambda) = \frac{(\lambda/\beta)^{\alpha-1}}{\Gamma(\alpha)\beta} \exp(-\lambda/\beta), \quad \lambda > 0, \ \alpha > 0, \beta > 0. \tag{10}$$

Posterior. Denote $\mathbf{N} = (N_1, \ldots, N_K)$. Given $\Lambda = \lambda$, under Model Assumptions 5.1, $N_1, \ldots, N_K$ are independent and their likelihood is given by

$$h(\mathbf{n}|\lambda) = \prod_{i=1}^{K} e^{-\lambda} \frac{\lambda^{n_i}}{n_i!}. \tag{11}$$

Thus, using formula (6), the posterior density is

$$\pi(\lambda|\mathbf{n}) \;\propto\; \frac{(\lambda/\beta)^{\alpha-1}}{\Gamma(\alpha)\beta}\exp(-\lambda/\beta)\prod_{i=1}^{K} e^{-\lambda}\frac{\lambda^{n_i}}{n_i!} \;\propto\; \lambda^{\alpha_K-1}\exp(-\lambda/\beta_K), \tag{12}$$

which is $\Gamma(\alpha_K, \beta_K)$ with updated parameters α_K and β_K given by:

$$\begin{aligned} \alpha \to \alpha_K &= \alpha + \sum_{i=1}^{K} n_i, \\ \beta \to \beta_K &= \tfrac{\beta}{1+K\beta}. \end{aligned} \tag{13}$$

Improper constant prior. It is easy to see that if the prior is improper constant (for improper prior, see Chapter 1 of this volume), i.e. $\pi(\lambda|\mathbf{n}) \propto h(\mathbf{n}|\lambda)$, then the posterior is $\Gamma(\alpha_K, \beta_K)$ with

$$\alpha_K = 1 + \sum_{i=1}^{K} n_i, \quad \beta_K = \frac{1}{K}. \tag{14}$$

In this case, the mode of the posterior density $\pi(\lambda|\mathbf{n})$, referred to as the maximum a posteriori (MAP) estimate, is

$$\widehat{\lambda}_K^{\text{MAP}} = (\alpha_K - 1)\beta_K = \frac{1}{K}\sum_{i=1}^{K} n_i, \tag{15}$$

which is the same as the maximum likelihood estimate (MLE) $\widehat{\lambda}_K^{\text{MLE}}$ of λ.

Full predictive distribution. Now we can calculate the predictive distribution of the loss frequency of the next period, e.g. the forthcoming year, given the observed loss frequencies $\mathbf{N}$ and the prior information. Given $\mathbf{N} = (N_1, \ldots, N_K)$, the full predictive distribution for N_{K+1} is negative binomial, $NegBin(\alpha_K, 1/(1+\beta_K))$,

$$\begin{aligned} \Pr[N_{K+1} = m|\mathbf{N} = \mathbf{n}] &= \int f(m|\lambda)\pi(\lambda|\mathbf{n})d\lambda \\ &= \frac{\Gamma(\alpha_K + m)}{\Gamma(\alpha_K)m!}\left(\frac{1}{1+\beta_K}\right)^{\alpha_K}\left(\frac{\beta_K}{1+\beta_K}\right)^{m}, \end{aligned} \tag{16}$$

see Shevchenko and Wüthrich (2006). This is valid for both gamma and improper constant priors. The following gives an intuitive interpretation:

$$\mathrm{E}[N_{K+1}|\mathbf{N} = \mathbf{n}] = \mathrm{E}[\Lambda|\mathbf{N} = \mathbf{n}] = \alpha_K\beta_K = w_K\widehat{\lambda}_K^{\text{MLE}} + (1 - w_K)\lambda_0, \tag{17}$$

where

- $\widehat{\lambda}_K^{\text{MLE}} = \frac{1}{K}\sum_{i=1}^{K} n_i$ is the MLE estimate of λ using the observations only;
- $\lambda_0 = \alpha\beta$ is the estimate of λ using a prior knowledge only (e.g. specified by expert);

- $w_K = \frac{\beta K}{1+\beta K}$ is the credibility weight used to combine λ_0 and $\widehat{\lambda}_K^{\text{MLE}}$.

Remarks 5.2

- As the number of years K increases, the credibility weight w_K increases and vice versa. That is, the more observations we have, the larger credibility weight we assign to the estimator based on the observed counts, while the lesser credibility weight is attached to the expert opinion estimate. Also, the larger the volatility of the expert opinion (larger β), the larger credibility weight is assigned to the observations.

- Recursive calculation of the posterior distribution is very simple. The update of information from $n_1, n_2, \ldots, n_{k-1}$ to $n_1, n_2, \ldots, n_k$ is given by

$$\alpha_k = \alpha_{k-1} + n_k, \quad \beta_k = \frac{\beta_{k-1}}{1+\beta_{k-1}}. \tag{18}$$

  This leads to an efficient recursive scheme, where the calculation of posterior distribution parameters is based on the most recent observation and parameters of posterior distribution calculated just before this observation.

5.2 Pure Bayesian Approach

One now has to determine the hyper-parameters α and β of the prior distribution used in Model Assumptions 5.1. The expert may estimate the expected number of events $\mathrm{E}[\Lambda] = \lambda_0 = \alpha\beta$ but can not be certain in the estimate. If the expert specifies $\mathrm{E}[\Lambda]$ and an uncertainty that the "true" λ for next year is within the interval $[a,b]$ with probability $\Pr[a \le \Lambda \le b] = p$, then the equations

$$\begin{aligned} &\mathrm{E}[\Lambda] = \alpha\beta, \\ &\Pr[a \le \Lambda \le b] = p = \int_a^b \pi(\lambda)d\lambda = F^{(\Gamma)}_{\alpha,\beta}(b) - F^{(\Gamma)}_{\alpha,\beta}(a) \end{aligned} \tag{19}$$

can be solved numerically to estimate the hyper-parameters α and β. Here, $F^{(\Gamma)}_{\alpha,\beta}(\cdot)$ is the gamma distribution $\Gamma(\alpha,\beta)$.

This is not the only way to assess the hyper-parammeters. For instance, in the insurance industry, the uncertainty for the "true" λ is often measured in terms of the coefficient of variation, $\mathrm{Vco}[\Lambda] = \sqrt{\mathrm{Var}(\Lambda)}/\mathrm{E}[\Lambda]$. Hence, instead of using (19), the hyper-parameters α and β can alternatively be computed given the expert estimates for $\mathrm{E}[\Lambda] = \alpha\beta$ and $\mathrm{Vco}[\Lambda] = 1/\sqrt{\alpha}$.

Example 5.3 (Numerical illustration)

If the expert specifies $\mathrm{E}[\Lambda] = 0.5$ and $\Pr[0.25 \le \Lambda \le 0.75] = 2/3$, then we can fit a prior distribution, $\Gamma(\alpha = 3.407,\ \beta = 0.147)$ by solving (19). Assume now that the bank experienced no losses over the first year (after the prior distribution was estimated). Then, using formulas (18), the posterior distribution parameters are $\widehat{\alpha}_1 = 3.407 + 0 = 3.407$, $\widehat{\beta}_1 = 0.147/(1 + 0.147) = 0.128$ and the estimated arrival rate using the posterior distribution is $\widehat{\lambda}_1 = \widehat{\alpha}_1\widehat{\beta}_1 = 0.436$. If during the next year no losses are observed again, then the posterior distribution parameters are $\widehat{\alpha}_2 = \widehat{\alpha}_1 + 0 = 3.407$, $\widehat{\beta}_2 = \widehat{\beta}_1/(1 + \widehat{\beta}_1) = 0.113$ and $\widehat{\lambda}_2 = \widehat{\alpha}_2\widehat{\beta}_2 = 0.385$. Subsequent observations will update the arrival rate estimator correspondingly using formulas (18). Thus, starting from the expert specified prior, observations regularly update (refine) the posterior distribution.

In Figure 1, we compare the Bayesian method for estimating the arrival rate of the losses (using two different prior distributions) with the classical maximum likelihood approach. In Figure 1a, we show the posterior mean for the arrival rate $\widehat{\lambda}_k = \widehat{\alpha}_k\widehat{\beta}_k$, $k = 1, \ldots, 15$ (with the prior distribution as specified in the above example), when the annual number of events N_k, $k = 1, \ldots, 15$, are simulated from $Poisson(\lambda = 0.6)$ and are given in Table 1.

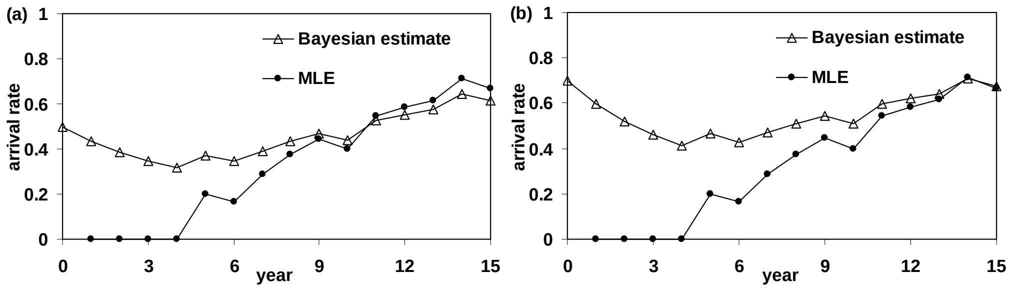

Figure 1: The Bayesian and the standard maximum likelihood estimates of the arrival rate versus the observation year. The Bayesian estimate is a mean of the posterior distribution when the prior distribution is gamma with: a) $\mathrm{E}[\Lambda] = 0.5$ and $\Pr[0.25 \le \Lambda \le 0.75] = 2/3$; b) $\mathrm{E}[\Lambda] = 0.7$ and $\mathrm{Vco}[\Lambda] = 0.5$. The annual counts were sampled from the Poisson(0.6) and are given in Table 1. See Example 5.3 for details.

Table 1: The annual number of losses simulated from $Poisson(0.6)$.

i	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
n_i	0	0	0	0	1	0	1	1	1	0	2	1	1	2	0

On the same figure, we show the standard maximum likelihood estimate of the arrival rate as a function of the year k, $\widehat{\lambda}_k^{\mathrm{MLE}} = \frac{1}{k}\sum_{i=1}^{k} n_i$. After approximately 8 years, the estimators are very close to each other. However, for a small number of observed years, the Bayesian estimate is more accurate as it takes the prior information into account. After 12 years, both estimators are close to the true value of 0.6. Note that for this example we assumed the prior distribution with a mean equal to 0.5, which is different from the true arrival rate 0.6. Thus, this example shows that an initially incorrect prior estimator is corrected by the observations as they become available. It is interesting to observe that, in year 14, the estimators become slightly different again. This is because the bank was unlucky to experience event counts 1, 1, and 2 in the years 12, 13, and 14 respectively. As a result, the maximum likelihood estimate becomes higher than the true value, while the Bayesian estimate is more stable (smooth) in respect to the unlucky years. If this example is repeated with different sequences of random numbers, then one would observe quite different maximum likelihood estimates (for small k) and more stable Bayesian estimates. For comparison, Figure 1b shows the example where expert specifies $\mathrm{E}[\Lambda] = 0.7$ and $\mathrm{Vco}[\Lambda] = 0.5$ that gives the prior parameters $\alpha = 4$ and $\beta = 0.175$.

5.3 Empirical Bayesian Approach

It is not difficult to include a priori known differences (for example, exposure indicators, expert opinions on the differences, etc) between the risk cells from the different banks. As an example, we consider the case when the annual frequency of the events is modelled by the Poisson distribution with the gamma prior and estimate the hyper-parameters, α and β, using the industry data with differences between the banks taken into account.

Model Assumptions 5.4 *Consider a risk cell in J banks with the annual frequencies $N_{j,k}$, $k = 1, \ldots, K_j{+}1$, $j = 1, \ldots, J$, where $K_j{+}1$ refers to the next year. Set $\mathbf{N}_j = (N_{j,1}, \ldots, N_{j,K_j+1})$ and assume that:*

- *Given bank specific intensities $\Lambda_j = \lambda_j$, $\mathbf{N}_j$ satisfies Model Assumptions 5.1 with $N_{j,k}$ from Poisson$(\lambda_j V_{j,k})$, where $V_{j,k} > 0$ is a fixed volume measure, which can be used to account for a priori known inhomogeneities of the loss frequencies between different years as well as different banks.*

- *$\Lambda_1, \ldots, \Lambda_J$ are assumed to be independent and identically distributed from $\Gamma(\alpha, \beta)$, see (10), and $(\Lambda_1, \mathbf{N}_1), \ldots, (\Lambda_J, \mathbf{N}_J)$ are independent.*

Denote $N_j = \sum_{k=1}^{K_j} N_{j,k}$, $V_j = \sum_{k=1}^{K_j} V_{j,k}$. Then, the likelihood of all available data (over all J banks), $\mathbf{N} = \{N_{j,k} : j = 1, \ldots, J;\ k = 1, \ldots, K_j\}$, can be written as

$$\ell(\mathbf{n}) = \prod_{j=1}^{J} \int \left[\prod_{k=1}^{K_j} f(n_{j,k}|\lambda_j) \right] \pi(\lambda_j) d\lambda_j \propto \prod_{j=1}^{J} \frac{\Gamma(\alpha + n_j)}{\Gamma(\alpha)\beta^{\alpha}(V_j + 1/\beta)^{\alpha + n_j}}. \tag{20}$$

The hyper-parameters, α and β can now be estimated by maximizing the log-likelihood

$$\ln \ell \propto \sum_{j=1}^{J} \left\{ \ln \Gamma(\alpha + n_j) - \ln \Gamma(\alpha) - \alpha \ln \beta - (\alpha + n_j) \ln \left(\frac{1}{\beta} + V_j \right) \right\}. \tag{21}$$

To fit the parameters but avoid the use of numerical optimization required for maximizing (21), one could also use a method of moments using the following proposition.

Proposition 5.5 (Method of moments) *Given Model Assumptions 5.4, denote* $\lambda_0 = \mathrm{E}[\Lambda_j] = \alpha\beta$, $\sigma_0^2 = \mathrm{Var}[\Lambda_j] = \alpha\beta^2$. *Then estimates* $\widehat{\lambda}_0$ *and* $\widehat{\sigma}_0^2$ *for* λ_0 *and* σ_0^2 *respectively are*

$$\begin{aligned} \widehat{\lambda}_0 &= \frac{1}{J} \sum_{j=1}^{J} \widehat{\lambda}_j, \quad \widehat{\lambda}_j = \frac{1}{K_j} \sum_{k=1}^{K_j} \frac{n_{j,k}}{V_{j,k}}, j = 1, \ldots, J. \\ \widehat{\sigma}_0^2 &= \max \left\{ \frac{1}{J-1} \sum_{j=1}^{J} (\widehat{\lambda}_j - \widehat{\lambda}_0)^2 - \frac{\widehat{\lambda}_0}{J} \sum_{j=1}^{J} \frac{1}{K_j^2} \sum_{k=1}^{K_j} \frac{1}{V_{j,k}},\ 0 \right\}. \end{aligned} \tag{22}$$

The above proposition can easily be used to estimate α and β as $\widehat{\alpha} = \widehat{\lambda}_0 / \widehat{\beta}$ and $\widehat{\beta} = \widehat{\sigma}_0^2 / \widehat{\lambda}_0$, correspondingly. For a proof, see Shevchenko and Wüthrich (2006). Alternative unbiased moment estimators are given in Bühlmann and Gisler (2005, Section 4.10).

Once the prior distribution parameters α and β are estimated, then the posterior distributions of λ_j of the different banks are $\Gamma(\widehat{\alpha}_j, \widehat{\beta}_j)$, calculated similarly to (12) with parameters

$$\widehat{\alpha}_j = \alpha + \sum_{k=1}^{K_j} n_{j,k}, \ \widehat{\beta}_j = \beta \left(1 + \beta \sum_{k=1}^{K_j} V_{j,k} \right)^{-1}. \tag{23}$$

The predictive distributions of the annual frequency N_{j,K_j+1} for the next year are negative binomial, $NegBin(\widehat{\alpha}_j, \widehat{p}_j = 1/(1 + V_{j,K_j+1} \widehat{\beta}_j))$, see (16). Here, V_{j,K_j+1} is a volume measure for the next year in the jth bank.

Remarks 5.6 Observe that in Model Assumptions 5.4, we have scaled the parameters λ_j by $V_{j,k}$ for considering a priori differences. This leads to a linear volume relation for the variance function. To obtain different functional relations, it might be better to scale the actual observations. For example, given observations $X_{j,k}, j = 1, \ldots, J,\ k = 1, \ldots, K_j$ (these could be frequencies or severities), consider variables $Y_{j,k} = X_{j,k}/V_{j,k}$. Assume that, for given $\boldsymbol{\theta}_j$, $Y_{j,k}$, $k = 1, \ldots, K_j$, are independent and identically distributed from $f(\cdot|\boldsymbol{\theta}_j)$. Also, assume that $\boldsymbol{\Theta}_1, \ldots, \boldsymbol{\Theta}_J$ are independent and identically distributed from $\pi(\cdot)$. Then one can construct the likelihood of $Y_{j,k}$ using (8) to fit parameters of $\pi(\cdot)$ or try to use the method of moments.

6 Combining Expert Opinions, External and Internal Data

In order to estimate the risk capital of a bank to fulfill the Basel II requirements, risk managers have to take into account internal data, relevant external data (industry data) and expert opinions. It was shown in previous sections how to combine two data sources: internal data and the prior information (obtained from either external data or expert opinions), using the classical Bayesian approach. The aim of this section is to provide an example of Bayesian methodology that can be used to combine three sources of information: internal data, external data and expert opinions. Here, we follow the approach suggested in Lambrigger et al. (2007). As in the previous section, we consider one risk cell only. From the methodological side we go through the following steps:

- In any risk cell, we model the loss frequency and the loss severity by a parametric distribution (e.g., Poisson distribution for the frequency and Pareto distribution for the severity). For the bank considered, the unknown parameters $\boldsymbol{\theta}_0$ (e.g., the Poisson parameter or the Pareto tail index) of these distributions have to be quantified.

- A priori, before we have any company specific information, only industry data is available. Hence, the best estimate of our bank specific parameter $\boldsymbol{\theta}_0$ is given by the belief in the available industry data. The unknown parameter of interest is modelled by a prior distribution corresponding to a random vector $\boldsymbol{\Theta}$. The parameters of the prior distribution are estimated using industry data by, e.g., maximum likelihood estimation, as described in Section 4.2. If no industry data are available, the prior distribution could come from a "super expert" that has an overview over all banks or we choose a non-informative prior distribution with high variance.

- In our terminology, we treat the true bank specific parameter $\boldsymbol{\theta}_0$ as a realization of $\boldsymbol{\Theta}$. The prior distribution of a random vector $\boldsymbol{\Theta}$ corresponds to the whole banking industry sector, whereas $\boldsymbol{\theta}_0$ stands for the unknown underlying parameter set of the bank being considered. Due to the variability amongst banks, it is natural to model $\boldsymbol{\Theta}$ by a probability distribution. Note that $\boldsymbol{\Theta}$ is random with known distribution, whereas $\boldsymbol{\theta}_0$ is deterministic but unknown.

- Over time, internal data $\mathbf{X} = (X_1, \ldots, X_K)$ as well as expert opinions $\boldsymbol{\Delta} = (\Delta_1, \ldots, \Delta_M)$ about the underlying parameter $\boldsymbol{\theta}_0$ become available. This affects our belief in the distribution of $\boldsymbol{\Theta}$ coming from external data only and adjusts the prediction of $\boldsymbol{\theta}_0$. The more information $\mathbf{X}$ and $\boldsymbol{\Delta}$ we have, the better we are able to predict $\boldsymbol{\theta}_0$. That is, we replace the prior density $\pi(\boldsymbol{\theta})$ by a conditional density of $\boldsymbol{\Theta}$ given $\mathbf{X} = \mathbf{x}$ and $\boldsymbol{\Delta} = \boldsymbol{\delta}$, denoted by $\pi(\boldsymbol{\theta}|\mathbf{x}, \boldsymbol{\delta})$, which is also called the posterior density.

In order to determine the posterior density $\pi(\boldsymbol{\theta}|\mathbf{x}, \boldsymbol{\delta})$ we have to introduce some notation. The joint conditional density of observations and expert opinions given the parameter vector $\boldsymbol{\theta}$

is denoted by

$$h(\mathbf{x}, \boldsymbol{\delta}|\boldsymbol{\theta}) = h_1(\mathbf{x}|\boldsymbol{\theta})h_2(\boldsymbol{\delta}|\boldsymbol{\theta}), \tag{24}$$

where h_1 and h_2 are the conditional densities (given $\boldsymbol{\Theta} = \boldsymbol{\theta}$) of $\mathbf{X}$ and $\boldsymbol{\Delta}$, respectively. Thus $\mathbf{X}$ and $\boldsymbol{\Delta}$ are assumed to be conditionally independent given $\boldsymbol{\theta}$.

Remarks 6.1

- Notice that, in this way, we naturally combine external data information $\pi(\boldsymbol{\theta})$ with internal data $\mathbf{X}$ and expert opinion $\boldsymbol{\Delta}$.

- Formula (24) is quite a reasonable assumption: Assume that the true bank specific parameter is $\boldsymbol{\theta}_0$. Then (24) says that the experts in this bank estimate $\boldsymbol{\theta}_0$ (by their opinion $\boldsymbol{\delta}$) independently of the internal observations. This makes sense if the experts specify their opinions regardless of the data observed. Otherwise we should work with the joint distribution $h(\mathbf{x}, \boldsymbol{\delta}|\boldsymbol{\theta})$.

We further assume that different observations as well as expert opinions are conditionally independent and identically distributed, given $\boldsymbol{\Theta} = \boldsymbol{\theta}$, so that

$$h_1(\mathbf{x}|\boldsymbol{\theta}) = \prod_{k=1}^{K} f_1(x_k|\boldsymbol{\theta}), \tag{25}$$

$$h_2(\boldsymbol{\delta}|\boldsymbol{\theta}) = \prod_{m=1}^{M} f_2(\delta_m|\boldsymbol{\theta}), \tag{26}$$

where f_1 and f_2 are the marginal densities of a single observation and a single expert opinion, respectively. We have assumed that all expert opinions are iid, but this can be generalized easily to expert opinions having different distributions or experts whose opinions are dependent.

Let $h(\mathbf{x}, \boldsymbol{\delta})$ denote the unconditional joint density of the data $\mathbf{X}$ and expert opinions $\boldsymbol{\Delta}$. Then it follows from Bayes' Theorem that

$$h(\mathbf{x}, \boldsymbol{\delta}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) = \pi(\boldsymbol{\theta}|\mathbf{x}, \boldsymbol{\delta})h(\mathbf{x}, \boldsymbol{\delta}). \tag{27}$$

Note that the unconditional density $h(\mathbf{x}, \boldsymbol{\delta})$ does not depend on $\boldsymbol{\theta}$. Thus, using (24)-(26), the posterior density is given by

$$\pi(\boldsymbol{\theta}|\mathbf{x}, \boldsymbol{\delta}) \propto \pi(\boldsymbol{\theta}) \prod_{k=1}^{K} f_1(x_k|\boldsymbol{\theta}) \prod_{m=1}^{M} f_2(\delta_m|\boldsymbol{\theta}). \tag{28}$$

For the purposes of operational risk it is used to estimate the full predictive distribution of future losses. Hereafter, we assume that the parameters of the prior $\pi(\cdot)$ are known and we look at a single risk cell in one bank.

6.1 Conjugate Prior Extension

Equation (28) can be used in a general set-up, but it is convenient to find conjugate prior distributions such that the prior and the posterior distribution have a similar type, or where, at least, the posterior distribution can be calculated analytically. For the case of (28), the standard definition of the conjugate prior distributions can be extended as follows.

Definition 6.2 (Conjugate Prior Distribution) *Let F denote the class of density functions $h(\mathbf{x}, \boldsymbol{\delta}|\boldsymbol{\theta})$, indexed by $\boldsymbol{\theta}$. A class U of prior densities $\pi(\boldsymbol{\theta})$ is said to be a* conjugate family *for F if the posterior density $\pi(\boldsymbol{\theta}|\mathbf{x}, \boldsymbol{\delta}) \propto \pi(\boldsymbol{\theta})h(\mathbf{x}, \boldsymbol{\delta}|\boldsymbol{\theta})$ also belongs to the class U for all $h \in F$ and $\pi \in U$.*

Remarks 6.3 We work with conjugate priors because in such situations posterior distribution can be calculated analytically. In general, this is not the case and one applies simulation methods, such as the Markov chain Monte Carlo (MCMC), Gibbs and Importance Sampling methods, see Gilks et al. (1996) or Chapter 2 of this volume.

6.2 Modelling Loss Frequency: Poisson Model

To model the loss frequency for operational risk in a risk cell, consider the following model:

Model Assumptions 6.4 (Poisson-gamma-gamma) *Assume that a risk cell in a bank has a volume factor V for the frequency in a specified risk cell.*

a) Assume that the prior distribution for the loss frequency is given by a gamma distribution $\Lambda \sim \Gamma(\alpha_0, \beta_0)$ with shape parameter $\alpha_0 > 0$ and scale parameter $\beta_0 > 0$.

b) The number of losses N_k in year k, $1 \le k \le K+1$, are assumed to be conditionally, given $\Lambda = \lambda$, iid Poisson distributed with intensity $V\lambda$, i.e. $N_k|\Lambda \sim Poisson(V\lambda)$.

c) The financial company has M expert opinions Δ_m, $1 \le m \le M$. Given $\Lambda = \lambda$, $\Delta_m|\Lambda \overset{iid}{\sim} \Gamma(\xi, \lambda/\xi)$, for $\xi > 0$ fixed.

d) Given Λ, frequencies $N_1, \ldots, N_{K+1}$ and expert opinions $\Delta_1, \ldots, \Delta_M$ are independent.

Remarks 6.5

- The hyper-parameters α_0 and β_0 in Model Assumptions 6.4 a) can be estimated using the maximum likelihood method or the method of moments from external data, see Section 5.3.
- In Model Assumptions 6.4 c) we assume

$$\mathrm{E}[\Delta_m|\Lambda] = \Lambda, \quad 1 \le m \le M, \tag{29}$$

that is, expert opinions are unbiased. A possible bias might be recognized by the regulator, as he alone has the overview of the whole market.

- Note that
$$\mathrm{Vco}[\Delta_m|\Lambda] = (\mathrm{Var}[\Delta_m|\Lambda)])^{1/2}/\mathrm{E}[\Delta_m|\Lambda] = 1/\sqrt{\xi}$$
is independent of Λ. Therefore, ξ is the parameter for the relative expert opinion uncertainty which can be estimated from the observed sample of expert opinions by
$$\widehat{\xi} = (\widehat{\mu}/\widehat{\sigma})^2, \tag{30}$$
where
$$\widehat{\mu} = \frac{1}{M}\sum_{m=1}^{M}\delta_m \quad \text{and} \quad \widehat{\sigma}^2 = \frac{1}{M-1}\sum_{m=1}^{M}(\delta_m - \widehat{\mu})^2, \quad M \geq 2.$$
In the insurance practice ξ is often specified by the regulator denoting a lower bound for expert opinion uncertainty; e.g. Swiss Solvency Test, see Swiss Financial Market Supervisory Authority (2006, Appendix 8.4).

We now calculate the posterior density of Λ, given two types of information, namely the losses up to year K and the expert opinion of M experts. We introduce the following notation for the loss database and the experts:
$$\begin{aligned}\mathbf{N} &= (N_1, \ldots, N_K),\\ \boldsymbol{\Delta} &= (\Delta_1, \ldots, \Delta_M).\end{aligned}$$
Here and in what follows, we denote arithmetic means by
$$\overline{n} = \frac{1}{K}\sum_{k=1}^{K} n_k, \quad \overline{\delta} = \frac{1}{M}\sum_{m=1}^{M}\delta_m, \quad \text{etc.} \tag{31}$$
The posterior density is given by the following theorem.

Theorem 6.6

Under Model Assumptions 6.4, given loss information $\mathbf{N} = \mathbf{n}$ and expert opinion $\boldsymbol{\Delta} = \boldsymbol{\delta}$, the posterior density of Λ is
$$\pi(\lambda|\mathbf{n}, \boldsymbol{\delta}) = \frac{(\omega/\phi)^{(\nu+1)/2}}{2K_{\nu+1}(2\sqrt{\omega\phi})}\lambda^{\nu} e^{-\lambda\omega - \lambda^{-1}\phi}, \tag{32}$$
with
$$\begin{aligned}\nu &= \alpha_0 - 1 - M\xi + K\overline{n},\\ \omega &= VK + \frac{1}{\beta_0},\\ \phi &= \xi M\overline{\delta},\end{aligned} \tag{33}$$

and

$$K_{\nu+1}(z) = \frac{1}{2}\int_0^\infty u^\nu e^{-z(u+1/u)/2}\mathrm{d}u. \tag{34}$$

$K_\nu(z)$ is called a modified Bessel function of the third kind; see for instance Abramowitz and Stegun (1965, p. 375).

Proof 1 Set $\alpha = \xi$ and $\beta = \lambda/\xi$. Model Assumptions 6.4 applied to (28) yield

$$\begin{aligned}
\pi(\lambda|\mathbf{n},\boldsymbol{\delta}) &\propto \lambda^{\alpha_0-1}e^{-\lambda/\beta_0}\prod_{k=1}^{K} e^{-V\lambda}\frac{(V\lambda)^{n_k}}{n_k!}\prod_{m=1}^{M}\frac{(\delta_m)^{\alpha-1}}{\beta^\alpha}e^{-\delta_m/\beta} \\
&\propto \lambda^{\alpha_0-1}e^{-\lambda/\beta_0}\prod_{k=1}^{K} e^{-V\lambda}\lambda^{n_k}\prod_{m=1}^{M}(\xi/\lambda)^\xi e^{-\delta_m\xi/\lambda} \\
&\propto \lambda^{\alpha_0-1-M\xi+K\overline{n}}\exp\left(-\lambda\left(VK+\frac{1}{\beta_0}\right)-\frac{1}{\lambda}\xi M\overline{\delta}\right).
\end{aligned}$$

□

Remarks 6.7

- A distribution with density (32) is referred to as the generalized inverse Gaussian distribution GIG(ω, ϕ, ν). This is a well-known distribution with many applications in finance and risk management; see McNeil et al. (2005, p.75 and p.497). The GIG has been analyzed by many authors. A discussion is found, e.g., in Jørgensen (1982). The algorithm for generating realizations from a GIG is provided in see Dagpunar (1989), also see Appendix A in Lambrigger et al. (2007).
- In compare to the classical Poisson-gamma case of combining two sources of information (considered in Section 5), where the posterior is a gamma distribution, the posterior $\pi(\lambda|\cdot)$ in (35) is more complicated. In the exponent, it involves both λ and $1/\lambda$.
- Observe that the classical exponential dispersion family with associated conjugate priors (see Bühlmann and Gisler (2005), Chapter 2.5) allows for a natural extension to GIG-like distributions. In this sense the GIG distributions enlarge the classical Bayesian inference theory on the exponential dispersion family.

It is interesting to observe how the posterior density transforms when new data from a newly observed year arrive. Let ν_k, ω_k and ϕ_k denote the parameters for the data $(N_1, \ldots, N_k)$ after k accounting years. Implementation of the update processes is then given by the following equalities (assuming that expert opinions do not change).

Information update process. Year $k \to$ year $k+1$:

$$\begin{aligned} \nu_{k+1} &= \nu_k + n_{k+1}, \\ \omega_{k+1} &= \omega_k + V, \\ \phi_{k+1} &= \phi_k. \end{aligned} \tag{35}$$

Obviously, the information update process has a very simple form and only the parameter ν is affected by the new observation N_{k+1}. The posterior density (35) does not change its type every time new data arrive and hence, is easily calculated.

The moments of a GIG are not available in a closed form through elementary functions but can be expressed in terms of Bessel functions. In particular, the posterior expected number of losses is

$$\mathrm{E}[\Lambda|\mathbf{N}=\mathbf{n},\boldsymbol{\Delta}=\boldsymbol{\delta}] = \sqrt{\frac{\phi}{\omega}}\frac{K_{\nu+2}(2\sqrt{\omega\phi})}{K_{\nu+1}(2\sqrt{\omega\phi})}. \tag{36}$$

The mode of a GIG has a simple expression

$$\mathrm{mode}(\Lambda|\mathbf{N}=\mathbf{n},\boldsymbol{\Delta}=\boldsymbol{\delta}) = \frac{1}{2\omega}(\nu + \sqrt{\nu^2+4\omega\phi}). \tag{37}$$

It can be used as an alternative point estimator instead of the mean. Also, the mode of a GIG differs only slightly from the expected value for large $|\nu|$.

We are clearly interested in a robust estimator of the bank specific Poisson parameter Λ and thus the Bayesian estimator (36) is a promising candidate. The examples below show that, in practice, (36) outperforms other classical estimators. To interpret (36) in more detail, we make use of asymptotic properties. Using properties of Bessel functions, one can show that

$$R_{\nu^2}(2\nu) \to \nu, \quad \text{as} \quad \nu \to \infty,$$

where $R_\nu(z) = K_{\nu+1}(z)/K_\nu(z)$, see e.g. Lambrigger et al. (2007, Lemma B.1 in Appendix B). Using this result, a full asymptotic interpretation of the Bayesian estimator (36) can be found as follows:

Theorem 6.8 (Lambrigger et al. (2007, Theorem 3.6))
Under Model Assumptions 6.4, the following asymptotic relations hold, P-almost surely,

a) For $K \to \infty : \mathrm{E}[\Lambda|\mathbf{N},\boldsymbol{\Delta}] \to \mathrm{E}[N_k|\Lambda=\lambda]/V = \lambda$.

b) For $\mathrm{Vco}[\Delta_m|\Lambda] \to 0 : \mathrm{E}[\Lambda|\mathbf{N},\boldsymbol{\Delta}] \to \Delta_m,\ m=1,\ldots,M$.

c) For $M \to \infty : \mathrm{E}[\Lambda|\mathbf{N},\boldsymbol{\Delta}] \to \mathrm{E}[\Delta_m|\Lambda=\lambda] = \lambda$.

d) For $\mathrm{Vco}[\Delta_m|\Lambda] \to \infty,\ m=1,\ldots,M$:
$\mathrm{E}[\Lambda|\mathbf{N},\boldsymbol{\Delta}] \to \frac{1}{VK\beta_0+1}\mathrm{E}[\Lambda] + \left(1-\frac{1}{VK\beta_0+1}\right)\overline{N}/V$.

e) For $\mathrm{E}[\Lambda] =$ *constant and* $\mathrm{Vco}[\Lambda] \to 0$: $\mathrm{E}[\Lambda|\mathbf{N}, \boldsymbol{\Delta}] \to \mathrm{E}[\Lambda]$.

Remarks 6.9 The GIG mode and mean are asymptotically the same for $\nu \to \infty$; also $4\omega\phi/\nu^2 \to 0$ for $K \to \infty$, $M \to \infty$, $M \to 0$ or $\xi \to 0$. Then one can approximate the posterior mode as

$$\mathrm{mode}(\Lambda|\mathbf{N} = \mathbf{n}, \boldsymbol{\Delta} = \boldsymbol{\delta}) \approx \frac{\nu}{2\omega} 1_{\{\nu \geq 0\}} + \frac{\phi}{|\nu|} \tag{38}$$

and obtain the results of Theorem 6.8 in an elementary manner avoiding Bessel functions.

Theorem 6.8 shows that the model behaves as we would expect and require in practice, namely that credibility is given to reliable sources of information. Thus, there are good reasons to believe that it provides an adequate model to combine internal observations with relevant external data and expert opinions, as required by many risk managers. One can even go further and generalize the results from this section in a natural way to a Poisson-gamma-GIG model, i.e., where the prior distribution is a GIG. Then the posterior distribution is again a GIG (see also Remarks 6.15 below).

Example 6.10

A simple example, taken from Lambrigger et al. (2007, Example 3.7), illustrates the above methodology combining three data sources. It also extends Example 5.3 displayed in Figure 1, where two data sources are combined using classical Bayesian inference approach. Assume that:

- External data (e.g., provided by an external database or the regulator) estimate the intensity of the loss frequency by a prior gamma distribution $\Lambda \sim \Gamma(\alpha_0, \beta_0)$, as $\mathrm{E}[\Lambda] = \alpha_0\beta_0 = 0.5$ and $\Pr[0.25 \leq \Lambda \leq 0.75] = 2/3$. Then, the parameters of the prior are $\alpha_0 = 3.407$ and $\beta_0 = 0.147$; see Example 5.3.
- One expert gives an estimate of the intensity as $\delta = 0.7$. For simplicity, we consider in this example one single expert only and hence, the coefficient of variation is given a priori, (e.g., by the regulator) by $\mathrm{Vco}[\Delta|\Lambda] = \sqrt{\mathrm{Var}[\Delta|\Lambda]}/\mathrm{E}[\Delta|\Lambda] = 0.5$, i.e., $\xi = 4$.
- The observations of the annual number of losses $N_1, N_2, \ldots$ are sampled from $Poisson(0.6)$ and are the same as in the Example 5.3, i.e. given in Table 1.

This means that a priori we have a frequency parameter Λ distributed with mean $\alpha_0\beta_0 = 0.5$. The true value of the parameter λ for this risk in our bank is 0.6, i.e., it does worse than the average institution. However, our expert has an even worse opinion of his institution, namely $\delta = 0.7$. Now, we compare:

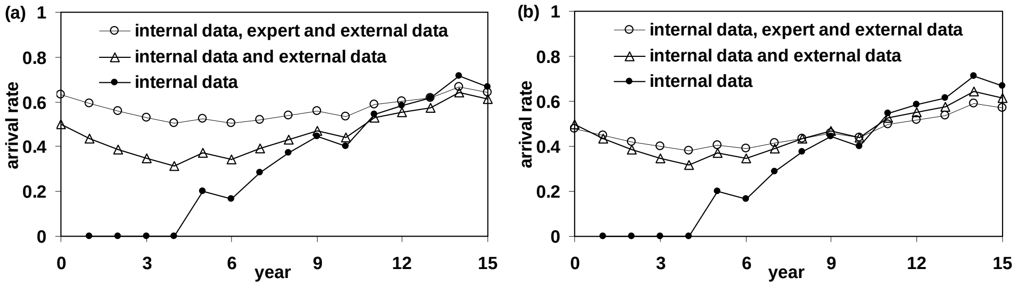

Figure 2: ($\circ$) The Bayes estimate $\widehat{\lambda}_k^{(3)}$, $k = 1, \ldots, 15$, combines the internal data simulated from $Poisson(0.6)$, external data giving $\mathrm{E}[\Lambda] = 0.5$, and expert opinion δ. It is compared with the Bayes estimate $\widehat{\lambda}_k^{(2)}$ ($\triangle$) that combines external data and internal data; and the classical maximum likelihood estimate $\widehat{\lambda}_k^{\mathrm{MLE}}$ ($\bullet$). (a) is the case of expert opinion $\delta = 0.7$ and (b) is the case of expert opinion $\delta = 0.4$. See Example 6.10 for details.

- The pure maximum likelihood estimate

$$\widehat{\lambda}_k^{\mathrm{MLE}} = \frac{1}{k}\sum_{i=1}^{k} n_i;$$

- The Bayesian estimate (17),

$$\widehat{\lambda}_k^{(2)} = \mathrm{E}[\Lambda | N_1 = n_1, \ldots, N_k = n_k], \tag{39}$$

  that combines internal data with the prior distribution. The latter in this example is assumed to be estimated from external data or provided by the regulator. Note that in Example 5.3, the prior was assumed to be determined from expert opinion only;

- The Bayesian estimate derived in formula (36):

$$\widehat{\lambda}_k^{(3)} = \mathrm{E}[\Lambda | N_1 = n_1, \ldots, N_k = n_k, \Delta = \delta], \tag{40}$$

  that combines internal data and expert opinions with the prior.

The results are plotted in Figure 2a. The estimator $\widehat{\lambda}_k^{(3)}$ shows a much more stable behavior around the true value $\lambda = 0.6$, due to the use of the prior information (market data) and the expert opinions. Given adequate expert opinions, $\widehat{\lambda}_k^{(3)}$ clearly outperforms the other estimators, particularly if only a few data points are available.

One could think that this is only the case when the experts' estimates are appropriate. However, even if experts fairly under- (or over-) estimate the true parameter λ, the method presented in here performs better for our dataset than the other mentioned methods, when a few data are available. Figure 2b displays the same estimators, but where the expert's opinion is $\delta = 0.4$, which clearly underestimates the true expected value 0.6.

The above example yields a typical picture observed in numerical experiments that demonstrates that the Bayes estimator (36) is often more suitable and stable than maximum likelihood estimators based on internal data only.

Remarks 6.11 Note that in this example the prior distribution as well as the expert opinion do not change over time. However, as soon as new information is available or when new risk management tools are in place, the corresponding parameters of $\pi(\lambda|\mathbf{N}, \boldsymbol{\Delta})$ may be easily adjusted.

Remarks 6.12 (Poisson with stochastic intensity) It is more realistic to assume that the Poisson intensity is not only different for different banks and different risks but also changes from year to year for the same risk in the same bank. Consider the annual number of events for a risk in one bank in year t modelled as random variable from a Poisson distribution $Poisson\,(\Lambda_t = \lambda_t)$. Conditional on Λ_t, the expected number of events per year is Λ_t and satisfies Model Assumptions 6.4b. In general, $(\Lambda_t)_{t\geq 0}$ is a stochastic process that can be modelled having deterministic (trend, seasonality) and stochastic components. A simple case when $(\Lambda_t)_{t\geq 0}$ is purely stochastic and distributed according to a gamma distribution was considered in Peters et al. (2009).

6.3 Modelling Loss Severities: Pareto Model

We now turn to the quantification of the severity distribution for operational risk. In general, one can use the methodology summarized by equation (28) to develop a model combining external data, internal data and expert opinion for the estimation of the severity. Here we consider the Pareto severity example studied in Lambrigger et al. (2007).

Consider modelling severities $X_1, \ldots, X_K$ using a Pareto distribution $Pareto(\theta, L)$ with threshold $L > 0$, tail parameter $\theta > 0$ and density

$$f(x) = \frac{\theta}{L}\left(\frac{x}{L}\right)^{-\theta-1}, \quad x \geq L. \tag{41}$$

Note that if $\theta > 1$, then the mean is $L\theta/(\theta-1)$, otherwise the mean does not exist. Here, we take an approach, where θ is unknown and the threshold L is known. The unknown θ is treated under the Bayesian approach as a random variable Θ. Then combining external data, internal data and expert opinions can be accomplished using the following model.

Model Assumptions 6.13 (Pareto-gamma-gamma) *Assume that for a risk cell in one bank the following holds:*

a) Let the tail parameter of the Pareto distribution $\Theta \sim \Gamma(\alpha_0, \beta_0)$ be a gamma distributed random variable with hyper-parameters α_0, β_0.

b) Given, $\Theta = \theta$, the losses $k = 1, \ldots, K+1$ in the risk cell are assumed to be conditionally iid $Pareto(\theta, L)$ distributed, where the threshold $L > 0$ is assumed to be non-random and given.

c) *We assume that the bank has M experts with opinions Δ_m, $1 \le m \le M$. Given $\Theta = \theta$, $\Delta_m \overset{iid}{\sim} Gamma(\xi, \theta/\xi)$, for $\xi > 0$ fixed.*

d) *Given Θ, $(X_1, \dots, X_{K+1})$ and $(\Delta_1, \dots, \Delta_M)$ are independent.*

Set $\mathbf{X} = (X_1, \dots, X_K)$ and $\boldsymbol{\Delta} = (\Delta_1, \dots, \Delta_M)$.

Theorem 6.14 (Lambrigger et al. (2007, Theorem 4.5))
Under Model Assumptions 6.13, given loss severities $\mathbf{X} = \mathbf{x}$ and expert opinions $\boldsymbol{\Delta} = \boldsymbol{\delta}$, the posterior distribution of Θ is $GIG(\omega, \phi, \nu)$ with the density

$$\pi(\theta|\mathbf{x}, \boldsymbol{\delta}) = \frac{(\omega/\phi)^{(\nu+1)/2}}{2K_{\nu+1}(2\sqrt{\omega\phi})}\theta^{\nu} e^{-\theta\omega - \theta^{-1}\phi}, \tag{42}$$

where

$$\begin{aligned} \nu &= \alpha_0 - 1 - M\xi + K, \\ \omega &= \frac{1}{\beta_0} + \sum_{k=1}^{K} \ln \frac{x_k}{L}, \\ \phi &= \xi M \overline{\delta}. \end{aligned} \tag{43}$$

Theorems 6.6 and 6.14 show that frequencies and severities can be handled similarly under our model assumptions. Remarks similar to the case of frequencies apply; see Remarks 6.7, 6.9, 6.11 and 6.12, and Theorem 6.8.

Remarks 6.15 It seems natural to generalize this result to the case of the GIG prior distribution. In particular, changing the assumption a) in Model Assumptions 6.13 to: $\Theta \sim GIG(\omega_0, \phi_0, \nu_0)$, with the parameters ν_0, ω_0, ϕ_0, the posterior $\pi(\theta|\mathbf{x}, \boldsymbol{\delta})$ is again $GIG(\omega, \phi, \nu)$ with

$$\begin{aligned} \nu &= \nu_0 - M\xi + K, \\ \omega &= \omega_0 + \sum_{k=1}^{K} \ln(x_k/L), \\ \phi &= \phi_0 + \xi M \overline{\delta}. \end{aligned} \tag{44}$$

Note that for $\phi_0 = 0$, the prior GIG is a gamma distribution and hence we are in the Pareto-gamma-gamma situation of Model Assumptions 6.13.

Remarks 6.16 The mean of the Pareto distributed severities is infinite for $\theta \le 1$, see (41). Thus, in the case of Model Assumptions 6.13 or assumptions in Remarks 6.15, the mean of the predictive distribution of X_{K+1} is infinite because the Pareto parameter Θ can be less than one with positive probability. For finite mean models, the range of possible θ has to be restricted to $\theta > 1$, see Shevchenko and Wüthrich (2006, Sections 3.4 and 4.3).

The update processes of (43) and (44) has again a simple linear iterative form.

Information update process. Loss $k \to$ loss $k+1$:

$$\begin{aligned} \nu_{k+1} &= \nu_k + 1, \\ \omega_{k+1} &= \omega_k + \ln(x_{k+1}/L), \\ \phi_{k+1} &= \phi_k. \end{aligned} \tag{45}$$

Example 6.17

To illustrate the simplicity and robustness of the posterior mean estimator, consider the following example. Assume that a bank would like to model its risk severity by a Pareto distribution with tail index Θ. Also, assume that the prior distribution is identified from external data (or provided by the regulator) as $\Theta \sim \Gamma(\alpha_0, \beta_0)$ with $\alpha_0 = 4$ and $\beta_0 = 9/8$, i.e., $\mathrm{E}[\Theta] = 4.5$ and $\mathrm{Vco}[\Theta] = 0.5$. The bank has one expert opinion δ with $\mathrm{Vco}[\Delta|\Theta = \theta] = 0.5$, i.e., $\xi = 4$. We then observe the losses given in Table 6.17 (sampled from a $Pareto(4, 1)$ distribution).

Table 2: Loss severities $X_i, i = 1, 2, \ldots, 15$, sampled from a $Pareto(4, 1)$ distribution.

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
x_i	1.089	1.181	1.145	1.105	1.007	1.451	1.187	1.116	1.753	1.383	2.167	1.180	1.334	1.272	1.123

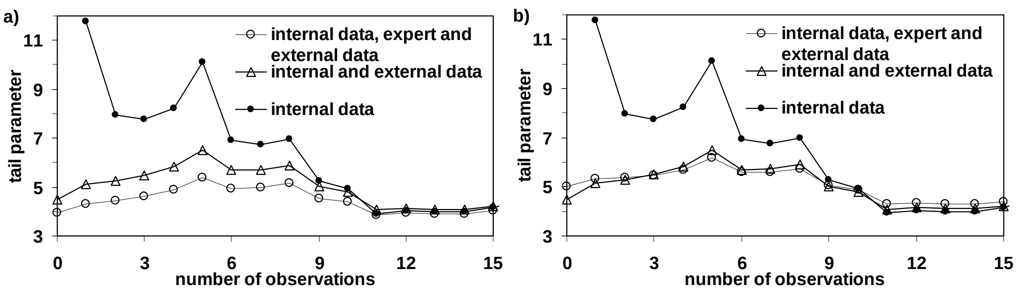

Figure 3: ($\circ$) The Bayes estimate $\widehat{\theta}_k^{(3)}$, $k = 1, \ldots, 15$, combines the internal data simulated from $Pareto(4, 1)$, external data giving $\mathrm{E}[\Theta] = 4.5$, and expert opinion δ. It is compared with the Bayes estimate $\widehat{\theta}_k^{(2)}$ ($\triangle$) that combines external data and internal data; and the classical maximum likelihood estimate $\widehat{\theta}_k^{\mathrm{MLE}}$ ($\bullet$). (a) is the case of expert opinion $\delta = 3$ and (b) is the case of expert opinion $\delta = 5$. See Example 6.17 for details.

In Figure 3 we compare:

- The classical maximum likelihood estimate

$$\widehat{\theta}_k^{\mathrm{MLE}} = \frac{k}{\sum_{i=1}^{k} \ln(x_i/L)}; \tag{46}$$

- The Bayesian posterior mean estimate

$$\theta_k^{(2)} = \mathrm{E}[\Theta|X_1 = x_1, \ldots, X_k = x_k], \tag{47}$$

that combines the internal loss data with the prior distribution, where the latter is assumed to be identified from external data or provided by the regulator;

- The Bayesian posterior mean estimate

$$\widehat{\theta}_k^{(3)} = \mathrm{E}[\Theta|X_1 = x_1, \ldots, X_k = x_k, \Delta = \delta], \tag{48}$$

that combines the internal loss data with expert opinions and the prior distribution.

Figure 3 shows the high volatility of the maximum likelihood estimator, for small numbers k. It is very sensitive to newly arriving losses. The estimator $\widehat{\theta}_k^{(3)}$ shows a much more stable behavior around the true value, most notably when a few data points are available. This example also shows that consideration of the relevant external data and well-specified expert opinions stabilizes and smoothes the estimator in an appropriate way.

7 Capital Charge Accounting for Parameter Uncertainty

According to the Basel II requirements (see BCBS (2006)) the final bank capital should be calculated as a sum of the risk measures in the risk cells if the bank's model cannot account for correlations between risks accurately. If this is the case, then one needs to calculate Value-at-Risk (VaR) for each risk cell separately and sum the VaRs over the risk cells to estimate the total bank risk capital. It is equivalent to the assumption of perfect dependence between risks. In this section we consider one risk cell but note that adding quantiles over the risk cells to find the quantile of the total loss distribution is not necessarily conservative. In fact it can underestimate the capital in the case of heavy tailed distribution as discussed in Embrechts et al. (2009), and Böcker and Klüppelberg (2009).

Consider the annual loss in a bank (or the annual loss in a risk cell where the 0.999 quantile is quantified) over the next year, Z_{T+1}. Denote the density of the annual loss, conditional on parameters $\boldsymbol{\theta}$, as $f(z|\boldsymbol{\theta})$. Typically, given observations, the MLEs $\widehat{\boldsymbol{\theta}}$ are used as the "best fit" point estimators for $\boldsymbol{\theta}$. Then, the annual loss distribution for the next year is estimated as $f(z|\widehat{\boldsymbol{\theta}})$ and its 0.999 quantile, $Q_{0.999}(\widehat{\boldsymbol{\theta}})$, is used for the capital charge calculation.

However, the parameters $\boldsymbol{\theta}$ are unknown and it is important to account for this uncertainty when the capital charge is estimated, especially for small datasets. As discussed in Shevchenko (2008), Bayesian methods are particularly useful and convenient to quantify this uncertainty because model parameters are modelled by the random variable $\boldsymbol{\Theta}$ following the posterior density $\pi(\boldsymbol{\theta}|\mathbf{y})$, where $\mathbf{Y} = \mathbf{y}$ are all available data (empirical frequencies and severities, expert opinions).

In this case, the full predictive density (accounting for parameter uncertainty) of Z_{T+1}, given all data $\mathbf{Y}$ used in the estimation procedure, is

$$f(z|\mathbf{y}) = \int f(z|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}, \tag{49}$$

also see (5). Here, it is assumed that, given $\boldsymbol{\Theta}$, Z_{T+1} and $\mathbf{Y}$ are independent. The quantile of the full predictive distribution (49),

$$Q_q^P = \inf\{z : \Pr[Z_{T+1} > z|\mathbf{Y}] \leq 1 - q\}, \tag{50}$$

at the level $q = 0.999$, can be used as a risk measure for capital calculations. Here, "P" in the upper script is used to emphasize that this is a quantile of the full predictive distribution. Another approach under a Bayesian framework to account for parameter uncertainty is to consider a quantile $Q_{0.999}(\boldsymbol{\Theta})$ of the conditional annual loss density $f(\cdot|\boldsymbol{\Theta})$:

$$Q_q(\boldsymbol{\Theta}) = \inf\{z : \Pr[Z_{T+1} > z|\boldsymbol{\Theta}] \leq 1 - q\}. \tag{51}$$

Then, given that $\boldsymbol{\Theta}$ is distributed as $\pi(\boldsymbol{\theta}|\mathbf{y})$, one can find the associated distribution of $Q_{0.999}(\boldsymbol{\Theta})$ and form a predictive interval to contain the true quantile value with some probability. Under this approach, one can argue that the conservative estimate of the capital charge accounting for parameter uncertainty should be based on the upper bound of the constructed predictive interval. Note that a specification of the confidence level for the predictive interval is required and it might be difficult to argue that the commonly used confidence level 0.95 is good enough for the estimation of the 0.999 quantile.

In operational risk, it seems that the objective should be to estimate the full predictive distribution (49) for the annual loss Z_{T+1} over the next year, conditional on all available information, and then estimate the capital charge as a quantile $Q_{0.999}^P$ of this distribution given in (50).

Consider a risk cell in the bank. Assume that the frequency $p(\cdot|\boldsymbol{\alpha})$ and severity $f(\cdot|\boldsymbol{\beta})$ densities are chosen. Also, suppose that the posterior density $\pi(\boldsymbol{\theta}|\mathbf{y})$, $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta})$ is estimated. Then, the full predictive annual loss distribution (49) in the cell can be calculated using the Monte Carlo (MC) procedure with the following logical steps, under the assumption of having a compound annual loss model for given $\boldsymbol{\Theta}$:

Algorithm 7.1 (Full predictive loss distribution via MC)

1. For $s = 1, \ldots, S$

   (a) For a given risk cell simulate the risk parameters $\boldsymbol{\theta}_s = (\boldsymbol{\alpha}_s, \boldsymbol{\beta}_s)$ from the posterior $\pi(\boldsymbol{\theta}|\mathbf{y})$. If the posterior is not known in closed form then this simulation can be done using MCMC, Gibbs and Importance Sampling methods, see Chapter 2 of this volume.

   (b) Given $\boldsymbol{\theta}_s = (\boldsymbol{\alpha}_s, \boldsymbol{\beta}_s)$, simulate the annual number of events N_s from $p(\cdot|\boldsymbol{\alpha}_s)$ and severities $X_1^{(s)}, \ldots, X_{N_s}^{(s)}$ from $f(\cdot|\boldsymbol{\beta}_s)$; and calculate the annual loss $Z^{(s)} = \sum_{n=1}^{N_s} X_n^{(s)}$.

2. Next s

The obtained annual losses $Z^{(1)}, \ldots, Z^{(S)}$ are samples from the full predictive density (49) and provide an empirical approximation. Extending the above procedure to the case of many risk cells is easy but requires specification of the dependence model, see Peters et al. (2009). In general, all model parameters (including the dependence parameters) should be simulated from their joint posterior in Step (a). Then, given these parameters, Step (b) should simulate all risks with a chosen dependence structure. In general, sampling from the joint posterior of all model parameters can be accomplished via MCMC, see e.g. Peters et al. (2009) and the Chapter of Dalla Valle in this volume. The 0.999 quantile $Q_{0.999}^P$ and other distribution characteristics can be estimated using the empirical distribution $(Z^{(s)})_{s=1,\ldots,S}$.

Remarks 7.2 Note that in the above MC procedure the risk profile $\boldsymbol{\theta}$ is simulated from its posterior distribution for each simulation. Thus, we model both the *process uncertainty*, which comes from the fact that frequencies and severities are random variables, and the parameter risk (*parameter uncertainty*), which comes from the fact that we do not know the true values of $\boldsymbol{\theta}$. The parameter uncertainty is ignored by the estimator $Q_{0.999}(\widehat{\boldsymbol{\theta}})$ but is taken into account by $Q_{0.999}^P$. For high frequency low impact risks, where a large amount of data is available, the impact is certainly expected to be small. However, for low frequency high impact risks, where the data are very limited, the impact can be significant, see Shevchenko (2008).

Also, in Step (b) of Algorithm 7.1, one can calculate the quantile $Q_{0.999}(\boldsymbol{\theta})$ of the conditional density $f(z|\boldsymbol{\theta})$, using MC or more numerically efficient methods such as FFT or Panjer recursion. Then, the obtained S samples of the quantile can be used to estimate the distribution of $Q_{0.999}(\boldsymbol{\theta})$ implied by the posterior $\pi(\boldsymbol{\theta}|\mathbf{y})$, also see (51).

Algorithm 7.3 (Posterior distribution of quantile via MC)

1. For $s = 1, \dots, S$

   (a) For a given risk simulate the risk parameters $\boldsymbol{\theta}_s = (\boldsymbol{\alpha}_s, \boldsymbol{\beta}_s)$ from the posterior $\pi(\boldsymbol{\theta}|\mathbf{y})$.

   (b) Given $\boldsymbol{\theta}_s = (\boldsymbol{\alpha}_s, \boldsymbol{\beta}_s)$, calculate the quantile $Q_q^{(s)} = Q_q(\boldsymbol{\theta}_s)$ of $f(z|\boldsymbol{\theta}_s)$, e.g. by MC, FFT or Panjer recursion.

2. Next s

Obtained $Q_q^{(1)}, \dots, Q_q^{(S)}$ provide an empirical approximation for the distribution of $Q_q(\boldsymbol{\Theta})$ defined in (51). Chapter of Dalla Valle in this volume provides some numerical examples and some graphical illustrations of this approach in the case of several risk cells.

8 General Remarks

In this chapter we have described how the parameters of the frequency and severity distributions are estimated using internal data, external data and expert opinion. Then calculation of VaR (accounting for parameter uncertainty) for each risk cell can easily be done using a simulation approach as described in Section 7.

The main motivation for the use of Bayesian approach is that, typically, the bank's internal data of the large losses in risk cells are limited, so that the standard maximum likelihood estimates are not reliable. Overall, the use of Bayesian inference methods for the quantification of the frequency and severity distributions of operational risk is very promising. The method is based on specifying the prior distributions for the parameters of the frequency and severity distributions using industry data. Then, the prior distributions are weighted with the actual observations in the bank and internal expert opinions to estimate the posterior distributions of the model parameters. These are used to estimate the predictive annual loss distribution for the next accounting year. The estimation of low frequency risks using this method has several appealing features such as: stable estimators, simple calculations (in the case of conjugate priors), and the ability to take into account expert opinions and industry data. The approach allows for combining all three data sources: internal data, external data and expert opinions required by Basel II.

The models presented in this chapter give illustrative examples that can be extended to a full scale application. The approach has a simple structure which is beneficial for the practical use and can engage the bank risk managers, statisticians and regulators in productive model development and risk assessment.

Several general remarks on the described Bayesian method for operational risk are worth to mention:

- Validation of the models in the case of small data sets are problematic. Formally, justification of the model assumptions (such as conditional independence between the losses or common distribution for the risk profiles across the risks) can be based on the analysis of the unconditional properties (e.g. unconditional means, covariances) of the losses and should be addressed during model implementation.
- Presented examples have a simplistic dependence on time but can be extended to the case of more realistic time components.
- Adding extra levels to the considered hierarchical structure may be required to model the actual risk cell structure in a bank.
- One of the features of the described method is that the variance of the posterior distribution $\pi(\theta|\cdot)$ will converge to zero for a large number of observations. This means that the true value of the risk profile will be known exactly. However, there are many factors (for example: political, economical, legal, etc.) changing over time that should not allow for the precise knowledge of the risk profiles. One can model this by limiting the variance of the posterior distribution by some lower level (e.g. 5%). This has been done in many solvency approaches for the insurance industry, e.g. in Swiss Solvency Test, see Swiss Financial Market Supervisory Authority (2006, formulas (25)-(26)).
- For convenience, we have assumed that expert opinions are iid but all formulas can be generalized to the case of expert opinions modelled by other distributions and dependence structures.
- It would be ideal if the industry risk profiles (prior distributions for frequency and severity parameters in risk cells) are calculated and provided by the regulator to ensure consistency across the banks. Unfortunately, this may not be realistic at the moment. Banks might thus estimate the industry risk profiles using industry data available through external databases from vendors and consortia of banks. The data quality, reporting and survival biases in external databases are the issues that should be considered in practice.

Finally, in this chapter we consider modelling operational risk but the use of similar Bayesian models is also useful in other areas (such as credit risk, insurance risk, environmental risk, ecology etc.) where, mainly due to lack of internal observations, a combination of internal data with external data and expert opinions is required.

References

[1] Abramowitz, M. and Stegun, I. A. (1965) *Handbook of Mathematical Functions*. Dover Publications, New York.

[2] Basel Committee on Banking Supervision (BCBS) (2006) *International Convergence of Capital Measurement and Capital Standards: a Revised Framework*. Bank for International Settlements, Basel, Switzerland.

[3] Berger, J. O. (1985) *Statistical Decision Theory and Bayesian Analysis*. Springer, New York, 2nd edn.

[4] Böcker, K. and Klüppelberg, C. (2009) Multivariate models for Operational Risk. *Quantitative Finance* To appear.

[5] Bühlmann, H. and Gisler, A. (2005) *A Course in Credibility Theory and its Applications*. Springer, Berlin.

[6] Bühlmann, H., Shevchenko, P. V. and Wüthrich, M. V. (2007) A “toy” model for operational risk quantification using credibility theory. *The Journal of Operational Risk* **2**(1), 3–19.

[7] Cruz, M. G. (2002) *Modeling, Measuring and Hedging Operational Risk*. Wiley.

[8] Dagpunar, J. S. (1989) An easily implemented generalised inverse Gaussian generator. *Communications in Statistics, Simulation and Computation* **18**, 703–710.

[9] Embrechts, P., Lambrigger, D. D. and Wüthrich, M. V. (2009) Multivariate extremes and the aggregation of dependent risks: examples and counter-examples. *Extremes* **12**(2), 107–127.

[10] Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (1996) *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London.

[11] Jørgensen, B. (1982) *Statistical Properties of the Generalized Inverse Gaussian Distribution*. Springer, New York.

[12] King, J. L. (2001) *Operational Risk: Measurements and Modelling*. Wiley.

[13] Lambrigger, D. D., Shevchenko, P. V. and Wüthrich, M. V. (2007) The quantification of operational risk using internal data, relevant external data and expert opinions. *The Journal of Operational Risk* **2**(3), 3–27.

[14] McNeil, A. J., Frey, R. and Embrechts, P. (2005) *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press, Princeton.

[15] Panjer, H. H. (2006) *Operational Risk: Modeling Analytics*. Wiley.

[16] Peters, G. W., Shevchenko, P. V. and Wüthrich, M. V. (2009) Dynamic operational risk: modeling dependence and combining different data sources of information. *The Journal of Operational Risk* **4**(2), 69–104.

[17] Robert, C. P. (2001) *The Bayesian Choice*. Springer, New York.

[18] Shevchenko, P. V. (2008) Estimation of operational risk capital charge under parameter uncertainty. *The Journal of Operational Risk* **3**(1), 51–63.

[19] Shevchenko, P. V. (2009) Implementing loss distribution approach for operational risk. *Applied Stochastic Models in Business and Industry* DOI: 10.1002/asmb.812.

[20] Shevchenko, P. V. (2010) *Modelling Operational Risk Using Bayesian Inference*. Springer, forthcoming.

[21] Shevchenko, P. V. and Wüthrich, M. V. (2006) The structural modeling of operational risk via Bayesian inference: combining loss data with expert opinions. *The Journal of Operational Risk* **1**(3), 3–26.

[22] Swiss Financial Market Supervisory Authority (FINMA) (2006) *Swiss Solvency Test, Technical Document*. Bern, Switzerland.

[23] Wüthrich, M. V. and Merz, M. (2008) *Stochastic Claims Reserving Methods in Insurance*. Wiley.