

# The Quantification of Operational Risk using Internal Data, Relevant External Data and Expert Opinions

Dominik D. Lambrigger<sup>1</sup> Pavel V. Shevchenko<sup>2</sup> Mario V. Wüthrich<sup>1</sup>

First version: April 13, 2007

This version: July 4, 2007

## Abstract

To quantify an operational risk capital charge under Basel II, many banks adopt a Loss Distribution Approach. Under this approach, quantification of the frequency and severity distributions of operational risk involves the bank's internal data, expert opinions and relevant external data. In this paper we suggest a new approach, based on a Bayesian inference method, that allows for a combination of these three sources of information to estimate the parameters of the risk frequency and severity distributions.

Keywords: Operational Risk, Basel II, Loss Distribution Approach, Bayesian inference, Advanced Measurement Approach, Quantitative Risk Management, generalized inverse Gaussian distribution.

## 1 Introduction

To meet the Basel II requirements, BIS [6], many banks adopt a Loss Distribution Approach (LDA). Under this approach, banks quantify distributions for the frequency and severity of operational losses for each risk cell over a one year time horizon; see, e.g., Cruz [7], McNeil et al. [19], Panjer [22]. Banks can use their own risk cell structure but they must be able to map the losses to the relevant Basel II risk cells (eight business lines times

---

<sup>1</sup>ETH Zurich, Department of Mathematics, CH-8092 Zurich, Switzerland.

<sup>2</sup>CSIRO Mathematical and Information Sciences, Sydney, Locked Bag 17, North Ryde, NSW, 1670, Australia; e-mail: Pavel.Shevchenko@csiro.au.

seven risk types). The commonly used LDA model for an annual loss in a single risk cell is the sum of individual losses

$$L = \sum_{k=1}^N X_k, \quad (1.1)$$

where  $N$  is the annual number of events (frequency) and  $X_k$ ,  $k = 1, \dots, N$ , are the severities of these events.

Several studies, e.g., Moscadelli [20] and Dutta and Perry [11], analyzed operational risk data collected over many banks by Basel II business line and event type; see Degen et al. [10] for a discussion and analysis of these studies. While analyses of collective data may provide a picture for the whole banking industry, estimation of frequency and severity distributions of operational risks for each risk cell is a challenging task for a single bank. The bank's internal data are typically collected over several years. On the one hand, there might be some cells with few internal data only. On the other hand, industry data available through external databases (from vendors and consortia of banks) are often difficult to adapt to internal processes, due to different volumes, thresholds etc.

Therefore, it is important to have expert judgments incorporated into the model. These judgments may provide valuable information for forecasting and decision making, especially for risk cells lacking internal loss data. In the past, quantification of operational risk was based on such expert judgments only. A quantitative assessment of risk frequency and severity distributions can be obtained from expert opinions; see, e.g., Alderweireld et al. [2]. By itself, this assessment is very subjective and should be combined with (supported by) the analysis of actual loss data. In practice, due to the absence of a sound mathematical framework, ad-hoc procedures are often used to combine the three sources of data: internal observations, external data and expert opinions. For example, the frequency distribution is estimated using internal data only, while the severity distribution is fitted to a sample combining internal and external data.

On several occasions, risk executives have emphasized that one of the main challenges in operational risk management is to combine internal data and expert opinion with relevant external data in an appropriate way; see, e.g., Davis [9], an interview with four industry's top risk executives in September 2006: *"[A] big challenge for us is how to mix the internal data with external data; this is something that is still a big problem because I don't think anybody has a solution for that at the moment."* Or: *"What can we do when we don't have enough data [...] How do I use a small amount of data when I can have external data with scenario generation? [...] I think it is one of the big challenges for operational risk*

*managers at the moment*".

A "Toy" model, based on hierarchical credibility theory, was proposed by Bühlmann et al. [5] for low frequency high impact operational risk losses exceeding some high threshold. However, this model can be too sensitive to expert opinions used to estimate scaling factors for distribution parameters. In the present framework we introduce a model that is more robust towards expert opinions.

We use Bayesian inference as the statistical technique to incorporate expert opinions into data analysis. There is a broad literature covering Bayesian inference and its applications to the insurance industry and other areas. The method allows for structural modeling of different sources of information. Shevchenko and Wüthrich [24] described the use of the Bayesian inference approach, in the context of operational risk, for estimation of frequency/severity distributions in a risk cell, where expert opinion *or* external data are used to estimate prior distributions. This allows the combining of *two* data sources: either expert opinion and internal data or external data and internal data.

The novelty in this paper is that we develop a Bayesian inference model that allows for combining three sources (internal data, external data and expert opinions) simultaneously. To the best of our knowledge, we have not seen any similar model that copes comprehensively with this task. Moreover, one should note that our framework enlarges the classical Bayesian inference models belonging to the exponential dispersion family with its associated conjugates; see, e.g., Bühlmann and Gisler [4], Chapter 2.

In Section 2 we develop a suitable method to combine the three types of knowledge in the context of operational risk. In Sections 3 and 4, this framework is used to quantify loss frequency and severity, respectively. Several examples illustrate the quality and the robustness of this quantitative approach for operational risk. In Section 5 we briefly discuss open challenges when aggregating risk cells and estimating risk capital.

## 2 Bayesian Inference

In order to estimate the risk capital of a bank and to fulfill the Basel II requirements, risk managers have to take into account information beyond the (often rare) internal data. This includes relevant external data (industry data) and expert opinions. The aim of this section is to provide some well-founded background to combining these three sources of information. Hereafter we consider one risk cell only.

In any risk cell, we model the loss frequency and the loss severity by a distribution (e.g., Poisson for the frequency or Pareto, lognormal etc. for the severity). For the considered bank, the unknown parameters  $\gamma_0$  (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 are available. Hence, the best prediction of our bank specific parameter  $\gamma_0$  is given by the belief in the available external knowledge such as the provided industry data. This unknown parameter of interest is modeled by a prior distribution (also called structural distribution or risk profile) corresponding to a random vector  $\gamma$ . The parameters of the prior distribution (so-called hyper-parameters) are estimated using data from the whole industry by, e.g., maximum likelihood estimation, as described in Shevchenko and Wüthrich [24]. If no industry data are available, the prior distribution could come from a “super expert” that has an overview over all banks.

In our terminology, we treat the true company specific parameter  $\gamma_0$  as a realization of  $\gamma$ . The random vector  $\gamma$  plays the role of the underlying parameter set of the whole banking industry sector, whereas  $\gamma_0$  stands for the unknown underlying parameter set of the bank being considered. Note that  $\gamma$  is random with known distribution, whereas  $\gamma_0$  is deterministic but unknown. Due to the variability amongst banks, it is natural to model  $\gamma$  by a probability distribution.

As time passes, internal observations  $\mathbf{X} = (X_1, \dots, X_K)$  as well as expert opinions  $\boldsymbol{\vartheta} = (\vartheta^{(1)}, \dots, \vartheta^{(M)})$  about the underlying parameter  $\gamma_0$  become available. This affects our belief in the distribution of  $\gamma$  coming from external data only and adjust the prediction of  $\gamma_0$ . The more information on  $\mathbf{X}$  and  $\boldsymbol{\vartheta}$  we have, the better we are able to predict  $\gamma_0$ . That is, we replace the prior density  $\pi(\gamma)$  by a conditional density of  $\gamma$  given  $\mathbf{X}$  and  $\boldsymbol{\vartheta}$ .

The natural question that arises at this point is: How does this company specific information  $\mathbf{X}$  and  $\boldsymbol{\vartheta}$  change our view of the underlying parameter  $\gamma$ , i.e., what is the distribution of  $\gamma|\mathbf{X}, \boldsymbol{\vartheta}$ ?

The Bayesian inference approach yields the canonical theory answering questions of the above type. In order to determine  $\gamma|\mathbf{X}, \boldsymbol{\vartheta}$  we have to introduce some notation. The joint conditional density of observations and expert opinions given the parameter vector  $\gamma$  is denoted by

$$h(\mathbf{X}, \boldsymbol{\vartheta}|\gamma) = h_1(\mathbf{X}|\gamma)h_2(\boldsymbol{\vartheta}|\gamma), \quad (2.1)$$

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

REMARKS 2.1:

- Notice that, in this way, we naturally combine external data  $\gamma$  with internal data  $\mathbf{X}$  and expert opinion  $\boldsymbol{\vartheta}$ .
- In classical Bayesian inference (as it is used, e.g., in actuarial science), one usually combines only two sources of information. The novelty in this paper is that we combine three sources simultaneously using an appropriate structure, i.e., equation (2.1).
- (2.1) is quite a reasonable assumption: Assume that the true bank specific parameter is  $\gamma_0$ . Then (2.1) says that the experts in this bank estimate  $\gamma_0$  (by their opinion  $\boldsymbol{\vartheta}$ ) independently of the internal observations. This makes sense if the experts specify their opinions regardless of the data observed.  $\square$

We further assume that observations as well as expert opinions are conditionally independent and identically distributed (i.i.d.), given  $\gamma$ , so that

$$h_1(\mathbf{X}|\gamma) = \prod_{k=1}^K f_1(X_k|\gamma), \quad (2.2)$$

$$h_2(\boldsymbol{\vartheta}|\gamma) = \prod_{m=1}^M f_2(\vartheta^{(m)}|\gamma), \quad (2.3)$$

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 identically distributed, but this can be generalized easily to expert opinions having different distributions.

The unconditional parameter density  $\pi(\gamma)$  is called the *prior* density, whereas the conditional parameter density  $\hat{\pi}(\gamma|\mathbf{X}, \boldsymbol{\vartheta})$  is called the *posterior* density. Let  $h(\mathbf{X}, \boldsymbol{\vartheta})$  denote the unconditional joint density of observations  $\mathbf{X}$  and expert opinions  $\boldsymbol{\vartheta}$ . Then it follows from Bayes' Theorem that

$$h(\mathbf{X}, \boldsymbol{\vartheta}|\gamma)\pi(\gamma) = \hat{\pi}(\gamma|\mathbf{X}, \boldsymbol{\vartheta})h(\mathbf{X}, \boldsymbol{\vartheta}). \quad (2.4)$$

Note that the unconditional density  $h(\mathbf{X}, \boldsymbol{\vartheta})$  does not depend on  $\gamma$  and, thus, the posterior density is given by

$$\hat{\pi}(\gamma|\mathbf{X}, \boldsymbol{\vartheta}) \propto \pi(\gamma) \prod_{k=1}^K f_1(X_k|\gamma) \prod_{m=1}^M f_2(\vartheta^{(m)}|\gamma), \quad (2.5)$$

where “ $\propto$ ” stands for “is proportional to” with the constant of proportionality independent of the parameter vector  $\boldsymbol{\gamma}$ . For the purposes of operational risk it is used to estimate the full predictive distribution of future losses.

Equation (2.5) can be used in a general set-up, but it is convenient to find some *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.

**DEFINITION 2.2 (CONJUGATE PRIOR DISTRIBUTION)** Let  $F$  denote the class of density functions  $h(\mathbf{X}, \boldsymbol{\vartheta}|\boldsymbol{\gamma})$ , indexed by  $\boldsymbol{\gamma}$ . A class  $U$  of prior densities  $\pi(\boldsymbol{\gamma})$  is said to be a *conjugate family* for  $F$  if the posterior density  $\hat{\pi}(\boldsymbol{\gamma}|\mathbf{X}, \boldsymbol{\vartheta}) \propto \pi(\boldsymbol{\gamma})h(\mathbf{X}, \boldsymbol{\vartheta}|\boldsymbol{\gamma})$  also belongs to the class  $U$  for all  $h \in F$  and  $\pi \in U$ .  $\square$

Conjugate distributions are very useful in practice and will be used consistently throughout this paper. At this point, we also refer to Bühlmann and Gisler [4], Section 2.5. In general, the posterior distribution cannot be calculated analytically but can be estimated numerically for instance by the Markov Chain Monte Carlo method; see, e.g., Peters and Sisson [23] or Gilks et al. [17].

### 3 Loss Frequency

#### 3.1 Combining internal data and expert opinions with external data

##### Model Assumptions 3.1 (Poisson-Gamma-Gamma)

Assume that bank  $i$  has a scaling factor  $V_i$ ,  $1 \leq i \leq I$ , for the frequency in a specified risk cell (e.g., it can be a product of economic indicators such as the gross income, the number of transactions, the number of staff, etc.). We choose the following model for the loss frequency for operational risk of a risk cell in bank  $i$ :

- a) Let  $\Lambda_i \sim \Gamma(\alpha_0, \beta_0)$  be a Gamma distributed random variable with shape parameter  $\alpha_0 > 0$  and scale parameter  $\beta_0 > 0$ , which are estimated from (external) market data. That is, the density of  $\Gamma(\alpha_0, \beta_0)$ ,  $x^{\alpha_0-1}e^{-x/\beta_0}/(\beta_0^{\alpha_0}\Gamma(\alpha_0))$  ( $x > 0$ ), plays the role of  $\pi(\boldsymbol{\gamma})$  in (2.5).
- b) The number of losses of bank  $i$  in year  $k$ ,  $1 \leq k \leq K_i$ , are assumed to be conditionally i.i.d., given  $\Lambda_i$ , Poisson distributed with frequency  $V_i\Lambda_i$ , i.e.,  $N_{i,1}, \dots, N_{i,K_i} | \Lambda_i \stackrel{\text{i.i.d.}}{\sim}$

$\text{Pois}(V_i \Lambda_i)$ . That is,  $f_1(\cdot | \Lambda_i)$  in (2.5) corresponds to the density of a  $\text{Pois}(V_i \Lambda_i)$  distribution.

- c) We assume that bank  $i$  has  $M_i$  experts with opinions  $\vartheta_i^{(m)}$ ,  $1 \leq m \leq M_i$ , about the company specific intensity parameter  $\Lambda_i$  with  $\vartheta_i^{(m)} | \Lambda_i \stackrel{\text{i.i.d.}}{\sim} \Gamma(\xi_i, \frac{\Lambda_i}{\xi_i})$ , where  $\xi_i$  is a known parameter. That is,  $f_2(\cdot | \Lambda_i)$  corresponds to the density of a  $\Gamma(\xi_i, \frac{\Lambda_i}{\xi_i})$  distribution.  $\square$

REMARKS 3.2:

- In the sequel, we only look at a single bank  $i$  and therefore we could drop the index  $i$ . However, we refrain from doing so in order to highlight the fact that we do not consider the whole banking industry, but only a single bank.
- The parameters  $\alpha_0$  and  $\beta_0$  in Model Assumptions 3.1 a) are called hyper-parameters (parameters for parameters); see, e.g., Bühlmann and Gisler [4], p. 38. These parameters are estimated using the maximum likelihood method or the method of moments; see for instance Shevchenko and Wüthrich [24], Section 5 and Appendix B.
- In Model Assumptions 3.1 c) we assume

$$\mathbb{E}[\vartheta_i^{(m)} | \Lambda_i] = \Lambda_i, \quad 1 \leq m \leq M_i, \quad (3.1)$$

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

Note that the *coefficient of variation* of the conditional expert opinion  $\vartheta_i^{(m)} | \Lambda_i$  of company  $i$  is  $\text{Vco}(\vartheta_i^{(m)} | \Lambda_i) = (\text{var}(\vartheta_i^{(m)} | \Lambda_i))^{1/2} / \mathbb{E}[\vartheta_i^{(m)} | \Lambda_i] = 1/\sqrt{\xi_i}$ , and thus is independent of  $\Lambda_i$ . This means that  $\xi_i$ , which characterizes the uncertainty in the expert opinions, is independent of the true bank specific  $\Lambda_i$ . For simplicity, we have assumed that all experts have the same conditional coefficient of variation and thus have the same credibility. Moreover, this allows for the estimation of  $\xi_i$  within each company  $i$ , e.g., by  $\hat{\xi}_i = (\hat{\mu}_i / \hat{\sigma}_i)^2$  with

$$\hat{\mu}_i = \frac{1}{M_i} \sum_{m=1}^{M_i} \vartheta_i^{(m)} \quad \text{and} \quad \hat{\sigma}_i^2 = \frac{1}{M_i - 1} \sum_{m=1}^{M_i} (\vartheta_i^{(m)} - \hat{\mu}_i)^2, \quad M_i \geq 2. \quad (3.2)$$

In a more general framework the parameter  $\xi_i$  can be estimated, e.g., by maximum likelihood. If the credibility differs among the experts, then  $\vartheta_i^{(m)}$  and  $\text{Vco}(\vartheta_i^{(m)} | \Lambda_i)$  should be estimated for all  $m$ ,  $1 \leq m \leq M_i$ . This may often be a (too) challenging issue in practice.

REMARKS 3.3:

- $\Lambda_i$  is the risk characteristic of a risk cell in bank  $i$ . A priori, before we have any observations, the banks are all the same, i.e.,  $\Lambda_i$  is i.i.d. Observations and expert opinions modify this characteristic  $\Lambda_i$  according to the actual experience in company  $i$ , which gives different posteriors  $\Lambda_i|N_{i,1}, \dots, N_{i,K_i}, \vartheta_i^{(1)}, \dots, \vartheta_i^{(M_i)}$ .
- This model can be extended to a model where one allows for more flexibility in the expert opinions. For convenience, we prefer that experts are conditionally i.i.d., given  $\Lambda_i$ . This has the advantage that there is only one parameter,  $\xi_i$ , that needs to be estimated.  $\square$

Using the notation from Section 2, we calculate the posterior density of  $\Lambda_i$ , given the losses up to year  $K_i$  and the expert opinion of  $M_i$  experts. We introduce the following notation for the loss database and the expert knowledge of bank  $i$ :

$$\begin{aligned} \mathbf{N}_i &= (N_{i,1}, \dots, N_{i,K_i}), \\ \boldsymbol{\vartheta}_i &= (\vartheta_i^{(1)}, \dots, \vartheta_i^{(M_i)}). \end{aligned}$$

Here and in what follows, we denote arithmetic means by

$$\bar{N}_i = \frac{1}{K_i} \sum_{k=1}^{K_i} N_{i,k}, \quad \bar{\vartheta}_i = \frac{1}{M_i} \sum_{m=1}^{M_i} \vartheta_i^{(m)}, \quad \text{etc.} \quad (3.3)$$

The posterior density  $\hat{\pi}$  is given by the following theorem.

**Theorem 3.4**

*Under Model Assumptions 3.1, the posterior density of  $\Lambda_i$ , given loss information  $\mathbf{N}_i$  and expert opinion  $\boldsymbol{\vartheta}_i$ , is given by*

$$\hat{\pi}_{\Lambda_i}(\lambda_i|\mathbf{N}_i, \boldsymbol{\vartheta}_i) = \frac{(\omega/\phi)^{(\nu+1)/2}}{2K_{\nu+1}(2\sqrt{\omega\phi})} \lambda_i^\nu e^{-\lambda_i\omega - \lambda_i^{-1}\phi}, \quad (3.4)$$

with

$$\begin{aligned} \nu &= \alpha_0 - 1 - M_i\xi_i + K_i\bar{N}_i, \\ \omega &= V_iK_i + \frac{1}{\beta_0}, \\ \phi &= \xi_iM_i\bar{\vartheta}_i, \end{aligned} \quad (3.5)$$

and

$$K_{\nu+1}(z) = \frac{1}{2} \int_0^\infty u^\nu e^{-z(u+1/u)/2} du. \quad (3.6)$$

$K_\nu(z)$  is called a modified Bessel function of the third kind; see for instance Abramowitz and Stegun [1], p. 375.

PROOF: Set  $\alpha_i = \xi_i$  and  $\beta_i = \lambda_i/\xi_i$ . Model Assumptions 3.1 applied to (2.5) yield

$$\begin{aligned}
\widehat{\pi}_{\Lambda_i}(\lambda_i | \mathbf{N}_i, \boldsymbol{\vartheta}_i) &\propto \lambda_i^{\alpha_0-1} e^{-\lambda_i/\beta_0} \prod_{k=1}^{K_i} e^{-V_i \lambda_i} \frac{(V_i \lambda_i)^{N_{i,k}}}{N_{i,k}!} \prod_{m=1}^{M_i} \frac{(\vartheta_i^{(m)}/\beta_i)^{\alpha_i-1}}{\beta_i} e^{-\vartheta_i^{(m)}/\beta_i} \\
&\propto \lambda_i^{\alpha_0-1} e^{-\lambda_i/\beta_0} \prod_{k=1}^{K_i} e^{-V_i \lambda_i} \lambda_i^{N_{i,k}} \prod_{m=1}^{M_i} (\xi_i/\lambda_i)^{\xi_i} e^{-\vartheta_i^{(m)} \xi_i/\lambda_i} \\
&\propto \lambda_i^{\alpha_0-1-M_i \xi_i+K_i \bar{N}_i} \exp\left(-\lambda_i \left(V_i K_i + \frac{1}{\beta_0}\right) - \frac{1}{\lambda_i} \xi_i M_i \bar{\vartheta}_i\right). \quad (3.7) \quad \square
\end{aligned}$$

REMARKS 3.5:

- A distribution with density (3.4) is referred to as the generalized inverse Gaussian distribution  $\text{GIG}(\omega, \phi, \nu)$ . This is a well-known distribution with many applications in finance and risk management; see McNeil et al. [19]. The GIG has been analyzed by many authors. A discussion is found, e.g., in Jørgensen [18]. The GIG belongs to the popular class of subexponential distributions; see Embrechts [12] for a proof and Embrechts et al. [13] for a detailed treatment of subexponential distributions. The GIG with  $\nu \leq 1$  is a first hitting time distribution for certain time-homogeneous processes; see for instance Jørgensen [18], Chapter 6. In particular, the (standard) inverse Gaussian (i.e., the GIG with  $\nu = -3/2$ ) is known by financial practitioners as the distribution function determined by the first passage time of a Brownian motion. Algorithms for generating realizations from a GIG are provided by Atkinson [3] and Dagpunar [8]; see also McNeil et al. [19] and Appendix A below.
- Unlike in the classical Poisson-Gamma case of combining two sources of information (see Shevchenko and Wüthrich [24], Bühlmann and Gisler [4]), we obtain in (3.7) a more complicated posterior distribution  $\widehat{\pi}$ , which involves in the exponent both  $\lambda_i$  and  $1/\lambda_i$ . Note that expert opinions enter via the term  $1/\lambda_i$  only. We give some basic properties of the GIG distribution below.
- Observe that the classical exponential dispersion family (EDF) with associated conjugates (see Bühlmann and Gisler [4], 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.  $\square$

For our purposes it is interesting to observe how the posterior density transforms when new data from a newly observed year arrive. Let  $\nu_k$ ,  $\omega_k$  and  $\phi_k$  denote the parameters for the observations  $(N_{i,1}, \dots, N_{i,k})$  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  $k \rightarrow$  year  $k + 1$ :

$$\begin{aligned}\nu_{k+1} &= \nu_k + N_{i,k+1}, \\ \omega_{k+1} &= \omega_k + V_i, \\ \phi_{k+1} &= \phi_k.\end{aligned}\tag{3.8}$$

Obviously, the information update process has a very simple form and only the parameter  $\nu$  is affected by the new observation  $N_{i,k+1}$ . The posterior density (3.7) does not change its type every time new data arrive and hence, is easily calculated.

The moments of a GIG cannot be given in a closed form by elementary functions. However, for  $\alpha \geq 1$ , all moments are given in terms of Bessel functions:

$$\mathbb{E}[\Lambda_i^\alpha | \mathbf{N}_i, \boldsymbol{\vartheta}_i] = \left(\frac{\phi}{\omega}\right)^{\alpha/2} \frac{K_{\nu+1+\alpha}(2\sqrt{\omega\phi})}{K_{\nu+1}(2\sqrt{\omega\phi})}.\tag{3.9}$$

A useful notation is the following:

$$R_\nu(z) = \frac{K_{\nu+1}(z)}{K_\nu(z)}.\tag{3.10}$$

Then it follows for the posterior expected number of losses

$$\mathbb{E}[\Lambda_i | \mathbf{N}_i, \boldsymbol{\vartheta}_i] = \sqrt{\frac{\phi}{\omega}} R_{\nu+1}(2\sqrt{\omega\phi}),\tag{3.11}$$

and for the higher moments

$$\mathbb{E}[\Lambda_i^\alpha | \mathbf{N}_i, \boldsymbol{\vartheta}_i] = \left(\frac{\phi}{\omega}\right)^{\alpha/2} \prod_{k=1}^{\alpha} R_{\nu+k}(2\sqrt{\omega\phi}), \quad \alpha = 2, 3, \dots\tag{3.12}$$

We are clearly interested in robust prediction of the bank specific Poisson parameter and thus the Bayesian estimator (3.11) is a promising candidate within this operational risk framework. The examples below show that, in practice, (3.11) outperforms other classical estimators. To interpret (3.11) in more detail, we make use of asymptotic properties. Here and throughout the paper,  $f(x) \sim g(x)$ ,  $x \rightarrow a$ , means that  $\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = 1$ . Lemma

B.1 in Appendix B basically says that  $R_{\nu^2}(2\nu) \sim \nu$  is asymptotically linear for  $\nu \rightarrow \infty$ . This is the key in the proof of Theorem 3.6 and yields a full asymptotic interpretation of the Bayesian estimator (3.11).

**Theorem 3.6**

*Under Model Assumptions 3.1, the following asymptotic relations hold  $\mathbb{P}$ -almost surely:*

- a) *Assume, given  $\Lambda_i = \lambda_i$ ,  $N_{i,k} \stackrel{\text{i.i.d.}}{\sim} \text{Pois}(V_i \lambda_i)$  and  $\vartheta_i^{(m)} \stackrel{\text{i.i.d.}}{\sim} \Gamma(\xi_i, \lambda_i / \xi_i)$ .  
For  $K_i \rightarrow \infty$  :  $\mathbb{E}[\Lambda_i | \mathbf{N}_i, \boldsymbol{\vartheta}_i] \rightarrow \mathbb{E}[N_{i,k} | \Lambda_i = \lambda_i] / V_i = \lambda_i$ .*
- b) *For  $\text{Vco}(\vartheta_i^{(m)} | \Lambda_i) \rightarrow 0$  :  $\mathbb{E}[\Lambda_i | \mathbf{N}_i, \boldsymbol{\vartheta}_i] \rightarrow \vartheta_i^{(m)}$ ,  $m = 1, \dots, M_i$ .*
- c) *Assume, given  $\Lambda_i = \lambda_i$ ,  $N_{i,k} \stackrel{\text{i.i.d.}}{\sim} \text{Pois}(V_i \lambda_i)$  and  $\vartheta_i^{(m)} \stackrel{\text{i.i.d.}}{\sim} \Gamma(\xi_i, \lambda_i / \xi_i)$ .  
For  $M_i \rightarrow \infty$  :  $\mathbb{E}[\Lambda_i | \mathbf{N}_i, \boldsymbol{\vartheta}_i] \rightarrow \mathbb{E}[\vartheta_i^{(m)} | \Lambda_i = \lambda_i] = \lambda_i$ .*
- d) *For  $\text{Vco}(\vartheta_i^{(m)} | \Lambda_i) \rightarrow \infty$ ,  $m = 1, \dots, M_i$  :  
$$\mathbb{E}[\Lambda_i | \mathbf{N}_i, \boldsymbol{\vartheta}_i] \rightarrow \frac{1}{V_i K_i \beta_0 + 1} \mathbb{E}[\Lambda_i] + \left(1 - \frac{1}{V_i K_i \beta_0 + 1}\right) \bar{N}_i / V_i$$
.*
- e) *For  $\mathbb{E}[\Lambda_i] = \text{constant}$  and  $\text{Vco}(\Lambda_i) \rightarrow 0$  :  $\mathbb{E}[\Lambda_i | \mathbf{N}_i, \boldsymbol{\vartheta}_i] \rightarrow \mathbb{E}[\Lambda_i]$ .*

PROOF: See Appendix C. □

Theorem 3.6 yields a natural interpretation of the posterior density (3.4) and its expected value (3.11). As the number of observations increases, we give more weight to them and in the limit  $K_i \rightarrow \infty$  (case a) we completely believe in the observations  $N_{i,k}$  and we neglect a priori information and expert opinion. On the other hand, the more the coefficient of variation of the expert opinions decreases, the more weight is given to them (case b). In Model 3.1, we assume experts to be conditionally independent. In practice, however, even for  $\text{Vco}(\vartheta_i^{(m)} | \Lambda_i) \rightarrow 0$ , the variance of  $\bar{\vartheta}_i | \Lambda_i$  cannot be made arbitrarily small when increasing the number of experts, as there is always a positive covariance term due to positive dependence between experts. Since we predict random variables, we never have “perfect diversification”, that is, in practical applications we would probably question property c.

Conversely, if experts become less credible in terms of having an increasing coefficient of variation, our model behaves as if the experts do not exist (case d). The Bayes estimator is then a weighted sum of prior and posterior information with appropriate credibility weights. This is the classical credibility result obtained from Bayesian inference on

the exponential dispersion family with two sources of information; see Shevchenko and Wüthrich [24], Formula (12).

Of course, if the coefficient of variation of the prior distribution (i.e., of the whole banking industry) vanishes, the external data are not affected by internal data and expert opinion (case e).

In this sense, Theorem 3.6 shows that our model behaves exactly as we would expect and require in practice. Thus, we have 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.

Note that 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 Model Assumptions 4.6 below).

### 3.2 Implementation and practical application

In this section we apply the above theory to a concrete example. The Bayesian estimator (3.11) derived above is easily implemented in practice. The following example extends the example displayed in Figure 1 in Shevchenko and Wüthrich [24].

**EXAMPLE 3.7** Assume that external data (e.g., provided by external databases or regulator) estimate the parameter of the loss frequency (i.e., the Poisson parameter  $\Lambda$ ) which has a Gamma distribution  $\Lambda \sim \Gamma(\alpha_0, \beta_0)$  as  $\mathbb{E}[\Lambda] = \alpha_0\beta_0 = 0.5$  and  $\mathbb{P}[0.25 \leq \Lambda \leq 0.75] = 2/3$ . Then, the parameters of the prior Gamma distribution are  $\alpha_0 \approx 3.407$  and  $\beta_0 \approx 0.147$ ; see Shevchenko and Wüthrich [24], Section 4.1.

Now, we consider one particular bank  $i$ :

- i) One expert says that  $\vartheta$  is estimated by  $\hat{\vartheta} = 0.7$ . For simplicity, we consider in this example one single expert only and hence, the coefficient of variation is not estimated using (3.2), but given a priori, e.g., by the regulator:  $\text{Vco}(\vartheta|\Lambda) = (\text{var}(\vartheta|\Lambda))^{1/2}/\mathbb{E}[\vartheta|\Lambda] = 0.5$ , i.e.,  $\xi = 4$ .
- ii) The observations of the annual number of losses are given as follows (sampled from a Poisson distribution with parameter  $\lambda = 0.6$ ; this is the dataset used in Shevchenko and Wüthrich [24]):

Year $i$	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

This means that a priori we have a frequency parameter distributed as  $\Gamma(\alpha_0, \beta_0)$  with mean  $\alpha_0\beta_0 = 0.5$ . The true parameter for this institution is  $\lambda = 0.6$ , i.e., it does worse than the average institution. However, our expert has an even worse opinion of his institution, namely  $\hat{\vartheta} = 0.7$ .

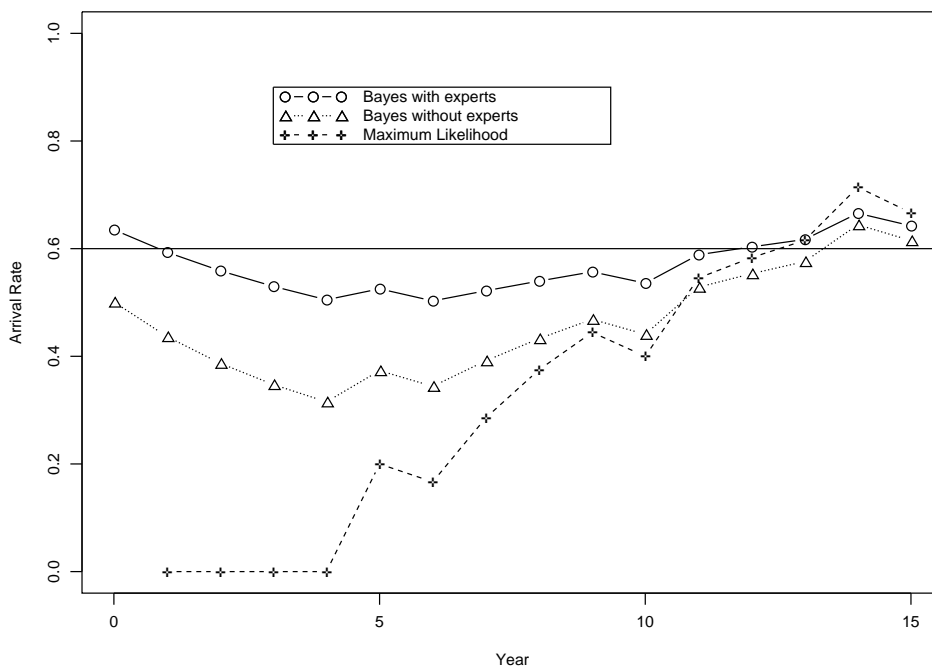


Figure 1: The Bayes estimator  $(\lambda_k)_{k=0,\dots,15}$  includes internal data simulated from  $\text{Poisson}(0.6)$ , external data  $\Lambda$  with  $\mathbb{E}[\Lambda] = 0.5$  and expert opinion  $\hat{\vartheta} = 0.7$  ( $\circ$ ). It is compared with the Bayes estimator  $\lambda_k^{\text{SW}}$  proposed in Shevchenko and Wüthrich [24] ( $\triangle$ ) and the classical maximum likelihood estimator ( $+$ ).

We compare the pure maximum likelihood estimator  $\lambda_k^{\text{MLE}} = \frac{1}{k} \sum_{i=1}^k N_i$  and the Bayesian estimator

$$\lambda_k^{\text{SW}} = \mathbb{E}[\Lambda | N_1, \dots, N_k], \quad (3.13)$$

proposed in Shevchenko and Wüthrich [24] (without expert opinion) with the Bayesian

estimator derived in formula (3.11), including expert opinion:

$$\lambda_k = \mathbb{E}[\Lambda | N_1, \dots, N_k, \vartheta]. \quad (3.14)$$

The results are plotted in Figure 1. The estimator (3.11) 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, the Bayesian estimator (3.11) clearly outperforms the other estimators, particularly if only a few data points are available.

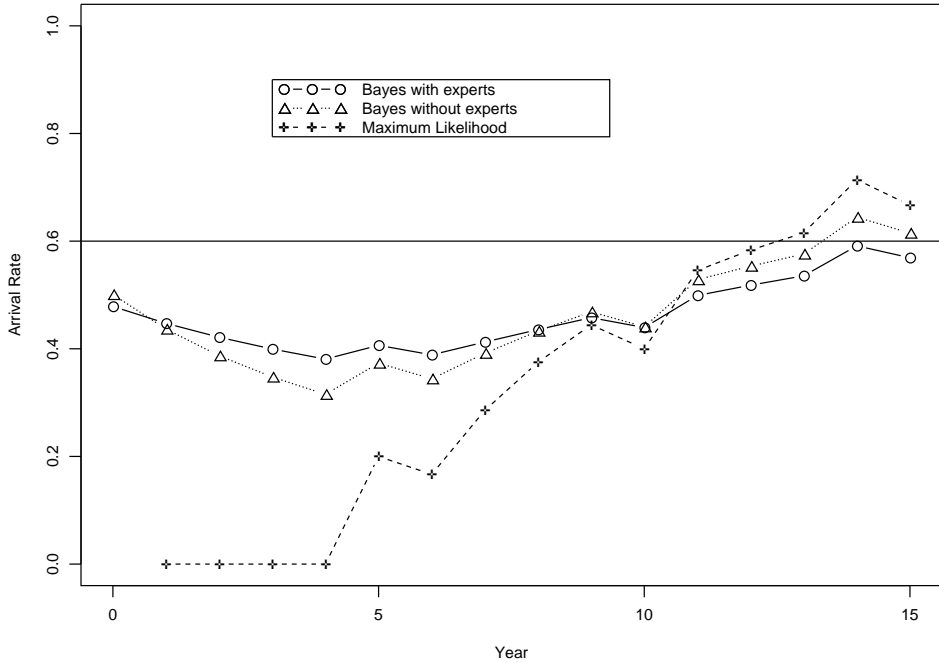


Figure 2: The same estimators as in Figure 1 are displayed, but where the expert underestimates the true  $\lambda = 0.6$  by  $\hat{\vartheta} = 0.4$ .

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  $\lambda$ , the method presented in this paper performs better for our dataset than the other mentioned methods, when a few data are available. In Figure 2 we display the same estimators, but where the experts' opinion is  $\hat{\vartheta} = 0.4$ , which clearly underestimates the true expected value 0.6.

In Figure 1  $\lambda_k$  gives better estimates when compared to  $\lambda_k^{\text{SW}}$ . Observe that also in Figure 2  $\lambda_k$  gives more appropriate estimates than  $\lambda_k^{\text{SW}}$ . Though the expert is too optimistic,  $\lambda_k$  manages to correct  $\lambda_k^{\text{MLE}}$  ( $k \leq 10$ ), which is clearly too low.  $\square$

This example yields a typical picture observed in numerical experiments that demonstrates that the Bayes estimator (3.11) is often more suitable and stable than maximum likelihood estimators based on internal data only.

REMARK 3.8 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 may be easily adjusted.  $\square$

### 3.3 Alternative estimator using the mode

Instead of calculating the mean of the  $\text{GIG}(\nu, \omega, \phi)$  as we did in the estimator (3.11), we could use the *mode* of the distribution, i.e., the point where the density function is maximum. The mode of a GIG differs only slightly from the expected value for large  $|\nu|$ . In particular, one proves, e.g., that for  $X \sim \text{GIG}(\nu, \omega, \phi)$  we have

$$\text{mode}(X) \sim \mathbb{E}[X] \quad \text{for } \nu \rightarrow \infty. \quad (3.15)$$

The mode of a  $\text{GIG}(\nu, \omega, \phi)$  is easily calculated by

$$\frac{\partial}{\partial x} x^\nu e^{-(\omega x + \phi/x)} = 0. \quad (3.16)$$

Hence,

$$\text{mode}(X) = \frac{1}{2\omega} (\nu + \sqrt{\nu^2 + 4\omega\phi}), \quad (3.17)$$

which gives us a good approximation to the mean for large  $\nu$ . Thus, we have

$$\text{mode}(\Lambda_i | \mathbf{N}_i, \boldsymbol{\vartheta}_i) = \frac{\nu}{2\omega} \left( 1 + \text{sign}(\nu) \sqrt{1 + \frac{4\omega\phi}{\nu^2}} \right), \quad (3.18)$$

where  $\nu$ ,  $\omega$ , and  $\phi$  are given by equations (3.5). Due to  $\frac{4\omega\phi}{\nu^2} \rightarrow 0$  for  $K_i \rightarrow \infty$ ,  $M_i \rightarrow \infty$ ,  $M_i \rightarrow 0$  or  $\xi \rightarrow 0$ , we approximate  $\sqrt{1+2x} \approx 1+x$ ,  $x \rightarrow 0$ , and hence

$$\text{mode}(\Lambda_i | \mathbf{N}_i, \boldsymbol{\vartheta}_i) \approx \frac{\nu}{2\omega} \mathbf{1}_{\{\nu \geq 0\}} + \frac{\phi}{|\nu|}. \quad (3.19)$$

With (3.19) we again get the results from Theorem 3.6 in an elementary manner avoiding Bessel functions.

## 4 Loss Severities

In the previous section we presented a method to quantify the operational risk loss frequency. We now turn to quantification of the severity distribution for operational risk. This is done in this section for different types of subexponential models.

### 4.1 Lognormal model (Model 1 for severities)

#### Model Assumptions 4.1 (Lognormal-normal-normal)

Let us assume the following severity model for operational risk of a risk cell in bank  $i$ ,  $1 \leq i \leq I$ :

- a) Let  $\Delta_i \sim \mathcal{N}(\mu_0, \sigma_0)$  be a normally distributed random variable with parameters  $\mu_0, \sigma_0$ , which are estimated from (external) market data, i.e.,  $\pi(\gamma)$  in (2.5) is the density of  $\mathcal{N}(\mu_0, \sigma_0)$ .
- b) The losses  $k = 1, \dots, K_i$  from institution  $i$  are assumed to be conditionally (on  $\Delta_i$ ) i.i.d. lognormally distributed:  $X_{i,1}, \dots, X_{i,K_i} | \Delta_i \stackrel{\text{i.i.d.}}{\sim} \text{LN}(\Delta_i, \sigma_i)$ , where  $\sigma_i$  is assumed known. That is,  $f_1(\cdot | \Delta_i)$  in (2.5) corresponds to the density of a  $\text{LN}(\Delta_i, \sigma_i)$  distribution.
- c) We assume that bank  $i$  has  $M_i$  experts with opinions  $\vartheta_i^{(m)}$ ,  $1 \leq m \leq M_i$ , about the parameter  $\Delta_i$  with  $\vartheta_i^{(m)} | \Delta_i \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_i = \Delta_i, \tilde{\sigma}_i = \xi_i)$ , where  $\xi_i$  is a parameter estimated using expert opinion data. That is,  $f_2(\cdot | \Delta_i)$  corresponds to the density of a  $\mathcal{N}(\Delta_i, \xi_i)$  distribution.  $\square$

REMARKS 4.2:

- For  $M_i \geq 2$ , the parameter  $\xi_i$  is, e.g., estimated by the standard deviation of  $\vartheta_i^{(m)}$ :

$$\xi_i = \left( \frac{1}{M_i - 1} \sum_{m=1}^{M_i} (\vartheta_i^{(m)} - \overline{\vartheta_i})^2 \right)^{1/2}. \quad (4.1)$$

- The hyper-parameters  $\mu_0$  and  $\sigma_0$  are estimated from market data, e.g., by maximum likelihood estimation or by the method of moments.
- In practice one often uses an ad hoc estimate for  $\sigma_i$ ,  $1 \leq i \leq I$ , which usually is based on expert opinion only. However one could think of a Bayesian approach for  $\sigma_i$ , but then an analytical formula for the posterior distribution in general does not

exist. The posterior distribution needs then to be calculated for example by the Markov Chain Monte Carlo method; see again Peters and Sisson [23] or Gilks et al. [17].  $\square$

Under Model Assumption 4.1, the posterior density is given by

$$\begin{aligned}
\hat{\pi}_{\Delta_i}(\delta_i | \mathbf{X}_i, \boldsymbol{\vartheta}_i) &\propto \frac{1}{\sigma_0 \sqrt{2\pi}} \exp\left(-\frac{(\delta_i - \mu_0)^2}{2\sigma_0^2}\right) \prod_{k=1}^{K_i} \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left(-\frac{(\log X_{i,k} - \delta_i)^2}{2\sigma_i^2}\right) \\
&\quad \prod_{m=1}^{M_i} \frac{1}{\tilde{\sigma}_i \sqrt{2\pi}} \exp\left(-\frac{(\vartheta_i^{(m)} - \delta_i)^2}{2\tilde{\sigma}_i^2}\right) \\
&\propto \exp\left[-\left(\frac{(\delta_i - \mu_0)^2}{2\sigma_0^2} + \sum_{k=1}^{K_i} \frac{1}{2\sigma_i^2} (\log X_{i,k} - \delta_i)^2\right.\right. \\
&\quad \left.\left.+ \sum_{m=1}^{M_i} \frac{1}{2\xi_i^2} (\vartheta_i^{(m)} - \delta_i)^2\right)\right] \\
&\propto \exp\left[-\frac{(\delta_i - \hat{\mu})^2}{2\hat{\sigma}^2}\right], \tag{4.2}
\end{aligned}$$

with

$$\hat{\sigma}^2 = \left(\frac{1}{\sigma_0^2} + \frac{K_i}{\sigma_i^2} + \frac{M_i}{\xi_i^2}\right)^{-1}, \tag{4.3}$$

and

$$\hat{\mu} = \hat{\sigma}^2 \cdot \left(\frac{\mu_0}{\sigma_0^2} + \frac{1}{\sigma_i^2} \sum_{k=1}^{K_i} \log X_{i,k} + \frac{1}{\xi_i^2} \sum_{m=1}^{M_i} \vartheta_i^{(m)}\right). \tag{4.4}$$

In summary we have the following theorem.

**Theorem 4.3**

*Under Model Assumptions 4.1 and with the notation  $\overline{\log X_i} = \frac{1}{K_i} \sum_{k=1}^{K_i} \log X_{i,k}$ , the posterior distribution of  $\Delta_i$ , given loss information  $\mathbf{X}_i$  and expert opinion  $\boldsymbol{\vartheta}_i$ , is a normal distribution  $\mathcal{N}(\hat{\mu}, \hat{\sigma})$  with*

$$\hat{\sigma}^2 = \left(\frac{1}{\sigma_0^2} + \frac{K_i}{\sigma_i^2} + \frac{M_i}{\xi_i^2}\right)^{-1}, \tag{4.5}$$

and

$$\hat{\mu} = \mathbb{E}[\Delta_i | \mathbf{X}_i, \boldsymbol{\vartheta}_i] = \omega_1 \mu_0 + \omega_2 \overline{\log X_i} + \omega_3 \overline{\vartheta_i}. \tag{4.6}$$

*The credibility weights are  $\omega_1 = \hat{\sigma}^2 / \sigma_0^2$ ,  $\omega_2 = \hat{\sigma}^2 K_i / \sigma_i^2$  and  $\omega_3 = \hat{\sigma}^2 M_i / \xi_i^2$ .*

This theorem yields a natural interpretation of the considered model. The estimator  $\hat{\mu}$  in (4.6) weights the internal and external data as well as the expert opinion in an appropriate manner. Observe that under Model Assumptions 4.1 we can explicitly calculate the mean

of the posterior distribution. This is different from the frequency model in Section 3. That is, we have an exact calculation and for the interpretation of the terms we do not rely on an asymptotic theorem as in Theorem 3.6. However, interpretation of the terms is exactly the same as in Theorem 3.6. The more credible the information, the higher is the credibility weight in (4.6). Hence, again, this theorem shows that our model is appropriate for combining internal observations, relevant external data and expert opinions.

## 4.2 Pareto model (Model 2 for severities)

### Model Assumptions 4.4 (Pareto-Gamma-Gamma)

Let us assume the following severity model for a particular operational risk cell of bank  $i$ ,  $1 \leq i \leq I$ :

- a) Let  $\Gamma_i \sim \Gamma(\alpha_0, \beta_0)$  be a Gamma distributed random variable with parameters  $\alpha_0, \beta_0$ , which are estimated from (external) market data, i.e.,  $\pi(\gamma)$  in (2.5) is the density of a  $\Gamma(\alpha_0, \beta_0)$  distribution.
- b) The losses  $k = 1, \dots, K_i$  from institution  $i$  are assumed to be conditionally (on  $\Gamma_i$ ) i.i.d. Pareto distributed:  $X_{i,1}, \dots, X_{i,K_i} | \Gamma_i \stackrel{\text{i.i.d.}}{\sim} \text{Pareto}(\Gamma_i, L_i)$ , where the threshold  $L_i \geq 0$  is assumed to be known and fixed. That is,  $f_1(\cdot | \Gamma_i)$  in (2.5) corresponds to the density of a  $\text{Pareto}(\Gamma_i, L_i)$  distribution.
- c) We assume that bank  $i$  has  $M_i$  experts with opinions  $\vartheta_i^{(m)}$ ,  $1 \leq m \leq M_i$ , about the parameter  $\Gamma_i$  with  $\vartheta_i^{(m)} | \Gamma_i \stackrel{\text{i.i.d.}}{\sim} \Gamma(\alpha_i = \xi_i, \beta_i = \Gamma_i / \xi_i)$ , where  $\xi_i$  is a parameter estimated using expert opinion data; see (3.2). That is,  $f_2(\cdot | \Gamma_i)$  corresponds to the density of a  $\Gamma(\xi_i, \Gamma_i / \xi_i)$  distribution.  $\square$

Under Model Assumptions 4.4, the posterior density is given by

$$\begin{aligned} \hat{\pi}_{\Gamma_i}(\gamma_i | \mathbf{X}_i, \boldsymbol{\vartheta}_i) &\propto \gamma_i^{\alpha_0-1} e^{-\gamma_i/\beta_0} \prod_{k=1}^{K_i} \frac{\gamma_i}{L_i} \left( \frac{X_{i,k}}{L_i} \right)^{-(\gamma_i+1)} \prod_{m=1}^{M_i} \frac{(\vartheta_i^{(m)}/\beta_i)^{\alpha_i-1}}{\beta_i} e^{-\vartheta_i^{(m)}/\beta_i} \\ &\propto \gamma_i^{\alpha_0-1-M_i\xi_i+K_i} \exp \left[ -\gamma_i \left( \frac{1}{\beta_0} + \sum_{k=1}^{K_i} \log \frac{X_{i,k}}{L_i} \right) - \frac{1}{\gamma_i} \xi_i M_i \bar{\vartheta}_i \right]. \end{aligned} \quad (4.7)$$

Hence, again, the posterior distribution is a GIG and it has the nice property that the term  $\gamma_i$  in the exponent in (4.7) is only affected by the internal observations, whereas the term  $1/\gamma_i$  is driven by the expert opinions.

**Theorem 4.5**

Under Model Assumptions 4.4, the posterior density of  $\Gamma_i$ , given loss information  $\mathbf{X}_i$  and expert opinion  $\boldsymbol{\vartheta}_i$ , is given by

$$\widehat{\pi}_{\Gamma_i}(\gamma_i | \mathbf{X}_i, \boldsymbol{\vartheta}_i) = \frac{(\omega/\phi)^{(\nu+1)/2}}{2K_{\nu+1}(2\sqrt{\omega\phi})} \gamma_i^\nu e^{-\gamma_i\omega - \gamma_i^{-1}\phi}, \quad (4.8)$$

with

$$\begin{aligned} \nu &= \alpha_0 - 1 - M_i \xi_i + K_i, \\ \omega &= \frac{1}{\beta_0} + \sum_{k=1}^{K_i} \log \frac{X_{i,k}}{L_i}, \\ \phi &= \xi_i M_i \overline{\vartheta}_i. \end{aligned} \quad (4.9)$$

It seems natural to generalize this result by substituting the prior Gamma distribution by a GIG as follows.

**Model Assumptions 4.6 (Pareto-Gamma-GIG)**

Let us assume the following severity model for a particular operational risk cell of bank  $i$ ,  $1 \leq i \leq I$ :

- a) Let  $\Gamma_i \sim \text{GIG}(\nu_0, \omega_0, \phi_0)$  be a generalized inverse Gaussian distributed random variable with parameters  $\nu_0, \omega_0, \phi_0$ , which are estimated from (external) market data, i.e.,  $\pi(\gamma)$  in (2.5) is the density of a  $\text{GIG}(\nu_0, \omega_0, \phi_0)$  distribution.
- b) The losses  $k = 1, \dots, K_i$  from bank  $i$  are assumed to be conditionally (on  $\Gamma_i$ ) i.i.d. Pareto distributed:  $X_{i,1}, \dots, X_{i,K_i} | \Gamma_i \stackrel{\text{i.i.d.}}{\sim} \text{Pareto}(\Gamma_i, L_i)$ , where the threshold  $L_i \geq 0$  is assumed to be known and fixed. That is,  $f_1(\cdot | \Gamma_i)$  in (2.5) corresponds to the density of a  $\text{Pareto}(\Gamma_i, L_i)$  distribution.
- c) We assume that bank  $i$  has  $M_i$  experts with opinions  $\vartheta_i^{(m)}$ ,  $1 \leq m \leq M_i$ , about the parameter  $\Gamma_i$  with  $\vartheta_i^{(m)} | \Gamma_i \stackrel{\text{i.i.d.}}{\sim} \Gamma(\alpha_i = \xi_i, \beta_i = \Gamma_i / \xi_i)$ , where  $\xi_i$  is a parameter estimated using expert opinion data. That is,  $f_2(\cdot | \Gamma_i)$  corresponds to the density of a  $\Gamma(\xi_i, \Gamma_i / \xi_i)$  distribution.  $\square$

Under Model Assumptions 4.6, the a posteriori density  $\widehat{\pi}_{\Gamma_i}(\gamma_i|\mathbf{X}_i, \boldsymbol{\vartheta}_i)$  is given by (4.8) with

$$\begin{aligned}\nu &= \nu_0 - M_i\xi_i + K_i, \\ \omega &= \omega_0 + \sum_{k=1}^{K_i} \log \frac{X_{i,k}}{L_i}, \\ \phi &= \phi_0 + \xi_i M_i \bar{\vartheta}_i.\end{aligned}\tag{4.10}$$

Hence, again, the posterior distribution is given by a GIG. Note that for  $\phi_0 = 0$ , the GIG is a Gamma distribution and hence we are in the Pareto-Gamma-Gamma situation of Model 4.4.

The following theorem gives us a natural interpretation of the Bayesian estimator

$$\mathbb{E}[\Gamma_i|\mathbf{X}_i, \boldsymbol{\vartheta}_i] = \sqrt{\frac{\phi}{\omega}} R_{\nu+1}(2\sqrt{\omega\phi}).\tag{4.11}$$

Denote the maximum likelihood estimator of the Pareto tail index  $\Gamma_i$  by

$$\gamma_i^{\text{MLE}} = \frac{K_i}{\sum_{k=1}^{K_i} \log \frac{X_{i,k}}{L_i}}.\tag{4.12}$$

Then, completely analogous to Theorem 3.6 we obtain the following theorem.

**Theorem 4.7**

*Under Model Assumptions 4.4 and 4.6, the following asymptotic relations hold  $\mathbb{P}$ -almost surely:*

- a) *Assume, given  $\Gamma_i = \gamma_i$ ,  $X_{i,k} \stackrel{\text{i.i.d.}}{\sim} \text{Pareto}(\gamma_i, L_i)$  and  $\vartheta_i^{(m)} \stackrel{\text{i.i.d.}}{\sim} \Gamma(\xi_i, \gamma_i/\xi_i)$ .  
For  $K_i \rightarrow \infty$ :  $\mathbb{E}[\Gamma_i|\mathbf{X}_i, \boldsymbol{\vartheta}_i] \rightarrow \mathbb{E}[X_{i,k}|\Gamma_i = \gamma_i]/V_i = \gamma_i$ .*
- b) *For  $\text{Vco}(\vartheta_i^{(m)}|\Gamma_i) \rightarrow 0$ :  $\mathbb{E}[\Gamma_i|\mathbf{X}_i, \boldsymbol{\vartheta}_i] \rightarrow \vartheta_i^{(m)}$ ,  $m = 1, \dots, M_i$ .*
- c) *Assume, given  $\Gamma_i = \gamma_i$ ,  $X_{i,k} \stackrel{\text{i.i.d.}}{\sim} \text{Pareto}(\gamma_i, L_i)$  and  $\vartheta_i^{(m)} \stackrel{\text{i.i.d.}}{\sim} \Gamma(\xi_i, \gamma_i/\xi_i)$ .  
For  $M_i \rightarrow \infty$ :  $\mathbb{E}[\Gamma_i|\mathbf{X}_i, \boldsymbol{\vartheta}_i] \rightarrow \mathbb{E}[\vartheta_i^{(m)}|\Gamma_i = \gamma_i] = \gamma_i$ .*
- d) *For  $\text{Vco}(\vartheta_i^{(m)}|\Gamma_i) \rightarrow \infty$ ,  $m = 1, \dots, M_i$ :  
 $\mathbb{E}[\Gamma_i|\mathbf{X}_i, \boldsymbol{\vartheta}_i] \rightarrow \left(1 - \frac{K_i\beta_0}{\gamma_i^{\text{MLE}} + K_i\beta_0}\right) \mathbb{E}[\Gamma_i] + \frac{K_i\beta_0}{\gamma_i^{\text{MLE}} + K_i\beta_0} \gamma_i^{\text{MLE}}$ .*
- e) *For  $\mathbb{E}[\Gamma_i] = \text{constant}$  and  $\text{Vco}(\Gamma_i) \rightarrow 0$ :  $\mathbb{E}[\Gamma_i|\mathbf{X}_i, \boldsymbol{\vartheta}_i] \rightarrow \mathbb{E}[\Gamma_i]$ .*

REMARKS 4.8:

- Theorem 4.7 basically says that the higher the precision of a particular source of risk information, the higher its corresponding credibility weight. This means that we obtain the same interpretations as for Theorem 3.6 and Formula (4.6).

- Observe that in Section 3 and Section 4.1 we have applied Bayesian inference to the expected values of the Poisson and the normal distribution, respectively. However, Bayesian inference is much more general, and basically, can be applied to any reasonable parameter. In this Section 4.2 it is, e.g., applied to the Pareto tail index.
- Observe that Model Assumptions 4.4 and 4.6 lead to an infinite mean model because the Pareto parameter  $\Gamma_i$  can be less than one with positive probability. For finite mean models, the range of possible  $\Gamma_i$  has to be restricted to  $\Gamma_i > 1$ . This does not impose difficulties; for more details we refer the reader to Shevchenko and Wüthrich [24], Section 3.4.  $\square$

### 4.3 Implementation and practical application

Note that the update process of (4.9) and (4.10) has again a simple linear form when new information arrives. The posterior density (4.8) does not change its type every time a new observation arrives. In particular, only the parameter  $\omega$  is affected by a new observation.

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

$$\begin{aligned}
 \nu_{k+1} &= \nu_k + 1, \\
 \omega_{k+1} &= \omega_k + \log \frac{X_{i,k+1}}{L_i}, \\
 \phi_{k+1} &= \phi_k.
 \end{aligned} \tag{4.13}$$

The following example shows the simplicity and robustness of the estimator developed.

**EXAMPLE 4.9** Assume that a bank would like to model its risk severity by a Pareto distribution with tail index  $\Gamma$ . The regulator provides external prior data, saying that  $\Gamma \sim \Gamma(\alpha_0, \beta_0)$  with  $\alpha_0 = 4$  and  $\beta_0 = 9/8$ , i.e.,  $\mathbb{E}[\Gamma] = 4.5$  and  $\text{Vco}(\Gamma) = 0.5$ . The bank has one expert opinion  $\hat{\vartheta} = 3.5$  with  $\text{Vco}(\vartheta|\Gamma) = 0.5$ , i.e.,  $\xi = 4$ . We then observe the following losses (sampled from a  $\text{Pareto}(\alpha = 4, L = 1)$  distribution); see also Figure 3:

Loss index $i$	1	2	3	4	5	6	7	8	9	10
Severity $X_i$	1.17	1.29	1.00	1.55	2.66	1.02	1.28	1.10	1.06	1.02
Loss index $i$	11	12	13	14	15					
Severity $X_i$	1.59	1.35	1.91	1.23	1.03					

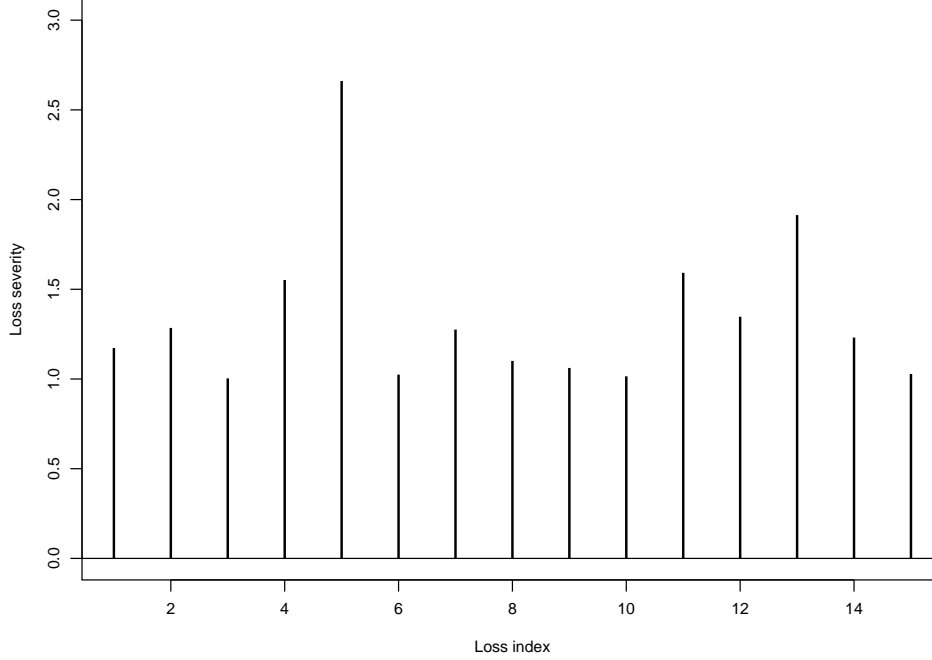


Figure 3: 15 loss severities sampled from a Pareto( $\alpha = 4, L = 1$ ) distribution.

In Figure 4 we compare the Bayes estimator

$$\gamma_k = \mathbb{E}[\Gamma | X_1, \dots, X_k, \vartheta], \quad (4.14)$$

given by (4.11) with the estimator proposed in Shevchenko and Wüthrich [24] without expert opinions

$$\gamma_k^{\text{SW}} = \mathbb{E}[\Gamma | X_1, \dots, X_k], \quad (4.15)$$

and the classical maximum likelihood estimator

$$\gamma_k^{\text{MLE}} = \frac{k}{\sum_{i=1}^k \log \frac{X_i}{L}}. \quad (4.16)$$

Figure 4 shows the high volatility of the maximum likelihood estimator, for small numbers  $k$ . It is very sensitive to newly arriving losses. However, the estimator proposed in this paper shows a much more stable behavior around the true value  $\alpha = 4$ , most notably when a few data points are available.  $\square$

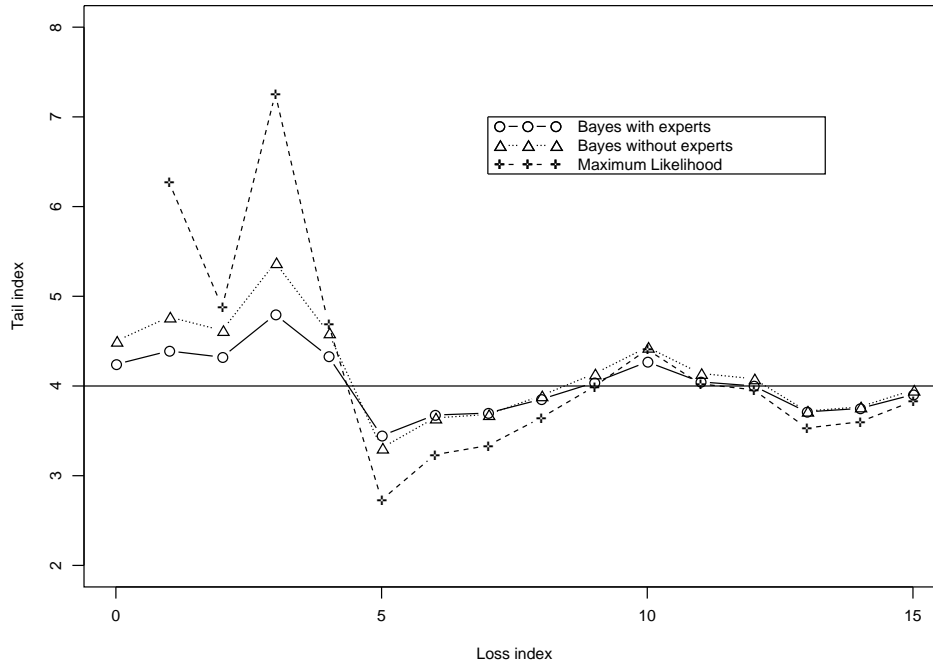


Figure 4: The Bayes estimator including expert opinions ( $\circ$ ) is compared with the Bayes estimator without expert opinions ( $\Delta$ ) and with the maximum likelihood estimator ( $+$ ).

This example also shows that when modeling severities of operational risk, Bayesian inference is a suitable method to combine different sources of information. The consideration of relevant external data and well-specified expert opinions stabilizes and smoothens the estimator in an appropriate way.

## 5 Total loss distribution and risk capital estimates

In the preceding sections we have described how the parameters of the distributions are estimated. According to the Basel II requirements (see BIS [6]) 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 VaR for each risk cell separately and sum VaRs over risk cells to estimate

the total bank capital. Adding quantiles over the risk cells to find the quantile of the total loss distribution is sometimes too conservative. It is equivalent to the assumption of perfect dependence between risks.

The calculation of VaR (taking into account parameter uncertainty) for each risk cell can, in view of the previous sections, easily be done using a simulation approach described in Shevchenko and Wüthrich [24], Section 6. Simulation procedures for independent risk cells and in the case of dependence between risks are also described in Shevchenko and Wüthrich [24] and thus we refrain from commenting further on this issue.

However, reasonable aggregation is still an open challenging problem that needs further investigation. The choice of appropriate dependence structures is crucial and determines the amount of diversification. In the general case, when no information about the dependence structure is available, Embrechts and Puccetti [15] work out bounds for aggregated operational risk capital; for further issues regarding aggregation we would like to refer to Embrechts et al. [14].

## 6 Conclusion

In this paper we propose a novel approach that allows for combining three data sources: internal data, external data and expert opinions. The approach is based on the Bayesian inference method. It is applied to the quantification of the frequency and severity distributions in operational risk, where there is a strong need for such a method to meet the Basel II regulatory requirements.

The method is based on specifying prior distributions for the parameters of the frequency and severity distributions using industry data. Then, the prior distributions are weighted by the actual observations and expert opinions from the bank to estimate the posterior distributions of the model parameters. These are used to estimate the annual loss distribution for the next reporting year. 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 expert opinions and industry data into account. This method also allows for calculation of VaR with parameter uncertainty taken into account.

For convenience we have assumed that expert opinions are i.i.d. but all formulas can easily be generalized to the case of expert opinions modeled by different distributions.

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 regulators 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 but go beyond the purposes of this paper.

The approach described is not too complicated and is well suited for operational risk quantification. It has a simple structure, which is beneficial for practical use and can engage the bank risk managers, statisticians and regulators in productive model development and risk assessment. The model provides a framework that can be developed further by considering other distribution types, dependencies between risks and dependence on time.

One of the features of the described method is that the variance of the posterior distribution  $\hat{\pi}(\gamma|\cdot)$  will converge to zero for a large number of observations. That is, the true values of the risk parameters will be known exactly. However, there are many factors (for example, political, economical, legal, etc.) changing in time that will not permit for the precise knowledge of the risk parameters. One can model this by limiting the variance of the posterior distribution by some lower levels (say, e.g., 5%). This has been done in many solvency approaches for the insurance industry; see, e.g., the Swiss Solvency Test, FOPI [16], formulas (25)-(26).

Although the main impetus motivation for the present paper is an urgent need from operational risk practitioners, the proposed method is also useful in other areas (such as credit risk, insurance, 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.

## A Generating realizations from a GIG random variable

For practical purposes it is required to generate realizations of a random variable  $X \sim \text{GIG}(\omega, \phi, \nu)$  with  $\omega, \phi > 0$ . Observe that we need to construct a special algorithm since we can not invert the distribution function analytically. The following algorithm can be found in Dagpunar [8]; see also McNeil et al. [19]:

**Algorithm A.1 (Generalized inverse Gaussian)**

1.  $\alpha = \sqrt{\omega/\phi}$ ;  $\beta = 2\sqrt{\omega\phi}$ ,  
 $m = \frac{1}{\beta} \left( \nu + \sqrt{\nu^2 + \beta^2} \right)$ ,  
 $g(y) = \frac{1}{2}\beta y^3 - y^2(\frac{1}{2}\beta m + \nu + 2) + y(\nu m - \frac{\beta}{2}) + \frac{1}{2}\beta m$ .
2. Set  $y_0 = m$ ,  
While  $g(y_0) \leq 0$  do  $y_0 = 2y_0$ ,  
 $y_+$ : root of  $g$  in the interval  $(m, y_0)$ ,  
 $y_-$ : root of  $g$  in the interval  $(0, m)$ .
3.  $a = (y_+ - m) \left( \frac{y_+}{m} \right)^{\nu/2} \exp \left( -\frac{\beta}{4} \left( y_+ + \frac{1}{y_+} - m - \frac{1}{m} \right) \right)$ ,  
 $b = (y_- - m) \left( \frac{y_-}{m} \right)^{\nu/2} \exp \left( -\frac{\beta}{4} \left( y_- + \frac{1}{y_-} - m - \frac{1}{m} \right) \right)$ ,  
 $c = -\frac{\beta}{4} \left( m + \frac{1}{m} \right) + \frac{\nu}{2} \log(m)$ .
4. Repeat  $U, V \sim \text{Unif}(0, 1)$ ,  $Y = m + a\frac{U}{V} + b\frac{1-V}{U}$ ,  
until  $Y > 0$  and  $-\log U \geq -\frac{\nu}{2} \log Y + \frac{1}{4}\beta(Y + \frac{1}{Y}) + c$ ,  
Then  $X = \frac{Y}{\alpha}$  is GIG( $\omega, \phi, \nu$ ); see Dagpunar [8].

To generate a sequence of  $n$  realizations from a GIG random variable, step 4 is repeated  $n$  times. □

## B Asymptotic results for modified Bessel functions

Let  $K_\gamma(z)$  denote the modified Bessel function of the third kind as defined in (3.6).

**Lemma B.1** *With notation (3.10), we have the following asymptotic relation for  $\nu \rightarrow \infty$ , for all  $a, b > 0$ :*

$$R_{b\nu}(a\sqrt{\nu}) \sim \frac{2b\sqrt{\nu}}{a}. \quad (\text{B.1})$$

PROOF: From Abramowitz and Stegun [1], Paragraph 9.7.8 and Olver [21], Chapter 4, we may deduce for large  $\nu$  and  $z \geq 0$

$$K_\nu(\nu z) = \sqrt{\frac{\pi}{2\nu}} \frac{\exp(-\nu\sqrt{1+z^2})}{(1+z^2)^{1/4}} \left( \frac{z}{1+\sqrt{1+z^2}} \right)^{-\nu} (1 + \varepsilon(\nu, z)), \quad (\text{B.2})$$

where the error term  $\varepsilon(\nu, z)$  is bounded by

$$|\varepsilon(\nu, z)| \leq \frac{1}{\nu - \nu_0} \int_0^{1/\sqrt{1+z^2}} |u_1'(s)| ds, \quad (\text{B.3})$$

with  $\nu_0 = \frac{1}{6\sqrt{5}} + \frac{1}{12}$  and  $u_1(s) = (3s - 5s^3)/24$ ; see Abramowitz and Stegun [1] for details. In (B.2) we replace  $\nu$  by  $b\nu$  and  $z$  by  $z_1 = a/\sqrt{\nu}$ . The error term  $\varepsilon(b\nu, a/\sqrt{\nu})$  in (B.2) is then vanishing for  $\nu \rightarrow \infty$ , because the right-hand side of (B.3) tends to 0. Analogously, we replace  $\nu$  by  $b\nu + 1$  and  $z$  by  $z_2 = a\sqrt{\nu}/(b\nu + 1)$  and observe that  $\varepsilon(b\nu + 1, a\sqrt{\nu}/(b\nu + 1))$  tends to 0. Thus, (B.2) gives us asymptotic expressions for  $K_{b\nu}(a\sqrt{\nu})$  and  $K_{b\nu+1}(a\sqrt{\nu})$ . Straightforward calculations then yield

$$R_{b\nu}(a\sqrt{\nu}) = \frac{K_{b\nu+1}(a\sqrt{\nu})}{K_{b\nu}(a\sqrt{\nu})} \sim \frac{2}{z_2} \sim \frac{2b\sqrt{\nu}}{a}, \quad \nu \rightarrow \infty. \quad (\text{B.4})$$

This completes the proof.  $\square$

## C Proof of Theorem 3.6

PROOF: With (3.11) the proof of this theorem is straightforward, using Lemma B.1 in Appendix B. The following statements hold  $\mathbb{P}$ -almost surely.

- a)  $\sqrt{\frac{\phi}{\omega}} R_{\nu+1}(2\sqrt{\omega\phi}) \sim \sqrt{\frac{\phi}{\omega}} R_{K_i \bar{N}_i}(2\sqrt{V_i K_i \phi}) \sim \sqrt{\frac{K_i}{V_i \omega}} \bar{N}_i \sim \mathbb{E}[N_{i,k} | \Lambda_i] / V_i.$
- b,c)  $\sqrt{\frac{\phi}{\omega}} R_{\nu+1}(2\sqrt{\omega\phi}) \sim \sqrt{\frac{\phi}{\omega}} R_{-M_i \xi_i}(2\sqrt{\omega \xi_i M_i \vartheta_i}) \sim \sqrt{\frac{\phi}{\omega}} \frac{1}{R_{M_i \xi_i}(2\sqrt{\omega \xi_i M_i \vartheta_i})}$   
 $\sim \sqrt{\frac{\phi \vartheta_i}{\xi_i M_i}} = \bar{\vartheta}_i = \vartheta_i^{(m)}, m = 1, \dots, M_i.$
- d) If  $\xi = 0$ , we are in the Gamma case  $\Gamma(\alpha, \beta)$  with  $\alpha = \alpha_0 + K_i \bar{N}_i$  and  $\beta = \beta_0 / (V_i K_i \beta_0 + 1)$ . Hence,  
 $\mathbb{E}[\Lambda_i | \mathbf{N}_i, \boldsymbol{\vartheta}_i] = \alpha\beta = \frac{1}{V_i K_i \beta_0 + 1} \mathbb{E}[\Lambda_i] + \left(1 - \frac{1}{V_i K_i \beta_0 + 1}\right) \bar{N}_i / V_i.$
- e)  $\sqrt{\frac{\phi}{\omega}} R_{\nu+1}(2\sqrt{\omega\phi}) \sim \sqrt{\frac{\phi \mathbb{E}[\Lambda_i]}{\alpha_0}} R_{\alpha_0}(2\sqrt{\frac{\alpha_0 \phi}{\mathbb{E}[\Lambda_i]}}) \sim \mathbb{E}[\Lambda_i].$   $\square$

### Acknowledgments

The authors would like to thank Isaac Meilijson for several fruitful discussions and Paul Embrechts for his useful comments.

## References

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

- [2] Alderweireld, T., Garcia, J. and Léonard, L. (2006) A practical operational risk scenario analysis quantification. *Risk Magazine* **19**(2), 93–95.
- [3] Atkinson, A. C. (1982) The simulation of generalized inverse Gaussian and hyperbolic random variables. *SIAM Journal of Scientific and Statistical Computation* **3**, 502–515.
- [4] Bühlmann, H. and Gisler, A. (2005) *A Course in Credibility Theory and its Applications*. Springer, Berlin.
- [5] Bühlmann, H., Shevchenko, P. V. and Wüthrich, M. V. (2007) A “toy” model for operational risk quantification using credibility theory. *Journal of Operational Risk* **2**(1), 3–19.
- [6] BIS (2005) Basel II: International Convergence of Capital Measurement and Capital Standards: a revised framework. Bank for International Settlements (BIS), [www.bis.org](http://www.bis.org).
- [7] Cruz, M. G. (2002) *Modeling, Measuring and Hedging Operational Risk*. Wiley, Chichester.
- [8] Dagpunar, J. S. (1989) An easily implemented generalised inverse Gaussian generator. *Communications in Statistics, Simulation and Computation* **18**, 703–710.
- [9] Davis, E. (2006) Theory vs Reality. *OpRisk and Compliance*. 1 September 2006, [www.opriskandcompliance.com/public/showPage.html?page=345305](http://www.opriskandcompliance.com/public/showPage.html?page=345305).
- [10] Degen, M., Embrechts, P. and Lambrigger, D. D. (2007) The quantitative modeling of operational risk: between g-and-h and EVT. *ASTIN Bulletin*, to appear.
- [11] Dutta, K. and Perry, J. (2006) A tale of tails: an empirical analysis of loss distribution models for estimating operational risk capital. Federal Reserve Bank of Boston, Working Paper No 06-13.
- [12] Embrechts, P. (1983) A property of the generalized inverse Gaussian distribution with some applications. *Journal of Applied Probability* **20**, 537–544.
- [13] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) *Modelling Extremal Events for Insurance and Finance*. Springer, Berlin.

- [14] Embrechts, P., Nešlehová, J. and Wüthrich, M. V. (2007) Additivity properties for Value-at-Risk under archimedean dependence and heavy-tailedness. Preprint, ETH Zurich.
- [15] Embrechts, P. and Puccetti, G. (2006) Aggregating risk capital, with an application to operational risk. *The Geneva Risk and Insurance Review* **31**(2), 71–90.
- [16] FOPI (2006) Swiss Solvency Test, Technical Document. Federal Office of Private Insurance, Bern. [www.bpv.admin.ch/themen/00506/00552](http://www.bpv.admin.ch/themen/00506/00552).
- [17] Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (1996) *Markov Chain Monte Carlo in practice*. Chapman & Hall, London.
- [18] Jørgensen, B. (1982) *Statistical Properties of the Generalized Inverse Gaussian Distribution*. Springer, New York.
- [19] McNeil, A. J., Frey, R. and Embrechts, P. (2005) *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press, Princeton.
- [20] Moscadelli, M. (2004) The modelling of operational risk: experiences with the analysis of the data collected by the Basel Committee. Bank of Italy, Working Paper No 517.
- [21] Olver, F. W. J. (1962) *Mathematical Tables, vol. 6, Tables for Bessel functions of moderate or large orders*. National Physical Laboratory. Her Majesty's Stationery Office, London.
- [22] Panjer, H. H. (2006) *Operational Risks: Modeling Analytics*. Wiley, New York.
- [23] Peters, G. W. and Sisson, S. A. (2006) Bayesian inference, Monte Carlo sampling and operational risk. *Journal of Operational Risk* **1**(3), 27–50.
- [24] Shevchenko, P. V. and Wüthrich, M. V. (2006) The structural modeling of operational risk via Bayesian inference: combining loss data with expert opinions. *Journal of Operational Risk* **1**(3), 3–26.