

SIMULATING TOTAL OPERATIONAL RISK USING RETROSPECTIVE INFORMATION AND SUBJECTIVE ASSESSMENTS - A BAYESIAN APPROACH

Arne Bang Huseby, University of Oslo, Norway
 Jan Thomsen, Norges Bank Investment Management, Norway

Abstract

Quantifying operational risk exposure typically involves gathering information from several sources, including historical data as well as subjective assessments. Using historical data one can estimate both an incident frequency distribution, as well as an incident consequence distribution. Based on these two distributions a simulation model can be established. However, by limiting the focus to data related to incidents which may reappear in the future, one is often left with a relatively short incident history. In order to improve the risk quantification, it is often necessary to include subjective risk assessments as well. In the present paper we propose a model for how to combine these two sources of information. The model can be used to represent situations ranging from cases where the two sources are disjoint, overlap completely as well as intermediate cases where the two sources are partially overlapping. The model is illustrated by considering a numerical example. In this example we vary the degree of overlap between the sources of information.

1 Introduction

When quantifying operational risk exposure one often needs to gather information from several sources. Such sources include statistical data based on historical observations as well as assessments made by a panel of experts. The statistical data represent observations of actual incidents collected through a suitable number of years. By counting the number of incidents per unit of time one can estimate an incident frequency distribution. Moreover, by looking at the occurred consequences an incident consequence distribution can be estimated. Together these two distributions enable the analyst to build a well-defined simulation model. In the present paper we refer to this as the *incident database model*. While this model may represent an adequate summary of the past, it may not work quite as well for the future. Due to changes in the overall risk picture, older incidents may sometimes be less likely to occur in the future. By limiting the focus to data related to incidents which may reappear in the future, one is often left with a relatively short incident history. Moreover, there may be important risk factors missing in the incident database. Such risk factors could arise from recent threats or rare incidents with possibly severe consequences. Thus,

in order to improve the risk quantification, it is often necessary to include subjective risk assessments as well. Typically, this is done by a panel of experts identifying a set of potential risk factors. For all these factors the experts assess the frequency of occurrence as well as the risk factor consequence distribution. Based on these assessments another simulation model can be established, referred to here as the *risk factor model*.

Obviously the incident database model and the risk factor model may produce results which differ significantly. Thus, there is a need to handle these differences by developing some way of combining the two models. In this paper we shall discuss how to combining incident database model and risk factor model. Underlying this approach is the assumption that the incident data and the subjectively assessed risk factor model can be viewed as two different but possibly overlapping sources of information. The proposed model is flexible and covers a range of cases with respect to the degree of overlap.

A similar approach can be found in [Huseby 2015]. The model presented in the present paper, however, is much more flexible with respect to the way the subjective assessments are handled. Moreover, a fully Bayesian approach is applied in the updating part.

Before introducing the models we review results related to compound processes and show how to incorporate parameter uncertainty into the models. The proposed models are illustrated by considering a numerical example.

2 Compound processes

A common approach to modelling the accumulated consequences from a series of events is using a *compound process*. Such processes are used extensively in e.g., insurance mathematics. For an excellent introduction to this field see [Bølviken 2014]. This topic of compound processes is also covered in [McNeil et al. 2005].

According to this approach the accumulated consequences in a time interval $[0, t]$, denoted $Z(t)$ can be expressed as:

$$Z(t) = \sum_{i=1}^{N(t)} X_i, \quad (2.1)$$

where $N(t)$ denotes the number of events in $[0, t]$, while X_i denotes the consequence of the i th event, $i = 1, 2, \dots$. Here $N(t)$ is a non-decreasing stochastic process with non-negative integer values such that $N(0) = 0$, also referred to as a *counting process*. We refer to this as *the event generating process*. Typically, one assumes that $X_1, X_2, \dots$ are stochastically independent and identically distributed with distribution F , which we refer to as the *consequence distribution*.

Unless one needs to include time dependence or non-stationarity into the model, a stationary *Poisson process* is a common choice for the event generating process. This type of process is also recommended in [Adachi et al. 2011], and will be used throughout the present paper. A compound process where the events are generated

from such a Poisson process, is referred to as a *compound Poisson process*. See e.g., [Adelson 1966] or [McNeil et al. 2005] for more details. We refer to a compound Poisson process by the process corresponding to the accumulated consequences, i.e., $Z(t)$.

When the event generating process is a stationary Poisson process, it follows that:

$$P(N(t) = n) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}, \quad n = 0, 1, 2, \dots, \quad (2.2)$$

where $\lambda > 0$ is a parameter denoting the (constant) *event rate* per time unit. Thus, for $t > 0$, $N(t)$ has a Poisson distribution with parameter λt , i.e., $N(t) \sim \text{Po}(\lambda t)$. In this context we define a time unit to be *one year*.

Using standard results from the theory of compound Poisson processes it is very easy to show that:

$$E[Z(t)] = E[N(t)] \cdot E[X_i], \quad (2.3)$$

$$\text{Var}[Z(t)] = E[N(t)] \cdot (\text{Var}[X_i] + (E[X_i])^2). \quad (2.4)$$

A well known property of compound Poisson processes is that if we sum the consequences of a set of independent compound Poisson processes, we get another compound Poisson process. More specifically, let $N_1(t), \dots, N_r(t)$ denote r independent Poisson processes with event rates $\lambda_1, \dots, \lambda_r$ respectively. The consequences associated with the s th process are denoted $X_{s1}, X_{s2}, \dots$, $s = 1, \dots, r$. We assume that the consequences associated with the s th process are stochastically independent and identically distributed with distribution F_s , $s = 1, \dots, r$. Moreover, we introduce the accumulated consequences of the s th process in a time interval $[0, t]$, and denote this by $Z_s(t)$. Thus, we have:

$$Z_s(t) = \sum_{i=1}^{N_s(t)} X_{si}. \quad (2.5)$$

Finally, we introduce $N(t) = N_1(t) + \dots + N_r(t)$ and $Z(t) = Z_1(t) + \dots + Z_r(t)$. Then $N(t)$, is a Poisson process with event rate $\lambda = \lambda_1 + \dots + \lambda_r$, while $Z(t)$ is a compound Poisson process. The consequences contributing to the sum $Z(t)$, conveniently denoted $X_1, X_2, \dots$, will be a mixture of consequences from the r individual compound Poisson processes, and we have:

$$Z(t) = \sum_{s=1}^r Z_s(t) = \sum_{s=1}^r \sum_{i=1}^{N_s(t)} X_{si} = \sum_{i=1}^{N(t)} X_i. \quad (2.6)$$

By using e.g., moment generating functions, it is very easy to show that the consequence distribution of $Z(t)$ is a mixture of the distributions of the consequence distributions $F_1, \dots, F_r$. More specifically, if F denotes the consequence distribution of $Z(t)$, then:

$$F(x) = \sum_{s=1}^r \frac{\lambda_s}{\lambda} F_s(x). \quad (2.7)$$

When running Monte Carlo simulations this property can be used to generate samples from the compound Poisson process $Z(t)$. The most obvious way of doing this is to sample the number of events and consequences from the sum of the individual processes:

Algorithm 2.1 *For each iteration in the Monte Carlo simulation:*

STEP 1. *Sample $N(t)$ from a Poisson distribution with parameter λt .*

STEP 2. *Sample $X_1, \dots, X_{N(t)}$ from the mixture distribution $F(x)$.*

STEP 3. *Calculate:*

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

Alternatively, one may sample the number of events and consequences from each individual compound Poisson process and add the results:

Algorithm 2.2 *For each iteration in the Monte Carlo simulation:*

STEP 1. *Sample $N_s(t)$ from a Poisson distribution with parameter $\lambda_s t$, $s = 1, \dots, r$.*

STEP 2. *Sample $X_{s1}, \dots, X_{s,N_s(t)}$ from the distribution $F_s(x)$, $s = 1, \dots, r$.*

STEP 3. *Calculate:*

$$Z(t) = \sum_{s=1}^r \sum_{i=1}^{N_s(t)} X_{si}$$

The latter method is often more convenient as one avoids sampling from the mixture distribution. In the present paper we will use this approach.

2.1 *Uncertainty about the event rates*

Realistically, the parameter of the Poisson process is typically an uncertain quantity. In order to include this uncertainty into the model, we assess a probability distribution representing our prior knowledge about the parameter. A convenient choice is the natural conjugate distribution, i.e., a Gamma distribution with positive parameters α and β . The density of this distribution, denoted by π , is given by:

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

In order to assess the hyperparameters α and β , one may imagine the prior distribution as a result of observing the process in β time units. (See [Huseby 1989].) Thus, the β -parameter can be considered as a measure of the *strength* of the prior knowledge. Thus, in cases where the prior knowledge is considered to be weak, the value of β should be chosen to be small, typically less than or equal to 1. The value of α is assessed indirectly using the fact that the prior expected value of λ is:

$$E[\lambda] = \frac{\alpha}{\beta}.$$

Thus, by assessing a prior point estimate for λ , denoted λ_π , the parameter α is typically chosen so that:

$$\alpha = \lambda_\pi \beta. \quad (2.8)$$

Note that if the prior point estimate λ_π is moderate, then a small value of β reflecting a weak prior knowledge, will imply that α is small as well.

As one gets observations from the Poisson process, the prior uncertainty about λ has to be updated using Bayes' theorem. See e.g., [Berger 2010]. In particular, if we have observed the process in τ units of time, during which we have recorded ν events, then the resulting uncertainty distribution, referred to as the *posterior distribution* is:

$$\pi(\lambda|\tau, \nu) = \frac{(\beta + \tau)^{\alpha+\nu}}{\Gamma(\alpha + \nu)} \lambda^{\alpha+\nu-1} e^{-(\beta+\tau)\lambda}, \quad \lambda > 0.$$

Thus the posterior distribution is another Gamma distribution with parameters $\alpha' = \alpha + \nu$ and $\beta' = \beta + \tau$.

2.2 Uncertainty about the consequence distribution

In this subsection we consider the problem of assessing the consequence distribution of a compound process. The true distribution is denoted by F . We then assume that we have collected a set of data, i.e., a set of observed consequences, denoted by $W_1, \dots, W_\nu$. Using these data we may update our prior assessments using some sort of Bayesian method. If we know that F belongs to a specific parametric class, this may be done using standard parametric methods. In this case, however, we assume that this is not the case. Instead we proceed in a more non-parametric fashion, and introduce the following function:

$$S_\nu(x) = \sum_{j=1}^{\nu} \mathbf{I}(W_j \leq x), \quad x > 0,$$

where $\mathbf{I}(\cdot)$ denotes the indicator function. It is easy to see, that given the true value of $F(x)$, $S_\nu(x)$ is binomially distributed, that is:

$$S_\nu(x)|F(x) \sim \text{Bin}(\nu, F(x)).$$

For each $x > 0$ we have assessed a prior estimate of $F(x)$ as well, denoted $\phi(x)$. In order to combine this with the data, we need to extend this to a full prior. Since obviously $F(x) \in [0, 1]$, a natural and convenient choice is a beta distribution with $\phi(x)$ as its mean value. Thus, we assume that:

$$F(x) \sim \text{Beta}(c\phi(x), c(1 - \phi(x))), \quad x > 0,$$

where $c > 0$ is a suitably chosen value. It then follows that:

$$\begin{aligned} \mathbb{E}[F(x)] &= \frac{c\phi(x)}{c\phi(x) + c(1 - \phi(x))} = \phi(x), \\ \text{Var}[F(x)] &= \frac{c\phi(x) \cdot c(1 - \phi(x))}{(c\phi(x) + c(1 - \phi(x)))^2(c\phi(x) + c(1 - \phi(x)) + 1)} = \frac{\phi(x)(1 - \phi(x))}{c + 1}. \end{aligned}$$

Note that high values of c implies that $\text{Var}[F(x)]$ is small. If this is the case, this indicates that the prior knowledge is strong.

Using Bayes theorem, we then obtain that:

$$F(x)|S_\nu(x) \sim \text{Beta}(c\phi(x) + S_\nu(x), c(1 - \phi(x)) + \nu - S_\nu(x)), \quad x > 0.$$

Moreover, the posterior mean of this distribution is:

$$\mathbb{E}[F(x)|S_\nu(x)] = \frac{c\phi(x) + S_\nu(x)}{c + \nu} = \frac{c}{c + \nu}\phi(x) + \frac{\nu}{c + \nu}\hat{F}_\nu(x), \quad (2.9)$$

where $\hat{F}_\nu(x)$ denotes the empirical distribution function:

$$\hat{F}_\nu(x) = \frac{1}{\nu}S_\nu(x) = \frac{1}{\nu} \sum_{j=1}^{\nu} \mathbb{I}(Y_j \leq x).$$

For convenience, we denote $\mathbb{E}[F(x)|S_\nu(x)]$ by $\phi'(x)$, and use this as our updated consequence distribution. We observe that $\phi'(x)$ is a mixture of the prior estimate $\phi(x)$ and the empirical distribution function $\hat{F}_\nu(x)$. If c is high, this indicates that we have a lot of confidence in our prior estimate, $\phi(x)$, while if c is low, most of the weight is put on the empirical distribution function, $\hat{F}_\nu(x)$.

Now, as mentioned in Subsection 2, the prior estimate of the consequence distribution may itself be a mixture of distributions:

$$\phi(x) = \sum_{s=1}^r \frac{\lambda_s}{\lambda} \phi_s(x), \quad (2.10)$$

where $\lambda = \lambda_1 + \dots + \lambda_r$. Inserting this into (2.9), we get:

$$\phi'(x) = \frac{c}{c + \nu} \sum_{s=1}^r \frac{\lambda_s}{\lambda} \phi_s(x) + \frac{\nu}{c + \nu} \hat{F}_\nu(x). \quad (2.11)$$

We observe that we end up with a posterior estimate of the consequence distribution being a mixture of all the prior distributions and the empirical distribution function, which turns out to be handy during Monte Carlo simulations.

2.3 Combining uncertainty

We observe that both the prior estimate (2.10) and the posterior estimate (2.11) of the consequence distribution depend on the event rates. Since these rates are uncertain as well, we need to include all these uncertainties in order to have a complete Monte Carlo simulation model for the consequences. In order to take a closer look at this, we consider a set of r compound Poisson processes. The event rate for the s th process is denoted λ_s , $s = 1, \dots, r$. We assume that $\lambda_s \sim \text{Gamma}(\alpha_s, \beta_s)$, apriori $s = 1, \dots, r$. Moreover, we assume that the prior consequence distribution for the s th process is assessed to be ϕ_s , $s = 1, \dots, r$.

For the prior estimate (2.10) the uncertainty about the event rates is easy to include. Instead of using fixed rates, $\lambda_1, \dots, \lambda_r$ are sampled from their respective prior distributions. That is, we use an extended version of Algorithm 2.2:

Algorithm 2.3 *For each iteration in the Monte Carlo simulation:*

STEP 1. *Sample λ_s from $\text{Gamma}(\alpha_s, \beta_s)$, $s = 1, \dots, r$.*

STEP 2. *Sample $N_s(t)$ from a Poisson distribution with parameter $\lambda_s t$, $s = 1, \dots, r$.*

STEP 3. *Sample $X_{s1}, \dots, X_{s, N_s(t)}$ from the distribution $\phi_s(x)$, $s = 1, \dots, r$.*

STEP 4. *Calculate:*

$$Z(t) = \sum_{s=1}^r \sum_{i=1}^{N_s(t)} X_{si}.$$

For the posterior estimate we consider two different cases. In the first case we assume that the data can be partitioned with respect to the individual compound processes. For the s th compound process, ν_s events have been observed during a period of length τ , and the corresponding consequences are denoted $W_{s1}, \dots, W_{s\nu_s}$, $s = 1, \dots, r$. Then it follows that the posterior distribution of λ_s is $\text{Gamma}(\alpha_s + \nu_s, \beta_s + \tau)$. Moreover, for $s = 1, \dots, r$, the s th posterior consequence distribution given λ_s is:

$$\phi'_s(x) = \frac{c}{c + \nu_s} \phi_s(x) + \frac{\nu_s}{c + \nu_s} \hat{F}_{\nu_s}(x),$$

where $\hat{F}_{\nu_s}(x) = \nu_s^{-1} \sum_{j=1}^{\nu_s} \mathbf{I}(W_{sj} \leq x)$ denotes the empirical distribution function, and c is a suitable constant¹ representing the strength of the prior estimate ϕ_s . In order to generate consequences, we could of course sample directly from the posterior distribution $\phi'_s(x)$. However, by utilizing the mixture property of compound Poisson processes, it is easier to do the following. Let $\tilde{\lambda}_s = \lambda_s c / (c + \nu_s)$, $\tilde{\mu}_s = \lambda_s \nu_s / (c + \nu_s)$, and note that $\tilde{\lambda}_s + \tilde{\mu}_s = \lambda_s$, $s = 1, \dots, r$. We then use the following algorithm:

Algorithm 2.4 *For each iteration in the Monte Carlo simulation:*

STEP 1. *Sample λ_s from $\text{Gamma}(\alpha_s + \nu_s, \beta_s + \tau)$ and calculate $\tilde{\lambda}_s$ and $\tilde{\mu}_s$, $s = 1, \dots, r$.*

STEP 2. *Sample $N_s(t)$ from a Poisson distribution with parameter $\tilde{\lambda}_s t$, $s = 1, \dots, r$.*

STEP 3. *Sample $M_s(t)$ from a Poisson distribution with parameter $\tilde{\mu}_s t$, $s = 1, \dots, r$.*

STEP 4. *Sample $X_{s1}, \dots, X_{s, N_s(t)}$ from the distribution $\phi_s(x)$, $s = 1, \dots, r$.*

STEP 5. *Sample $Y_{s1}, \dots, Y_{s, M_s(t)}$ from the distribution $\hat{F}_{\nu_s}(x)$, $s = 1, \dots, r$.*

STEP 6. *Calculate:*

$$Z(t) = \sum_{s=1}^r \left(\sum_{i=1}^{N_s(t)} X_{si} + \sum_{i=1}^{M_s(t)} Y_{si} \right).$$

This procedure works since $N_s(t) + M_s(t)$ has a Poisson distribution with parameter $(\tilde{\lambda}_s + \tilde{\mu}_s)t = \lambda_s t$, $s = 1, \dots, r$. The resulting consequence distribution mixes the consequences $X_{s1}, \dots, X_{s, N_s(t)}$ with the consequences $Y_{s1}, \dots, Y_{s, M_s(t)}$, and thus gets exactly the desired mixture distribution $\phi'_s(x)$, $s = 1, \dots, r$.

In the second case we consider the perhaps more realistic case where it is not possible to identify which of the individual compound processes that have produced the observed events and consequences. Thus, all that is known is the total number of events during the period of length τ , denoted ν and the corresponding consequences

¹The constant c may be different for each of the r compound processes. In order to simplify the model assessments, however, we have chosen to use the same c for all processes.

$W_1, \dots, W_\nu$. In order to proceed with this case, we make the simplifying assumption that $\beta_1 = \dots = \beta_r$ and let β_R denote this common value. Using this assumption it is easy to show that the prior distribution for the total event rate $\lambda = \lambda_1 + \dots + \lambda_r$ is $\text{Gamma}(\alpha_R, \beta_R)$, where we have introduced $\alpha_R = \alpha_1 + \dots + \alpha_r$. Moreover, the posterior distribution for λ is $\text{Gamma}(\alpha_R + \nu, \beta_R + \tau)$.

The posterior distributions for the individual event rates, $\lambda_1, \dots, \lambda_s$ must be constructed so that the convolution of these distributions is the same as the posterior distribution for λ . A reasonable choice is to let λ_s be $\text{Gamma}((\alpha_R + \nu)\alpha_s/\alpha_R, \beta_R + \tau)$ -distributed aposteriori, $s = 1, \dots, r$. By sampling $\lambda_1, \dots, \lambda_r$ from these distributions and calculating λ as the sum of these rates, we have all the quantities needed to sample the number of events as well as the resulting consequences. However, by utilizing once again the mixture property of compound Poisson processes, it is easier to proceed in a similar fashion as in the previous case. Thus, we introduce $\tilde{\lambda}_s = \lambda_s c / (c + \nu)$, $s = 1, \dots, r$, $\tilde{\mu} = \lambda \nu / (c + \nu)$, and note that $\sum_{s=1}^r \tilde{\lambda}_s + \tilde{\mu} = \lambda$. We then use the following algorithm:

Algorithm 2.5 *For each iteration in the Monte Carlo simulation:*

STEP 1. *Sample λ_s from $\text{Gamma}((\alpha_R + \nu)\alpha_s/\alpha_R, \beta_R + \tau)$ and calculate $\tilde{\lambda}_s$, $s = 1, \dots, r$ and $\tilde{\mu}$.*

STEP 2. *Sample $N_s(t)$ from a Poisson distribution with parameter $\tilde{\lambda}_s t$, $s = 1, \dots, r$.*

STEP 3. *Sample $M(t)$ from a Poisson distribution with parameter $\tilde{\mu} t$.*

STEP 4. *Sample $X_{s1}, \dots, X_{s, N_s(t)}$ from the distribution $\phi_s(x)$, $s = 1, \dots, r$.*

STEP 5. *Sample $Y_1, \dots, Y_{M(t)}$ from the distribution $\hat{F}_\nu(x)$.*

STEP 6. *Calculate:*

$$Z(t) = \sum_{s=1}^r \sum_{i=1}^{N_s(t)} X_{si} + \sum_{i=1}^{M(t)} Y_i.$$

Similar to the previous case this procedure works since $\sum_{s=1}^r N_s(t) + M(t)$ has a Poisson distribution with parameter $(\sum_{s=1}^r \tilde{\lambda}_s + \tilde{\mu})t = \lambda t$. The resulting consequence distribution mixes the consequences $X_{s1}, \dots, X_{s, N_s(t)}$ with the consequences $Y_1, \dots, Y_{M(t)}$, and thus gets exactly the desired mixture distribution $\phi'(x)$.

3 The main models

In this section we will introduce the main models for describing the total operational risk exposure. More specifically, we introduce the *incident database model* used to summarize data from observed incidents. Moreover, we introduce the risk factor model which is based on subjective assessments from a panel of experts. Finally, we explain how these two models can be combined into a unified model.

3.1 The incident database model

The incident database model consists of a single compound Poisson process. The model is fitted by using all available data from incidents that have occurred. We

assume that we have observed the event process for a period of τ units of time. In this period we have observed ν events. For each of these events we have also recorded their resulting consequences, denoted $W_1, \dots, W_\nu$ respectively. The number of events in an upcoming period of length $t = 1$ year is denoted N_I , and the corresponding consequences are denoted $X_{I1}, \dots, X_{I,N_I}$, where the subscript I indicates that we are considering the incident database model.

For a given λ_I , we assume that $N_I \sim \text{Po}(\lambda_I)$, while $\lambda_I \sim \text{Gamma}(\alpha_I, \beta_I)$ a priori. Following the results from Subsection 2.1 the posterior distribution for λ_I is $\text{Gamma}(\alpha_I + \nu, \beta_I + \tau)$. Using the observed consequences, $W_1, \dots, W_\nu$ we compute the empirical cumulative distribution function:

$$\hat{F}_I(x) = \nu^{-1} \sum_{i=1}^{\nu} \mathbf{I}(W_i \leq x).$$

For the incident database model we assume that the prior knowledge is weak. Thus, we assume that α_I and β_I are small compared to ν and τ . As a result the posterior distribution for λ_I is approximately $\text{Gamma}(\nu, \tau)$. With a weak prior knowledge, we also do not include a prior estimate of the consequence distribution. Instead we simply use $\hat{F}_I(x)$ as our consequence distribution. Under these assumptions we may use the following simplified algorithm:

Algorithm 3.1 *For each iteration in the Monte Carlo simulation:*

STEP 1. *Sample λ_I from $\text{Gamma}(\nu, \tau)$.*

STEP 2. *Sample N_I from a Poisson distribution with parameter λ_I .*

STEP 3. *Sample $X_{I1}, \dots, X_{I,N_I}$ from the distribution $\hat{F}_I(x)$.*

STEP 4. *Calculate:*

$$Z(t) = \sum_{i=1}^{N_I} X_{Ii}.$$

3.2 The risk factor model

One of the reasons for including subjectively assessed operational risks in the calculation is the relatively short history of relevant incidents. Moreover, recent changes in both internal and external conditions may significantly change the various risk factors and even introduce new and completely unknown factors. Such changes will typically not be represented in the incident database. As a result, the observed incidents alone may give an acceptable description of the impact distribution's body, but are in many cases not sufficient to describe the tail.

Through the use of a panel of experts a list of r potential operational risk factors is identified. A common approach to risk factor modelling is to assess the probability that the risk factor occurs within a unit of time, say a year. A weakness with this approach is that it does not allow repeated events within the same unit of time. In order to allow the risk factor to occur more than once, we will be using a Poisson model for the number of occurrences of a risk factor.

Note that with a Poisson distribution with rate λ the probability of zero events is $e^{-\lambda}$. If λ is small, it follows by a Taylor expansion that this probability is approximately $1 - \lambda$, and hence, the probability of at least one event is approximately λ . If λ is small, the probability of at least one event is approximately equal to the probability of exactly one event. Thus, in such cases the difference between a Poisson model and a binary model is small, and the rate of the Poisson model is close to the probability that the risk factor occurs in the binary model.

In the light of this initial discussion, the chosen risk factor model consists of a set of r compound Poisson processes, one for each risk factor. As for the incident database model, we consider events in an upcoming period of length $t = 1$ year, and we let N_s and λ_s denote respectively the number of events in this period and the event rate for the s th process $s = 1, \dots, r$.

The expert panel assesses priors to each rate, and we let $\lambda_s \sim \text{Gamma}(\alpha_s, \beta_s)$, $s = 1, \dots, r$. We recall from Subsection 2.1 that the β_s -parameters can be interpreted as a measure of the strength of the prior knowledge, while the α_s -parameters can be assessed indirectly using (2.8).

The expert panel also assesses prior estimates for consequence distributions. As before we denote the assessed prior estimate of the consequence distribution for the s th risk factor by ϕ_s , $s = 1, \dots, r$.

At this stage the risk factor model does not include the use of any data. Thus, the model may be viewed as a pure prior model, and simulated using Algorithm 2.3.

3.3 *The combined model*

In this section we turn to the problem of combining the incident database model and the risk factor model. We recall that the main purpose of including the risk factor model is to obtain a better estimate of the tail of the accumulated consequence distribution. Typically, the incident database will include a lot of minor events with small consequences. Such events are not likely to contain relevant information about the risk factors.

If we use a model with a *full overlap* between the incident model and the risk factor model, the consequence distribution becomes strongly dominated by the observed consequences. At the same time the posterior distributions of the event rates of the risk factors are shifted towards much higher values. Thus, we end up with a model with a relatively high number of events with low consequences.

On the other hand, if we use a model with *no overlap*, the consequence distributions for the risk factors remain the same as assessed by the experts. At the same time the posterior distributions of the event rates of the risk factors are shifted towards much lower values. The reason for this is that we have to take into account that no overlap also implies that none of the risk factors have occurred during the observation period.

The model we propose, however, tries to balance these issues by using some but not all of the incident data to update the occurrence rate uncertainty and the consequence distributions in the risk factor model. In order to choose the set of incident data to be included in the updating of the risk factor model, we observe that the consequences associated with the risk factors typically are high. Thus, we use a random criterion

where the probability of being included in the updating of the risk factor model increases by the size of the consequence. As before we assume that we have observed the consequences $W_1, \dots, W_\nu$ over a period of length τ . We then let $R_i = 1$ if the i th observed incident is included in the updating of the risk factor model, and 0 otherwise, $i = 1, \dots, \nu$. The R_i s are sampled using the following probability:

$$P(R_i = 1) = 1 - e^{-\rho W_i}, \quad i = 1, \dots, \nu, \quad (3.1)$$

where ρ is a suitable constant. We observe that $P(R_i = 1)$ is an increasing function of ρ . Hence, a high value of ρ implies that a lot of the incidents will be selected, while a small value of ρ implies that only a few incidents will be selected. Note, however, that the choice of ρ depends strongly on the scale of the consequences. Converting the monetary consequences e.g., between different currencies implies that ρ must be changed accordingly. We also introduce the sets $\Omega_R = \{i : R_i = 1\}$ and $\Omega_I = \{i : R_i = 0\}$. Thus, the sets Ω_R and Ω_I is a partition of the index set of observations such that Ω_R contains the indices of the observations included in the updating of the risk factor model, while Ω_I contains the indices of the observations used in the incident database model.

If the observations included in the updating of the risk factor model can be partitioned with respect to the individual risk factors, the updated risk factor model can be simulated using Algorithm 2.4. If this is not possible, we make the assumption that $\beta_1 = \dots = \beta_r = \beta_R$ and use Algorithm 2.5 instead.

The following algorithm summarizes how the simulations are done for the combined model:

Algorithm 3.2 *For each iteration in the Monte Carlo simulation:*

STEP 1. *Sample R_i so that $P(R_i = 1) = 1 - e^{-\rho W_i}$, $i = 1, \dots, \nu$, and determine the index sets Ω_R and Ω_I .*

STEP 2. *Simulate the incident database model based on the observations with indices in Ω_I using Algorithm 3.1.*

STEP 3. *Simulate the risk factor model based on the observations with indices in Ω_R using either Algorithm 2.4 or Algorithm 2.5.*

STEP 4. *Calculate the total consequence by adding the results from both models.*

4 A numerical example

In this section we illustrate the proposed methods by considering a numerical example². In this example the incident database consists of data from a period of $\tau = 5$ years. During this period $\nu = 460$ incidents are recorded. The estimated cumulative distribution function, $\hat{F}_\nu(x)$, is shown in Figure 1. We observe that most of the observed consequences are relatively minor. The mean consequence is 2.18, while the standard deviation is 8.61. The average number of incidents pr. year is $\nu/\tau = 460/5 = 92$. Thus, the mean total consequence pr. year is $92 \cdot 2.18 = 200.56$.

²All the incident data and risk factors used in the example are fictive. Moreover, for simplicity we have skipped the monetary unit throughout the paper.

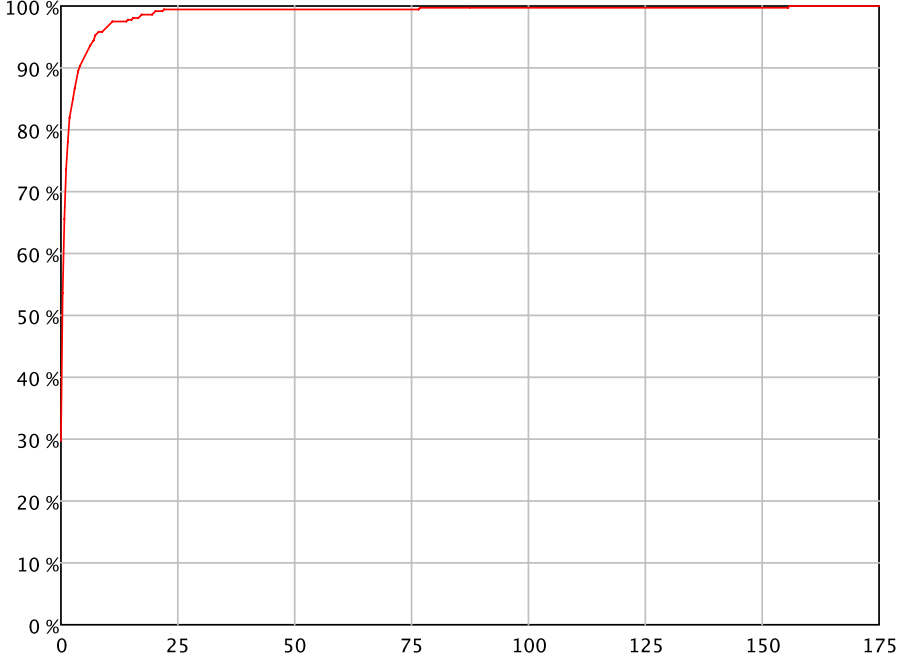


Figure 1: The estimated cumulative distribution function, $\hat{F}_\nu(x)$, for the consequences of $\nu = 460$ observed incidents.

Using Algorithm 3.1 we have run a simulation of the incident database model. The results consisting of 20000 iterations are shown in Figure 2. The mean value of the estimated distribution is 200.44, which is close to the value computed above, i.e., 200.56, while the standard deviation is 86.32.

In the risk factor model 30 risk factors are included. Their respective parameters are shown in Table 1. In the table the ratio α_s/β_s , i.e., the prior mean value of λ_s , is given for each of the risk factors. In order to obtain the value of α_s , this ratio has to be multiplied with β_s . We assume that $\beta_1 = \dots = \beta_{30}$, and denote this common value by β_R . We recall that this parameter indicates the strength in the prior estimates of the event rates. In order to study the effect of the β_R -parameter, we consider three different values for this: $\beta_R = 0.2$ (base case), $\beta_R = 0.5$ and $\beta_R = 1.0$. The resulting mean values and standard deviations are shown in Table 2, while the corresponding estimated prior cumulative distribution functions for the risk factor model are shown in Figure 3. We observe that the mean values of the three prior distributions are approximately equal, while the standard deviations vary a lot. When β_R is low, the uncertainties about the event rates are large. As β_R grows, so does the strength of the prior knowledge about the event rates. As a result the standard deviation shrinks from 418 for $\beta_R = 0.2$ to 247 for $\beta_R = 1.0$.

Now, we turn to the posterior combined model. We assume that it is not possible to partition the data with respect to the individual risk factors. This means that we use Algorithm 2.5 in Step 3 of Algorithm 3.2 in order to generate results for the risk factor model. However, before we get to this step, we start out by partitioning the index set of the consequence observations into the two sets Ω_I and Ω_R using the

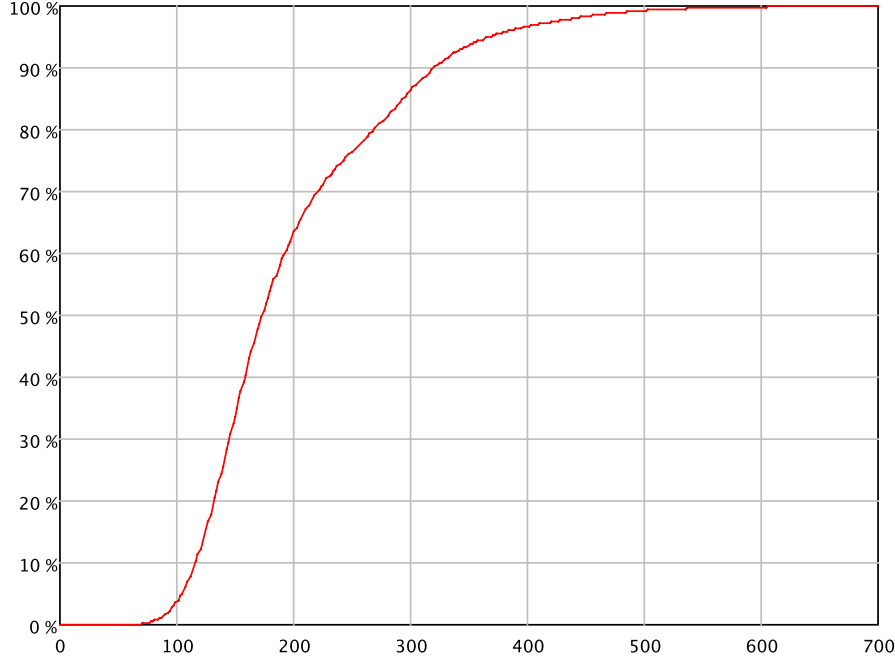


Figure 2: The estimated prior cumulative distribution function for the incident database model.

Table 1: Risk factor model parameters

s	α_s/β_s	ξ_s	σ_s	s	α_s/β_s	ξ_s	σ_s
1	0.1	136	34	16	0.2	72	18
2	0.1	124	31	17	0.1	72	18
3	0.1	120	30	18	0.2	72	18
4	0.2	112	28	19	0.2	60	15
5	0.2	108	27	20	0.1	60	15
6	0.1	96	24	21	0.1	56	14
7	0.1	96	24	22	0.2	52	13
8	0.1	96	24	23	0.1	52	13
9	0.2	80	20	24	0.3	48	12
10	0.3	80	20	25	0.3	48	12
11	0.1	80	20	26	0.3	44	11
12	0.1	80	20	27	0.3	44	11
13	0.2	76	19	28	0.3	40	10
14	0.2	76	19	29	0.3	40	10
15	0.1	76	19	30	0.2	36	9

random partition method explained in the previous section. Thus, the probability that index i with consequence W_i ends up in the set Ω_R is given by (3.1). Given all the consequences $W_1, \dots, W_{460}$, we can estimate the expected size of the set Ω_R as a function of the parameter ρ . The result is shown in Figure 4. By considering this curve the final judgement about the parameter ρ can be made.

Three different values for the parameter ρ are chosen for the simulation study. The

Table 2: Mean and standard deviations for the estimated prior distribution of the risk factor model.

	$\beta_R = 0.2$	$\beta_R = 0.5$	$\beta_R = 1.0$
Mean	367.15	365.33	369.94
Stdev	417.98	297.87	246.58

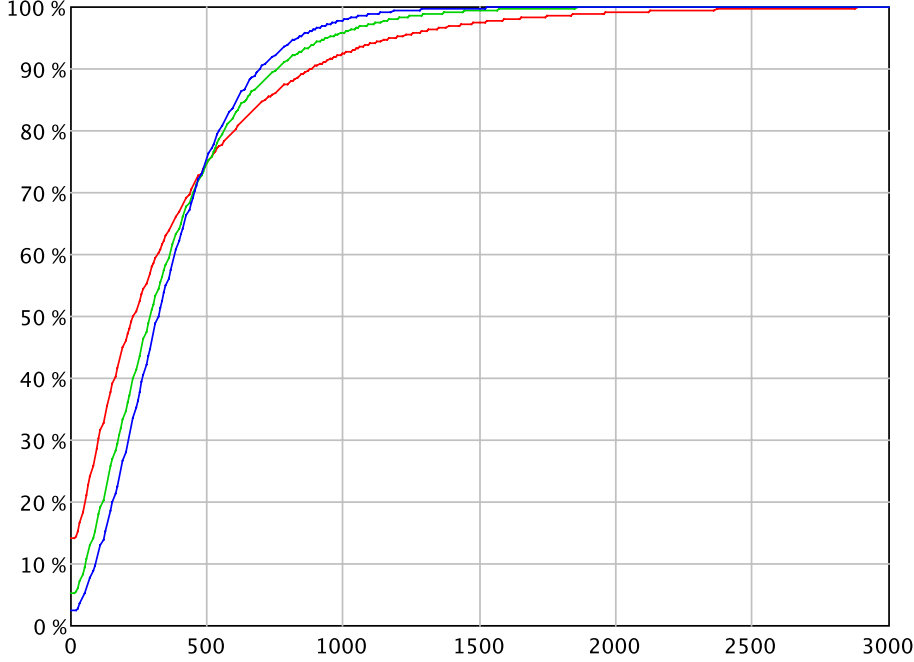


Figure 3: The estimated prior cumulative distribution function for the risk factor model using $\beta_R = 0.2$ (red curve), $\beta_R = 0.5$ (green curve) and $\beta_R = 1.0$ (blue curve).

base case value is $\rho = 0.50$. With this value the expected size of Ω_R is 160 which is considered to be a reasonable value in this context. In addition to the base case we also include two more extreme values: $\rho = 0.005$ and $\rho = 50$. When the smallest ρ -value is used, the expected size of the set Ω_R is about 5, while the expected size of the set Ω_I is about 445. Conversely, when the largest ρ -value is used, the expected size of the set Ω_I is close to zero, while the expected number of indices in the set Ω_R is almost 450.

We also need to specify the constant c used in (2.11). This parameter indicates the strength of the prior knowledge about the consequence distributions. While the parameter β_R is a measure of the strength of the knowledge about the event rates, the sum of the α_s -parameters indicates the strength of the prior knowledge about the consequence distribution. Thus, a reasonable choice is to let c be approximately equal to the sum of the α_s -parameters. Note, however, that in our example the values of the α_s -parameters are linked to the value of the β_R -parameter. As a result the parameter c is linked to β_R as well. The resulting values for c are given in Table 3.

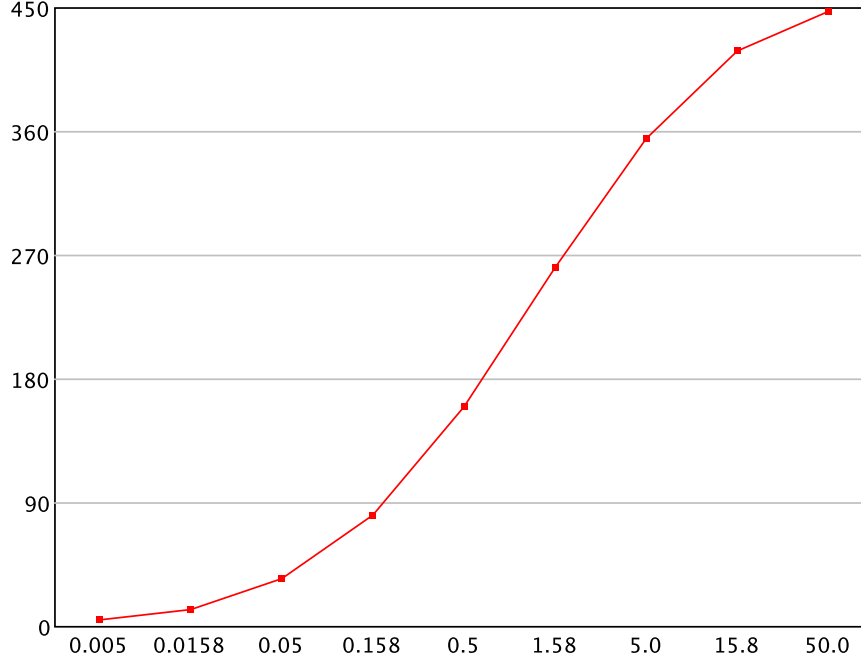


Figure 4: Expected number of observations included in the updating of the risk factor model as a function of ρ .

Table 3: The values of the c -parameter used in the simulations.

	$\beta_R = 0.2$	$\beta_R = 0.5$	$\beta_R = 1.0$
c	1.08	2.70	5.40

We are now ready to run the Monte Carlo simulations. We run 20000 iterations for all combinations of the three values of β_R , i.e., 0.2, 0.5 and 1.0 and the three values of ρ , i.e., 0.005, 0.5 and 50. The results for $\beta_R = 0.2$ are shown in Table 4 and in Figure 5. The results for $\beta_R = 0.5$ are shown in Table 5 and in Figure 6. Finally, the results for $\beta_R = 1.0$ are shown in Table 6 and in Figure 7.

Table 4: Mean and standard deviations for the case when $\beta_R = 0.2$.

		$\rho = 0.005$	$\rho = 0.5$	$\rho = 50$
Incidence model	Mean	172.07	32.72	0.05
	St.dev.	67.14	7.81	0.05
Risk factor model	Mean	40.47	174.43	207.1
	St.dev.	68.71	91.45	91.48
Combined model	Mean	212.54	207.15	207.17
	St.dev.	93.23	91.76	91.48

We recall that when we ran separate simulations on the incident database model and the risk factor model, the resulting mean values were about 200 and 367 respec-

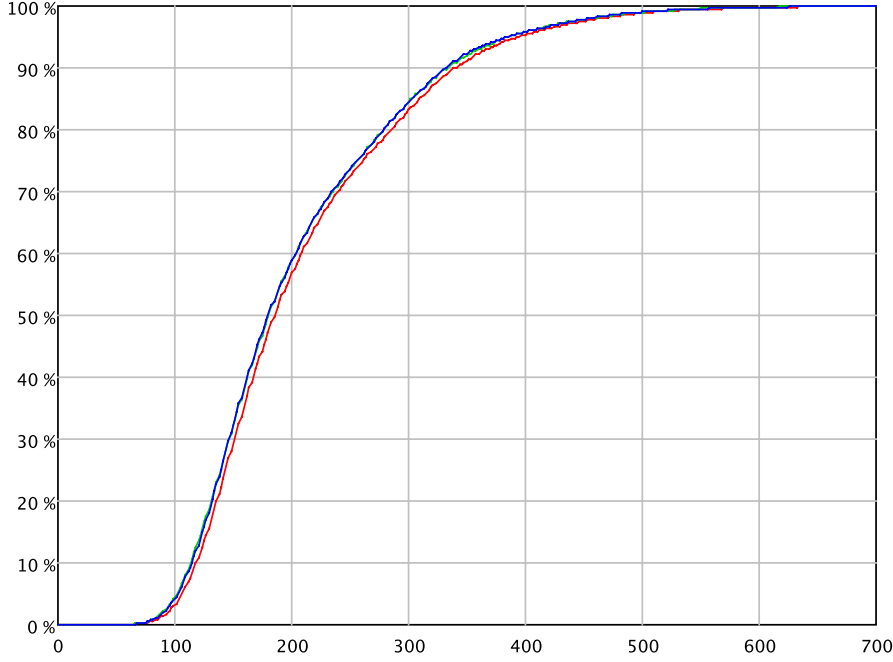


Figure 5: Estimated cumulative distributions for the accumulated consequences using $\rho = 0.005$ (red curve), $\rho = 0.5$ (green curve) and $\rho = 50$ (blue curve), $\beta_R = 0.2$.

Table 5: Mean and standard deviations for the three models for the case when $\beta_R = 0.5$.

		$\rho = 0.005$	$\rho = 0.5$	$\rho = 50$
Incidence model	Mean	172.07	32.72	0.05
	St.dev.	67.14	7.81	0.05
Risk factor model	Mean	57.95	184.69	215.62
	St.dev.	79.41	97.47	97.57
Combined model	Mean	230.02	217.40	215.67
	St.dev.	101.51	97.75	97.57

Table 6: Mean and standard deviations for the three models for the case when $\beta_R = 1.0$

		$\rho = 0.005$	$\rho = 0.5$	$\rho = 50$
Incidence model	Mean	172.07	32.72	0.05
	St.dev.	67.14	7.81	0.05
Risk factor model	Mean	84.06	199.40	227.87
	St.dev.	92.67	105.86	105.86
Combined model	Mean	256.13	232.11	227.93
	St.dev.	112.23	106.10	105.86

tively. Thus, if we had combined the two models simply by adding all the consequences, we would have ended up with a mean total consequence of about 567. On

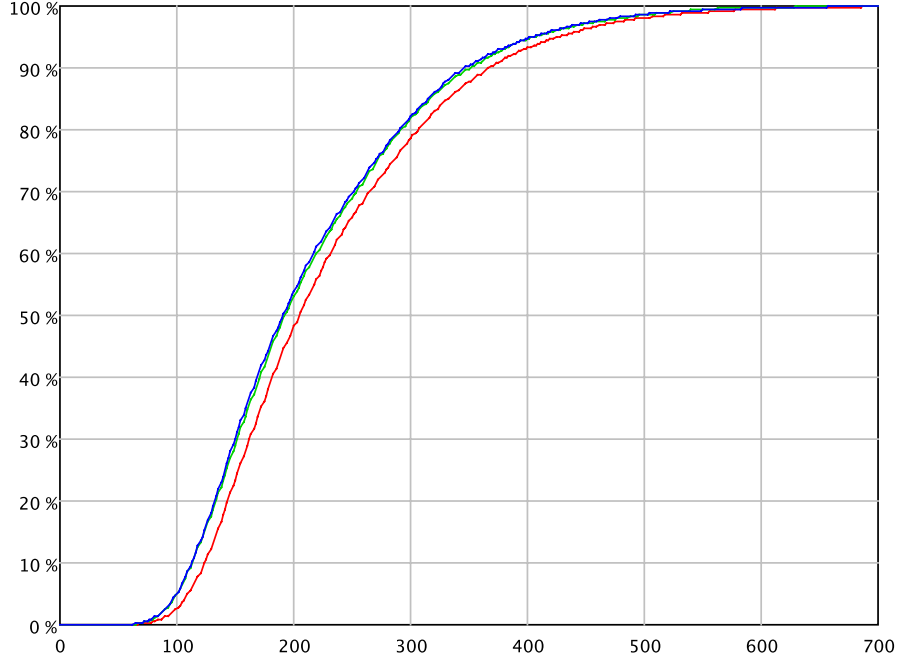


Figure 6: Estimated cumulative distributions for the accumulated consequences using $\rho = 0.005$ (red curve), $\rho = 0.5$ (green curve) and $\rho = 50$ (blue curve), $\beta_R = 0.5$.

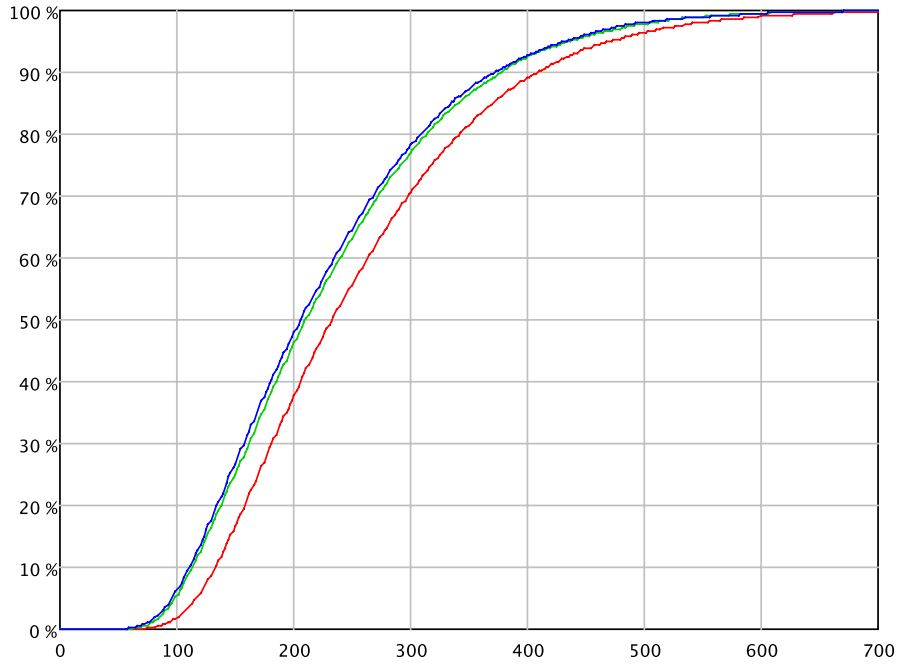


Figure 7: Estimated cumulative distributions for the accumulated consequences using $\rho = 0.005$ (red curve), $\rho = 0.5$ (green curve) and $\rho = 50$ (blue curve), $\beta_R = 1.0$.

the other hand, if we had combined the two by averaging, the result would have been about half of this, i.e., 284. Comparing this the results obtained by the proposed

model, we see that even in the case where $\beta_R = 1.0$ and $\rho = 0.005$ the mean value is just 256. The reason for this reduction depends on the value of ρ . In the cases where ρ is small, there is almost no overlap between the incident database model and the risk factor model. In such cases the consequence distributions of the risk factors are not affected by the data. The posterior distributions of the event rates, however, are shifted towards lower values. Thus, the smaller consequences sampled from the incident database model dominate the total risk. On the other hand, in cases where ρ is large, there is a large degree of overlap between the incident database model and the risk factor model. In these cases the consequence distribution of the risk factors are strongly affected by the smaller observed consequences. At the same time the event rates for the risk factors are actually increased considerably.

The value of the parameter β_R mainly affects how much the event rates of the risk factors are changed as a consequence of the observations. A high value of β_R implies that the prior assessments of the event rates get more weight, and thus, the effects of the observations are reduced. Recalling that in our model the parameter c is linked to the parameter β_R , we see from Table 3 that c increases proportionally with β_R . Hence, a high value of β_R also implies a high value of c , and when c is high, the prior consequence distributions also get more weight compared to the observed consequences. The combined effects of the value of β_R is that the mean total consequence increases as β_R increases. In the base case where $\rho = 0.5$ the mean total consequence increases from 207 when $\beta_R = 0.2$ up to 232 when $\beta = 1.0$, i.e., a 12% increase. The corresponding numbers for the other cases are 20% when $\rho = 0.005$ and 9% when $\rho = 50$. We observe that the sensitivity to the value of β_R is greater when there is little overlap between the two models, which is reasonable.

5 Conclusions

In the present paper we have studied how one can combine incident data with subjective assessments. A very important issue in relation to this is the degree of overlap between these two sources of information. The models we have proposed cover the full range from no overlap to full overlap. In the case of no overlap the consequence data for the incidents is irrelevant with respect to the consequence distribution for the risk factors. Still having observed the processes a number of years without recording any events related to the risk factors, is vital information which need to be taken into account. In the case of full overlap the incident data is used both to update the rate distribution and the consequence distribution.

The focus of this paper has partly been on the rates of occurrence of the various risk factors, and how these can be updated in a consistent way. Moreover, we have shown how to fit and update the consequence distributions. The proposed model for the consequences is very flexible as it works for all types of prior consequence distributions, including non-parametric distributions, and includes a full scale Bayesian updating procedure.

In the numerical example presented in the paper we have studied the sensitivity of the proposed model with respect to two important parameters: β_R which indicates the

strength of the prior knowledge, and ρ which reflects the degree of overlap between the two models. The calculated sensitivities indicate that in a real-life situations, one needs to have a clear understanding of these parameters, and how they influence the results. One should always examine these assessments and preferably perform a sensitivity analysis in order to obtain robust conclusions.

Throughout the paper we have used *stationary models* for all the compound processes. In a real-life case the risk picture may change dynamically as time goes by. Such changes make it even more difficult to use historical data as part of the assessments. In the proposed models, changes enter into the picture through the subjectively assessed risk factor models. A more advanced approach may include dynamics more explicitly.

References

- [Adachi et al. 2011] Adachi, M. et al. *Operational Risk - Supervisory Guidelines for the Advanced Measurement Approaches* Basel Committee on Banking Supervision, 2011.
- [Adelson 1966] Adelson, R. M. Compound Poisson distributions. *Operational Research Quarterly* (17): 73–75, 1966.
- [Berger 2010] Berger, J. O. *Statistical Decision Theory and Bayesian Analysis*. Springer Series in Statistics, 2010.
- [Bølviken 2014] Bølviken, E. *Computation and Modelling in Insurance and Finance* Cambridge University Press, 2014.
- [Huseby 1989] Huseby, A. B. Combining Opinions in a Predictive Case In *Bayesian Statistics 3*.: 641–651, Oxford University Press, 1989.
- [Huseby 2015] Huseby, A. B., Thomsen, J. Quantifying operational risk exposure by combining incident data and subjective risk assessments In Podofillini et al. (Eds) *Safety and Reliability of Complex Engineered Systems*: 443–451, Taylor & Francis Group, London, 2015.
- [McNeil et al. 2005] McNeil, A. J., Frey, R., Embrechts, P. *Quantitative Risk Management, Concepts, Techniques and Tools* Princeton University Press, 2005.