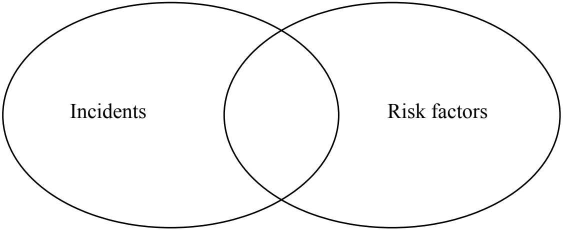
Quantifying operational risk exposure by combining incident data and subjective risk assessments

Arne Bang Huseby [1] and Jan Thomsen [2]

[1]University of Oslo, Norway

[2]Norges Bank Investment Management, Norway

Combining incident data and subjective risk assessments



- INCIDENTS: Observations of incidents that have occurred in the past and their respective consequences (i.e., economic loss)
- RISK FACTORS: A panel of experts identify a set of potential risk factors. For each risk factor the experts assess the *frequency of occurrence* as well as a *consequence distribution*.

Main challenges

- Only relatively recent incidents are relevant when predicting future events. Thus, we only have data from a few years of operation.
- The main purpose of including the subjectively identified risk factors is to cover potential risk events that have not yet occurred.
- There will typically be some overlap between the types of incidents that have occurred, and the identified risk factors.
- The incident data does not allow the analyst to match observed incidents and identified risk factors. Thus, the amount of overlap is uncertain.
- There are major differences between the observed incidents and the identified risk factors.

Compound Poisson process

$$N(t) = \text{The number of incidents in the interval } [0, t)$$

$$X_i = \text{The consequence of the } i\text{th incident}$$

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

Main assumptions

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

$X_1, X_2, \ldots$ are independent and identically distributed variables with common distribution F_X.

Parameter uncertainty

Prior distribution for λ: $[\lambda \sim \mathrm{Gamma}(\alpha, \beta)]$

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

Posterior distribution for λ: $[\lambda|\tau, \nu \sim \mathrm{Gamma}(\alpha+\nu, \beta+\tau)]$

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

where τ is the length of the interval the process has been observed in and ν is the number of observed incidents in this interval.

Note: The parameter β can be interpreted as a measure of the *strength* of the prior knowledge. More specifically, the prior knowledge is comparable to having observed the process over a time interval of length β.

Parameter uncertainty (cont.)

PROOF: Assuming that the process has been observed in τ units of time, the joint distribution of λ and $N(\tau)$ is:

$$\begin{aligned}\pi(\lambda) \cdot P(N(\tau) = \nu) &= \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} \cdot \frac{(\lambda\tau)^\nu}{\nu!} e^{-\lambda\tau} \\ &= C \cdot \lambda^{\alpha+\nu-1} e^{-(\beta+\tau)\lambda},\end{aligned}$$

where C is a suitable constant not depending on λ, and from this it easily follows that:

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

Thus, $\lambda|\tau, \nu \sim \mathrm{Gamma}(\alpha+\nu, \beta+\tau)$ as stated.

The incident database model

Data: The process is observed for a period of τ units of time during which ν incidents occurred. The corresponding observed consequences are $X_1, \ldots, X_\nu$.

N_I = Number of incidents in the upcoming year

X_I = The consequence of an incident in the upcoming year

Frequency model: $N_I | \lambda_I \sim \mathrm{Po}(\lambda_I)$, $\lambda_I \sim \mathrm{Gamma}(\alpha_I + \nu, \beta_I + \tau)$, where α_I and β_I are the parameters of the prior for λ_I.

Consequence model: $X_I \sim F_I$, where the distribution F_I is a lognormal distribution with mean value ξ_I and standard deviation σ_I estimated using the observed consequences.

The risk factor frequency model

We let r denote the number of risk factors, N_s denote the number of times the sth risk factor occurs in the upcoming year, and assume that:

$$N_s|\lambda_s \sim \mathrm{Po}(\lambda_s), \quad \lambda_s \sim \mathrm{Gamma}(\alpha_s, \beta_s), \quad s = 1, \ldots, r.$$

We also introduce:

$$\begin{aligned} N_R &= N_1 + \cdots + N_r \\ \lambda_R &= \lambda_1 + \cdots + \lambda_r \\ \alpha_R &= \alpha_1 + \cdots + \alpha_r \end{aligned}$$

Assuming independence between the risk factors, and making the simplifying assumption that $\beta_1 = \beta_2 = \cdots = \beta_s = \beta_R$, we get:

$$\begin{aligned} N_R|\lambda_R &\sim \mathrm{Po}(\lambda_R) \\ \lambda_R &\sim \mathrm{Gamma}(\alpha_R, \beta_R) \end{aligned}$$

The risk factor consequence model

Assume that there has been an event related to one of the risk factors. The actual risk factor that caused this event, however, is unknown. The unknown index of this risk factor is denoted by S and the corresponding consequence is X_R.

Then it can be shown that:

$$P(S = s) = \frac{\alpha_s}{\sum_{i=1}^{r} \alpha_i} = \frac{\alpha_s}{\alpha_R}.$$

By the assumptions we also have that:

$$\mathrm{E}[X_R | S = s] = \xi_s$$

$$\mathrm{Var}[X_R | S = s] = \sigma_s.$$

The risk factor consequence model (cont.)

We denote the distribution of X_R by F_R and fit this to a lognormal distribution with mean value ξ_R and standard deviation σ_R. Then:

$$\xi_R = \mathrm{E}\,\mathrm{E}[X_R|S] = \frac{\sum_{s=1}^{r} \xi_s \alpha_s}{\alpha_R},$$

$$\begin{aligned}
\sigma_R^2 &= \mathrm{E}\,\mathrm{Var}[X_R|S] + \mathrm{Var}\,\mathrm{E}[X_R|S] \\
&= \frac{\sum_{s=1}^{r} \sigma_s^2 \alpha_s}{\alpha_R} + [\frac{\sum_{s=1}^{r} \xi_s^2 \alpha_s}{\alpha_R} - \xi_R^2] \\
&= \frac{\sum_{s=1}^{r} (\xi_s^2 + \sigma_s^2) \alpha_s}{\alpha_R} - \xi_R^2.
\end{aligned}$$

No overlap between the models

Incident database model: The prior distribution for λ_R is updated to:

$$\lambda_I | \text{data} \sim \mathrm{Gamma}(\alpha_I + \nu, \beta_I + \tau)$$

The consequence model is estimated using *all* the observed consequences.

Risk factor model: The prior distribution for λ_R is updated to:

$$\lambda_R | \text{data} \sim \mathrm{Gamma}(\alpha_R, \beta_R + \tau)$$

No change in the consequence model.

Total consequence: The two models are simulated independently, and the total consequence is the sum of the consequences from the two models.

Full overlap between the models

Incident database model: Merged into the risk factor model.

Risk factor model: The distribution for λ_R is updated to:

$$\lambda_R|\text{data} \sim \mathrm{Gamma}(\alpha_R + \nu, \beta_R + \tau)$$

The consequence model is updated using *all* the observed consequences.

Total consequence: Only the risk factor model is simulated, and the total consequence is the sum of the consequences from the risk factor model.

Partial overlap between the models

Incident database model: The prior distribution for λ_R is updated to:

$$\lambda_I | \text{data} \sim \mathrm{Gamma}(\alpha_I + \nu_I, \beta_I + \tau)$$

The consequence model is estimated using the ν_I observed consequences which are *not* associated with any risk factor.

Risk factor model: The prior distribution for λ_R is updated to:

$$\lambda_R | \text{data} \sim \mathrm{Gamma}(\alpha_R + \nu_R, \beta_R + \tau)$$

The consequence model is estimated using the ν_R observed consequences which are associated with some risk factor.

Total consequence: The two models are simulated independently, and the total consequence is the sum of the consequences from the two models.

Data: $\tau = 5$, $\nu = 460$, (Partial overlap: $\nu_I = 292$, $\nu_R = 168$)

Base case: $\beta_R = 0.2$ **Alternative case:** $\beta_R = 1.0$

Incident database model (base case):

	Prior	No overlap	Full overlap	Partial overlap
$\mathrm{E}[\lambda_I]$	-	91.82	-	58.29
$\mathrm{E}[X_I]$	-	2.2	-	0.5

Risk factor model with $r = 30$ **risk factors (base case):**

	Prior	No overlap	Full overlap	Partial overlap
$\mathrm{E}[\lambda_R]$	5.40	0.21	88.67	32.52
$\mathrm{E}[X_R]$	37.9	37.9	2.3	5.3

Simulated results when $\beta_R = 0.2$

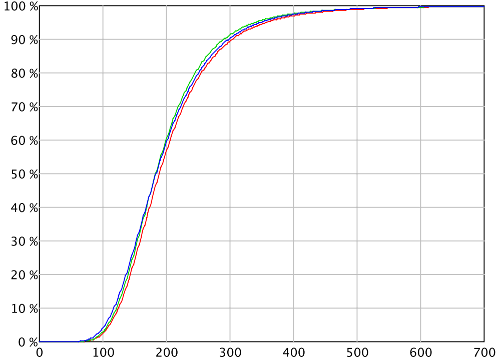


Figure: No overlap (red), Full overlap (green), Partial overlap(blue).

	No overlap	Full overlap	Partial overlap
Mean	207.674	200.369	201.472
St.dev.	89.761	84.732	85.608

Simulated results when $\beta_R = 1.0$

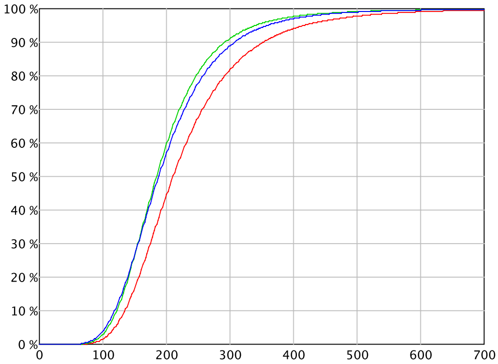


Figure: No overlap (red), Full overlap (green), Partial overlap(blue).

	No overlap	Full overlap	Partial overlap
Mean	233.930	200.960	205.794
St.dev.	103.703	83.240	86.753

Conclusions

- Models for how to combine incident data with subjective assements have been developed
- The models cover the full range from no overlap to full overlap
- Main focus has been on the *rates of occurrence* of the various risk factors, and how these can be updated and combined in a consistent way
- In a more refined model, wider classes of consequence distribution should be considered
- A more flexible approach should include fitting specific distributions for each individual risk factor
- In order to combine the various consequence distributions and observations, a full scale Bayesian updating approach should be developed