A Online Appendix to the paper "A Comprehensive Model for Cyber Risk based on Marked Point Processes and its Application to Insurance"

A.1 Background on Point Processes

Proposition 4 (Superposition, ([49], p.16)). *Let $\{N_i\}_{i\in\mathbb{N}}$ be a countable collection of point processes, then their superposition $\bigcup_{i=1}^{\infty} N_i$ also forms a point process. If $N_1, N_2, \ldots$ are independent Poisson processes with mean measures $\Lambda_1, \Lambda_2, \ldots$, then their superposition will also be a Poisson process with mean measure $\Lambda = \sum_{i=1}^{\infty} \Lambda_i$.*

Proposition 5 (Thinning ([28], p.34)). *Let $N(\cdot)$ be a (simple, inhomogeneous) Poisson process with rate $\lambda(\cdot)$. Let $p(\cdot)$ be a measurable function on $[0,\infty)$ such that $0 \le p(x) \le 1$ holds $\forall x \in [0,\infty)$. Let a new process $\tilde{N}(\cdot)$ be formed by independently looking at each point of a realization $\{t_i\}$ of $N(\cdot)$ and retaining it with probability $p(x_i)$ (thus deleting it with probability $1-p(x_i)$). Then $\tilde{N}(\cdot)$ is a Poisson process with rate $p(x)\lambda(x)$.*

Definition 1 (Marked Point Process ([28], 6.4.I)). *A marked point process (MPP) with locations in $\mathcal{X}$ and marks in $\mathcal{K}$ is a point process $\{x_i, k_i\}$ on $\mathcal{X} \times \mathcal{K}$ with the additional property that the ground process $N_g(\cdot)$, meaning the process of locations $\{x_i\}$ is itself a point process, i.e. for bounded $A \in \mathcal{B}_{\mathcal{X}}$, $N_g(A) = N(A \times \mathcal{K}) < \infty$.*

Proposition 6 ([28], Prop. 6.4.IV). *Let N be a MPP with independent marks. Then the probability structure of N is completely defined by the distribution of N_g and the mark kernel $\{F(k|x) : k \in \mathcal{B}_{\mathcal{K}}, x \in \mathcal{X}\}$ representing the conditional distribution of the mark, given location x.*

Definition 2 (Compound Poisson process). *Let $N := (N(t))_{t\ge 0}$ be a Poisson process with mean measure $\Lambda(t) > 0$. Let $\{Z_i\}_{i\in\mathbb{N}}$ be a sequence of iid. random variables independent of N. Then the process $R := (R(t))_{t\ge 0}$ defined as*

$$R(t) := \sum_{i=1}^{N(t)} Z_i, \; t \ge 0,$$

is called a compound Poisson process.

A.2 Characteristics of Compound Poisson Distribution

Theorem 1 (Wald equation ([78])). *Let $\{X_i\}$ be a sequence of real-valued, iid. random variables and let $N(t) \ge 0$ be an integer-valued r.v. independent of the sequence $\{X_i\}$. Suppose $\mathbb{E}[N(\cdot)] < \infty$ and $\mathbb{E}[X_i] < \infty$. Then*

$$\mathbb{E}\left[\sum_{i=1}^{N(t)} X_i\right] = \mathbb{E}[X_1]\mathbb{E}[N(t)].$$

Theorem 2 (Law of total variance ([15], p. 401)). *Let X and Y be random variables on the same probability space and assume $\mathbb{V}ar[Y] < \infty$. Then*

$$\mathbb{V}ar[Y] = \mathbb{E}[\mathbb{V}ar(Y|X)] + \mathbb{V}ar(\mathbb{E}[Y|X]).$$

The last two results imply that if $\{X_i\}$ is a sequence of iid. random variables and $N(t) \ge 0$ an integer-valued random variable independent of the sequence $\{X_i\}$, then it holds

$$\begin{aligned}\mathbb{V}ar\left(\sum_{i=1}^{N(t)} X_i\right) &=: \mathbb{V}ar(Y(t)) = \mathbb{E}[\mathbb{V}ar(Y(t)|N(t)] + \mathbb{V}ar(\mathbb{E}[Y(t)|N(t)]) \\ &= \mathbb{E}[N(t)\mathbb{V}ar(X_1)] + \mathbb{V}ar(N(t)\mathbb{E}[X_1]) \\ &= \mathbb{V}ar(X_1)\mathbb{E}[N(t)] + \mathbb{E}[X_1]^2\mathbb{V}ar(N(t)).\end{aligned}$$

Proposition 7 ([57], Prop.3.3.4). *Consider the independent compound Poisson sums*

$$L_j = \sum_{i=1}^{N_j} X_i^{(j)}, \qquad j = 1, \ldots, K,$$

where $N_j \sim Poi(\lambda_j)$ for some $\lambda_j > 0$ and, for every fixed j, $(X_i^{(j)})_{i=1,2,\dots}$ is an iid. sequence of claim sizes. Then the sum

$$\tilde{L} = L_1 + \ldots + L_K$$

is again compound Poisson with representation

$$\tilde{L} \stackrel{d}{=} \sum_{i=1}^{N_\lambda} Y_i, \qquad N_\lambda \sim Poi\Big(\sum_{j=1}^{K} \lambda_j\Big),$$

and (Y_i) is an iid. sequence, independent of N_λ, with mixture distribution given by

$$F_{Y_1}(x) = \sum_{j=1}^{K} \frac{\lambda_j}{\sum \lambda_j} F_{X_1^{(j)}}(x), \qquad x \in \mathbb{R}.$$

A.3 Calculations and Proofs from Chapter 4

Proof of Proposition 1. Note that based on *(A4)* for generating S_i and generally m_i being distributed according to cdf. F_M, S_i^* can be thought of as generated analogously to S_i by drawing a realisation of m_i first and then letting

$$\mathbb{P}(Z_{ij} = 1 \mid G_i = 0,\ m_i) = \begin{cases} p_{gen} & \text{iid. } \forall j \in \{1,\dots,K\} \ s.t.\ c_j < m_i, \\ 0 & \text{else} \end{cases}$$

$$\mathbb{P}(Z_{ij} = 1 \mid G_i = 1, B_i = \hat{b},\ m_i) = \begin{cases} p_{sec} & \text{iid. } \forall j \in \{1 + \sum_{\ell=1}^{\hat{b}-1} K_\ell, \dots, \sum_{\ell=1}^{\hat{b}} K_\ell\} \ s.t.\ c_j < m_i, \\ 0 & \text{else} \end{cases}$$

i.e. one adds $Z_{ij} \equiv 0$ for all $j : c_j \geq m_i$ in each case, by effectively drawing only on the subset of the portfolio of size $K^* = \max_{k \in \{0,\dots,K\}} c_{[k]} < m_i$ (resp. the subset of one industry sector $\hat{b}$ of size $K_{\hat{b}}^* = \max_{k \in \{0,\dots,K_{\hat{b}}\}} c_{[k]}^{\hat{b}} < m_i$).
Conditioning on the realisation of $G_i \in \{0,1\},\ B_i \in \{1,\dots,B\},\ m_i \in [0,1]$ (in particular, for m_i distinguishing the cases of falling in any of the intervals $[c_{[K^*]}, c_{[K^*+1]}]$ resp. $[c_{[k_{\hat{b}}^*]}^{\hat{b}}, c_{[k_{\hat{b}}^*+1]}^{\hat{b}}])$ yields

$$\mathbb{P}(|S_i^*| = k) = \mathbb{P}(|S_i^*| = k \mid G_i = 0)\, \mathbb{P}(G_i = 0) + \sum_{\hat{b}=1}^{B} \mathbb{P}(|S_i^*| = k \mid G_i = 1,\ B_i = \hat{b})\, \mathbb{P}(B_i = \hat{b} \mid G_i = 1)\, \mathbb{P}(G_i = 1)$$

$$= \underbrace{(1 - p_G) \int_0^1 \mathbb{P}(|S_i^*| = k \mid G_i = 0, m_i = m) \mathrm{d}F_M(m)}_{(I)} + \underbrace{p_G \sum_{\hat{b}=1}^{B} p_{\hat{b}} \int_0^1 \mathbb{P}(|S_i^*| = k \mid G_i = 1, B_i = \hat{b}, m_i = m) \mathrm{d}F_M(m)}_{(II)},$$

where

$$\begin{aligned}(I) &= (1 - p_G) \int_0^1 \binom{K^*}{k} p_{gen}^k (1 - p_{gen})^{K^*-k} \mathrm{d}F_M(m) \\ &= (1 - p_G) \sum_{K^*=0}^{K} \int_0^1 \mathbb{1}_{[c_{[K^*]}, c_{[K^*+1]}]}(m) \binom{K^*}{k} p_{gen}^k (1 - p_{gen})^{K^*-k} \mathrm{d}F_M(m) \\ &= (1 - p_G) \sum_{K^*=0}^{K} \binom{K^*}{k} p_{gen}^k (1 - p_{gen})^{K^*-k} \big(F_M(c_{[K^*+1]}) - F_M(c_{[K^*]})\big),\end{aligned}$$

and

$$\begin{aligned}(II) &= p_G \sum_{\hat{b}=1}^{B} p_{\hat{b}} \int_0^1 \binom{K_{\hat{b}}^*}{k} p_{sec}^k (1 - p_{sec})^{K_{\hat{b}}^*-k} \mathrm{d}F_M(m) \\ &= p_G \sum_{\hat{b}=1}^{B} p_{\hat{b}} \sum_{K_{\hat{b}}^*=0}^{K_{\hat{b}}} \int_0^1 \mathbb{1}_{[c_{[K_{\hat{b}}^*]}^{\hat{b}}, c_{[K_{\hat{b}}^*+1]}^{\hat{b}}]}(m) \binom{K_{\hat{b}}^*}{k} p_{sec}^k (1 - p_{sec})^{K_{\hat{b}}^*-k} \mathrm{d}F_M(m) \\ &= p_G \sum_{\hat{b}=1}^{B} p_{\hat{b}} \sum_{K_{\hat{b}}^*=0}^{K_{\hat{b}}} \binom{K_{\hat{b}}^*}{k} p_{sec}^k (1 - p_{sec})^{K_{\hat{b}}^*-k} \big(F_M(c_{[K_{\hat{b}}^*+1]}^{\hat{b}}) - F_M(c_{[K_{\hat{b}}^*]}^{\hat{b}})\big).\end{aligned}$$

This implies that $\left|S_i^*\right|$ follows a Binomial mixture distribution, i.e. $f_{\left|S_i^*\right||n,p}(k) = Binom(n,p,k)$ with parameters and weights ($2K + B + 1$ cases):

$$(n,p) = \begin{cases} (K^*, p_{gen}) & \text{with weight} \quad (1-p_G)\Big(F_M(c_{[K^*+1]}) - F_M(c_{[K^*]})\Big), \quad K^* \in \{0,\dots,K\}, \\ (k_{\hat{b}}^*, p_{sec}) & \text{with weight} \quad p_G\, p_{\hat{b}}\, \Big(F_M(c^{\hat{b}}_{[k^*_{\hat{b}}+1]}) - F_M(c^{\hat{b}}_{[k^*_{\hat{b}}]})\Big), \quad k^*_{\hat{b}} \in \{0,\dots,k_{\hat{b}}\},\ \hat{b} \in \{1,\dots,B\}. \end{cases}$$

Again, intuitively this means that, depending on G_i, B_i, and m_i, one draws from a set of different size of potentially affected firms to suffer a loss. As on the respective set, the draws are conditionally iid. Bernoulli draws, the number of "successes" of interest is of course Binomially distributed.
Equation (6) in Proposition 1 follows immediately from above using *(A3)*, i.e. $m_i \sim Unif([0,1])$, thus $F_M(c) = c$, $\forall c \in [0,1]$. Likewise, Equation (5) follows immediately from *(A3)* and by considering the case $c_{[K^*]} = c^{\hat{b}}_{[k^*_{\hat{b}}]} = 0$, $\forall K^* \in \{1,\dots,K\}, \forall k^*_{\hat{b}} \in \{1,\dots,k_{\hat{b}}\}$, $\hat{b} \in \{1,\dots,B\}$. □

Corollary 1 (Moments of number of incidents and losses per event)**.**

$$\mathbb{E}\big[|S_i|\,\big] = (1-p_G)\ K\ p_{gen} + p_G\ p_{sec} \sum_{\ell=1}^{B} p_\ell\ K_\ell,$$

$$\mathbb{E}\big[|S_1|^2\,\big] = (1-p_G)\ \big(K^2 p_{gen}^2 + K\ p_{gen}\ (1-p_{gen})\big) + p_G \sum_{\ell=1}^{B} p_\ell\ \big(K_\ell^2\ p_{sec}^2 + K_l\ p_{sec}\ (1-p_{sec})\big),$$

$$\mathbb{E}\big[|S_i^*|\,\big] = (1-p_G)\ p_{gen} \sum_{k^*=0}^{K} k^*\ (c_{[k^*+1]} - c_{[k^*]}) + p_G\ p_{sec} \sum_{\ell=1}^{B} \sum_{k_\ell^*=0}^{K_\ell} p_\ell\ k_\ell^*\ (c^\ell_{[k^*_\ell+1]} - c^\ell_{[k^*_\ell]}),$$

$$\begin{aligned}\mathbb{E}\big[|S_i^*|^2\,\big] = & \sum_{k^*=0}^{K} (1-p_G)\ (c_{[k^*+1]} - c_{[k^*]})\ \big((k^*)^2 p_{gen}^2 + k^* p_{gen}(1-p_{gen})\big) \\ & + \sum_{\ell=1}^{B} \sum_{k_\ell^*=0}^{K_\ell} p_G\ p_\ell\ \big(c^\ell_{[k^*_\ell+1]} - c^\ell_{[k^*_\ell]}\big)\ \big((k_\ell^*)^2 p_{sec}^2 + k_\ell^* p_{sec}(1-p_{sec})\big).\end{aligned}$$

Proof of Corollary 1. By Proposition 1, $|S_i|$ and $\left|S_i^*\right|$ follow a Binomial mixture distribution.
For $X_i \sim Binom(n,p)$, it holds of course that

$$\begin{aligned}\mathbb{E}[X_i] &= n\ p, \\ \mathbb{E}[X_i^2] &= n\ p\ (1-p) + n^2 p^2, \\ \mathbb{V}ar[X_i] &= n\ p\ (1-p).\end{aligned}$$

For a general mixture X of r.v. $\{X_i\}$ with weights $\{w_i\}$, means $\{\mu_i\}$, and variances $\{\sigma_i^2\}$, it holds that

$$\begin{aligned}\mathbb{E}[X] &= \sum_i w_i \mu_i, \\ \mathbb{E}[X^2] &= \sum_i w_i \mathbb{E}[X_i^2], \\ \mathbb{V}ar[X] &= \Big(\sum_i w_i(\mu_i^2 + \sigma_i^2)\Big) - \mu^2.\end{aligned}$$

The claims follow directly. □

Lemma 1 (Joint incident and loss probability)**.** *The probability for two firms* $j_1, j_2 \in \{1,\dots,K\}$ *(given their covariates) to register an incident / loss simultaneously from an event is given by*
Case 1: $b_{j_1} = b_{j_2}$ *(same industry sector)*

$$\begin{aligned}\mathbb{P}(j_1, j_2 \in S_i) &= p_{sec}^2\ p_{b_{j_1}}\ p_G + p_{gen}^2\ (1-p_G), \\ \mathbb{P}(j_1, j_2 \in S_i^*) &= \bar{F}_M\big(\max(c_{j_1}, c_{j_2})\big)\ \big(p_{sec}^2\ p_{b_{j_1}}\ p_G + p_{gen}^2\ (1-p_G)\big).\end{aligned}$$

Case 2: $b_{j_1} \neq b_{j_2}$ *(different industry sector)*

$$\begin{aligned}\mathbb{P}(j_1, j_2 \in S_i) &= p_{gen}^2\ (1-p_G), \\ \mathbb{P}(j_1, j_2 \in S_i^*) &= \bar{F}_M\big(\max(c_{j_1}, c_{j_2})\big)\ p_{gen}^2\ (1-p_G).\end{aligned}$$

Proof of Lemma 1. The statements follow immediately by conditioning and using conditional independence:
Case 1: $b_{j_1} = b_{j_2}$

$$\begin{aligned}
\mathbb{P}(j_1, j_2 \in S_i) &= \mathbb{P}(j_1, j_2 \in S_i \mid G_i = 1, B_i = b_{j_1})\, \mathbb{P}(B_i = b_{j_1} \mid G_i = 1)\, \mathbb{P}(G_i = 1) + \mathbb{P}(j_1, j_2 \in S_i \mid G_i = 0)\, \mathbb{P}(G_i = 0) \\
&= p_{sec}^2\, p_{b_{j_1}}\, p_G + p_{gen}^2\, (1 - p_G), \\
\mathbb{P}(j_1, j_2 \in S_i^*) &= \mathbb{P}\big(j_1, j_2 \in S_i^* \mid G_i = 1, m_i > \max(c_{j_1}, c_{j_2})\big)\, \mathbb{P}\big(G_i = 1 \mid m_i > \max(c_{j_1}, c_{j_2})\big)\, \mathbb{P}\big(m_i > \max(c_{j_1}, c_{j_2})\big) \\
&\quad + \mathbb{P}\big(j_1, j_2 \in S_i^* \mid G_i = 0, m_i > \max(c_{j_1}, c_{j_2})\big)\, \mathbb{P}\big(G_i = 0 \mid m_i > \max(c_{j_1}, c_{j_2})\big)\mathbb{P}\big(m_i > \max(c_{j_1}, c_{j_2})\big) \\
&= p_{sec}^2\, p_{b_{j_1}}\, p_G\, \bar{F}_M\big(\max(c_{j_1}, c_{j_2})\big) + p_{gen}^2\, (1 - p_G)\, \bar{F}_M\big(\max(c_{j_1}, c_{j_2})\big) \\
&= \bar{F}_M\big(\max(c_{j_1}, c_{j_2})\big)\, \big(p_{sec}^2\, p_{b_{j_1}}\, p_G + p_{gen}^2\, (1 - p_G)\big).
\end{aligned}$$

Case 2: $b_{j_1} \neq b_{j_2}$

$$\begin{aligned}
&\mathbb{P}(j_1, j_2 \in S_i) = \mathbb{P}(j_1, j_2 \in S_i \mid G_i = 1)\, \mathbb{P}(G_i = 1) + \mathbb{P}(j_1, j_2 \in S_i \mid G_i = 0)\, \mathbb{P}(G_i = 0) = p_{gen}^2\, (1 - p_G), \\
&\mathbb{P}(j_1, j_2 \in S_i^*) = \bar{F}_M\, \big(\max(c_{j_1}, c_{j_2})\big)\, p_{gen}^2\, (1 - p_G).
\end{aligned}$$

□

Proof of Proposition 2. It follows immediately using Lemma 1

$$\mathbb{P}(j_1 \in S_i \mid j_2 \in S_i) = \frac{\mathbb{P}(j_1, j_2 \in S_i)}{\mathbb{P}(j_2 \in S_i)} = \begin{cases} \frac{p_{sec}^2\, p_{b_{j_2}}\, p_G + p_{gen}^2\, (1-p_G)}{\tilde{p}(b_{j_2})}, & b_{j_1} = b_{j_2}, \\ \frac{p_{gen}^2\, (1-p_G)}{\tilde{p}(b_{j_2})}, & b_{j_1} \neq b_{j_2}, \end{cases}$$

$$\mathbb{P}(j_1 \in S_i^* \mid j_2 \in S_i^*) = \frac{\mathbb{P}(j_1, j_2 \in S_i^*)}{\mathbb{P}(j_2 \in S_i^*)} = \begin{cases} \frac{p_{sec}^2\, p_{b_{j_2}}\, p_G + p_{gen}^2\, (1-p_G)}{\tilde{p}(b_{j_2})} & b_{j_1} = b_{j_2},\ c_{j_1} \leq c_{j_2}, \\ \frac{\bar{F}_M(c_{j_1})}{\bar{F}_M(c_{j_2})} \Big(\frac{p_{sec}^2 \cdot p_{b_{j_2}}\, p_G + p_{gen}^2\, (1-p_G)}{\tilde{p}(b_{j_2})}\Big) & b_{j_1} = b_{j_2},\ c_{j_1} > c_{j_2}, \\ \frac{p_{gen}^2\, (1-p_G)}{\tilde{p}(b_{j_2})} & b_{j_1} \neq b_{j_2},\ c_{j_1} \leq c_{j_2}, \\ \frac{\bar{F}_M(c_{j_1})}{\bar{F}_M(c_{j_2})} \Big(\frac{p_{gen}^2\, (1-p_G)}{\tilde{p}(b_{j_2})}\Big) & b_{j_1} \neq b_{j_2},\ c_{j_1} > c_{j_2}. \end{cases}$$

□

Proof of remark about conditional vs. unconditional probabilities. We have remarked that for firms of the same industry sector, the knowledge about an incident for a firm in the same sector always has a non-negative effect on the incident probability, i.e. for $j_1, j_2 \in \{1, \ldots, K\}$ with $b_{j_1} = b_{j_2} =: b_j$

$$\begin{aligned}
&\mathbb{P}(j_1 \in S_i \mid j_2 \in S_i) \geq \mathbb{P}(j_1 \in S_i \mid b_{j_1}) \\
\overset{\text{Prop.2}}{\iff}\ & p_{sec}^2\, p_{b_j}\, p_G + p_{gen}^2\, (1 - p_G) \geq \big(\tilde{p}(b_j)\big)^2 \\
\iff\ & p_{sec}^2\, p_{b_j}\, p_G + p_{gen}^2\, (1 - p_G) \geq \big(p_G\, p_{b_j}\, p_{sec} + (1 - p_G)\, p_{gen}\big)^2. \qquad (14)
\end{aligned}$$

Generally, for any $\mathbf{x} = (x_1, \ldots, x_n)', \mathbf{y} = (y_1, \ldots, y_n)' \in \mathbf{R}^n$ $(n \in \mathbb{N})$, the Cauchy–Schwarz inequality states that

$$\Big(\sum_{i=1}^n x_i\, y_i\Big)^2 \leq \Big(\sum_{i=1}^n x_i^2\Big)\Big(\sum_{i=1}^n y_i^2\Big).$$

Let $\mathbf{a} = (a_1, \ldots, a_n)' \in (0, \infty)^n, \mathbf{b} = (b_1, \ldots, b_n)' \in \mathbf{R}^n$ $(n \in \mathbb{N})$, and assume $\sum_{i=1}^n a_i \leq 1$. Substituting above $\mathbf{x} = \sqrt{\mathbf{a}}\, \mathbf{b}, \mathbf{y} = \sqrt{\mathbf{a}}$ yields

$$\Big(\sum_{i=1}^n a_i\, b_i\Big)^2 \leq \Big(\sum_{i=1}^n a_i\, b_i^2\Big)\, \underbrace{\Big(\sum_{i=1}^n a_i\Big)}_{\leq 1} \leq \sum_{i=1}^n a_i\, b_i^2.$$

Substituting for $n = 2$

$$\begin{aligned}
\mathbf{a} &= (a_1, a_2)' = (p_G\, p_{b_j}, (1 - p_G))', \\
\mathbf{b} &= (b_1, b_2)' = (p_{sec}, p_{gen})',
\end{aligned}$$

yields (14). □

Lemma 2 (Moments of cumulative incident and loss numbers)**.** *It holds that*

$$\mathbb{E}\big[\bar{N}^{\cdot,syst}(T)\big] = \Lambda^{\cdot,g}(T)\, \Big((1-p_G)\; K\; p_{gen} + p_G\; p_{sec}\; \sum_{\ell=1}^{B} p_\ell\; K_\ell\Big),$$

$$\mathbb{V}ar\big[\bar{N}^{\cdot,syst}(T)\big] = \Lambda^{\cdot,g}(T)\, \Big((1-p_G)\; \big(K^2 p_{gen}^2 + K\; p_{gen}\; (1-p_{gen})\big) + p_G \sum_{\ell=1}^{B} p_\ell\; \big(K_\ell^2\; p_{sec}^2 + K_\ell\; p_{sec}\; (1-p_{sec})\big)\Big),$$

$$\mathbb{E}\big[N^{\cdot,syst}(T)\big] = \Lambda^{\cdot,g}(T)\, \Big((1-p_G)\; p_{gen} \sum_{k^*=0}^{K} k^*\; (c_{[k^*+1]} - c_{[k^*]}) + p_G\; p_{sec} \sum_{\ell=1}^{B} \sum_{k_\ell^*=0}^{K_\ell} p_\ell\; k_\ell^*\; (c^\ell_{[k_\ell^*+1]} - c^\ell_{[k_\ell^*]})\Big),$$

$$\begin{aligned}\mathbb{V}ar\big[N^{\cdot,syst}(T)\big] = \Lambda^{\cdot,g}(T)\, \Big(&\sum_{k^*=0}^{K} (1-p_G)\; (c_{[k^*+1]} - c_{[k^*]})\; ((k^*)^2 p_{gen}^2 + k^* p_{gen}(1-p_{gen})) \\ &+ \sum_{\ell=1}^{B} \sum_{k_\ell^*=0}^{K_\ell} p_G\; p_\ell\; (c^\ell_{[k_\ell^*+1]} - c^\ell_{[k_\ell^*]})\; ((k^*)^2 p_{sec}^2 + k^* p_{sec}(1-p_{sec}))\Big).\end{aligned}$$

Proof of Lemma 2. For the number of arrivals of the ground process on any interval $[0,T]$, it holds that

$$\mathbb{E}\big[N^{\cdot,g}(T)\big] = \mathbb{V}ar\big[N^{\cdot,g}(T)\big] = \Lambda^{\cdot,g}(T) = \int_0^T \lambda^{\cdot,g}(t)\mathrm{d}t.$$

By Wald's equation and the law of total variance (see Appendix A.2), it follows from Corollary 1

$$\begin{aligned}&\mathbb{E}\big[\bar{N}^{\cdot,syst}(T)\big] = \mathbb{E}\big[N^{\cdot,g}(T)\big]\; \mathbb{E}\big[|S_i|\big] = \Lambda^{\cdot,g}(T)\, \Big((1-p_G)\; K\; p_{gen} + p_G\; p_{sec} \sum_{\ell=1}^{B} p_\ell\; K_\ell\Big), \\ &\mathbb{V}ar\big[\bar{N}^{\cdot,syst}(T)\big] = \mathbb{E}\big[N^{\cdot,g}(T)\big]\; \mathbb{V}ar\big[|S_i|\big] + \mathbb{E}\big[|S_i|\big]^2\; \mathbb{V}ar[N^{\cdot,g}(T)] = \mathbb{E}\big[N^{\cdot,g}(T)\big]\; \mathbb{E}\big[|S_i|^2\big] \\ &= \Lambda^{\cdot,g}(T)\, \Big((1-p_G)\; \big(K^2 p_{gen}^2 + K\; p_{gen}\; (1-p_{gen})\big) + p_G \sum_{\ell=1}^{B} p_\ell\; \big(K_\ell^2\; p_{sec}^2 + K_\ell\; p_{sec}\; (1-p_{sec})\big)\Big).\end{aligned}$$

Likewise,

$$\begin{aligned}\mathbb{E}\big[N^{\cdot,syst}(T)\big] &= \mathbb{E}\big[N^{\cdot,g}(T)\big]\; \mathbb{E}\big[|S_i^*|\big] \\ &= \Lambda^{\cdot,g}(T)\, \Big((1-p_G)\; p_{gen} \sum_{k^*=0}^{K} k^*\; (c_{[k^*+1]} - c_{[k^*]}) + p_G\; p_{sec} \sum_{\ell=1}^{B} \sum_{k_\ell^*=0}^{K_\ell} p_\ell\; k_\ell^*\; (c^\ell_{[k_\ell^*+1]} - c^\ell_{[k_\ell^*]})\Big),\end{aligned}$$

$$\begin{aligned}\mathbb{V}ar\big[N^{\cdot,syst}(T)\big] &= \mathbb{E}\big[N^{\cdot,g}(T)\big]\; \mathbb{E}\big[|S_1^*|^2\big] \\ &= \Lambda^{\cdot,g}(T)\, \Big(\sum_{k^*=0}^{K} (1-p_G)\; (c_{[k^*+1]} - c_{[k^*]})\; ((k^*)^2 p_{gen}^2 + k^* p_{gen}(1-p_{gen})) \\ &+ \sum_{\ell=1}^{B} \sum_{k_\ell^*=0}^{K_\ell} p_G\; p_\ell\; (c^\ell_{[k_\ell^*+1]} - c^\ell_{[k_\ell^*]})\; ((k^*)^2 p_{sec}^2 + k^* p_{sec}(1-p_{sec}))\Big).\end{aligned}$$

□

Proof of Proposition 3. It follows immediately from Lemma 2 that

$$\begin{aligned}DI(\bar{N}^{\cdot,syst}(T)) &= \frac{\mathbb{V}ar\big[\bar{N}^{\cdot,syst}(T)\big]}{\mathbb{E}\big[\bar{N}^{\cdot,syst}(T)\big]} = \frac{\mathbb{E}\big[|S_i|^2\big]}{\mathbb{E}\big[|S_i|\big]} \\ &= \frac{(1-p_G)\; \big(K^2 p_{gen}^2 + K\; p_{gen}\; (1-p_{gen})\big) + p_G \sum_{\ell=1}^{B} p_\ell\; (K_\ell^2\; p_{sec}^2 + K_\ell\; p_{sec}\; (1-p_{sec}))}{(1-p_G)\; K\; p_{gen} + p_G\; p_{sec} \sum_{\ell=1}^{B} p_\ell\; K_\ell} \\ &= 1 + \frac{(1-p_G)\; p_{gen}^2\; (K^2-K) + p_G\; p_{sec}^2 \sum_{\ell=1}^{B} p_\ell (K_\ell^2 - K_\ell)}{(1-p_G)\; K\; p_{gen} + p_G\; p_{sec} \sum_{\ell=1}^{B} p_\ell\; K_\ell} > 1,\end{aligned}$$

and likewise

$$
\begin{aligned}
DI\big(N^{\cdot,syst}(T)\big) &= \frac{\mathbb{V}ar\big[N^{\cdot,syst}(T)\big]}{\mathbb{E}\big[N^{\cdot,syst}(T)\big]} = \frac{\mathbb{E}\big[|S_i^*|^2\big]}{\mathbb{E}\big[|S_i^*|\big]} \\
&= \frac{\sum_{k^*=0}^{K}(1-p_G)\,(c_{[k^*+1]}-c_{[k^*]})\,((k^*)^2p_{gen}^2+k^*p_{gen}(1-p_{gen}))}{(1-p_G)\,p_{gen}\sum_{k^*=0}^{K}k^*\,(c_{[k^*+1]}-c_{[k^*]})+p_G\,p_{sec}\sum_{\ell=1}^{B}\sum_{k_\ell^*=0}^{K_\ell}p_\ell\,k_\ell^*\,(c^\ell_{[k_\ell^*+1]}-c^\ell_{[k_\ell^*]})} \\
&\quad+\frac{\sum_{\ell=1}^{B}\sum_{k_\ell^*=0}^{K_\ell}p_G\,p_\ell\,(c^\ell_{[k_\ell^*+1]}-c^\ell_{[k_\ell^*]})\,((k_\ell^*)^2p_{sec}^2+k_\ell^*p_{sec}(1-p_{sec}))}{(1-p_G)\,p_{gen}\sum_{k^*=0}^{K}k^*\,(c_{[k^*+1]}-c_{[k^*]})+p_G\,p_{sec}\sum_{\ell=1}^{B}\sum_{k_\ell^*=0}^{K_\ell}p_\ell\,k_\ell^*\,(c^\ell_{[k_\ell^*+1]}-c^\ell_{[k_\ell^*]})} \\
&= 1+\frac{(1-p_G)\sum_{k^*=0}^{K}p_{gen}^2((k^*)^2-k^*)(c_{[k^*+1]}-c_{[k^*]})+p_G\sum_{\ell=1}^{B}\sum_{k_\ell^*=0}^{K_\ell}p_{sec}^2p_\ell((k_\ell^*)^2-k_\ell^*)(c^\ell_{[k_\ell^*+1]}-c^\ell_{[k_\ell^*]})}{(1-p_G)p_{gen}\sum_{k^*=0}^{K}k^*(c_{[k^*+1]}-c_{[k^*]})+p_Gp_{sec}\sum_{\ell=1}^{B}\sum_{k_\ell^*=0}^{K_\ell}p_\ell k_\ell^*(c^\ell_{[k_\ell^*+1]}-c^\ell_{[k_\ell^*]})} > 1.
\end{aligned}
$$

The fractions in the last lines of both equations are obviously non-negative (positive, under the additional assumptions in Proposition 3), as they contain only sums and products of non-negative (positive) quantities. □

A.4 Alternative Severity Distributions

DB: Use link between number of records and cost

While it is difficult to find reliable empirical data about the cost of a data breach, some data about the number of breached / stolen records is available. Thus, several authors have tried to find a link between the number of records and the cost of a data breach. The often cited *Jacob's formula* ([47]) suggests to link the log-transformed cost L of a data breach to the number of compromised records D according to

$$\log(L) = 7.68 + 0.7584\,\log(D). \tag{15}$$

An amendment to this formula was proposed in [41], who argue that [47] did not yet take into account the cost of *mega data breaches* observed in future years and thus alternatively propose

$$\log(L) = -1.998 + 7.503\,\log\big(\log(D)\big). \tag{16}$$

Therefore, an alternative to modelling the cost of a data breach directly using a combination of log-normal and GPD would be to first model the number of breached records using a log-normal (as suggested by the results in [32]) and then convert the number of records into monetary losses using (15) or (16).

In the context of this work, let D_{ij} be the number of lost / stolen records in a DB incident at time t_i affecting firm j (where $\{t_i\}_{i\in\mathbb{N}}$ only counts the event times at firm j). Then assume

$$
\begin{aligned}
D_{ij} &\sim LN\big(\mu_j^{DB}(t_i),\sigma_j^{DB}(t_i)\big), \qquad (17)\\
\mu_j^{DB}(t_i) &= \alpha_{\mu,DB}+f_{\mu,DB,3}(x_{j3})+f_{\mu,DB,4}(x_{j4})+g_{\mu,DB}(t_i),\\
\sigma_j^{DB}(t_i) &\equiv \sigma^{DB},
\end{aligned}
$$

where the functions $f_{\mu,DB,\cdot}$ and $g_{\mu,DB}$ are as usual. By (15), the number of records D_{ij} is converted into the cost of the breach L_{ij} according to

$$\log(L_{ij}) = 7.68 + 0.7584\,\log(D_{ij}),$$

which is equivalent to directly assuming that

$$
\begin{aligned}
L_{ij} &\sim LN(\hat{\mu}_j^{DB}(t_i),\hat{\sigma}^{DB}),\\
\hat{\mu}_j^{DB}(t_i) &= \alpha_{\hat{\mu},DB}+f_{\hat{\mu},DB,3}(x_{j3})+f_{\hat{\mu},DB,4}(x_{j4})+g_{\hat{\mu},DB}(t_i),\\
\hat{\sigma}_j^{DB}(t_i) &\equiv \hat{\sigma}^{DB}
\end{aligned}
$$

Likewise, using (16) to convert the number of records into the cost of the breach, i.e. assume D_{ij} to be distributed according to (17) and the breach cost L_{ij} then to be given by

$$\log(L_{ij}) = -1.998 + 7.503\log\big(\log(D_{ij})\big).$$

BI: Replace log-normal by PERT

Regarding the economic impact of BI incidents, some sources from the non-cyber domain are available ([21, 30, 44, 48, 85]). The only sources including indications of which distributions are useful to model economic loss from BI are [85], who finds the size of yearly BI insurance claims to follow a Pareto distribution with an extremely heavy tail and infinite expected claim size, and [44], who suggests modelling BI loss by a *PERT* distribution, a special case of the beta distribution with the three parameters minimum x_{min}, mode x_{mode}, and maximum x_{max} with density

$$\begin{aligned} f_{PERT}(x) &= \frac{(x - x_{min})^{v-1}(x_{max} - x)^{w-1}}{Beta(v,w)(x_{max} - x_{min})^{v+w-1}} 1_{[x_{min}, x_{max}]}, \\ v &= 1 + \gamma_P \Big(\frac{x_{mode} - x_{min}}{x_{max} - x_{min}} \Big), \\ w &= 1 + \gamma_P \Big(\frac{x_{max} - x_{mode}}{x_{max} - x_{min}} \Big), \end{aligned}$$

where $Beta(\cdot)$ is the Beta function and for the standard PERT $\gamma_P = 4$.
Thus, for BI incidents, one could suggest replacing the log-normal distribution for the body by a PERT distribution, i.e. assume for a BI loss L_{ij} at time t_i affecting firm j it holds

$$\begin{aligned} (L_{ij} \mid L_{ij} \leq u_{ij}^{BI}) &\sim PERT(x_{ij}^{\min}, x_{ij}^{mode}, x_{ij}^{\max}, 4), \\ x_{ij}^{\min} &= 0, \\ x_{ij}^{\max} &= u_{ij}^{BI}, \\ x_{ij}^{\text{mode}} &= \exp\big(\mu_j^{BI}(t_i) - \sigma_j^{BI}(t_i)^2\big), \end{aligned}$$

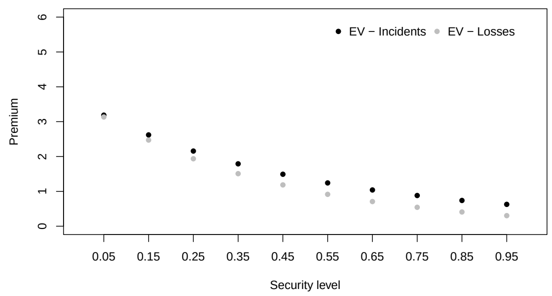
where $PERT(x^{\min}, x^{\text{mode}}, x^{\max}, 4)$ denotes the *PERT* distribution with minimum, mode, and maximum values $x^{\min}, x^{\text{mode}}, x^{\max}$ respectively and standard shape parameter $\gamma_p = 4$. The mode and threshold (maximum) are chosen such that they coincide with the ones from the underlying log-normal used to find the threshold between body and tail of the loss distribution.

A.5 Comparison of all Premium Calculation Results

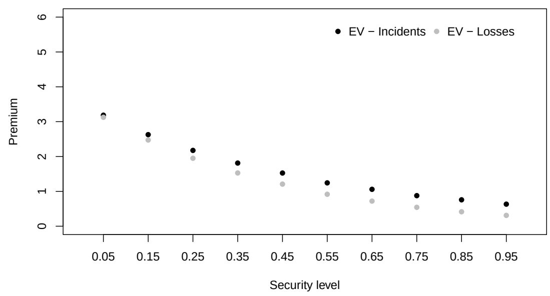
Below, we compare the premiums (for three individual firms and all sub-portfolios) obtained from the simulation with dependent losses, with independent losses, with cover limit, and with the premiums obtained from calculating the (discretized) loss distribution pdf. using Panjer recursion. We observe that they are very similar in all cases; as for the latter two cases (simulation with cover limit and Panjer recursion) loss severities are truncated from above, the application of premium principles that depend on more than just the first moment are feasible and the results for the expected value principle are slightly lower.

Premium Principle	Expected Value ($\rho = 0.2$)	Exponential ($\gamma = 10^{-3}$)	Standard Deviation ($\rho = 0.2$)
Premium based on dependent simulated losses (incidents)			
Firm 1	2.0814 (2.2338)	–	–
Firm 2	0.4451 (0.7746)	–	–
Firm 3	1.1732 (1.5164)	–	–
Premium based on independent simulated losses (incidents)			
Firm 1	2.1213 (2.2514)	–	–
Firm 2	0.4726 (1.3395)	–	–
Firm 3	1.2149 (1.5353)	–	–
Premium based on simulated losses (incidents) with cover limit			
Firm 1	2.1592 (2.3051)	1.8993 (2.0258)	4.5101 (4.7022)
Firm 2	0.4385 (0.7783)	0.3717 (0.6608)	1.0745 (1.6300)
Firm 3	1.1620 (1.5115)	0.9960 (1.2956)	2.4413 (2.9404)
Premium based on Panjer recursion			
Firm 1	1.7633 (1.8849)	2.1160 (2.2619)	1.8366 (1.9632)
Firm 2	0.3797 (0.662)	0.4557 (0.7944)	0.3861 (0.6731)
Firm 3	0.9605 (1.2643)	1.1526 (1.5171)	0.9874 (1.2997)

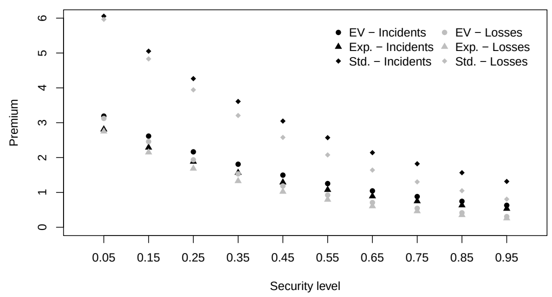
Table 10: Comparison of one-year cyber insurance premiums for three selected firms, based on 50.000 simulation runs (upper panels) and Panjer recursion for the given assumptions and parameter values (lower panel). Numbers in brackets indicate what the premium would have been if all incoming incidents had been counted.



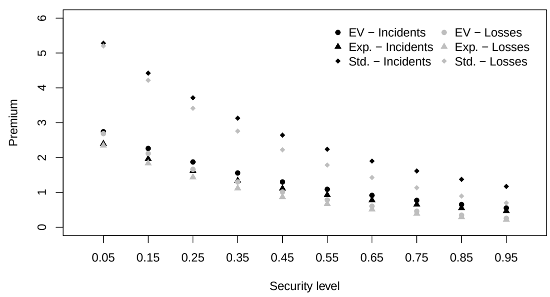
(a) Results from simulation (dependent).



(b) Results from simulation (independent).



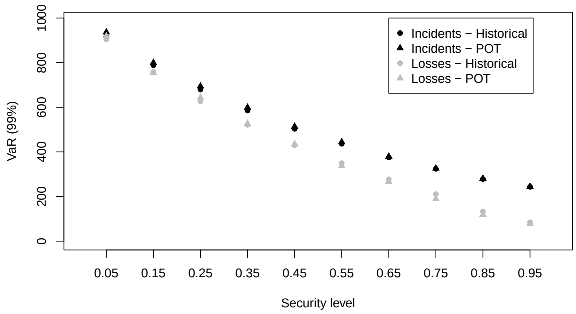
(c) Results from simulation (with cover limit).



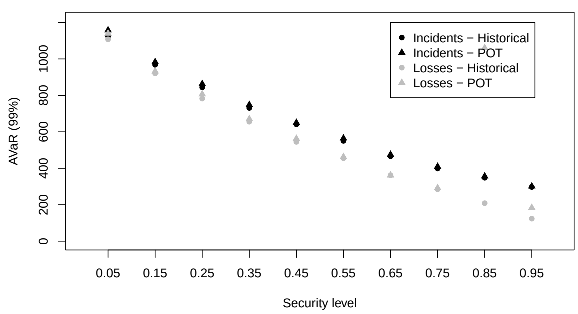
(d) Results from Panjer recursion.

Figure 8: We compare the premium that would be assigned to firms if they were grouped according to their IT security level, based on the three simulation studies and the implemented Panjer recursion scheme. Note that the results in Figure 8d are for firms with the given security level and otherwise baseline covariate levels, so should be expected slightly below results from simulations of the sub-portfolios with mixed covariate levels.

A.6 Risk Measures for Simulation with Cover Limit



(a) $VaR_{0.99}$; Cover Limit $\bar{M}_2$.



(b) $AVaR_{0.99}$; Cover Limit $\bar{M}_2$.

Figure 9: Comparison of $VaR_{0.99}$ and $AVaR_{0.99}$ for all sub-portfolios.

A.7 Details on Covariate Levels

Revenue	*Number of employees* small	medium	large
small	1	1	(2)
medium	1	2	3
large	(2)	3	3

Table 11: Factor levels of s by combinations of revenue and number of employees. This is in line with the classification of *SMEs* in the European Union. As revenue and number of employees are highly correlated, only very few companies should fall into the classifications in the upper right or lower left cell.

Sensitive Data	*Number of stored records* $\leq$ threshold	$>$ threshold
No	1	2
Yes	2	3

Table 12: Factor levels of d, given the number of stored records and sensitivity of data. Sensitive data includes, e.g. *Personally Identifiable Information (PII)*, *Protected Health Information (PHI)*, or classified government data. Despite the labels, this is not a numerical attribute and it is not clear whether the two cases labeled 2 (medium risk) are comparable, or, if considered not comparable, how they should be ordered.

Sector b	*Number of employees e* small	medium	large
HC, EDU, GOV	1	1	2
FI, BR, MAN	1	2	3

Table 13: Factor levels of *nsup* by combinations of sector and number of employees. The classification relies on expert judgment and is not founded by empirical evidence. An insurance company might simply obtain this information from its customers.