



CONTRIBUTED PAPER

Loss modelling from first principles

Pietro Parodi^{1,2}, Derek Thrumble³, Peter Watson¹, Zhongmei Ji¹, Alex Wang⁴, Ishita Bhatia⁵, Joseph Lees⁶, Sophia Mealy⁷, Rushabh Shah⁸, Param Dharamshi⁹ and Federica Gazzelloni¹⁰

¹SCOR Property & Casualty, London, UK; ²Bayes Business School, City University of London, London, UK; ³Adesco, London, UK; ⁴Gen Re, P&C Treaty, Paris, France; ⁵QBE, London, UK; ⁶Protector Forsikring, Manchester, UK; ⁷Hiscox Re, Bermuda; ⁸Marsh Advisory, London, UK; ⁹KPMG, London, UK and ¹⁰Freelance, Rome, Italy

Corresponding author: Pietro Parodi; Email: pparodi@scor.com

Abstract

A common statistical modelling paradigm used in actuarial pricing is (a) assuming that the possible loss model can be chosen from a dictionary of standard models; (b) selecting the model that provides the best trade-off between goodness of fit and complexity. Machine learning provides a rigorous framework for this selection/validation process. An alternative modelling paradigm, common in the sciences, is to prove the adequacy of a statistical model from first principles: for example, Planck's distribution, which describes the spectral distribution of blackbody radiation empirically, was explained by Einstein by assuming that radiation is made of quantised harmonic oscillators (photons). In this working party we have been exploring the extent to which loss models, too, can be derived from first principles. Traditionally, the Poisson, negative binomial, and binomial distributions are used as loss count models because they are familiar and easy to work with. We show how reasoning from first principles naturally leads to non-stationary Poisson processes, Lévy processes, and multivariate Bernoulli processes depending on the context. For modelling severities, we build on previous research that shows how graph theory can be used to model property-like losses. We show how the methodology can be extended to deal with business interruption/supply chain risks by considering networks with higher-order dependencies. For liability business, we show the theoretical and practical limitations of traditional models such as the lognormal distribution. We explore the question of where the ubiquitous power-law behaviour comes from, finding a natural explanation in random growth models. We also address the derivation of severity curves in territories where compensation tables are used. This research is foundational in nature, but its results may prove useful to practitioners by guiding model selection and elucidating the relationship between the features of a risk and the model's parameters.

Keywords: Random growth models; graph theory; Lévy process; Baremo tables; contingent business interruption; cyber business interruption.

1. Introduction

There is a story about Enrico Fermi that Freeman Dyson, a mathematical physicist who played a crucial role in the development of quantum electrodynamics, recounts during an interview¹ and that we think is the ideal introduction to what the *Loss modelling from first principles* working party is attempting to do.

Dyson and his team at Cornell had produced the first draft of a theory that aimed at explaining Fermi's experimental results on the behaviour of pions, and Dyson went on an expedition to Chicago to share these results with Fermi and seek encouragement to continue the team's work. After listening to Dyson and having a quick glance at the graph that displayed an amazing fit between the theory and Fermi's results, Fermi told Dyson "I'm not very impressed with what

¹<https://www.youtube.com/watch?edufilter=NULL&v=hV41QEKiMIM>.

you've been doing." He continued: "When one does a theoretical calculation, there are two ways of doing it: either you should have a clear physical model in mind, or you should have a rigorous mathematical basis. You have neither." When Dyson pressed him about the excellent agreement between the model and the data, Fermi asked: "How many free parameters are there in your method?" It turned out to be four. Fermi commented: "My friend von Neumann always used to say: 'with four parameters I can fit an elephant, and with five I can make him wiggle his trunk'. So, I don't find the numerical agreement very impressive either."²

This story illustrates the important point that there are essentially two ways of producing a model.

1. Start from an intuition about how the world works and/or using a rigorous mathematical process to derive the model from first principles, and then test the explanatory power of the model against reality (i.e., experimental data).
2. Start with any model – ignoring how it was derived – and check the significance of the fit using statistical tests, which basically puts von Neumann's comment about elephant fitting on a quantitative footing.

Both approaches, if done with a proper methodological framework, are viable. Actuarial practitioners have historically – and for the most part – followed the second approach, and not always in the most rigorous way. At one end of the spectrum of rigour, actuaries may just pick a simple distribution such as the lognormal distribution for loss amounts or use the distribution from a distribution-fitting tool that best fits the observed data, without regard to the number of parameters. At the other end of the spectrum, actuaries might use rigorous machine learning methodologies such as using training, selection and validation sets and be mindful of the bias-variance trade-off when doing rating factor selection and calibration (see Parodi (2023) for a discussion of these different approaches). The increased interest in machine learning in recent years has helped actuaries to use the proper model selection techniques.

This working party aims to investigate the extent to which it is possible to produce loss models according to the first approach, starting from our intuition and knowledge around the loss generation process and then developing a mathematically consistent model. The idea is that a model that fairly reflects the loss generation process will have a natural advantage and is likely to have fewer parameters. For example, an oscillating function might be helpfully modelled in a certain time interval with a polynomial or a set of splines, but if we have theoretical reasons to believe that an actual sinusoid of the form $x(t) = A \sin(\omega t + \phi)$ should be used, this will lead to a superior and more economic model.

The idea that a model should draw inspiration from how the world works is not, of course, a new concept in actuarial practice. Risk theory (Beard *et al.*, 1984), the theory that gave us the individual risk model (IRM) and the collective risk model, successfully framed the loss-generating process in terms of count (frequency) models and severity models and natural catastrophe modelling. It relied in part on historical loss experience and is heavily informed by a scientific understanding of the modelled perils. The technical actuarial standards for modelling (TAS-M) produced by the UK Institute and Faculty of Actuaries have this to say about models representing the real world:

[A]ctuarial information should be based on models that sufficiently represent those aspects of the real world that are relevant to the decisions for which the actuarial information will be used. This is, deliberately, a fairly general statement.

²Dyson wasn't bitter about this conversation, which marked the end of the project. Rather, he was grateful because it prevented his team from wasting further time working on this theory – which, as Fermi had predicted, turned out to be incorrect. The experimental results were later explained by Murray Gell-Mann's theory, according to which the physics of pions could be explained by assuming that pions are made of a quark and an anti-quark.

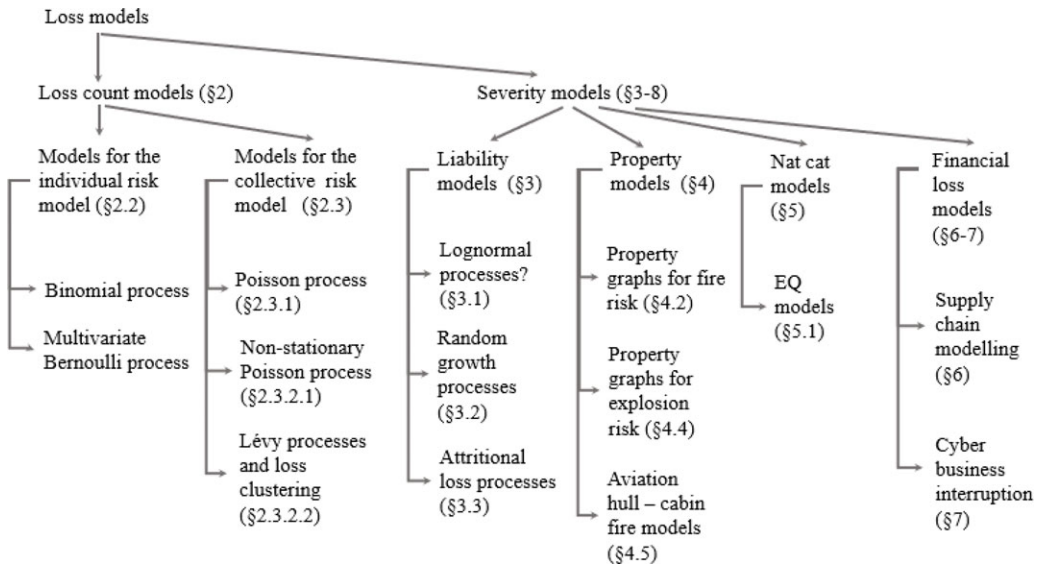


Figure 1. A conceptual view of the structure of this paper.

Despite this, in most cases “representing the relevant aspects of the world” is taken to mean “capturing the correct factors that affect a risk,” but there is a dearth of results in actuarial science that dictate the use of specific statistical models³.

1.1 The Structure of the Paper

The structure of this paper is summarised in Figure 1. Note that while we could have looked at any number of other models related to losses (e.g., payment pattern models), we have focused our attention on frequency and severity models, while noting that this separation is not always possible (Section 2.3.2.3). Section 2 is devoted to loss count (frequency) models, while Sections 3 to 8 are devoted to severity models for various types of business (property, liability, financial loss). These core sections are then followed by general conclusions (Section 8), acknowledgements (Section 9) and the bibliography (Section 10).

2. Modelling Loss Counts

2.1 The Traditional Approach

The traditional actuarial approach to modelling loss counts is to use one of the binomial, Poisson, and negative binomial distributions depending on how large the variance of the loss counts is believed to be relative to the mean. This choice is not typically made for fundamental reasons but because these distributions are familiar and easy to work with. It’s also difficult to argue for more complex distributions when there is such a low number of data points (typically the loss counts for 5–15 years) for calibration.

It is useful to distinguish between frequency models used in the IRM and those used in the collective risk model (Klugman et al., 2012) – two attempts to produce risk models that reflect the reality of the loss generating process under different circumstances.

³One notable exception is the use of the generalised Pareto distribution (GPD) for modelling the tail distribution of severities, which is demanded (asymptotically) by extreme value theory.

2.2 Loss Count Models for the Individual Risk Model

In the IRM, losses are assumed to come from a finite number of risks, each of which can have at most one loss. Although we are dealing here with non-life insurance, this is perhaps best understood in terms of insuring lives. In equine insurance, a payout is given if a horse in the establishment dies or needs to be put down – which can only happen once to the same horse. The modelling of defaults in credit risk follows a similar pattern, where one replaces “death” with “bankruptcy.”

If all probabilities of death/default are the same, and the losses are independent of one another, the number of losses can be modelled with a binomial distribution – the sum of independent Bernoulli variables. According to this distribution, the probability of having k losses from n risks in a given period is:

$$\Pr(N = k) = \binom{n}{k} p^k (1 - p)^{n-k} \tag{1}$$

The mean is np , the variance is $np(1 - p)$, and the variance to mean ratio is $1 - p$.

When the probabilities are different, this generalises to the Multivariate Bernoulli Distribution (a.k.a. the Poisson binomial distribution), which is (Parodi, 2023) an n -dimensional distribution with random variable $\vec{Y} = (Y_1, Y_2 \dots Y_n)$, where each component Y_j can take the values 0 (no loss) or 1 (loss). The probability of having k claims from n risks is given by:

$$\Pr(N = k) = \sum_{y_1, \dots, y_n \text{ such that } y_1 + \dots + y_n = k} \prod_{j=1}^n p_j^{y_j} (1 - p_j)^{1-y_j} \tag{2}$$

2.2.1 Modelling the dependency between losses

Equations (1) and (2) are valid when the losses are independent of one another. When that is not the case, the variance/mean ratio will increase and can even exceed 1. Rather than adopting a different distribution in that case, it is generally more productive to model the dependency.

Unless the mechanism by which correlations occur is clear, the easiest way to model dependencies is by assuming the presence of a systemic shock – an increase in the probability of loss that applies simultaneously to all risks. This increase will normally be temporary: an epidemic among horses or an economic recession for defaults, for example.

Using this framework, the underlying model can still be seen as a binomial or a multivariate Bernoulli, but with the probabilities of loss themselves being random variates.

2.3 Loss Count Models for the Collective Risk Model

In the case of the collective risk model, the underlying portfolio of risks is also finite, but more than one loss is possible for each risk, and the effect of risks being removed from the portfolio by a total loss (think property insurance) is small if the portfolio is large enough. Consequently, the occurrence of one loss does not significantly affect the likelihood of another loss, and the variance of the loss count is at least that of a Poisson distribution. The Poisson distribution is also justified as the limit of a binomial distribution when the probability p of a risk being affected by the loss is very small, the number of risks n is very large, and the product np converges to a finite (and non-zero) number.

Before proceeding further, let us introduce some key concepts. A **count process** $N(t)$ with $t \geq 0$ is a stochastic process where $N(t)$ represents the cumulative number of events happening between 0 and t (more specifically, in the interval $[0, t]$). The number of events in the interval $(t, t']$ is given by $N(t') - N(t)$ and is called an **increment**. A process is said to have **independent increments** if the numbers of events occurring in two non-overlapping intervals are independent.

A process is said to be **stationary** if the probability of a given number of events in a given interval $(t_0, t_0 + \tau]$ does not depend on t_0 but only on τ : assuming $N(0) = 0$, we have equality in distribution $N(t_0 + \tau) - N(t_0) = N(\tau)$.

2.3.1 The stationary Poisson process as the foundational count process for the collective risk model

The simplest count process used in the context of the collective risk model is arguably the stationary Poisson process (Ross, 2003). This is a count process with the following properties:

- (a) $N(0) = 0$
- (b) The increments are both independent and stationary
- (c) $\Pr(N(\epsilon) = 1) = \lambda\epsilon + o(\epsilon)$
- (d) $\Pr(N(\epsilon) = 2) = o(\epsilon)$

Below are some consequences of these assumptions:

- The number of events in an interval of length τ , $N(\tau)$, follows a Poisson distribution: $\Pr(N(\tau) = k) = \exp(-\lambda\tau)(\lambda\tau)^k/k!$, where λ is the event rate per unit of time.
- The expected number of events and the variance of the number of events for an interval of length τ are the same and are given by:

$$\mathbb{E}(N(\tau)) = \mathbb{V}(N(\tau)) = \lambda\tau$$

- The possibility of simultaneous events is excluded.

In practice, we notice that the assumptions that allow for the Poisson distribution to be used in the collective risk model often break down. This happens for example (note that the points below overlap to some extent):

1. Where events do not occur independently, whether via direct dependence or common shock
 - Fire – insurer covers a geographically close-knit community and, for example, an explosion in one flat may cause damage in all surrounding flats
 - Professional lines – higher PI and D&O activity following a recession
 - Healthcare – a disease is brought to a community, and increases the risk of everyone else getting the disease
 - In riot/terrorism losses following a larger political/social event
 - In cyber losses following a weakness emerging from an operating system.
2. Where the future average rate may be altered by future events/decisions:
 - Casualty – a new court award setting a precedent for future similar claims.
3. Where multiple events occur in clusters:
 - Catastrophic events (natural catastrophes and man-made catastrophes such as terrorism) often result in several risks being hit at the same time, for example, hail events leading to many property/agriculture losses
 - External factors driving catastrophic event frequency can also result in several events of the same peril happening at the same time because “conditions” are right (e.g., warm seas causing a high-frequency hurricane season, tectonic activity leading to increased earthquake frequency).

The following section will focus on the most important reasons for the departure from the simple Poisson frequency model.

2.3.2 Departures from the stationary Poisson process for the collective risk model

This section looks at the main reasons for the practical inadequacy of the stationary Poisson model (constant λ) in the context of the collective risk model, and for the widespread use of overdispersed models. We will see that overdispersion typically arises from the fact that the underlying process is a non-stationary Poisson process (Section 2.3.2.1) or a pure-jump (stationary or not) Lévy process that allows loss clustering (Section 2.3.2.2).

2.3.2.1 Non-stationary (non-homogeneous) Poisson processes. It is well known in actuarial practice that modelling the number of losses in a future policy period using a Poisson distribution will often underestimate the variance of possible outcomes. For this reason, models that exhibit overdispersion, that is, models where the variance exceeds the mean such as in the negative binomial model, are used routinely.

Surprisingly, using an overdispersed frequency model might be necessary even if the underlying process for generating losses in a given policy year is a Poisson process (stationary or not).

Let us assume that the process by which claims are generated is a Poisson process. When the Poisson rate is not stationary, the total number of losses in a given period (say, one year) will follow a Poisson distribution with Poisson rate:

$$\Lambda = \int_0^1 \theta(t) dt$$

where $\theta(t)$ represents the Poisson rate density (the number of expected losses per unit time) (Ross, 2003), and time t is expressed in years. The Poisson rate density $\theta(t)$ will in general be a stochastic variable itself, whose underlying distribution may or may not be known (analysis of historical loss experience may help in part). As a result, Λ will also be a stochastic variable. Such process is normally referred to as a Cox process.

Therefore, while the underlying process is fully Poisson and the distribution of the number of losses conditional on Λ follows a Poisson distribution (and its variance will be equal to the mean), since Λ is unknown at the time of pricing, the unconditional distribution used to model the future number of losses has a higher variance⁴. The overdispersion comes from ignorance about Λ and not from the loss-generating process itself.

Examples of stochastic processes affecting the frequency of losses to various lines of business are:

- Various physical quantities such as temperature, pressure, wind strength, rainfall, etc. are stochastic processes with deterministic seasonal trends that affect the number of losses in various lines of business such as motor and household/commercial insurance
- Economic indicators such as GDP growth in a given country may affect the number of professional indemnity claims or trade credit claims.

In an approach from first principles, when we determine that a non-stationary Poisson process is responsible for the number of claims, we should then investigate the cause of the volatility around the Poisson rate and model the rate intensity function $\theta(t)$ accordingly as a stochastic or

⁴The *unconditional* variance of the number of losses for a doubly stochastic Poisson process, $\text{Var}(N)$, can be written as $\text{Var}(N) = \text{E}(\text{Var}(N|\Lambda)) + \text{Var}(\text{E}[N|\Lambda])$. Since the conditional process is Poisson, $\text{Var}(N|\Lambda) = \text{E}(N|\Lambda)$. When Λ is constant (λ), $\text{Var}(N|\Lambda) = \lambda$ and $\text{Var}(\text{E}[N|\Lambda]) = 0$. In general, $\text{Var}(N) = \text{E}(\text{E}(N|\Lambda)) + \text{Var}(\text{E}[N|\Lambda]) = \text{E}(N) + \text{Var}(\text{E}[N|\Lambda]) \geq \text{E}(N)$.

deterministic variable. From that, we can derive the distribution of values of Λ . The loss count model corresponding to this case will therefore be a so-called **compound Poisson process**. There exist some special cases:

- While in general $\theta(t)$ is a stochastic variable and therefore Λ is also random, there are some special cases where $\theta(t)$ is fully deterministic. For example, if the underlying rate of car accidents only depended on the number of hours of daylight, $\theta(t)$ would be seasonal but fully predictable. In this case Λ would not be random and a Poisson model could be used
- In some formalisations the Poisson rate is not only non-stationary but the underlying process $\theta(t)$ depends on previous states through a Markov process. Specifically, in Avanzi et al. (2021) the process is formalised as a Markov-modulated non-homogeneous Poisson process. It is assumed that the system can be in one of the states $1, 2, \dots, r$ and that the probability of transition between state i and j is $q_{i,j}$. The duration of each state follows an exponential distribution (as in any Poisson process) with probability $\Pr(\text{duration in state } i > t) = \exp(-\sum_{i \neq j} q_{i,j} t)$. In a given state, events arrive at rate:

$$\lambda_M(t) = \lambda_{M(t)} \times \gamma(t)$$

where $\lambda_{M(t)}$ is the intensity corresponding to the state the system is in and is therefore constant for a given state; $\gamma(t)$ is a correction factor that takes exposure changes and seasonal effects into account, as in the previous point. The probability of a given number of events occurring in a given time interval (say, one year) is then given by a Poisson distribution with rate $\int_0^1 \lambda_{M(t)} \gamma(t) dt$

- A special case of the compound Poisson process is when $\theta(t)$ is stochastic and the behaviour of Λ – which being the integral of a stochastic variable is also in general a stochastic variable – can be approximately described as a gamma distribution. In this case, a well-known result says that the resulting distribution is a **negative binomial**. This has become the model of choice when there is need for additional volatility. However, the case where Λ comes from a gamma distribution is very special indeed and its use is mainly justified for practical reasons
- Actuaries are also quite familiar with a simplified but very useful special case of a compound Poisson model that is used in **common shock modelling** techniques (Meyers, 2007). Common shocks are a way in which the assumption of independence between events breaks, because the shock creates a correlation between the events. The basic idea is to consider that in most scenarios the Poisson distribution will have a “normal” Poisson rate λ_{normal} . Occasionally, however, there will be a shock to the system and the Poisson rate will be much higher, λ_{high} (obviously more than two scenarios are in general possible). One therefore needs to model the probability of a shock and its size (that is, λ_{high}) of the shock to the Poisson rate according to past empirical evidence or from first principles. Classical examples of shocks are a zoonosis (increasing number of livestock or bloodstock losses) or an economic recession (impacting the number of financial loss claims)
- An interesting example of a non-stationary Poisson process is the so-called **Weibull count process** – this is a non-stationary Poisson process where the waiting time between consecutive events is not exponential as in the stationary Poisson process, but follows a Weibull distribution, that is, a distribution whose survival probability is given by:

$$\bar{F}(t) = \exp\left(-\left(\frac{t}{\tau}\right)^\beta\right)$$

- The best-known use of the Weibull distribution is in *reliability engineering*, to model the failure rate of a manufactured product, and this can be used in insurance, for example in an extended warranty loss model. When $\beta < 1$, the failure rate decreases over time; when $\beta > 1$, it increases; when $\beta = 1$, it stays constant, and the Weibull reduces to an exponential

distribution. In the typical life cycle of a manufactured product, you have $\beta < 1$ at the beginning (when most production faults show up), $\beta = 1$ for most of the lifetime of the product, and $\beta > 1$ towards the end (when the product shows signs of wear). All this gives rise to the bathtub shape of the failure rate of products⁵

- When the waiting time can be modelled as a Weibull distribution, the resulting process will be a non-stationary Poisson process with Poisson rate $\lambda(t) = \frac{\beta}{\tau} \left(\frac{t}{\tau}\right)^{\beta-1}$, which depends on time unless $\beta = 1$
- Another interesting case of a non-stationary Poisson processes is given by the Hawkes process (see Laub et al., 2021, for a review) – a type of self-exciting process with applications in seismology, finance, epidemiology and neuroscience among others. The main idea behind the Hawkes process is that “the occurrence of any event increase[s] the probability of further events occurring” (Hawkes, 2018), as in the case of earthquakes, where an earthquake can trigger more earthquakes near the original location, causing the creation of clusters

2.3.2.2 Loss clustering and Lévy processes. Another way in which the independence between events breaks down causing an increase in volatility is when events are clustered. The Poisson distribution is often described informally as a distribution of *rare events*. What does this mean? It is true that the Poisson can be *derived* as the limit of a binomial distribution (see Equation (2)) when $p \rightarrow 0, n \rightarrow \infty$ and np tends to a finite non-zero number. However, in a temporal sense, whether events are rare or not depends on the scale at which you look at them. The only sensible, scale-independent definition of rare events is therefore that there should never be more than one event at a given time. Clustering is where this breaks down. Clustering may also involve claims that are not brought together by simultaneity or rather near-simultaneity, but for contractual reasons. This is the case for an *integrated occurrence* under the Bermuda form (a contract type for certain types of liability), claims may come in clusters of tens, hundreds or even thousands.

Let us consider a compound distribution where we model the number of clusters, and then the number of events in each cluster. If the process by which clusters are generated is a stationary process, then the whole process can be seen as a special case of a Lévy process (Appelbaum, 2009; Barndorff-Nielsen & Shephard, 2012; Kozubowski & Podgorski, 2009), and specifically a **pure-jump Lévy process**.

A Lévy process is a process with stationary and independent increments⁶. Brownian motion and Poisson processes are examples of Lévy process. Brownian motion describes the movement of particles under the assumption that their trajectory is not deterministic but is affected by Gaussian noise. The stochastic equation for generic Brownian motion (in integral form) is given by:

$$X_t = \mu t + \sigma B_t$$

where B_t is a Gaussian process with mean 0 and standard deviation 1, and μ is the drift. A typical application of this is to model stock prices: for that purpose, however, X_t is taken to be the logarithm of the stock price (which is never below zero): $X_t = \ln Y_t$ and therefore the equation can be rewritten as $\ln Y_t = \mu t + \sigma B_t$ ($dY_t = \mu Y_t dt + \sigma Y_t dB_t$ in differential form). Y_t is referred to as *geometric Brownian motion*.

⁵It is also relevant, even more intuitively, to human mortality, which tends to be higher among infants and among the elderly, while being roughly stationary in the middle: for this reason the regime $\beta < 1$ is also called the “infantile phase,” and the regime $\beta > 1$ is referred as the “aging phase.”

⁶Formally, a process $X = \{X_t : t \geq 0\}$ defined on a probability space (Ω, F, P) is said to be a Lévy process if it possesses the following properties: (a) The probability that $X_0 = 0$ is 1 (b) For $0 \leq s \leq t, X_t - X_s$ is equal in distribution to X_{t-s} (stationary increments) (c) For $0 \leq s \leq t \leq u \leq v, X_t - X_s$ is independent of $X_v - X_u$ (independent increments) (d) $\lim_{h \rightarrow 0} \Pr(|X_{t+h} - X_t| > \varepsilon) = 0$ for all values of t : that is, the probability of a finite jump at an arbitrarily chosen time is zero (probability continuity). Notably, (d) implicitly allows for jumps of random sizes at random times, as we will see later when we introduce the Lévy-Khintchine decomposition.

One problem with Brownian motion as a model of insurance losses is that the sample paths are continuous and sudden jumps are not possible: when Δt goes to zero, $\Delta X_t = X_{t+\Delta t} - X_t$ also *always* goes to zero. In reality, whether we speak about particles or stock prices, we see examples of the system undergoing sudden jumps, and any model that doesn't take that into account underestimates the volatility of the underlying phenomenon. A Lévy process explicitly allows the possibility of having a number of finite jumps over a given time interval. A Lévy process can be described (informally) by the following Lévy-Khintchine decomposition⁷ (Kozubowski & Podgorski, 2009):

$$X_t = \mu t + \sigma B_t + \sum_j h_j \mathbb{1}_{[\Gamma_j, \infty)}(t)$$

where the last term describes a *finite* sequence of jumps at random times Γ_j of size $h_j = X_{\Gamma_j} - X_{\Gamma_j^-}$. Note that the process can always be put in a form that is right-continuous and has limits from the left. The symbol $\mathbb{1}_{[\Gamma_j, \infty)}(t)$ stands for the random step function $\mathbb{1}_{[a, \infty)}(t) = 1$ for $t \geq a$, 0 for $t < a$, where $a = \Gamma_j(\omega)$ depends on the random sample.

A Lévy process has three components: a deterministic drift, Gaussian noise, and a collection of jumps. In the context of modelling loss counts we do not need the whole mathematical machinery of Lévy processes, and we can focus on the case where $\mu = \sigma = 0$ and only the jump components remain (a “pure jump” Lévy process). The Poisson process can be described as the case where $h_j = 1$ for all j .

An obvious generalisation of the Poisson process to describe losses can be obtained when h_j is itself a random variable with integer values larger than or equal to 1, or in other terms, the case where losses are not isolated rare events but can occur in clusters. This is of course an approximation – in the real world, the losses do not actually need to happen at exactly the same time, but *around* the same time.

An interesting special case occurs when the number of clusters N_c is Poisson-distributed with rate $\lambda = k \ln(1 + p)$, and the number of events N_e occurring within a cluster can be modelled with the logarithmic distribution whose probability $f(n_e)$ for $n_e \geq 1$ is given by:

$$f(n_e) = \frac{1}{n_e \ln(1 + p)} \left(\frac{p}{1 + p} \right)^{n_e}$$

This case will produce a negative binomial distribution with parameters p and k (Anscombe, 1950).

However, in general, the numbers of events (losses) within each cluster need not be logarithmic: see Anscombe (1950) for alternatives⁸ such as the Newman Type A distribution (Poisson distribution of events in each cluster) and the Pólya-Aeppli distribution (geometric distribution of events). The number of clusters need also not be Poisson. For example, if the underlying process for the number of clusters is a non-stationary Poisson, then the overall process will not be a Lévy process, but a so-called *non-stationary Lévy process*.

Modelling from first principles, we should use a non-stationary Poisson when there is local independence and we have some sensible model of the intensity function. However, where clustering of losses occurs, we should use other models such as pure-jump Lévy processes or their generalisation to non-stationary increments (non-stationary pure-jump Lévy processes).

⁷More formally, the Lévy-Khintchine decomposition states (Appelbaum, 2009) that a stochastic process X_t can be written as $X_t = \mu t + \sigma B_t + J_t$, where μt is the drift, B_t is a Brownian motion, and J_t is a pure-jump process, which can be further decomposed into a compound Poisson process L_t with a finite number of jumps larger or equal to 1, and a *compensated generalised Poisson process* S_t , a process with countably many jumps smaller than 1, but such that the integral of these jumps converges in distribution. In the case most relevant to us in this paper, that of count processes, the only relevant component is J_t , and specifically L_t , as all jumps have positive integer size.

⁸Some of these alternatives display a multimodal behaviour that might be appropriate in certain circumstances.

2.3.2.3 Does introducing Lévy processes bring any advantage? It may appear that formalising loss count processes as Lévy processes is overkill, as the Brownian component and the drift component are not used. A compound Poisson model would be a sufficient description. However, using the framework of Lévy processes gives us access to general mathematical results that are available **off the shelf** in this well-studied area of research. One such result is related to **infinite divisibility**.

A distribution for variable X is said to be infinitely divisible if for any positive integer n , there exist independent and identically distributed random variables X_1, X_2, \dots, X_n such that the distribution of X is the same as the distribution of the sum $X_1 + X_2 + \dots + X_n$. The distribution of each X_k depends on the original distribution⁹.

It can be shown that if $X(t)$ is a Lévy process then $X(1)$ is infinitely divisible, and conversely if X is infinitely divisible then there is a Lévy process such that $X(1) = X$ (Mildenhall, 2017; Gardiner, 2009).

The interest in this for loss modelling is that “[i]n an idealized world, insurance losses should follow an infinitely divisible distribution because annual losses are the sum of monthly, weekly, daily, or hourly losses. [...] The Poisson, normal, lognormal, gamma, Pareto, and Student- t distributions are infinitely divisible; the uniform is not infinitely divisible, nor is any distribution with finite support [...]” (Mildenhall, 2017). The extension to non-stationary Lévy processes, which better reflect the reality of a changing risk landscape, is then straightforward.

Just a few examples of infinitely divisible distributions that are useful for loss modelling purposes are:

- The Poisson distribution
- The negative binomial distribution
- The normal distribution
- The Pareto distribution
- The lognormal distribution
- All compound Poisson distributions.

In relation to the last item, it is also important to note that all infinitely divisible distributions can be constructed as limits of compound Poisson distributions.

Finally, Lévy processes can do much more than modelling loss counts. As shown in Mildenhall (2017), Lévy processes can be used to model aggregate losses. For that purpose, the jumps do not just represent the occurrence of an event but also its size. An interesting consequence is that **this gives us a first-principles reason for the separation between frequency and severity**. It turns out that in general it is not possible to split the aggregate loss process into a count process and a severity process: it is only possible when the number of jumps is finite! In other cases, there may be an infinite number of small jumps and this split is not possible. In that case, however, it is still possible to model the process as a Lévy process, using a loss frequency curve (basically the OEP used in catastrophe modelling), as in Patrik *et al.* (1999).

2.3.2.4 Generalising the Poisson process: illustrations. Figure 2 shows examples of the main ways of generalising the Poisson process: through non-stationarity (a non-stationary Poisson process), clustering (general pure-jump Lévy process), or both (a non-stationary Lévy process). All processes are represented in terms of the cumulative number of unitary jumps as a function of time.

⁹Infinitely divisible distributions of the variable X can also be defined in terms of their characteristic function $\Phi(t) = \mathbb{E}(e^{itX})$. If $X = \sum_{k=1}^n X_k$ where the X_k s are iid variables, the characteristic function can be written for any n as $\Phi(t) = (\Phi_n(t))^n$, where $\Phi_n(t)$ is the characteristic function of each of the X_k s.

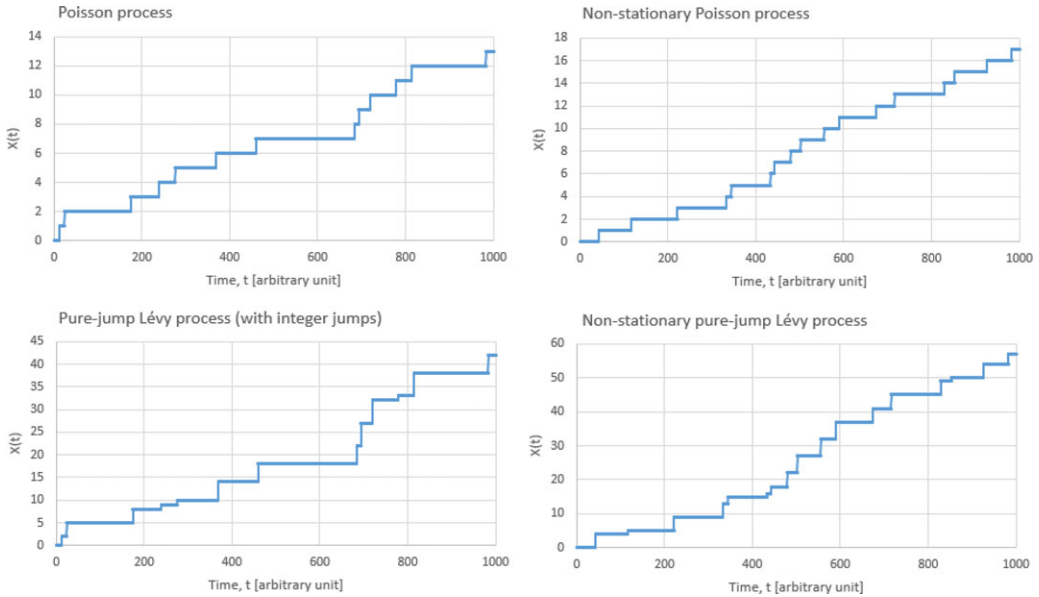


Figure 2. Top left: An example of a Poisson process. The x-axis shows the timing of the jumps (losses) while the y-axis shows the cumulative number of jumps, $X(t)$. Note how each jump has unitary size ($X(t_j) - X(t_j^-) = 1$ if there is a jump at t_j), as in a Poisson process there is no clustering of events (events are “rare”). The condition that the jumps have unitary size can be relaxed – as long as all the jumps have exactly the same size the process is still called Poisson. Top right: an example of a non-stationary Poisson process. Notice how the frequency of losses is higher for $400 < t < 600$ and $750 < t < 1000$. Bottom left: an example of a pure-jump Lévy process, where the jump values are integers (the negative binomial process falls under this category). Bottom right: An example of non-stationary pure-jump Lévy process (again with integer jump values). Note that the process at the bottom right shares with the one at the top right the timings of the jumps (hence the similarity), but the jumps do not necessarily have unitary size in the bottom right figure.

3. Modelling the Severity of Liability Losses

Liability losses arise from bodily injury or property damage to a third party. The amount of a liability loss can be determined by the courts in a variety of ways that also depend on the jurisdiction. As a result, a detailed derivation of a ground-up liability loss model from first principles for all cases is probably hopeless. However, it is possible to find constraints on the scaling behaviour of severity distributions and justify them based on reasoning from first principles. Before doing that, it is useful to look at a traditional severity models and improvements that actuaries have made on them over the years using statistical theory.

3.1 Our Misplaced Fascination with the Lognormal Model and the Evidence of a Power-Law Tail

Since the beginnings of risk theory, the lognormal distribution has been viewed as a sensible attempt at modelling the severity of losses: it produces losses that are always positive and may span over several orders of magnitude. Although it does not have a power-law behaviour, you cannot say that it is thin-tailed either: its behaviour is genuinely intermediate (Taleb, 2020). A well-known paper by Benckert (1962) provides some empirical evidence that the lognormal works well. Unfortunately, upon closer investigation with richer data sets, it becomes clear that the lognormal model is rarely a good model for either the body of the distribution or the tail.

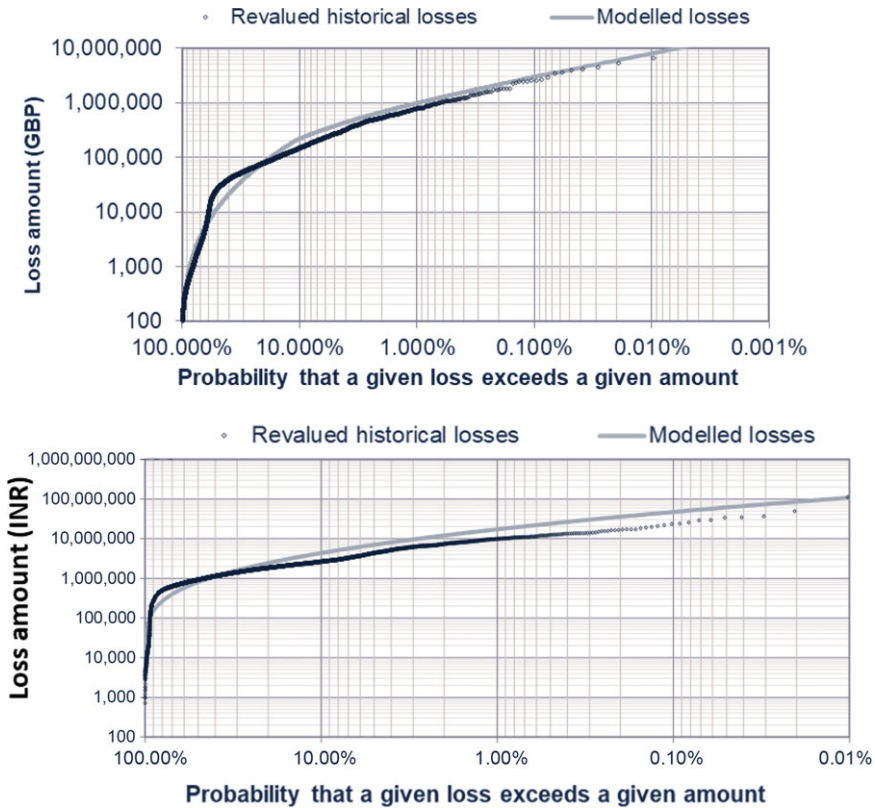


Figure 3. Top: A Lognormal fit to a large number (around 10,000) of employers’ liability claims in the UK market. ($\ll 0.1\%$). Bottom: A lognormal fit to a large number of motor bodily injury claims in the Indian market. In both cases, based on the value of the KS statistic, the probability that the data set comes from a lognormal distribution is negligible ($\ll 0.1\%$). The original losses have been masked by a scalar factor.

3.1.1 The empirical inadequacy of the lognormal model

We have looked at several data sets of casualty losses and the fit to a lognormal has always proved to be severely unsatisfactory¹⁰. Two examples are given in Figure 3, which shows that the lognormal model is inadequate, and this can be proven even more sharply using a metric to assess the distance between distributions, such as the KS statistic.

While the empirical inadequacy of the lognormal distribution is clear, we have also found that the tail of a portfolio of liability losses is almost always modelled well by a generalised Pareto distribution (GPD) with a power-law tail ($\xi > 0$), as predicted by extreme value theory (EVT).

3.1.2 Is the lognormal model theoretically adequate to model large losses?

According to EVT, the tail of the distribution can be asymptotically described – apart from unusual cases – by a GPD, that is, a (survival) distribution of the form:

¹⁰Note that it is difficult to evaluate goodness of fit visually if one uses a linear representation of the empirical CDF against the corresponding model: all fits look similar and they tend to look better than they are; also, the tail behaviour – which is the one that we are keenest to represent faithfully – is shrunk in the top right corner of the chart and it is very difficult to distinguish one tail behaviour from the other. For this reason, a representation which is now more commonly used adopts a log-log scale, which has the useful property that a Pareto distribution appears as a straight line in the chart.

$$\bar{F}(x) = \left(1 + \xi \frac{x - \mu}{\nu}\right)^{-\frac{1}{\xi}}$$

In turn, this distribution exhibits three possible types of behaviour depending on the value of ξ : a finite-support distribution (Beta distribution) for $\xi < 0$, an exponential distribution for $\xi = 0$ (a limiting case), and a power-law distribution for $\xi > 0$. The case $\xi > 0$ is observed in most practical cases. Note that a GPD with $\xi > 0$ is asymptotically equivalent to a single-parameter Pareto ($\bar{F}(x) = (\theta/x)^\alpha$) with $\alpha = 1/\xi$.

This does not immediately rule out lognormal behaviour. The lognormal distribution shows an intermediate behaviour between the exponential and the power law, but asymptotically it falls under the exponential case – or in technical terms, the lognormal distribution falls within Gumbel’s domain of attraction (Embrechts et al., 1997). This can be shown analytically using the concept of local Pareto alpha (Riegel, 2008):

$$\alpha(x) = \frac{-x\bar{F}'(x)}{\bar{F}(x)} = \frac{x F'(x)}{1 - F(x)}$$

It is easy to prove that (assuming F is a lognormal distribution) $\lim_{x \rightarrow \infty} \alpha(x) = \infty$, corresponding to $\xi = 0$.

Therefore, while the lognormal distribution could be theoretically consistent with EVT, in practice it should be discarded because portfolio losses tend to stabilise around a strictly positive value of ξ .

3.1.3 Is the lognormal model theoretically adequate to model attritional losses?

Having become cognizant of the fact that the GPD captures the behaviour of the tail better than the lognormal, actuaries now tend to use a GPD (or a simpler Pareto) for tail modelling and reserve the lognormal for modelling attritional (small) losses. This is the approach taken, for example, in Knecht & Küttel (2003) and in Fackler (2013), where new classes of spliced lognormal/Pareto and lognormal/GPD distributions are introduced to formalise this approach. This is a good workaround to ensure that a suitable model is used where it matters most: in the tail.

However, experience shows that the lognormal distribution is not generally a good match for attritional losses either. Indeed, Figure 3 shows two examples of a behaviour that we often found in our work as practitioners: empirical distributions often appear as the conjunction of two parts that are *broadly* linear in a log-log scale, with a concave elbow connecting the two. This behaviour is not one that a lognormal distribution is able to capture and is one reason why the lognormal performs poorly everywhere. This also suggests that the distinction between attritional losses and large losses is probably more than a convenient trick to simplify the analysis. It also points to the potential usefulness of using different types of distributions for the attritional losses, for example by using a Type II Pareto/GPD splicing.

3.2 Modelling Large Losses from First Principles: The Emergence of the Power-Law Behaviour

The asymptotic behaviour of large losses is mainly dominated by their scaling behaviour: namely, by how the DoublingRatio(x) = $\frac{\Pr(X > 2x)}{\Pr(X > x)} = \frac{\bar{F}(2x)}{\bar{F}(x)}$ behaves as a function of x . If the severity distribution is an exponential distribution ($\xi = 0$), then Doubling Ratio(x) = $\exp(-x)$: it decreases quickly and soon becomes indistinguishable from zero. If the severity distribution is a single-parameter Pareto with parameter α , then Doubling Ratio(x) = $2^{-\alpha}$: the ratio is the same at all x , or, equivalently, the ratio is the same at all scales and has a so-called fractal nature. If the severity is GDP with $\xi > 0$, the ratio will have a slight dependence on x that will asymptotically disappear: Doubling Ratio(x) = $\left(\frac{1 + \xi(2x - \mu)}{1 + \xi(x - \mu)}\right)^{-\frac{1}{\xi}} \rightarrow 2^{-\frac{1}{\xi}}$.

<i>Law and description</i>	<i>Power law</i>	<i>Ultimate reason for behaviour</i>
Champernowne's distribution of incomes (1953)	α between 1.6 and 2.4	Above a certain threshold, "the prospects of various amounts of percentage change of income are independent of the initial income"
Zipf's law, modelling the scaling law of city size (Zipf, 1949; Gabaix, 1999)	$\ln(\text{Rank}) = \beta - \alpha \times \ln(\text{Size})$ $\alpha \sim 1$ consistently across countries.	The standard deviation is the same across sizes
Wold and Whittle's distribution of wealth (1957)	$\alpha = \frac{\text{mortality rate}}{\text{asset growth rate}} \sim 1.3 - 2.0$	The asset growth rate is the same at all levels of wealth (>a certain threshold).
Simon and Bonini's distribution of the size of business firms (1958)	$\alpha \sim 1$ if the contribution %c of new firms to growth is negligible. Else, the predicted parameter is $\alpha = (1 - \%c)^{-1}$.	The expected percentage growth for firms is the same at all sizes

Figure 4. Examples of the emergence of the power law in economics.

Thanks to EVT and specifically to the Pickand–Balkema–de Haan (PBdH) theorem, we know that the exponential and power-law behaviour contains all possibilities with positive unbounded support, that is, in most practical cases. Observation also tells us that the most common behaviour is a power law. While the PBdH theorem constrains the asymptotic behaviour of the severity distribution, it does not say what drives a particular behaviour and specifically why the power-law behaviour is so ubiquitous. Where does this power-law behaviour come from?

3.2.1 *Random growth processes in economics*

A power law is often observed in statistical physics, especially in the context of phase transitions. It occurs in the presence of system-wide coherence. In economics, power laws have been observed for the distribution of incomes, wealth, the size of cities, the size of firms, and many other examples (Figure 4). In all these cases, the power-law behaviour (and specifically the Pareto law) have been explained as the steady-state distribution resulting from a random growth process (Gabaix *et al.*, 2016) where the growth rate is uniform across all sizes.

Let us expand on this. In all cases shown in Figure 4, the phenomenon can be described by a stochastic differential equation whose solution is a Pareto distribution. According to Gabaix (1999) and Gabaix *et al.* (2016), the basic insight is that these other power laws are the steady-state distribution arising from the scale invariance of the physical process of growth. Growth is homogeneous at all scales: therefore, the final distribution process should also be invariant, which means that it has fractal nature and must follow a power law.

In the most modern (and simplest) incarnation, this means that the process can be described by stochastic differential equations of this type (we follow Gabaix *et al.*, 2016):

$$dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dZ_t$$

where X_t is a stochastic variable (such as income, wealth, city size, firm size), $\mu(X_t, t)$ is the growth at time t , $\sigma(X_t, t)$ is the spread at time t , and Z_t is a Brownian motion. This equation can be interpreted as follows: the variable X_t increases in time because of a deterministic drift $\mu(X_t, t)dt$ and random effects that can be described as a Brownian motion. The case in which we are most

interested is the one where both the growth rate and the spread rate are constant at all sizes and are time-independent, that is, $\mu(X_t, t) = \mu X_t$ and $\sigma(X_t, t) = \sigma X_t$. This gives¹¹:

$$dX_t = \mu X_t dt + \sigma X_t dZ_t$$

The distribution of variable X_t at time t is¹² $f(x, t)$. We want to describe the evolution of the distribution given the distribution at time $t = 0, f(x, 0)$. The tool to do so is the forward Kolmogorov equation, which for our case can be written as:

$$\frac{\partial f}{\partial t} = -\frac{\partial(\mu x f(x, t))}{\partial x} + \frac{1}{2} \frac{\partial^2(\sigma^2 x^2 f(x, t))}{\partial x^2}$$

We are specifically interested in the steady state of distribution $f(x, t)$ – if it exists – because that will tell us what we need to know, the distribution of income, wealth etc in equilibrium. We are therefore interested to solve the FKE for the case $\frac{\partial f}{\partial t} = 0$. We are basically seeking the solution $f(x) = \lim_{t \rightarrow \infty} f(x, t)$ for what is now an ordinary differential equation:

$$-\mu \frac{d(xf(x))}{dx} + \frac{1}{2} \sigma^2 \frac{d^2(x^2 f(x))}{dx^2} = 0$$

This can be rewritten as:

$$\frac{\sigma^2}{2} x^2 \frac{d^2 f(x)}{dx^2} + (2\sigma^2 - \mu)x \frac{df(x)}{dx} + (\sigma^2 - \mu)f(x) = 0 \tag{3}$$

it becomes clear that one possible solution is a function of the form $f(x) = kx^\beta$, which turns Equation (3) into a second-degree equation in β . This has two roots: $\beta = -1$ and $\beta = 2(\frac{\mu}{\sigma^2} - 1)$. The first root is not “physical,” as it leads to a divergent equation. As for the other, it leads to a Pareto distribution with $\alpha = -(1 + \beta) = 1 - 2\frac{\mu}{\sigma^2}$

Assuming that the solution must make economic sense, which translates into a smoothness constraint, it can also be proved that this is the only solution (Gabaix et al., 2016).

We have however performed a sleight of hand: we have assumed that such an equilibrium distribution is possible, which is not the case unless we make other extra assumptions. If these extra assumptions are not included the solution to the forward Kolmogorov equation is a lognormal distribution with indefinitely increasing mean and variance, which never reaches a steady state (Gabaix et al., 2016). Researchers have resolved this issue in the past in two ways:

- (a) By assuming that the random variable has a minimum value X_{\min} (any value will do!) that acts as a reflective barrier (e.g., a minimum salary, or a minimum city size).
- (b) By assuming that the system has some type of attrition so that new elements enter and exit the system, for example, in the case of the income distribution, new people join the labour market and others retire/die.

Both assumptions are reasonable in the real world. The theory then shows that in these cases a steady-state solution is possible and therefore that solution is a Pareto law, at least asymptotically. This is consistent with the results of EVT, as a GPD with positive shape parameter asymptotically becomes a Single-Parameter Pareto law.

¹¹Equivalently, we can write $\frac{dX_t}{X_t} = d \log X_t = \mu dt + \sigma dZ_t$, which has perhaps a more familiar format.

¹²In the absence of other constraints, the solution of this equation is a standard lognormal distribution with indefinitely increasing mean and variance. As we will see shortly, however, under certain conditions an altogether different behaviour emerges.

3.2.2 *The random growth process and loss distributions*

We have now seen several examples of circumstances – mostly related to economic quantities – in which the Pareto law emerges because of an underlying random growth process. On the other hand, we have significant empirical evidence for the Pareto law applying – at least approximately – to the realm of large insurance losses. What can we say about whether such an empirical fact can be reasonably explained by a random growth process?

3.2.2.1 Mechanisms for the emergence of power laws. There appear to be two possible ways in which we can show that a Pareto law will eventually emerge for a (liability) loss distribution in most cases: an indirect mechanism and a direct one. The **indirect mechanism** is quite straightforward: liability losses comprise different components. Limiting ourselves to bodily injury claims, these components are:

- loss of earnings (or dependant’s loss of earnings in the case of death)
- cost of care including medical costs,
- pain and suffering
- other (e.g., punitive damages).

If at least one of these components follows a power law, the overall resulting distribution will be (asymptotically) a power law. The “loss of earnings” component of a loss is proportional to income (in a non-trivial way, as it is also affected by other factors such as age) and if we accept Champernowne’s analysis (1953) – which has indeed been tested successfully for many territories – this follows a Pareto in the tail of the distribution. If more than one power law is in play, the one with the thickest tail will eventually dominate the tail and will fully characterise the asymptotic behaviour.

Another example is given by D&O losses. These are proportional to the size of firms, which also follows a Pareto law according to Simon and Bonini’s analysis (1958).

The **direct mechanism** is derived by looking at the growth rate of losses themselves. While a full-blown theory for this is not available yet, there are clear indications that one or more such mechanisms exist. As an example, consider court inflation: compensations from previous cases are used as benchmarks for current cases and increases tend to be a percentage of the benchmark, leading to an exponential increase.

For the existence of a steady-state distribution and the emergence of a Pareto law, we also need to find one limiting mechanism: either a minimum loss or a factor of attrition. While the existence of an attrition factor for losses cannot be excluded (e.g., certain types of losses disappearing from the list of potential losses), the existence of a minimum loss seems to be the most straightforward way of demonstrating the existence of a steady-state solution: whether because of the existence of deductibles or because below a certain monetary amount a claim simply doesn’t make sense, we can safely assume that a minimum loss exists.

3.2.2.2 Attritional versus large losses. Champernowne’s article also points to an interesting distinction between low incomes and high incomes, which appear to follow different laws. The existence of these two different “regimes” is explained by the fact that lower incomes tend to grow less rapidly than higher incomes. The same distinction appears to be valid for losses. Traditionally, actuaries subdivide losses into attritional and large losses. The main reason behind this distinction is often one of convenience: small losses are treated as an aggregate, allowing actuaries to pay more attention to individual large losses. However, there may be more to the distinction than simply size. Attritional losses tend to go through a more streamlined claims settlement process with little or no court involvement, and therefore their increase in time is linked more to standard inflationary factors such as CPI and wage inflation and less to court-driven social inflation. Also, if this is the case, we should see a radically distinct behaviour in the log-log construct standardised curve (CDF) graph between attritional and large losses. This is indeed what we observe in practice

(see Figure 3): both attritional losses and large losses appear to be distributed on a straight line in log-log scale, although this is only a very rough approximation for attritional losses. This points to two different types of behaviour for the two different types of losses. These empirical observations and results from other economic examples both point to an intrinsic rather than convenience-driven difference between attritional and large losses.

An **alternative explanation** for the difference in regimes is that the growth for the low part of the distribution is not invariant and that attritional losses have not (yet?) achieved a steady-state distribution, in which case we could attempt to represent the lower part of the distribution as an intermediate stage in the transition between a lognormal distribution and a Pareto distribution, as it occurs in the simulation in Figure 5. That would be an interesting comeback for the lognormal distribution!

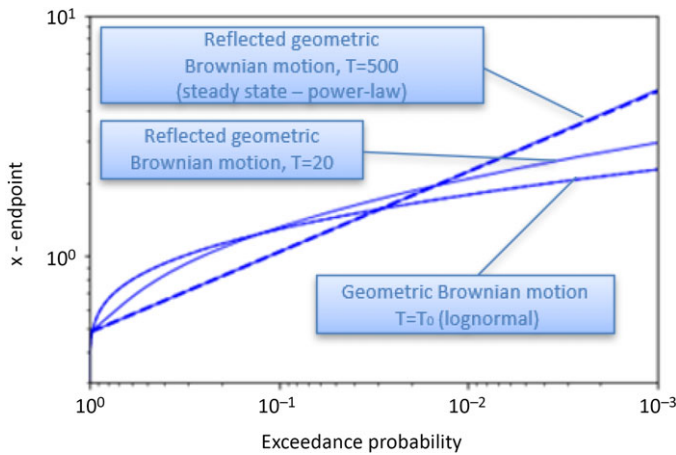


Figure 5. Survival function for increasing time periods, showing the increasing move away from the lognormal distribution towards the steady-state Pareto distribution.

3.2.2.3 Empirical demonstration of the emergence of a power law. We can demonstrate the emergence of a power law, that is, the steady-state solution to the forward Kolmogorov equation, for a system where the random variable has a minimum value empirically using a numerical simulation. Consider two types of stochastic random walk for the random variable $X(t)$. The first is a standard Geometric Brownian motion (GBM), which can be written as $X(t + dt) = X(t) + \epsilon X(t)$. The second is the reflected Geometric Brownian Motion (rGBM), which can be written as:

$$X(t + dt) = \begin{cases} X(t) + \epsilon X(t), & X(t) \geq x_{\min} \\ X(t) + \max(0, \epsilon)X(t), & X(t) < x_{\min} \end{cases}$$

The stochastic simulation is based on random increments ϵ drawn from a normal distribution with mean μdt and variance $\sigma^2 dt$:

$$\epsilon = \mu dt + \sigma \sqrt{dt}Z, \quad Z \sim N(0, 1).$$

The parameters are $\mu = -0.01$, $\sigma = 0.1$, and $dt = 0.1$ throughout.

In both walks, the starting point is $X(0) = 1.0$ and we take $x_{\min} = 0.5$. We generate a sample of 1,000,000 paths over a total time-period T , where each path is a set $[X(0), X(dt), X(2dt), \dots, X(T)]$, and study the distribution of the endpoints $X(T)$ for different values of T .

The assertion is that for large values of T , the reflected geometric Brownian motion will result in a power-law distribution for the endpoints, with a survival function:

$$S_{X(T)}(x) \sim ax^b, \quad b = -1 + \frac{2\mu}{\sigma^2}.$$

This distribution arises as the steady-state solution to the dispersion relation for the geometric Brownian motion. In the continuum limit $dt \rightarrow 0$, the coefficient a appearing in $S_{X(T)}(x)$ could be determined via:

$$E[X(T)] = \int_{x_{\min}} dx S_{X(T)}(x).$$

However, for numerical simulation, it is not possible to arbitrarily decrease dt to zero, since for a relatively long total time-period, the number of steps required to form the path becomes prohibitively large. This means that the normalisation coefficient a cannot be determined from simulation.

The empirical distribution of the endpoints is represented via the survival function (exceedance probability) for various time periods T in Figure 5.

For $T_0 = 10$, the geometric Brownian motion and reflected geometric Brownian motion result in lognormally distributed endpoints (the simulated curves both coincide with the dashed “lognorm” analytic curve), as expected. For $T = 20$, the reflected geometric Brownian motion endpoint distribution starts to deviate and straighten out. This continues up to $T = 500$ (the straight line), where the simulated curve lies on top of the power-law fit (the dashed straight line, using the power $-1 + 2\mu/\sigma^2$, and determining the coefficient a from the 95th percentile of the simulated curve).

3.3 Is an Approach from First Principles to Attritional Losses Even Needed?

Attritional losses are messy and we do not expect to capture them with a neat distribution derived from first principles, although when the empirical distribution displays a clear two-phase behaviour and it appears smooth for both attritional and large losses, it might be worth trying a Type II Pareto/GPD fit. On the positive side, that may not be necessary as we can avoid modelling the attritional losses altogether: attritional losses are by definition usually small and plentiful and therefore we can simply resample from the available data. One can also use interpolation, but this doesn’t add much value in most circumstances.

The theoretical underpinning for the resampling approach to attritional losses is Glivenko–Cantelli’s theorem, which states that the empirical distribution converges uniformly ($\sup|F_n(x) - F(x)| \rightarrow 0$ as $n \rightarrow \infty$) to the real distribution $F(x)$. The connection with the Kolmogorov Smirnov (KS) test is obvious and it is possible to put limits around the KS statistic with a certain probability. For example, we know that $\sup|F_n(x) - F(x)| < 1.358\sqrt{n}$ with probability 95% for sufficiently large values of n .

This theoretical result says almost everything we need to know but there are still interesting practical considerations and questions. For one thing, resampling has obvious limitations: for example, if the loss data set is $\{x_1, \dots, x_n\}$, resampling can never produce losses above $\max\{x_1, \dots, x_n\}$; also, losses tend to become more rare when they become large and therefore the distortions arising from interpolation become more significant. We can expect that resampling works better for attritional losses than for the general population of losses. Another practical consideration is, how much do we lose in practice by using resampling? To be more specific, if we know that the correct distribution is, say, a lognormal we want to compare (a) resampling from the data available with (b) using a lognormal model calibrated *based on the available data*. We can calculate the KS distance between the true underlying model and comparing the former versus the latter. Repeating many times gives an estimate of the average KS distance for the two approaches. Note that the KS distance must be evaluated for losses below some threshold, since simple resampling is not interesting for large losses, except in special cases.

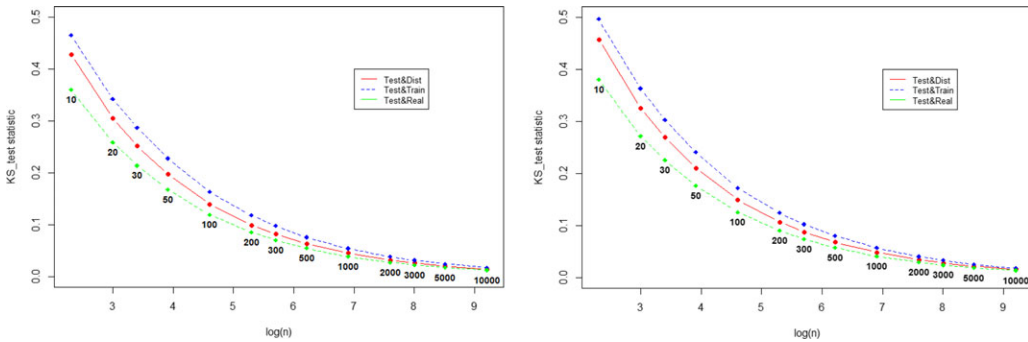


Figure 6. KS distance for test set versus (data-calibrated) model, test set versus training set, test set versus true model for different values of the sample size. Left: no threshold. Right: 90% threshold. Note that the KS distance (rather than the normalised KS distance, which will remain relatively flat) is shown here.

We have tested data coming from a lognormal distribution with mean equal to 25 and standard deviation equal to 5, for different sample sizes (10, 20, 30, 50, 100, 200, 300, 500, 1,000, 2,000, 3,000, 5,000, 10,000). The main message is that for small samples the model tends to perform better than resampling, but the difference rapidly shrinks as the sample increases in size, and eventually resampling performs in a way that is indistinguishable from modelling (Figure 6).

We have shown that even in the case where we *know* the true distribution and we generate losses from it, using resampling is almost as good as using a model calibrated based on existing data, apart from the case of very small samples (say, 100 if the threshold for attritional is at the 90th percentile of the distribution, in the case of a lognormal). We have also argued, however, that the lognormal model is usually not a good model. Indeed, given the messy character of attritional losses, it is unlikely that any standard continuous distribution is a very good model. Therefore, in practice, for data sets of sufficient size, resampling the empirical distribution will outperform any attritional loss model. Using software such as R or Python makes resampling as easy as working with a parametric model, essentially removing the only reason to use a parametric model.

4. Modelling the Severity of Property Losses

4.1 The Traditional Approach to Property Modelling

The standard method for producing a severity model for liability losses – that is, fitting a curve to historical loss data – is not straightforward to apply to property business because future losses depend critically on the maximum possible loss (MPL) profile of the portfolio, which is not necessarily the same as the MPL profile of past years. For this reason, it is more natural to model losses as percentages of the MPL for each building and construct standardised curves (CDFs and exposure curves), under the assumption that their shape does not depend on the MPL but only on the type of construction.

In Europe and other territories, Bernegger's curves (also called MBBEFD curves) are widely used for this purpose.

4.1.1 Bernegger's curves and their connection to statistical mechanics

Bernegger's curves (Bernegger, 1997) are curves borrowed from statistical mechanics. However, as Bernegger emphasised from the beginning, their use is purely practical and there is no suggestion of any deeper connection with statistical mechanics.

It is useful to state the formal connection between Bernegger curves and statistical mechanics, because it was not done in the original paper, and because we will need to refer back to it later when exploring the use of graph theory in the context of property pricing.

The *probability densities* for the different curves from statistical physics are given by:

$$f(x) = \begin{cases} \frac{1}{Z} e^{-\frac{x}{kT}} & x \geq 0, \text{ Maxwell – Boltzmann} \\ \frac{a}{ce^{kT}-1} & x \geq 0, \text{ Bose – Einstein} \\ \frac{a}{ce^{kT}+1} & x \geq 0, \text{ Fermi – Dirac} \end{cases}$$

In the most general format, the *survival probability* corresponding to Bernegger’s curves is given by:

$$\bar{F}(x) = \begin{cases} \frac{1-b}{(g-1)b^{1-x}+(1-bg)} & x < 1, b > 0, b \neq 1, bg \neq 1, g > 1 \\ b^x & x < 1, bg = 1, g > 1 \\ \frac{1}{1+(g-1)x} & x < 1, b = 1, g > 1 \\ 1 & x < 1, g = 1 \text{ or } b = 0 \\ 0 & x = 1 \end{cases}$$

$$= \begin{cases} \frac{1-b}{\frac{(g-1)b}{g^{b-1}}e^{-(\ln b)x}-1} & x < 1, b > 0, b \neq 1, bg \neq 1, g > 1 \\ e^{(\ln b)x} & x < 1, bg = 1, g > 1 \\ \frac{1}{1+(g-1)x} & x < 1, b = 1, g > 1 \\ 1 & x < 1, g = 1 \text{ or } b = 0 \\ 0 & x \geq 1 \end{cases}$$

where the curly bracket on the left shows the standard way of writing Bernegger’s curves, while the curly bracket on the right shows a rewriting that makes the connection with statistical physics more obvious.

The three different cases can therefore be obtained by changing the value of bg : $bg = 1$ corresponds to the Maxwell–Boltzmann distribution, $bg > 1$ to the Bose–Einstein distribution, $bg < 1$ to the Fermi–Dirac distribution. The case $b = 1$ is not physical. Setting the survival probability to zero at $x \geq 1$ is also a mathematical trick to produce a finite probability for a total loss and does not correspond to a physical situation.

It should also be mentioned that there is a special (“Swiss Re”) parameterisation dependent on a single parameter c formed from a particular configuration: $b = b(c) = \exp(3.1 - 0.15c(1 + c))$, $g = g(c) = \exp((0.78 + 0.12c)c)$. Based on empirical studies, Bernegger suggests that residential property fire damage corresponds to $c \approx 1.5$, whereas industrial properties may be described by values up to $c = 8.0$.

The probability of a total loss ($X = 1$) is given by $\Pr[X = 1] = \bar{F}(1^-) = \frac{1}{g}$ for any value of b , and for $g \geq 1$.

It should be noted that all Swiss Re c curves are solidly within the Bose–Einstein region of Bernegger’s curves.

As mentioned, Bernegger was careful not to suggest a deeper relationship with statistical physics. Interestingly, however, if one considers certain possible mechanisms for producing severity curves for fire one finds the three different regimes (MB, BE and FD) corresponding to three different ways in which fire propagates. To see why that is the case, we need to see how property losses can be modelled from first principles using graph theory.

4.2 Property Losses Using Graph (Network) Theory

4.2.1 The basic theory

Parodi & Watson (2019) presented an approach to modelling property losses from first principles based on graph theory, a branch of discrete mathematics.

A graph, which can also be a *network*, is an abstract mathematical object that can be thought of as a collection of nodes, some of which are connected via edges.

The paper presented two ways in which fire losses can be understood. In both cases, the approach relies on a representation of a property as a graph whose nodes represent rooms/units, with edges between two nodes if there is a passageway – that is, if direct fire propagation between the two corresponding rooms is possible.

In the simplest version, the value of each room/unit is the same, say 1, and each node reached by the fire is considered fully lost so that the loss for each node is either 0 or 1. In either approach, the fire is assumed to start at a randomly selected node, that node plus those nodes connected to it are assumed to have been reached by the fire and the loss is then simply equal to the number of nodes in that subset.

In this framework, the total insured value (TIV) is the number of nodes in the graph, while the MPL is taken to be the size of the largest connected component of the graph since a fire starting in one connected component cannot propagate to a separate connected component due to lack of passageways. This simple framework can of course be refined at will: the value of different nodes may vary, and partial losses at nodes allowed.

4.2.1.1 Static approach. The static approach starts from a weighted graph¹³ $G = (V, E, W)$, V being the (set of) nodes (or vertices), E the edges, and W the edges' weights. Losses are simulated by picking a node v^* at random as the origin of the fire, and then deleting each edge e in the graph with a given probability $(1 - w_e)$. The loss amount is given by the size of the connected component containing the node v^* in what remains of the graph G . Dividing by the MPL (the largest connected component of the graph *before* edge deletion) gives the damage ratio. The process is repeated for many simulations until the distribution of the damage ratios emerges. The algorithm is described in more detail in Parodi & Watson (2019) and a visual representation of it is shown in Figure 7. The simulation can be run for both a single property and for a portfolio of properties, in which case one must also pick properties at random.

4.2.1.2 Dynamic approach. An alternative approach is to view the fire as propagating through a building deterministically with the route defined by the graph connectivity, but for a random amount of time. In this approach, edge weights are related to the time it takes for fire to spread through a given passageway and the status of those passageways (e.g., open versus closed doors).

The fire will spread and burn for a time T ("delay"), until it is finally extinguished. In Parodi & Watson (2019), the delay distribution was assumed to be a simple exponential distribution $F_T(t) = 1 - \exp(-\lambda t)$, λ being the inverse of the expected time before the fire is put out, for lack of sufficient data to produce a data-driven or principle-based distribution.

As for the static approach, the simulation to produce a normalised severity curve starts by choosing a node where the fire starts. However, a random time t^* for the fire propagation is also selected. From the moment the fire starts, it traverses the graph deterministically. The time to

¹³A *graph* (Trudeau, 2003) is a mathematical structure comprised of a collection of *nodes* and a set of *edges* that link pairs of nodes. When every edge has a specific weight, the graph is known as a *weighted graph*. Unweighted graphs can be thought of as weighted graphs with all weights = 1. If there exists a *path* from node v to node v' through a sequence of connected edges, the two nodes are said to be connected. A *connected component of a graph* is a subset of the graph where every pair of nodes are connected. A directed graph, or *di-graph*, is a graph where the edges have a direction, meaning the order of the nodes the edge connects is crucial. The edges are represented by arrows. Directed graphs can also be weighted. One popular (if uneconomical) representation of graphs and di-graphs is as matrices where the entry at row j and column k is non-zero if and only if there is an edge connecting j to k , and the entry value is equal to the weight. The matrix will be symmetric for undirected graphs. Matrices will be used in Section 6 to represent supply chain networks.

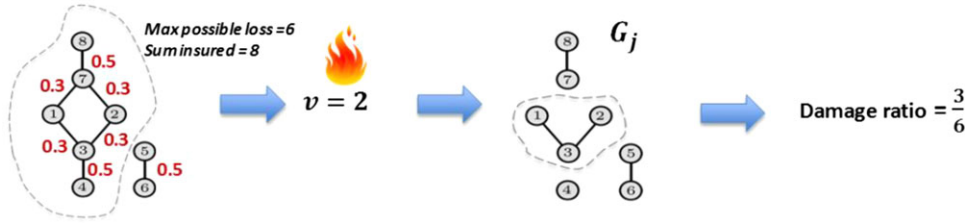


Figure 7. In this simulation example, we are assuming that each node represents a unit of worth equal to 1, so the total insured value (TIV) equals 8. The maximum possible loss (MPL) is equal to 6, the size of the largest connected component. The origin of the fire is selected at random (node 2 in the example), then the edges are removed with probability equal to 1 minus the weight of that edge. The size of the connected component that includes node 2 is then calculated, and the loss is divided by the MPL, giving the damage ratio. The process is then repeated for the desired number of scenarios. Figure taken from Parodi & Watson (2019).

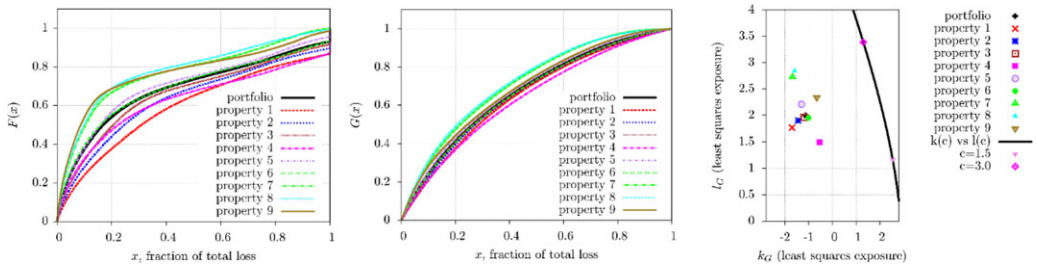


Figure 8. The result of the simulation with the random time approach with the following parameters: $\lambda = \frac{1}{2}$, $\text{Pr}(\text{open}) = 0.8$, $n_{\text{open}} = 1$, $n_{\text{closed}} = 10$, $n_{\text{wall}} = 30$. The severity (left) and exposure (centre) curves for different property structures and for the whole portfolio, obtained simulating 100,000 different scenarios. (Right) The parameters $k = \ln(b)$, $l = \ln(g - 1)$ of Bernegger’s curves for the various property structures and the portfolio and compared with the values of k and l corresponding to different values of c for the Swiss Re c curves (the black curve).

traverse an edge depends on the nature of the separation between two rooms: n_{open} , n_{closed} , n_{wall} represent the number of time steps it takes to traverse an edge representing an open door, a closed door, and a wall/floor respectively. For a particular simulation a door is assumed to be open with a certain probability $\text{Pr}(\text{open})$. The traversing progresses until time t^* is reached. See Parodi & Watson (2019) for details.

As for the static case, the simulation can be performed for a single property or for a portfolio. The results of the simulation with a given set of parameters are shown in Figure 8.

The Bernegger model gives a good approximation of the various curves produced by the simulation for both the individual properties and the portfolio, but the values of the parameters differ greatly from the Swiss Re c curves. This finding remains true for a wide range of different choices of the parameters.

Having summarised Parodi & Watson (2019), we next investigate the relationship between the MBBEFD parameters that approximate the severity curve of given graphs, and the type of temporal law for the spread of fire.

4.2.2 Bernegger (MBBEFD) curves and an interpretation of MB versus BE versus FD

Let us concentrate on fire damage and treat it as a physical process, using the random time approach of Section 4.2.1.2. Namely, a fire starts, spreads and is then put out (or at least brought under control so that no further damage occurs) after some time. If the fire spread is not stopped

within a certain period, there is a total loss. The spread of fire is not modelled as a random physical process here, but as a deterministic process. What is considered random is the cause of the fire, and how long it takes for the emergency services to react and stop the fire from doing further damage.¹⁴

Assume that the fractional loss is a deterministic function of a random time and further assume that the random time is drawn from the exponential distribution with parameter λ . We can write:

$$X = f(T), T \sim \text{Exp}(\lambda), \bar{F}_T(t) = e^{-\lambda t}, E[T] = \frac{1}{\lambda}$$

The as-yet unknown function f is called the evolution function and represents the fractional amount of damage done. It should be non-decreasing and has the boundary conditions

$$f(0) = 0, f(T > t_{\max}) = 1$$

for some maximum time t_{\max} . Given that the probability of a total loss $P[X = 1] = 1/g$, we can write:

$$\Pr[T > t_{\max}] = e^{-\frac{t_{\max}}{E[T]}} = \frac{1}{g}$$

such that

$$g = \exp\left(\frac{t_{\max}}{E[T]}\right) = e^{\lambda t_{\max}}$$

In other words, the MBBEFD parameter g can be described in terms of the time it would take a fire to spread and destroy the building completely and the mean time taken to respond and stop the fire from spreading further. This is a simple, intuitive, and natural interpretation¹⁵. Further, the parameter for the exponential distribution has been set (without loss of generality) to 1.

Now let us derive the evolution function f that corresponds to the CDF for the MBBEFD distribution that obeys the boundary conditions. This can be done by considering the probabilities (restricting to $0 < x < 1$ for now):

$$\bar{F}(x) = P[f(T) > x] = P[T > f^{-1}(x)] = e^{-\lambda f^{-1}(x)}$$

or

$$t = f^{-1}(x) = -\frac{\ln(\bar{F}(x))}{\lambda}$$

Using the definition of $\bar{F}(x)$, we can invert the expression above to yield x . Before doing that, however, we introduce new variables $\alpha = bg - 1, \beta = 1 - b$ to have an expression that covers all cases¹⁶.

$$x = f(t) = -\frac{\ln\left(\frac{\beta e^{\lambda t} + \alpha}{\beta + \alpha}\right)}{\ln(1 - \beta)}$$

¹⁴Here, we are considering a single building in isolation. If we were to consider a portfolio of properties, one might also treat which property is burning as a random variable.

¹⁵Moreover, recall that the Swiss Re curves are based on the parameter g being an exponential of parameter c - the typical scales in the two descriptions are similar.

¹⁶The case $bg = 1$ is equivalent to $\alpha = 0$ and the case $b = 1$ corresponds to the limit $\beta \rightarrow 0$. The denominator is always well-defined since $1 - \beta = b > 0$. Also, $\alpha + \beta = b(g - 1) \geq 0$.

We can calculate $f(t_{\max})$ using $e^{\lambda t_{\max}} = g$ and expanding the definitions of α and β :

$$f(t_{\max}) = -\frac{\ln\left(\frac{\beta g + \alpha}{\beta + \alpha}\right)}{\ln(1 - \beta)} = -\frac{\ln\left(\frac{1}{\beta}\right)}{\ln(b)} = 1$$

Having considered what happens for $0 < x < 1$, we can now apply our boundary conditions and write down the expression for $X \sim \text{MBBEFD}(b, g)$ rewritten in terms of the deterministic evolution of a process over a random time drawn from an exponential distribution:

$$X = f(T), T \sim \text{Exp}(\lambda), f(t) = \min\left(1, -\frac{\ln\left(\frac{\beta e^{\lambda t} + \alpha}{\beta + \alpha}\right)}{\ln(1 - \beta)}\right).$$

It should be emphasised that the above is formally *identical* to the MBBEFD distribution, the difference being in how it is written. The physical picture of a fire spreading over a random time is simply our motivation for choosing the form of the expression and how to interpret it.

4.2.2.1 Separating the MB, BE, and FD regimes. Using the above definitions, the three regimes correspond precisely to the sign of α : the MB regime corresponds to $\alpha = 0$, the BE regime to $\alpha > 0$ and the FD regime to $\alpha < 0$. We will show that the evolution function $f(t)$ exhibits three distinct behaviours in the three regimes.

By differentiating the evolution function with respect to time (for $0 < t < t_{\max}$), we can derive a differential equation form:

$$\frac{\ln(b)}{\lambda} f'(t) = \frac{\alpha}{\alpha + \beta} e^{\ln(b)f(t)} - 1$$

where we recall that $f(t)$ is an increasing function and that $\alpha + \beta \geq 0$. So, the three regimes are:

1. MB ($\alpha = 0$). Here we have already seen that $f(t) = -\frac{\lambda t}{\ln(b)}$ and the evolution function is *linear* in time.
2. FD ($\alpha < 0$). Here we use that $\alpha = bg - 1 < 0$, such that $0 < b < \frac{1}{g} < 1$ and $\ln(b) < 0$. Then (keeping only the signs and ignoring constants) $f'(t) \sim e^{-f(t)g}$ and we see that $f'(t)$ decreases with time and that the evolution function is *decelerating*.
3. BE ($\alpha > 0$). Now we must consider the values of b (since $b = 1$ is a special case):
 - o Case $\frac{1}{g} < b < 1$. For this case, $\ln(b) < 0$ and $f'(t) \sim \text{constant} - e^{-f(t)}$ is increasing with time such that the evolution function is *accelerating*.
 - o Case $b = 1$. Now $f(t) = \frac{e^{\lambda t} - 1}{g - 1}$ and the evolution function is exponentially increasing. This can also be written as $f'(t) = e^{\lambda t} \lambda / \alpha$ and again the evolution function is *accelerating*.
 - o Case $b > 1$. For this case, $\ln(b) > 0$ and $f'(t) \sim e^{X(t)}$ so again, the evolution function is *accelerating*.

In summary, the three different regimes correspond to whether the evolution function is decelerating (FD), linear (MB), or accelerating (BE).

4.2.2.2 Some examples. Figure 9 shows a few scenarios for the time evolution of fire, corresponding to Bernegger curves in different regions of the (b, g) space. The parameter for the exponential distribution is set to $\lambda = 1$ in all cases, without loss of generality because λ is simply a scaling factor.

4.2.2.3 Connection between the property graph topology and the parameters of the Bernegger distribution. We have rewritten the Bernegger distribution in terms of a system undergoing a deterministic evolution over an exponentially distributed random time. There are two parameters (g and b) and three types of behaviour. The parameter g dictates how long it takes relative to the

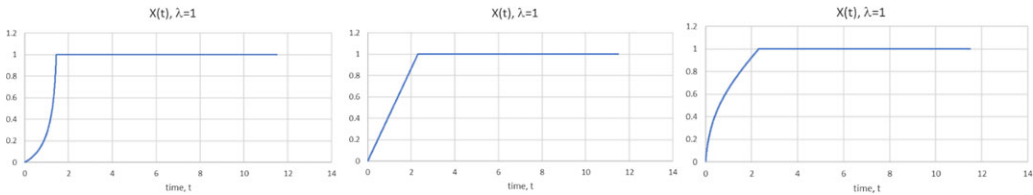


Figure 9. Left: example of the BE regime: time evolution for a Swiss Re curve with $c = 1.5$ (residential property). Increasing the value of c yields the same type of behaviour but with a lower acceleration/convexity; Centre: example of the MB regime: evolution for a Bernegger curve with $b = 0.1, g = 10$; Right: example of the FD regime: evolution for a Bernegger curve with $b = 0.01, g = 10$.

mean time taken to stop the process for there to be complete destruction. The parameter b governs how the process evolves: a larger value of b is associated with a system where the evolution is accelerating¹⁷. This provides an intuitive and appealing physical picture for the interpretation of Bernegger curves.

What is still missing is to make explicit the connection between a property's graph (its topology) and the type of statistic that best describes the evolution of the fire. Interestingly, all examples based on real-world topologies for household properties in Parodi & Watson (2019) led to parameters solidly belonging to the decelerating FD region. However, it is not difficult to imagine graphs that lead to an accelerating BE distribution. One such example is a graph where one *hallway node* is connected to a very large percentage of other nodes, and each node is separated from the hallway node by at most two degrees. In this case, if the fire originates in a node separated from another node by two degrees, after one step it will have affected a small number of other nodes, but one step later it will touch the hallway node and from there it will spread almost immediately almost everywhere, an obvious case of spread acceleration.

Another way of producing the BE behaviour is by considering a constructive total loss whenever the number of nodes exceeds a certain percentage of the whole graph. This will remove the effect of obstinate nodes that only burn after a very long time and provide an alternative mechanism for acceleration. It is like considering that alongside fire there are other related phenomena such as smoke that cause damage, so once a good percentage of the property is out of order the likelihood of salvaging the rest is reduced.

Similar considerations can be made for the static approach. The removal of certain edges with a given probability is unlikely to disconnect the hallway node from the rest of the graph, and therefore a total loss (or near-total loss) is more likely.

At the other end of the spectrum, we have graphs with poor connectivity. An example of this is a Cayley tree¹⁸ of degree m . If the depth of this graph is k , the total number of nodes is $1 + m + m^2 + \dots + m^k = \frac{m^{k+1}-1}{m-1}$. If a fire originates at a random point, it is more likely to originate close to the boundary, where most nodes are, and therefore to propagate slowly (of course if it originates at the root, it will expand quickly).

¹⁷Note that all Swiss Re curves are in the Bose–Einstein region provided the parameter c is between -0.48 and 21.48 .

¹⁸Some definitions may be helpful here. The **degree of a node** is the number of edges connected to the node. A **tree** is a graph with no loops. A **rooted tree** is a tree where one of the nodes has been designated as the **root**. The designation of a root establishes a hierarchical structure within the tree: every other node in the tree has exactly one path to the root node, and a parent-child relationship in the tree is created: every node except the root has **parent nodes**, and nodes may have **child nodes** (in which case they are called intermediate nodes) or not (in which case they are called leaf nodes or interface nodes). The **depth of a rooted tree** is the number of nodes in the longest path between the root node and any leaf node. A **Cayley tree of degree m** is a rooted tree that has a highly regular and symmetric structure: a root node of degree m , leaf nodes of degree 1 and intermediate nodes of degree $m + 1$. A tree structure is an idealised topology for a property where the removal of an edge (a fire route between two units) fully disconnects the property into two, then drastically limiting the spread of the fire. A Cayley tree has been chosen as an example because it is especially simple to describe mathematically.

4.3 Percolation Theory and the Spread of Fires

The spread of fire is akin to an epidemic and can be formalised in terms of percolation theory. Percolation theory describes how the addition of nodes or edges to a graph changes the behaviour of the graph. Historically, the motivation for this type of theory is explaining percolation, the movement of fluids through a porous material, but it has now developed into something much more general that sheds light on phase transition and critical phenomena.

Percolation theory has been applied to model the spread of forest fires (Rath and Toth, 2009; van den Berg, and Brouwer, 2006). Those results don't apply immediately to the types of graphs considered in Parodi & Watson (2019) and in this paper, because we mainly use small graphs while most of the interesting results appear when considering the asymptotic behaviour of networks.

Interestingly, the Bose–Einstein statistic and the Fermi–Dirac statistic emerge quite naturally in the study of percolation (Bianconi, 2001; 2015), suggesting that the Bernegger curves may provide more than a purely formal connection to the losses caused by the spread of fires.

To summarise, there appear to be two types of networks: power-law networks that give rise to a BE statistic and Cayley trees that give rise to an FD statistic. According to Bianconi and Barabási (2001), growing networks self-organise into complex networks in which some nodes have an outsized number of connections to the rest of the network. These networks can be mapped into an equilibrium Bose–Einstein gas, with energy levels corresponding to nodes and particles corresponding to edges. This gas exhibits condensation as well – that is, a single node captures a macroscopic fraction of edges. How can this be related to a fire? One way is for the nodes to represent property units already reached by the fire, and new nodes being added out of an existing network of inactive nodes representing the existing property units with their connections. However, currently this is only a research suggestion.

Bianconi (2015) showed that if the structure of the graph is a growing Cayley tree, the distribution of the energies at the interface (i.e., at the leaf nodes) converges to an FD distribution.

In both Bianconi & Barabási (2001) and Bianconi (2015) the key idea is to create a mapping between the graph model and a condensed state with various energy levels. Each node corresponds to an energy level, while an edge corresponds to two non-interacting particles at different energy levels.

4.4 Explosion Risk

So far, we have only looked at fire risk. However, property policies traditionally cover insurance against the so-called FLExA perils of fire, lightning, explosion, aircraft collision, and All Risks policies also cover natural catastrophes. Explosion risk was dealt with in Parodi & Watson (2019) using weighted directed graphs and we refer the interested reader to that paper.

4.5 Aviation Hull Losses Arising from Cabin Fires

As an alternative example of how it is possible to model fire losses from first principles, we have considered Aviation Hull and specifically losses arising from post-crash cabin fires, which result from ground collisions or high-speed landings. A study by NASA¹⁹ in 1982 on a B737 aircraft on how cabin fires develop showed a slow initial temperature rise, followed by a rapid increase, reaching flashover (a near-simultaneous ignition of most combustible material in an enclosed area). Subsequent studies by FAA²⁰ and NIST²¹ using modern aircraft material suggested the flashover point would occur *on average* 325 seconds after ignition. Once flashover is reached, the hull experiences a total loss.

¹⁹Full-scale flammability test data for validation of aircraft fire mathematical models <https://ntrs.nasa.gov/citations/19820013292>.

²⁰Background : FAA Fire Safety - <https://www.fire.tc.faa.gov/Research/Background>.

²¹Fire Dynamics – NIST - <https://www.nist.gov/el/fire-research-division-73300/firegov-fire-service/fire-dynamics>.

An independent analysis of the damage ratio for fire losses using a third-party database, analysing over 60 fire/explosion losses of wide-body aircraft between 1974 and 2020, reveals a 30–37% probability of total hull loss in the event of the cabin fire.

Combining these two types of analysis, we can assume that (a) the hull damage is a function of time from ignition, analogous to the approach we outlined in Section 4.2.1.2; (b) the distribution of time before the fire is extinguished is exponential: $F_T(t) = 1 - \exp(-\lambda t)$, where $\lambda = 1/325s$, putting the probability of reaching flashover at $\sim 37\%$; (c) the distribution of the damage ratio follows a Bernegger distribution (for example, a Swiss Re c curve) with a probability of total loss of 37%.

This analysis is of course very preliminary, as it is based on several conjectures on the relationship between hull damage and physical parameters and is affected by both model and parameter uncertainty, and more research is needed.

5. Modelling the Severity of Natural Catastrophe Losses

Natural catastrophe (cat) modelling relies on geophysical and engineering models rather than historical loss fitting, although historical losses obviously play a role. It is a clear example of where modelling from first principles already exists. We will therefore not dwell on this topic for long as it would mean mostly repeating results already well known in the literature, but will use a single example, earthquake modelling, for illustration purposes.

5.1 Modelling Losses Arising from Earthquakes

An example is provided by earthquake modelling. Loss data suggest that the severity of losses broadly follow a Pareto law with exponent α of around 1 (Mitchell-Wallace et al., 2012). This can be related to the Gutenberg–Richter law, which is an empirical relationship between the magnitude M of an earthquake and the total number N of earthquakes in any given region/time period of at least that magnitude: $\log_{10} N = a - bM$. The parameter b is typically between 0.8 and 1.5 in seismically active regions. This relationship needs to be paired up with that which gives the energy E released during the earthquake: $\log_{10} E = c - dM$. The combination of the equations for N and E leads to the following power law for the number of earthquakes releasing an energy larger than E :

$$N(E) \propto E^{-\frac{d}{b}}$$

Hence, the probability that an earthquake releases an energy larger than E is:

$$\bar{F}(E) \propto E^{-\frac{d}{b}}$$

which is a Pareto distribution with $\alpha = \frac{d}{b}$. The insured loss will then depend on the energy released and the vulnerability of the structures exposed to the earthquake.

There is no unique explanation from first principles as to why this empirically observed law holds, but various possible mechanisms have been put forward. The most interesting explanation (Bak & Tang, 1988) is perhaps the one based on the concept of **self-organised criticality**²², which suggests that the Earth's crust can be modelled as a dynamical system that is both dissipative (i.e., it releases energy) and is spatially extended with a number of degrees of freedom that is approximately infinite. The interactions between various physical, chemical, and mechanical processes in the Earth's crust lead to a critical state where earthquakes can occur spontaneously,

²²Self-organized criticality (SOC) is a concept in the field of complex systems that refers to the idea that some systems, without external tuning, can evolve towards a critical state, where small perturbations can cause large-scale effects. The classical example is that of a heap of sand, to which you slowly add one grain at a time. Eventually it will reach a state where a small nudge will cause an avalanche, sending many grains tumbling down. This is an example of a system that is self-organized into a critical state. In this state, the size of the avalanches follows a power-law distribution, meaning that small avalanches are much more frequent than large ones, but large ones are not impossible.

analogous to what is often cited as an example of self-organised criticality, the grain of sand that leads to an avalanche of the heap of sand it falls on. The Gutenberg–Richter relationship reflects the scaling properties of these interactions and the distribution of energy stored in the Earth’s crust and leads to a power-law distribution of earthquake sizes and ultimately payouts.

5.2 Modelling other Natural Catastrophes

Note that, as mentioned in Bak and Tang (1988), dynamical phenomena with power-law correlation function appear to be widespread in nature and are seen in weather, landscapes and other areas, so the pattern described here is likely to be found for other types of natural catastrophes (Mandelbrot, 1982); similar to Helmstetter’s approach of fractal scaling (2003).

5.3 Non-Geophysical Explanations

Apart from these geophysical explanations, it should be noted that another plausible mechanism to explain the distribution of loss amounts in earthquakes and catastrophes in general could be related to the distribution of city sizes – see Zipf’s law in Section 3.2.1. It makes sense that earthquakes of the same magnitude will have different impacts depending on the size of the population affected, and therefore an earthquake in a large city will cause much more damage than a similar earthquake in a small city.

6. Modelling Contingent Business Interruption (Supply Chain) Losses

The insurance industry has traditionally dealt with business interruption/loss of revenue impacts in a rather arbitrary manner. The Business Interruption premium rate is often set as a simple multiple of the Property Damage rate. The loss data available is often expressed in terms of financial amounts as opposed to physical delay times (numbers of days lost) and so often combines information on:

- the interruption period
- the financial impact per day (which will differ for partial losses and total losses²³)
- the basis of recovery (fixed costs only; full loss of profit etc.).

The risk can become even more complex where third-party revenues are impacted, as in contingent business interruption (CBI). Underwriters will often try to identify a “network” of potential loss scenarios where coverage is scheduled and offer a lower limit for any “unscheduled” losses that fall outside of these scenarios.

With the recent advances in mapping technology and real-time estimates of values at risk it is possible for underwriters to set up all their exposures and create a dynamic dashboard reflecting any peak concentrations, such as a terminal through which a significant amount of oil and gas production may flow or a critical pipeline to transport oil or gas from a producing field to a terminal or a processing facility/refinery. To truly estimate an expected maximum loss (EML) for such a network, probability distributions would be required to reflect:

- the chance of a significant loss event (property damage, terrorism/war, cyber, pandemic . . .)
- the likely downtime period
- the recovery period (considering that production may return on a partial basis before full production is restored)

²³See LMA bulletins 5607/5608 that intend to create a standard basic wording for calculation of business interruption losses and the application of daily (monthly) indemnity caps.

- any mitigation that may be available, for example, the ability to re-route production (at a cost) via an alternative pipeline or through use of an alternative processing facility.

The loss of revenue impact from a key distribution centre can be many times the direct value of the site/building. This revenue impact could also be triggered by other risks such as cyber attack, terrorist activity either directly at the location or in the immediate vicinity, or the impact of a pandemic impact on the workforce.

6.1 Modelling Supply Chain Risks using Networks

Most academic research tends to focus on Property Damage modelling and the interpretation of risk engineering surveys to evaluate the physical risk of loss (estimated maximum loss studies). There is far less research available on Business Interruption exposures.

Perhaps even more than in the case of property risk, supply chain risk lends itself naturally to being analysed through a directed graph²⁴ representation. The nodes of the supply chain represent suppliers/assets and there is a directed edge connecting A to B if A provides materials or services to B needed (or useful) for the overall functioning of the supply chain. The direction of the edge is of course essential as the role of supplier/receiver cannot be reversed²⁵. The edge from A to B will in general have a weight, reflecting the proportion (or absolute amount) of product that B receives from A. Nodes might belong to different classes: for example, one class of nodes might represent production sites and another class of nodes might represent warehouses. Edges might also be of different natures: for example, road routes versus sea routes or even internet routes. All of this suggests that an appropriate representation of complex supply chains may require a graph with a very rich structure. We will typically refer to such a complex graph as a “network” as it is a term more commonly used in the context of supply chains.

A more sophisticated analysis would also consider the time element, for example, to adjust the exposure for insuring conditions such as waiting periods, indemnity periods and maximum daily recovery rates etc. As a simple example consider a typical oil and gas exposure where the network can be described as in Figure 10. The Direct Business Interruption values at risk and the relative dependency is summarised in Table 1.

The critical dependencies can be summarised as follows:

- Assets 2 to 5 are routing production via Asset 1 (a gathering platform for the whole field) so are 50% dependent: for example, if Asset 1 is closed/lost then some routing/mitigation may be possible but Assets 2 to 5 would have to reduce production by 50%.
- Terminals 1 and 2 are each taking 50% of the production from Asset 1.
- Pipeline 1 gets 25% of its throughput from Terminal 1. Pipeline 2 gets 25% of its throughput from Terminal 2.
- Refinery 1 gets 25% of its feedstock from Pipeline 1. Refinery 2 gets 25% of its feedstock from Pipeline 2.

We can then calculate Business Interruption effects as a loss propagates through the network. For, example a loss of Asset 1 (a platform) would impact the direct BI value (\$100M) but also 50% of the value at platforms 2/3/4/5, that is, 50% of \$175M = \$87.5M. The total exposure from platform one could be considered as \$187.5M.

When considering insurance some attention may need to be paid to any sub-limits being applied for contingent business interruption. If, for example, the BI is being declared as \$100M at

²⁴A directed graph is a graph where the edges have specified directions and they are called directed edges.

²⁵Naturally, it might be the case that A supplies products/materials to B and B supplies other products/materials to A, in which case you will have two separate edges, one from A to B and another from B to A.

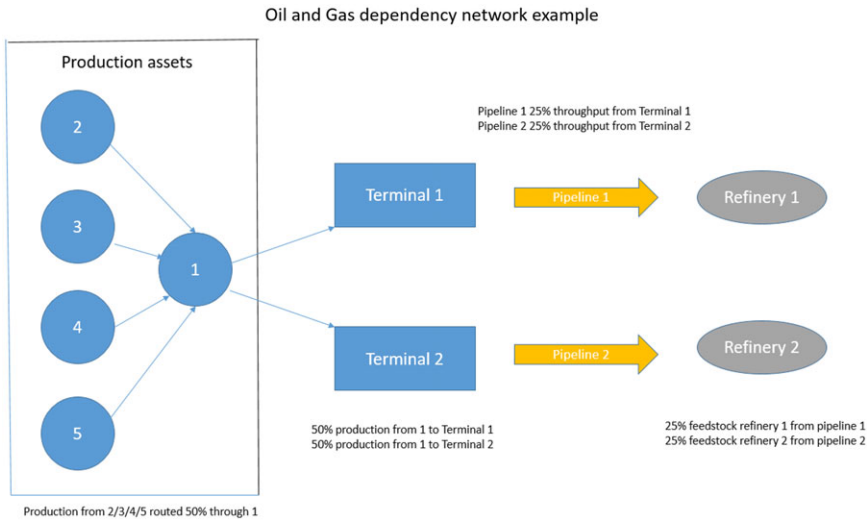


Figure 10. Oil and gas network example.

Asset 1 with a blanket \$50M available for contingent BI then the overall recovery may then be limited to \$150M etc.

As we move through the supply chain there are also lower-order effects to consider (third order, fourth order etc.). Consider, for example, Refinery 1. The direct BI exposure is declared as \$500M. The loss of Refinery 1 would, however, also trigger:

- a 25% loss at Pipeline 1 (\$37.5M – second order effect)
- a loss capacity at pipeline 1 would then trigger a loss at terminal 1 ($\$75M \times 25\% \times 25\% = \$4.69M$ – third order effect)
- a loss of capacity at Terminal 1 would trigger a loss at Asset 1 ($\$100M \times 50\% \times 25\% \times 25\% = \$3.13M$ – fourth order effect)
- a loss of capacity at Asset 1 would trigger a loss at assets 2/3/4/5 – total \$2.73M
 - Asset 2 $\$25M \times 50\% \times 50\% \times 25\% \times 25\% = 0.391M$
 - Asset 3 $\$25M \times 50\% \times 50\% \times 25\% \times 25\% = 0.391M$
 - Asset 4 $\$25M \times 50\% \times 50\% \times 25\% \times 25\% = 0.391M$
 - Asset 5 $\$100M \times 50\% \times 50\% \times 25\% \times 25\% = 1.563M$

The full lower-order effects were calculated as in Table 2.

As a sensitivity check let’s consider the impact if the direct values at risk at the five production assets were to double (e.g., because of higher oil and gas prices) and, due to supply constraints, we assume Pipelines 1 and 2 are now receiving 50% of their throughput from Terminals 1 and 2 respectively. The business interruption lower-order effects are now calculated as shown in Table 3.

This can be summarised by considering the relationship between direct, first-order BI and indirect lower-order BI effects as in Table 4.

When considering an overall business interruption limit, however, it would be unusual for the buyer (or the broker) to analyse the risk at this level of detail. Using the example above, it may be more common to declare a simple catch-all additional \$100M for CBI based on Table 4 (the original network), even though the knock-on impact of the factors described in Scenario 2 suggests this could lead to significant underinsurance.

It is possible to build a simple pricing tool that accounts for these factors and tests the impact of varying the coverage such as increased waiting periods or sub-limits on indirect losses.

Table 1. The dependency matrix for Figure 10. As mentioned in Section 4.2.1, such a matrix can be viewed as a representation of a weighted (directed) graph, with an edge from j to k if the value at (j, k) is > 0

ID	Name	Type	Direct BI (\$M)	Direct										
				1	2	3	4	5	6	7	8	9	10	11
1	Asset 1	Platform	100	100%	50%	50%	50%	50%	0%	0%	0%	0%	0%	0%
2	Asset 2	Platform	25	0%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%
3	Asset 3	Platform	25	0%	0%	100%	0%	0%	0%	0%	0%	0%	0%	0%
4	Asset 4	Platform	25	0%	0%	0%	100%	0%	0%	0%	0%	0%	0%	0%
5	Asset 5	Platform	100	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	0%
6	Terminal 1	Terminal	75	50%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%
7	Terminal 2	Terminal	75	50%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%
8	Pipeline 1	Pipeline	150	0%	0%	0%	0%	0%	25%	0%	100%	0%	0%	0%
9	Pipeline 2	Pipeline	150	0%	0%	0%	0%	0%	0%	25%	0%	100%	0%	0%
10	Refinery 1	Refinery	500	0%	0%	0%	0%	0%	0%	0%	25%	0%	100%	0%
11	Refinery 2	Refinery	500	0%	0%	0%	0%	0%	0%	0%	0%	25%	0%	100%

Table 2. Business interruption: Lower order effects

Name	Direct	2nd order	3rd order	4th order	5th order	Total BI
Asset 1	100	87.50	–	–	–	187.50
Asset 2	25	–	–	–	–	25.00
Asset 3	25	–	–	–	–	25.00
Asset 4	25	–	–	–	–	25.00
Asset 5	100	–	–	–	–	100.00
Terminal 1	75	50.00	43.75	–	–	168.75
Terminal 2	75	50.00	43.75	–	–	168.75
Pipeline 1	150	18.75	12.50	10.94	–	192.19
Pipeline 2	150	18.75	12.50	10.94	–	192.19
Refinery 1	500	37.50	4.69	3.13	2.73	548.05
Refinery 2	500	37.50	4.69	3.13	2.73	548.05

7. Modelling the Severity of Cyber Business Interruption Losses

Cyber Business Interruption (Cyber BI) is difficult to model via experience rating because of the paucity of data. For this reason, a risk engineering approach that focuses on the underlying risk mechanism and disruption pattern of Cyber BI incidents is both interesting from a theoretical point of view and practically helpful.

Cyber BI differs from conventional property damage-related BI in some important ways:

- The disruption period (which doesn't involve restoration of physical assets) is much shorter, translating into shorter contractual indemnity periods
- Cyber BI is not location-specific: a cyberattack can trigger business interruption losses in different territories

Table 3. Business interruption: Lower order effects, revised

Name	Direct	2nd order	3rd order	4th order	5th order	Total BI
Asset 1	200	175.00	–	–	–	375.00
Asset 2	50	–	–	–	–	50.00
Asset 3	50	–	–	–	–	50.00
Asset 4	50	–	–	–	–	50.00
Asset 5	200	–	–	–	–	200.00
Terminal 1	75	100.00	87.50	–	–	262.50
Terminal 2	75	100.00	87.50	–	–	262.50
Pipeline 1	150	37.50	50.00	43.75	–	281.25
Pipeline 2	150	37.50	50.00	43.75	–	281.25
Refinery 1	500	37.50	9.38	12.50	10.94	570.31
Refinery 2	500	37.50	9.38	12.50	10.94	570.31

Table 4. Direct BI versus Indirect BI: Original and revised network

Name	Original network			Revised network		
	Direct	Indirect	% Indirect	Direct	Indirect	% Indirect
Asset 1	100.00	87.50	87.5%	200.00	175.00	87.5%
Asset 2	25.00	–	0.0%	50.00	–	0.0%
Asset 3	25.00	–	0.0%	50.00	–	0.0%
Asset 4	25.00	–	0.0%	50.00	–	0.0%
Asset 5	100.00	–	0.0%	200.00	–	0.0%
Terminal 1	75.00	93.75	125.0%	75.00	187.50	250.0%
Terminal 2	75.00	93.75	125.0%	75.00	187.50	250.0%
Pipeline 1	150.00	42.19	28.1%	150.00	131.25	87.5%
Pipeline 2	150.00	42.19	28.1%	150.00	131.25	87.5%
Refinery 1	500.00	48.05	9.6%	500.00	70.31	14.1%
Refinery 2	500.00	48.05	9.6%	500.00	70.31	14.1%

To derive Cyber BI curves we will adopt a risk engineering approach focusing on the underlying risk mechanisms. Several loss scenarios are considered such as an IT service provider (cloud) outage, a local IT service outage, and a supplier’s outage due to Cyber risks.

7.1 Disruption Pattern for a Cyber BI Event

The extent of Cyber BI depends on the disruption pattern: the duration of the outage, the availability and effectiveness of loss mitigation measures, the repair/replacement of information technology (IT)/operational technology (OT) services, and eventually the time required to fully restore operations of the insured.

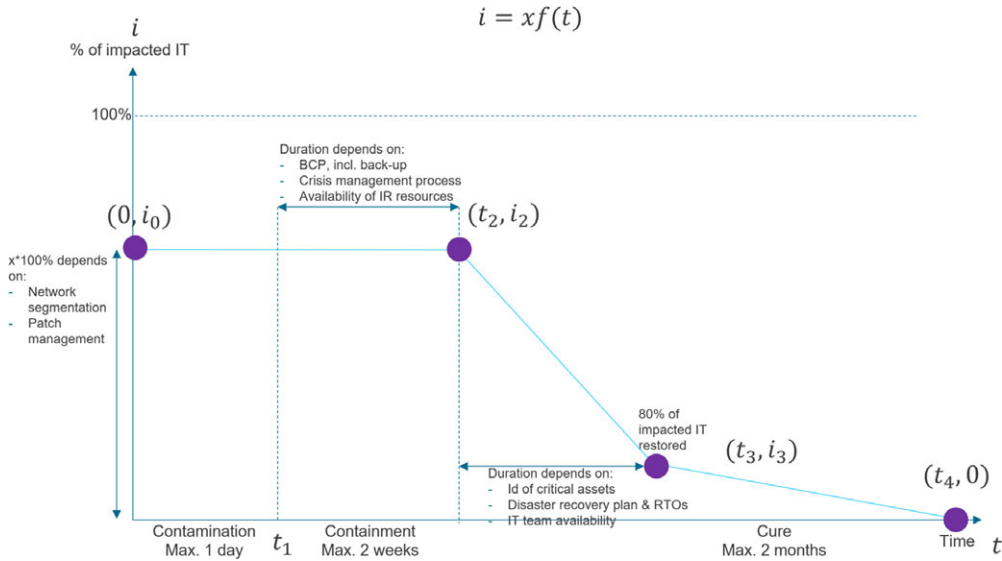


Figure 11. A simplified, linearised version of the cyber disruption pattern.

To model the disruption pattern, we assume the pattern is composed of three phases:

1. Contamination – cyber threats are starting to emerge and impacting the operation of the insured by up to $X\%$ depending on the insured’s cybersecurity posture²⁶ and business structure/organisation.
2. Containment – once the threat is discovered, the insured will contain the impact and reduce the adverse impact. The duration of this phase depends on several factors, for example, the insured’s business continuity plan, their crisis management process and third-party assistance, and the resources available for incident response.
3. Cure – in this phase, the insured’s IT-related business is restored up to a certain level. The time required for this will depend on the identification of critical cyber assets, disaster recovery plan, the recovery time objective (RTO), and the availability of the IT team. The full recovery of the business would further depend on some non-IT factors and may take considerably more time.

Figure 11 shows a simplified and linearised representation of the cyber disruption pattern. Some terminology is needed to explain the graph:

- x is the (estimated) maximum percentage of interruption that can be caused by a cyber-related issue. This number depends on the type of occupancy, network segmentation, patch management, and is normally below 100%: for example, if the reservation system is down, a hotel can still work to some extent without the system.
- $f(t)$ describes the estimated interruption pattern: hence, $xf(t)$ represents the percentage of interruption at time t , which will typically decline with time as the IT system is brought back to full functionality. The times t_1, t_2, t_3 can be considered realisations of random variates T_1, T_2, T_3 describing the onset of the three different phases.

²⁶Cybersecurity posture refers to the cyber defence measures of the insured, for example, authentication, employee training, deployed security system/solutions and the collective effectiveness of the combination of individual measures.

7.2 Calibration of Parameters

More detailed calibration work has been done on how to assess the maximum possible durations of each period, and the maximum level of interruption (*x*) of IT-related revenue disruption. Table 5 is an illustration of such an assessment for a company of a certain size in a given country, for 3 NACIS codes.

Table 5. Example of assessment of the parameters of a cyber disruption pattern.

	$\max(t_1)$	$\max(t_2-t_1)$		$\max(t_3-t_2)$	$\max(t_4-t_2)$	$\max(x)$
NAICS-based segmentation	Max contamination (days)	Max containment time (days)	RM process disruption (Cyrius)	Max time to restore 80% of impacted IT (days)	Max time to restore 100% of impacted IT (days)	Max(x)
Agriculture, forestry, fishing and hunting	1	14	1.53	38.45	115.34	0.38
Mining, quarrying, and oil & gas extraction	1	14	1.84	43.25	129.76	0.49
Utilities	1	14	2.00	45.66	136.97	0.55

7.3 Simulation of Loss Scenarios

The computation of insured BI losses is sometimes complex and differs significantly depending on the wording of each policy and local practices, notably loss of gross margin, loss mitigation cost and increased costs of working. However, our intention is to identify and quantify the main components and then build a model.

Let \bar{L}_d denote the average daily loss if the IT-related activity of the insured is completely interrupted ($X = 100\%$). We could then calculate the insured BI loss using:

$$L_{BI}(x, f(t)) = \int_0^{t_4} i(t)\bar{L}_d dt = \bar{L}_d x \int_0^{t_4} f(t) dt$$

Monte-Carlo simulations can then be performed to generate one million Cyber BI scenarios for each sector. A few random scenarios generated for the finance and insurance sector are shown in Figure 12.

Using this simulation, we can produce a distribution of the possible losses for a particular sector and derive from that the corresponding exposure curve. Figure 13 shows both for finance and insurance.

One advantage of this approach is that the MPL is approximately equal to the maximum simulated loss if the number of simulations is sufficient. For finance and insurance, this appears to be close to $48.5 \bar{L}_d x$. This can, however, also be calculated analytically assuming the estimated maximum time for each interval; the estimated maximum time will depend on the industry sector.

The empirically derived CDF for the damage ratio Y (loss as a percentage of MPL) can be denoted as $F(y) = P(Y \leq y)$. The damage ratio is:

$$Y = \frac{L_{BI}(x, f(t))}{MPL_{BI}} = \frac{\bar{L}_d \int_0^{t_4} xf(t) dt}{M \bar{L}_d x} = \frac{\int_0^{t_4} f(t) dt}{M}$$

where M is 48.47 according to the calculations above for the finance and insurance sector.

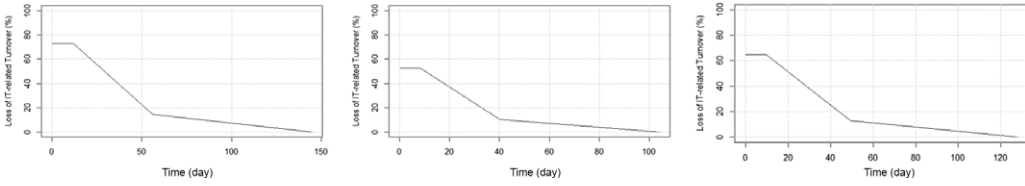


Figure 12. Three random scenarios generated by varying the parameters of the cyber disruption period.

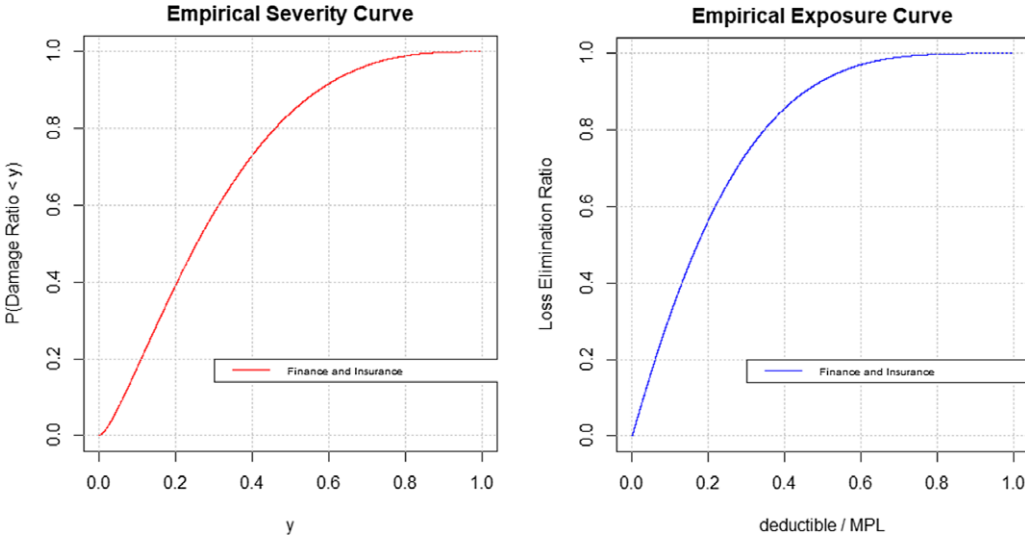


Figure 13. The empirically derived CDF plot for the damage ratio (left) and exposure curve (right) for the finance and insurance sector.

The exposure curve can then be derived from the corresponding (empirical) severity curve by the standard relationship $G(d) = d \int_0^{\bar{F}(y)} dy / \int_0^{\bar{F}(y)} dy$, where d is the normalised deductible $d = \frac{\text{deductible}}{MPL_{BI}}$ and $\bar{F}(y) = 1 - F(y)$.

8. Conclusions

Given the wide remit of our working party, and the fact that different covers will be modelled differently in terms of first principles, a variety of techniques and approaches was to be expected. Despite this, a few significant common threads have emerged.

Graph (network) theory appears to be a natural choice to model losses of properties that can be broken down into different units/components that are connected to one another. It also finds a natural use in modelling financial losses arising from failures in actual networks such as a supply chain, and in contingent business interruption.

Another broad strand is that of stochastic processes, which will be no surprise to actuaries since the loss generating process is itself a stochastic process. To be more specific, loss count processes are best viewed as examples of pure-jump processes, stationary or not. As for the severity of losses, especially in the context of casualty (liability) insurance, the ubiquitous presence of the Pareto behaviour for large losses can be explained by the presence of underlying random growth processes.

8.1 Limitations and Future Research

Each of the topics addressed in this paper is broad enough to merit a separate paper. In each case, more research is needed to flesh out the theoretical arguments fully and bring further empirical support.

In Section 2, more research is needed on modelling the causes of non-stationarity, whether natural or man-made, and the size of the clusters where such clusters are observed.

In Section 3, conjectured mechanisms leading to a Pareto behaviour need to be confirmed and articulated further.

In Section 4, more research is needed on the relationship between graph structure and exposure curve and why most curves used in practice appear to be in the Bose–Einstein region while most curves originating from graphs produce exposure curves that are in the Fermi–Dirac region. The connection with percolation theory needs to be investigated further. Research on other types of cover such as extended warranty, machinery breakdown, and satellite insurance can also be carried out through the lenses of risk engineering and network theory.

Further work would be needed to calibrate the dependency matrix approach in Section 6 by reviewing actual historical loss events and to consider (from engineering and supply chain data) a suitable probability distribution for downtime and the time to reinstate facilities.

Since cyber is a relatively new risk and is continually evolving, the details of the approach in Section 7 may quickly become obsolete and models will need to be updated accordingly to keep abreast of these changes.

Acknowledgments. A previous version of this paper was presented at the International Conference of Actuaries 2023 in Sydney, and was awarded the Best General Insurance/ASTIN paper prize. Our thanks to Frank Cuypers, Yuriy Krvavych and the ASTIN section for encouraging us to create the working party from which this paper originated; and to Christian Levac, Walter Neuhaus and Eric del Moro for their help through the process. We also thank our companies: SCOR, Gallagher Specialty, WTW, QBE, Hiscox, Acasta, and Marsh for their support in this research initiative. We are indebted to several people for advice on the actual content of this paper. Pierre-Louis Lions crucially pointed us in the fruitful direction of random growth models when exploring the severity of liability losses. Balázs Ráth and Ginestra Bianconi helped us with discussions around percolation. Stephen Mildenhall provided helpful feedback and material on Lévy processes and reviewed the rest of the paper. Jean-Stéphane Bodo pointed us to the emergence of power laws in the context of natural catastrophes. Massimo Vascotto pointed us to helpful information around the process for establishing liability payouts in Italy and other countries. Bernard Wong pointed us to useful material around non-stationary Poisson processes. One anonymous reviewer provided useful pointers to the Poisson binomial model and the Hawkes process. Finally, we would like to thank Shane Lenney for reviewing this paper and providing us with helpful advice

References

- Anscombe, F. J.** (1950). Sampling theory of the negative binomial and logarithmic series distributions. *Biometrika*, **37**(3/4), 358–382. <https://www.jstor.org/stable/2332388>
- Appelbaum, S.** (2009). *Lévy Processes*. Oxford University Press.
- Avanzi, B., Taylor, G., Wong, B., & Xian, A.** (2021). Modelling and understanding count processes through a Markov-modulated non-homogeneous Poisson process framework. *European Journal of Operational Research*, **290**(1), 177–195. <https://doi.org/10.1016/j.ejor.2020.07.022>.
- Bak, P., & Tang, C.** (1988). Earthquakes as a self-organized critical phenomenon. *Journal of Geological Research*, **94**(B11), 15635–15637.
- Barabási, A. L., & Bianconi, G.** (2001). Emergence of scaling in random networks. *Physical Review Letters*, **86**(24), 5632–5635. <https://doi.org/10.1103/PhysRevLett.86.5632>
- Barndorff-Nielsen, O., & Shephard, N.** (2012). Basics of Lévy processes. Lévy driven volatility models (Chapter 2), available at <https://pure.au.dk/ws/files/193874249/levybook.pdf>
- Beard, R.E., Pentikäinen, T., Pesonen, E.** (1984). *Risk Theory: The Stochastic Basis of Insurance* (3rd ed.). Chapman & Hall.
- Benckert, L.** (1962). The lognormal model for the distribution of one claim. *ASTIN Bulletin: the Journal of the IAA*, **2**(1), 9–23. <https://doi.org/10.1017/S0515036100007583>
- Bernegger, S.** (1997). The Swiss Re exposure curves and the MBBEFD distribution class. *ASTIN Bulletin*, **27**(1), 99–111.
- Bianconi, G.** (2001). Bose-Einstein condensation in complex networks. *Physical Review E*, **64**(016114), 1–4. <https://doi.org/10.1103/PhysRevE.64.016114>

- Bianconi, G.** (2015). Growing Cayley trees described by fermi distribution. *Physical Review E*, **92**(1), 012803. <https://doi.org/10.1103/PhysRevE.92.012803>
- Champernowne, D. G.** (1953). A contribution to the theory of economic growth. *Economic Journal*, **63**(250), 318–335.
- Embrechts, P., Klüppelberg, C., & Mikosch, T.** (1997). *Modelling Extremal Events: For Insurance and Finance*. Springer-Verlag.
- Fackler, M.** (2013). Reinventing Pareto: Fits for both small and large losses. ASTIN Colloquium, The Hague.
- Gabaix, X.** (1999). Zipf's law for cities: an explanation. *The American Economic Review*, **89**(2), 129–137.
- Gabaix, X., Lasry, J.-M., Lions, P.-L., & Moll, B.** (2016). The dynamics of inequality. *Econometrica*, **84**(6), 2071–2111.
- Gardiner, C.** (2009). *Stochastic Methods: A Handbook for the Natural and Social Sciences*. Springer.
- Hawkes, Alan G.** (2018). Hawkes processes and their applications to finance: A review. *Quantitative Finance*, **18**(2), 193–198.
- Helmstetter, A.** (2003). Is earthquake triggering driven by small earthquakes? *Physical Review Letters*, **91**(5). <https://doi.org/10.1103/PhysRevLett.91.058501>
- Klugman, S. A., Panjer, H. H., & Willmot, G. E.** (2012). *Loss Models: From Data to Decisions*. John Wiley & Sons.
- Knecht, M., & Küttel, S.** (2003). The Czeledin Distribution Function. In *XXXIV ASTIN Colloquium, Berlin*.
- Kozubowski, T., & Podgorski, K.** (2009). Distributional properties of the negative binomial Lévy process. *Probability and Mathematical Statistics*, **29**, 43–71.
- Laub, P. J., Lee, Y., & Taimre, T.** (2021). *The Elements of Hawkes Processes*. Springer.
- Mandelbrot, B. B.** (1982). *The Fractal Geometry of Nature*. New York: W. H. Freeman.
- Meyers, G.** (2007). The common shock model for correlated insurance losses. *Variance*, **1**(1), 40–52.
- Mildenhall, S. J.** (2017). Actuarial geometry. *Risks*, **5**(2), 31.
- Mitchell-Wallace, P., Wilson, D., & Peacock, N.** (Eds.). (2012). *Natural Catastrophe Risk Management and Modelling*. John Wiley & Sons.
- Parodi, P., & Watson, P.** (2019). Property graphs – a statistical model for fire and explosion losses based on graph theory. *ASTIN Bulletin*, **49**(2), 263–297. <https://doi.org/10.1017/asb.2019.4>
- Parodi, P.** (2023). *Pricing in General Insurance, 2nd Ed.* New York: CRC Press, New York.
- Patrik, G., Bernegger, S., & Rüegg, M. B.** (1999). The use of risk adjusted capital to support business decision-making, in *Casualty Actuarial Society Forum, Spring* (pp. 243–334). Casualty Actuarial Society.
- Rath, B. and Toth, B.** (2009). Erdos-Renyi random graphs %2B forest fires = self-organized criticality. *Electronic Journal of Probability* **14**, Paper no. 45, 1290–1327. <https://doi.org/10.1214/EJP.v14-653>. <https://projecteuclid.org/euclid.ejp/1464819506>.
- Riegel U** (2008). Generalizations of common ILF models. *Blätter der DGVFM* **29**(1), 45–71.
- Ross, S. M.** (2003). *Introduction to Probability Models, 8th Ed.*, Academic Press.
- Simon, H. A., & Bonini, C. P.** (1958). The size distribution of business firms. *The American Economic Review*, **48**(4), 607–617.
- Taleb, N. N.** (2020). *Statistical Consequences of Fat Tails: Real World Preasymptotics, Epistemology, and Applications*. Technical Incerto – The Technical Incerto Collection, Scribe Media.
- Trudeau, R. J.** (2003). *Introduction to Graph Theory*. New York: Dover Publications Inc.
- van den Berg, J., & Brouwer, R.** (2006). Self-organized forest fires near the critical time. *Communications in Mathematical Physics*, **267**, 265–277.
- Wold, H. O. A. & Whittle, P.** (1957). A model explaining the pareto distribution of wealth. *Econometrica*, **25**(4), 591–595.
- Zipf G. K.** (1949). *Human Behaviour and the Principles of Least Effort*. Addison-Wesley.

Pietro Parodi is Head of Pricing & Modelling Operations and Actuarial Methods at SCOR P&C, and a part-time lecturer at Bayes Business School. He has a PhD in Physics from the University of Genoa, Italy, and is a fellow of the Institute and Faculty of Actuaries in the UK. Prior to becoming an actuary, he worked in IT project management and prior to that as a research scientist in the fields of artificial intelligence, theoretical computer science and neuroscience. He is the author of the book “Pricing in General Insurance” (CRC Press), now in its second edition, and of several papers, including “Triangle-free reserving” which was awarded the Brian Hey prize in 2012. <https://www.linkedin.com/in/pietro-parodi-8ba89368/>

Derek Thrumble is Head of Analytics at Gallagher in London, UK, and a part-time lecturer at Glasgow Caledonian University (London) on Enterprise Risk Modelling. He has an MSc in Statistics from University College London and a degree in Mathematics from the University of Manchester. Derek has over 35 years’ experience in the London insurance market focusing on the use of statistical & actuarial modelling to deliver structured insurance programmes to the large corporate risks working across all specialty insurance lines. His work continues today with strategic and financial analysis of loss information, exposure data to optimise insurance programme design, especially through the use of captives and industry mutuals, parametric products and other risk financing solutions. <https://www.linkedin.com/in/derek-thrumble-566222b/>

Peter Watson is a Senior Quant Analyst in the Pricing & Modelling Operations and Actuarial Methods team at SCOR P&C in London. He is a student actuary, with an MSc in Actuarial Science from Bayes Business School (London). Prior to that, he was a theoretical physicist in the field of non-perturbative quantum chromodynamics, with a PhD from the University of Durham (UK) and many years' experience as a postdoctoral researcher at the Universities of Tuebingen and Giessen (Germany). <https://www.linkedin.com/in/peterwatson4/>

Zhongmei Ji is a Technical Actuary in the Pricing & Modelling Operations and Actuarial Methods team at SCOR P&C in London. She is an associate of the UK Institute and Faculty of Actuaries (AIA) with eight years' working experience. She has a PhD in Mathematics (Statistics Group) from the University of York. She worked in actuarial consulting for three years for clients across a wide range of industries and geographies in WTW before moving to specialty insurance and reinsurance pricing. <https://www.linkedin.com/in/zhongmei-ji-2666837b/>

Alex Wang is currently an Account Executive at Gen Re, based in France. He holds a Masters in Mathematics from EURIA, and has been a qualified actuary of the French Institute Des Actuaire since 2015. He worked on re/insurance pricing for 7 years, at both AXA and SCOR before switching to reinsurance underwriting. <https://www.linkedin.com/in/alex-wang-3b68b937/>

Ishita Bhatia is the Pricing Data Science Lead for Energy, Construction & IM Division at QBE in London, UK. She is currently undertaking CII Certification in London Market Insurance and is a Fellow of the Institute and Faculty of Actuaries in the UK and Singapore Actuarial Society. She has experience working across UK, Europe and Asia Pacific from her current and previous roles. <https://www.linkedin.com/in/ishita-bhatia-84263766/>

Joseph Lees is a Pricing Actuary at Protector Forsikring in Manchester, UK. He has a Masters in Theoretical Physics from Imperial College London, and is a fellow of the Institute and Faculty of Actuaries in the UK. Prior to becoming an actuary, he worked as a Mathematics teacher. <https://www.linkedin.com/in/joseph-lees-55526a258/>

Sophia Mealy is Head of Actuarial Pricing at Hiscox Re, and based in the Bermuda office. She has a BA in Mathematics, and a Masters in Statistics, both from Cambridge University. She has been a qualified Actuary of the UK Institute and Faculty of Actuaries since 2007. Having worked in Reinsurance Pricing in London, Zurich and Bermuda since 2004, she is passionate about understanding and challenging the pricing techniques and assumptions used in the industry, in order to best reflect the risk underwritten. <https://www.linkedin.com/in/sophia-mealy-1b7ba510a/>

Rushabh Shah is an Actuary with over 11 years of work experience across employee benefits, reinsurance pricing and general insurance pricing. He currently leads the UK & I Specialty Analytics team at Marsh based in London with a focus on using statistics to help large corporates in optimising their risk finance strategy by building actuarial loss models. He specialises in modelling risk classes such as Aviation, Energy, Credit, Cyber and Financial lines and enjoys writing code. He is an Investment Actuary by specialisation and enjoys working on modelling derivative & option-based products. <https://www.linkedin.com/in/rushabhshah1989/>

Param Dharamshi is has twelve years of experience working in the non-life actuarial domain. He has predominantly worked in direct non-life insurance companies in India and has experience across reserving, pricing and capital. He is a qualified actuary from the IFoA and the IAI. He also holds a PGDM in Insurance from National Insurance Academy. Currently he is working as a manager with KPMG, UK in their non-life actuarial team. In his free time he likes to play badminton, tennis or read autobiographies.

Federica Gazzelloni is a Statistician, Actuary, and enthusiastic R Data Modeller. She graduated in Statistics and Actuarial Science from the University "La Sapienza" in Rome and earned her senior actuarial qualification in 2007 from the Ordine degli Attuari in Italy. In 2015, she became affiliated with the Institute and Faculty of Actuaries in the UK. Federica also holds an Executive Master in AFC from Luiss Business School, IT. Her interests include machine learning and data visualisation. She has contributed to significant projects, including collaborations with the Institute for Health Metrics and Evaluation (IHME) on COVID-19 research and Global Burden of Disease (GBD) studies. <https://www.linkedin.com/in/federicagazzelloni/>

Cite this article: Parodi P, Thrumble D, Watson P, Ji Z, Wang A, Bhatia I, Lees J, Mealy S, Shah R, Dharamshi P, and Gazzelloni F. (2024). Loss modelling from first principles. *British Actuarial Journal*. <https://doi.org/10.1017/S1357321724000163>