



**2023 INTERNATIONAL
CONGRESS OF ACTUARIES**
28 MAY – 1 JUNE 2023 • SYDNEY



Loss Modelling from First Principles *Working Party Report*

*Prepared by Pietro Parodi (chair), Derek Thrumble, Peter Watson, Zhongmei Ji,
Alex Wang, Ishita Bhatia, Joseph Lees, Sophia Mealy, Rushabh Shah, Param
Dharamshi, Federica Gazzelloni*

Presented to the
2023 International Congress of Actuaries
28 May – 1 June 2023

*This paper has been prepared for the 2023 International Congress of Actuaries.
The Actuaries Institute Australia's Council wishes it to be understood that opinions put forward herein are not
necessarily those of the Actuaries Institute Australia and the Council is not responsible for those opinions.*

The Institute will ensure that all reproductions of the paper acknowledge the author(s) and include the above copyright statement.

Institute of Actuaries of Australia
ABN 69 000 423 656
Level 2, 50 Carrington Street, Sydney NSW Australia 2000
t +61 (0) 2 9239 6100
e actuaries@actuaries.asn.au w www.actuaries.asn.au

Loss Modelling from First Principles – Working Party Report

PIETRO PARODI¹ (CHAIR), DEREK THRUMBLE, PETER WATSON, ZHONGMEI JI, ALEX WANG,
ISHITA BHATIA, JOSEPH LEES, SOPHIA MEALY, RUSHABH SHAH, PARAM DHARAMSHI, FEDERICA
GAZZELLONI

Abstract

A common statistical modelling paradigm used in actuarial pricing is (a) assuming that the possible loss model can be chosen from a standard model dictionary; (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 and validation process.

An alternative modelling paradigm, common in the sciences, is to prove the adequacy of a statistical model from first principles: e.g., 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.

Loss count models traditionally used are the Poisson, negative binomial, and binomial distributions. They are used because they simplify the numerical calculation of the total loss distribution. 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 results from the paper by Parodi & Watson (2019) to show how graph (network) theory can be used to model property-like losses. We note a tantalising relationship between the fire-spreading behaviour and whether the relevant exposure curve is in the Maxwell-Boltzmann, Bose-Einstein, or Fermi-Dirac region of Bernegger's MBBEFD curves. 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, percolation, self-organised criticality, lognormal distribution, Lévy process, Poisson process, Weibull process, negative binomial, stochastic differential equation, cyber business interruption, contingent business interruption, supply chain

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 electrodynamics² – recounts during an interview³ 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

¹ Corresponding address: Pietro Parodi, SCOR, 10 Lime Street, London EC3M 7AA, pparodi@scor.com.

² Freeman Dyson showed that the two rival approaches to QED that physicists were weighing – the analytical approach of Schwinger and Tomonaga using Green functions and the more visual approach using Feynman's diagrams – were mathematically equivalent.

³ <https://www.youtube.com/watch?edufilter=NULL&v=hV41QEKiMIM>

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 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."⁴

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 simply pick a simple distribution such as the lognormal distribution for loss amounts or use the distribution from a distribution-fitting tool best fits the observed data, without regard to the number of parameters (for a critique, see Chhabra & Parodi, 2010; Parodi, 2023). At the other end of the spectrum, actuaries might use rigorous machine learning methodologies such as using training, selection and validation sets and being mindful of the bias-variance trade-off when doing rating factor selection and calibration (Hastie et al., 2001). The increased interest in machine learning in recent years has considerably helped actuaries in using the proper model selection techniques.

This working party aims instead to investigate the extent to which it is possible to produce loss models according to the first approach – that is, starting from our intuition and knowledge around the loss generation process and developing a mathematically consistent model based on that. 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 (e.g., 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. Popular risk models such as the individual risk model and the collective risk model successfully frame the loss-generating process in terms of count (frequency) models and severity models. The technical actuarial standards for modelling (TAS-M) of the UK Institute and Faculty of Actuaries have this to say about models representing the real world:

The first part of the proposed purpose of the modelling TAS, as set out in paragraph 2.9a, is that actuarial 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.

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 models. One 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.

1.1 The structure of the report

The content of this report is going to be structured as summarised in Figure 1.1. Section 2 is going to be devoted to loss count (frequency) models, while Sections 3 to 8 are going to be 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).

⁴ Dyson 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.

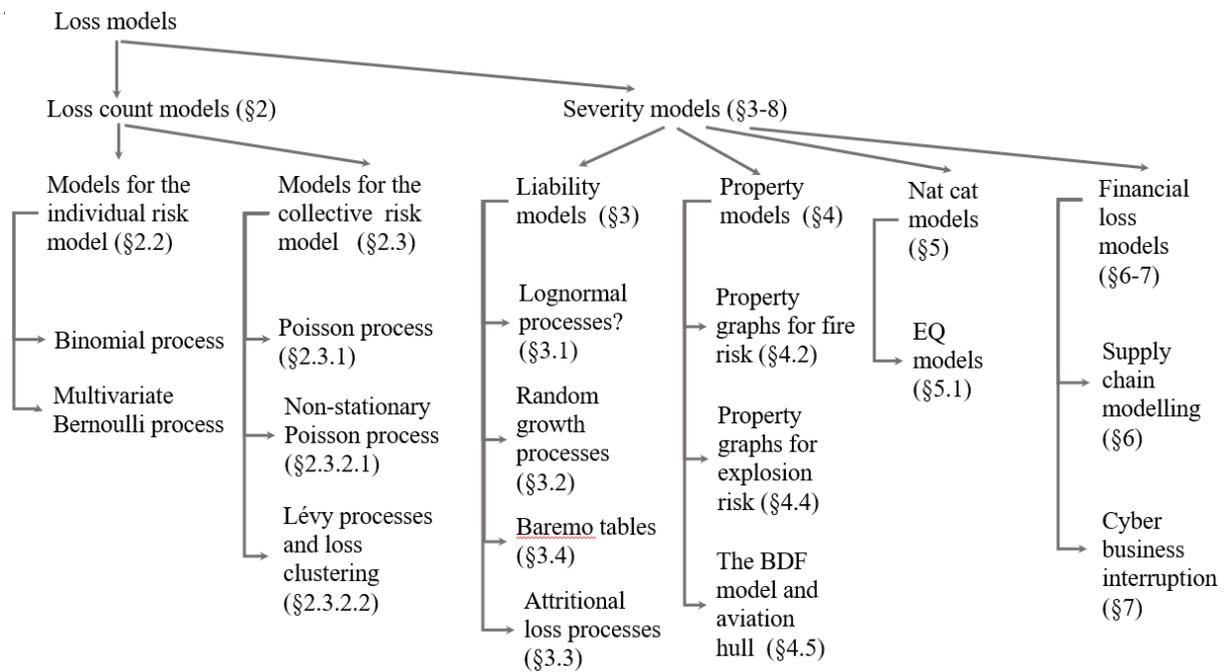


Figure 1.1. A conceptual view of the structure of this report.

2. Modelling Loss Counts.

2.1 The traditional approach

The traditional actuarial approach to modelling loss counts is to use one of three distributions (binomial, Poisson, negative binomial) depending on how large the volatility of the loss count is believed to be. This choice is not typically made for fundamental reasons but because it makes calculations easy: the three distributions mentioned above belong to the Panjer class of distributions (Klugman et al., 2012) for which the evaluation of aggregate loss models via numerical methods (such as the Fast Fourier Transform) is greatly simplified. It's also difficult to argue for more complex distributions when there are such a low number of data points (normally, the loss count for 5-15 years) for calibration.

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

2.2 Loss count models for the individual risk model

In the individual risk model (IRM), losses are assumed to come from a finite number of risks each of which can have only 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. According to this distribution, the probability of having k losses from n risks in a given period is:

$$\Pr(N = k) = \binom{n}{k} p^k (1 - p)^{n-k} \quad (2.1)$$

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, 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) and 1 (loss). The probability of having k claims from n risks is given by:

$$\Pr(N = k) = \sum_{y_1, \dots, y_n \text{ such that } y_1 + \dots + y_n = k} \prod_{j=1}^n p_j^{y_j} (1 - p_j)^{1 - y_j} \quad (2.2)$$

2.2.1 Modelling the dependency between losses

Equations 2.1 and 2.2 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 correlation.

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.

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 every risk and the effect of risks being removed from the portfolio by a total loss (think property insurance) is small while the portfolio is large enough. As a consequence, the occurrence of one loss doesn't significantly affect the likelihood of another loss, and the volatility of the loss count is at least that of a Poisson (where the volatility is equal to the square root of the mean). 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.

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, 2006). This is a count process with increments that are both **independent** (i.e., the number of events in an interval is independent from the number of events in another, disjoint interval) and **stationary** (i.e., the probability of a given number of events $N(t_0, t_0 + \tau)$ occurring in $[t_0, t_0 + \tau]$ does not depend t_0 but only on τ): $N(t_0, t_0 + \tau) = N(\tau)$. Furthermore, $N(\tau)$ follows a Poisson distribution: $\Pr(N(\tau) = k) = \exp(-\lambda\tau) (\lambda\tau)^k / k!$, where λ is the rate per unit of time.

Two important consequences of the assumptions above are:

- the variance of the number of events in a given time interval is equal to its mean;
- the stationary Poisson process (and the Poisson process in general) excludes the possibility of simultaneous events.

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 tight community, and e.g., 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 risk of everyone else getting it
 - 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 – e.g., 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 and 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 “natural” Poisson model in the context of the collective risk model, and for the widespread use of overdispersed models. We will see that overdispersion typically arise 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 (and specifically in pricing) that modelling the number of losses in a future policy period using a Poisson distribution (a distribution for which the variance equals the mean) will often lead to underestimating the volatility 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 routinely used.

Notice that – surprisingly, perhaps – *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 the integral covers one year and $\theta(t)$ represents the Poisson rate density (number of expected losses per unit of time) for the interval (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.

Therefore, while the underlying process is fully Poisson and the distribution of the number of losses follows a Poisson distribution (and its variance will be equal to the mean), since Λ is unknown at the time of pricing, the model used to predict the (range of the) number of losses in the future, e.g., for pricing purposes, will need to have extra variance. It should be stressed that the overdispersion here comes from our ignorance about what value Λ will have 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 and can be used to build indices for weather derivatives;
- economic indicators such as GDP growth in a given country may affect the number of professional indemnity claims or of 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 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**.

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. E.g., if the underlying rate of car accidents only depended on the number of hours of light, $\theta(t)$ would be seasonal but fully predictable. In this case Λ would not be random and a Poisson model could be used.
- 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 reasons⁵.
- 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 here 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, i.e., 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, e.g., 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 products⁶.

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 in general depends on time.

⁵ One such practical reason is that the negative binomial distribution (along with the Poisson and binomial distribution) belongs to the Panjer class, a class of distributions for which the calculation of the aggregate loss distribution for the collective risk model via Fast Fourier Transform is drastically simplified (Klugman et al., 2012; Parodi, 2023).

⁶ It is also relevant, even more intuitively, to human mortality, that 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 “infantile phase”, and the regime $\beta > 1$ is referred as the “aging phase”.

2.3.2.2 Loss clustering and Lévy processes

Another way in which the independence between events breaks causing an increase in volatility is when events are clustered. The Poisson statistics is often described informally as a statistic of *rare events* – but what does this mean? It is true that it can be *derived* as the limit of a binomial distribution (see Equation 2.1) 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 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 also may involve claims that are not brought together by simultaneity or rather near-simultaneity but for contractual reasons, as 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 (Barndorff-Nielsen & Shephard, 2012; Kozubowski & Podgorski, 2009), and specifically a **pure-jump Lévy process**.

A Lévy process is a process with stationary and independent increments⁷. It is a generalisation of the Brownian motion that encompasses – among other things – the Poisson process. Brownian motion describes the movement of particles under the assumption that their trajectory is not deterministic but 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 exchange 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 format). This is referred to as *geometric Brownian motion*.

One problem with this model is that the change is erratic but always “smooth” – i.e. 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 exchange returns – 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. To cater for those, a Lévy process adds 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 decomposition⁸ (Kozubowski & Podgorski, 2009):

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

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 $\mathbb{I}_{[\Gamma_j, \infty)}(t)$ stands for the step function $\mathbb{I}_{[a, \infty)}(t) = 1$ for $t \geq a, 0$ for $t < a$.

⁷Formally, 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) $\lim_{h \rightarrow 0} \Pr(|X_{t+h} - X_t| > \epsilon) = 0$ for all values of t : i.e. the probability of a finite jump at an arbitrarily chosen time is zero (probability continuity).

Notably, Condition (d) implicitly allows for jumps of random sizes at random times, as we will see later when we introduce the Lévy-Khintchine decomposition.

⁸ More 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 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.

So the Lévy process is made of three components: a deterministic drift, some 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 remains (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 is when h_j is itself a random variable with integer values larger than or equal to 1 – in other terms, the case where losses are not rare events that can be isolated but come as 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 alternatives⁹ such as the Newman Type A distribution (Poisson distribution of events in each cluster) and the Pólya-Aeppli distribution (geometric distribution of events). Also, the number of clusters need also not be Poisson – if the underlying process for the number of clusters is for example a non-stationary Poisson, then the overall process will not be a Lévy process, but a so-called *non-stationary Lévy process*.

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

It may appear that formalising loss count processes as Lévy processes is an 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 (such as infinite divisibility) that are available *off the shelf* in this well-studied area of research.

Figure 2.1 shows examples of the main ways of generalising the Poisson process: through non-stationarity (non-stationary Poisson process), clustering (general pure-jump Lévy process), or both (non-stationary Lévy process). All the processes are represented in terms of the cumulative number of unitary jumps as a function of time.

⁹ Some of these alternatives display a multi-model behaviour which might be appropriate in certain circumstances.

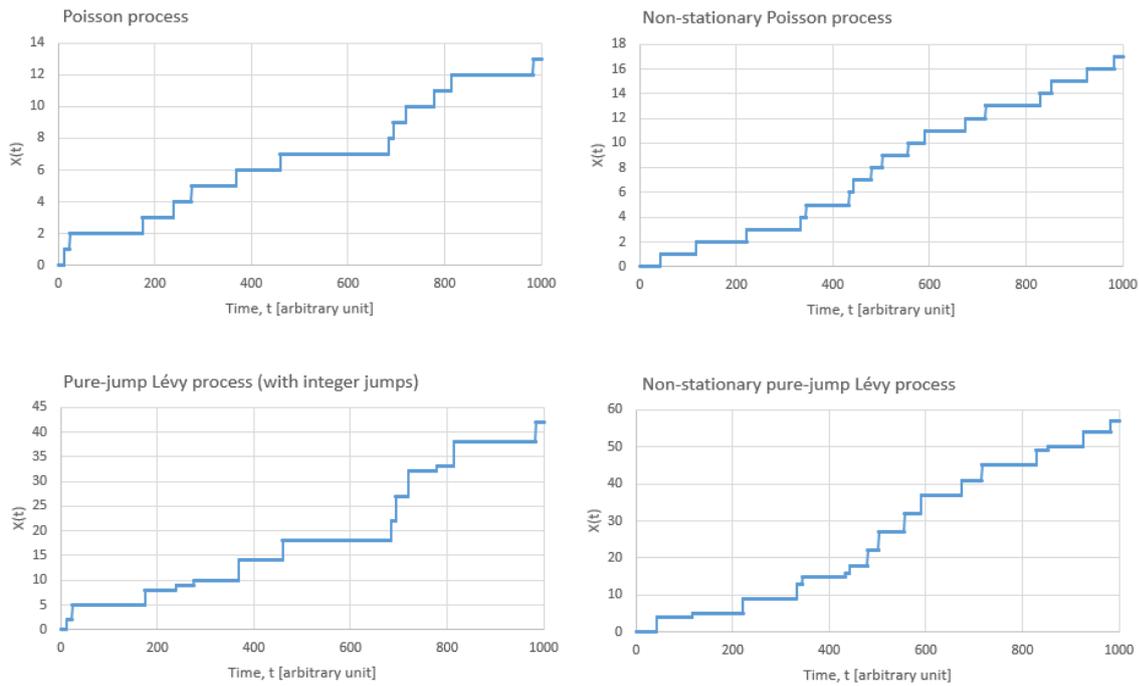


Figure 2.1. (Top left) An example of a Poisson process. The x -axis shows the timing of the jumps (losses) while the y -axis shows the cumulative number of jumps, $X(t)$. Note how each jump has unitary size ($X(t_j) - X(t_j^-) = 1$ if there is a jump at t_j), as in a Poisson process there is no clustering 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 compensate the insured for bodily injury or property damage caused to a third party, and this can be determined by the courts in a variety of ways that also depends on the jurisdiction. This is an indication that 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 model and improvements that actuaries have made on it 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 which vary over several of magnitude. Although it does not have a power-law behaviour, you can't say that it's a thin tail either: its behaviour is genuinely intermediate (Taleb, 2020). A well-known paper by Benckert (1962) provided some empirical evidence that the lognormal works well. Unfortunately, upon closer investigation with richer data sets, it becomes clear that the lognormal model is in most cases not a good model for either the body of the distribution or the tail.

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 3.1 (losses have been multiplied by a masking factor).

As it can be seen from the charts, the lognormal model is inadequate, and this can be proven even more sharply by using a metric to assess the distance between distributions, such as the KS statistic.

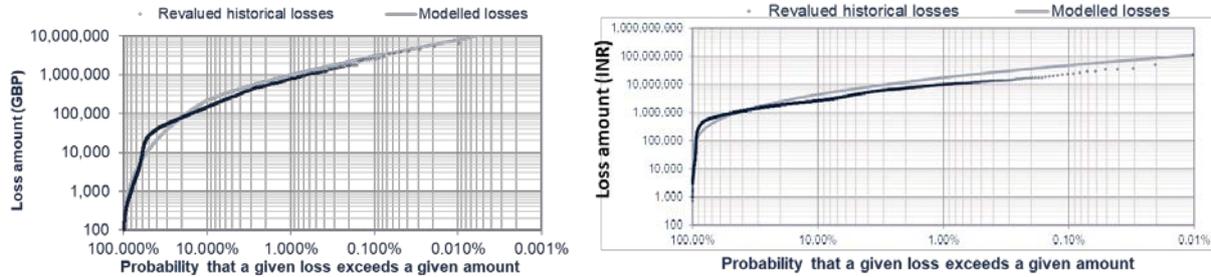


Figure 3.1. (Left) A Lognormal fit to a large number (around 10,000) of employers' liability claims in the UK market. ($\ll 0.1\%$). (Right) A lognormal fit to a large number of motor bodily injury claims in the Indian market. In both cases, based on the value of the KS statistic, the probability that the data set comes from a lognormal distribution is negligible ($\ll 0.1\%$).

While the empirical inadequacy of the lognormal distribution is clear, it has also been noticed ubiquitously that the tail of a portfolio of liability losses is modelled well by a Generalised Pareto Distribution (GPD) with a power-law tail ($\xi > 0$), as predicted by extreme value theory.

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

According to **extreme value theory**, the tail of the distribution can be asymptotically described – apart from pathological circumstances – by a generalised Pareto distribution (GPD), i.e., a (survival) distribution of the form

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

This in turn 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 $\xi > 0$ will be asymptotically equivalent to a single-parameter Pareto ($\bar{F}(x) = (\theta/x)^\alpha$) with $\alpha = 1/\xi$.

This doesn't immediately rule out a 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 $\lim_{x \rightarrow \infty} \alpha(x) = \infty$, corresponding to $\xi = 0$.

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

¹⁰ Note 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's 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.

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, e.g., in Knecht & Küttel (2003) and in Fackler (2013), where new classes of spliced Lognormal/GPD distributions were introduced to formalise this approach. This is a good workaround to ensure that a suitable model is used where it matters most – the tail.

However, experience shows that the lognormal distribution is not generally a good match for attritional losses either. Indeed, Figure 3.1 shows two examples of a behaviour that we have often found in our work as practitioners: empirical distributions often appear as the conjunction of two parts of the distribution that are *broadly* linear in a log-log scale, with a concave elbow to connect 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 – e.g., 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 DoublingRatio(x) = $\exp(-x)$ – it decreases quickly and soon becomes undistinguishable from zero. If the severity distribution is a single-parameter Pareto with parameter α , then DoublingRatio(x) = $2^{-\alpha}$: the ratio is the same at all x 's, or, equivalently, the ratio is the same at all scales. In other terms, the behaviour is of a fractal nature. If the severity is GDP with $\xi > 0$, the ratio will have a slight dependence on x that will asymptotically disappear: DoublingRatio(x) = $\left(\frac{1+\xi(2x-\mu)}{1+\xi(x-\mu)}\right)^{-\frac{1}{\xi}} \rightarrow 2^{-\frac{1}{\xi}}$.

Thanks to extreme value theory (EVT) and specifically to the Pickand–Balkema–de Haan (PBdH) theorem, we know that the exponential and power-law behaviour exhaust – alongside the case of a finite-support distribution ($\xi < 0$) – all possibilities in most practical cases. Observation also tells us that the most common behaviour is by and large 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 3.2). In all these cases, the power law behaviour (and specifically the Pareto law) has 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.

<i>Law and description</i>	<i>Power law</i>	<i>Ultimate reason for behaviour</i>
Champerowne'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 3.2. Examples of the emergence of the power law in economics.

Let us expand on this. In all cases shown in Figure 3.2, the phenomenon can be described by a stochastic differential equation whose solution is a Pareto distribution. According to Gabaix (1999; 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, i.e., $\mu(X_t, t) = \mu X_t$ 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 is $f(x, t)$. 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 so-called 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:

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

Once rewritten (after a few calculations) as:

$$\frac{\sigma^2}{2} x^2 \frac{d^2 f(x)}{dx^2} + (2\sigma^2 - \mu)x \frac{df(x)}{dx} + (\sigma^2 - \mu)f(x) = 0 \quad (3.1)$$

it becomes clear that one possible solution is a function of the form $f(x) = kx^\beta$, which turns Equation 3.1 into a second-degree equation in β . This has two roots: $\beta = -1$ and $\beta = 2\left(\frac{\mu}{\sigma^2} - 1\right)$. The first root is not “physical”, as it leads to a divergent equation. As for the other, it leads to a Pareto distribution with $\alpha = -(1 + \beta) = 1 - 2\frac{\mu}{\sigma^2}$

It can also be proven that – assuming that the solution must make economic sense, which translates into a smoothness constraint – this is the only solution (Gabaix et al., 2016).

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

We have however done 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 FKE is a lognormal distribution with indefinitely increasing mean and variance, which never reaches a steady state (Gabaix et al., 2016). There are mainly two ways in which researchers have resolved this issue in the past:

- (a) one is 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), and
- (b) that the system has some type of attrition so that new elements enter and exit the system (e.g., 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.

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 as a result 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? 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.

3.2.2.1 Indirect mechanisms for the emergence of power laws

The **indirect mechanism** is quite straightforward: liability losses are made of different components. Limiting ourselves to bodily injury claims, these components are:

- loss of earning (or dependant's loss of earning in the case of death)
- cost of care including medical costs,
- pain and suffering, and
- 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. 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. It so happens that we already have proof of that. 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 Champerowne's analysis (1953) – which has indeed been tested successfully for many territories – this follows a Pareto in the tail of the distribution.

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).

3.2.2.2 Direct mechanisms for the emergence of power laws

The **direct mechanism** is instead one that 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 this may be the case: (a) courts are increasingly generous for compensations for such things as pain and suffering, and that happens proportionally to the compensation themselves. This is what is normally referred to as court inflation: compensations from previous cases are used as benchmarks for current cases and increases tend to be a percentage of the benchmark. Punitive damages also are established by courts and again they increase exponentially in time. Medical costs are also likely to increase proportionally to their current cost, as medical technology of increasing sophistication is required, and the salary of medical professionals is likely to follow Champerowne's distribution of incomes (so this follows partly under the indirect mechanism).

For the existence of a steady state distribution, we 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 (remember that that threshold can be arbitrarily small), we can safely assume that a minimum loss exists.

Note that it may be slightly confusing that an exponential increase in costs should translate into a steady-state Pareto. It is therefore useful to formalise the types of losses as indexed by a continuous index ι , and write the stochastic differential equation as

$$dX_{\iota,t} = \mu X_{\iota,t} dt + \sigma X_{\iota,t} dZ_t$$

And imagine that each type drifts exponentially and the value for that type is replaced by other types of claims that previously had lower values. Eventually, the overall distribution reaches a steady state in which both time and type can be ignored.

Note that we may also have border conditions that prevent the value of the loss to go beyond a certain amount (this is common for other economic quantities: see Gabaix et al., 2016).

3.2.2.3 Attritional vs large losses

Champerowne's article also points to an interesting distinction – that 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: treat small losses as an aggregate and pay more individual attention to 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 less or no involvement from the courts, and therefore the increase in time will be linked more to standard inflationary factors (CPI, wage inflation) and less to social inflation (courts, etc.). Also, if this is the case, we should see a radically distinct behaviour in the log-log CDF graph between attritional and large losses. This is indeed what we observe in practice (see Figure 3.1): both attritional losses and large losses appear to be distributed on a straight line in log-log scale (for attritional losses, this is only a very rough approximation). This points to two different types of behaviour for the two different types of losses. Both these empirical observations and the results from other economic examples point to an intrinsic (and not purely 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 one 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 3.3. That would be an interesting comeback for the lognormal distribution!

3.2.2.4 Empirical demonstration of the emergence of a power law

We can demonstrate the emergence of a power law (i.e., the steady state solution to the FKE) for a system where the random variable has a minimum value empirically using a numerical simulation. Let us consider two types of stochastic random walk for the random variable $X(t)$. The first is the 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 $x_{\min} = 0.5$ for the latter. We generate a sample of 1,000,000 paths over a total time-period, T (each path is the 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 could be determined via the relationship

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

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 in the above cannot be determined from the simulation.

The empirical distribution of the endpoints is represented via the survival function (exceedance probability) for various time periods T in Figure 3.3.

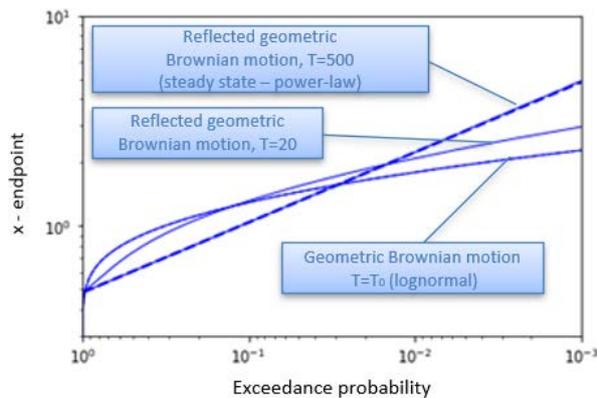


Figure 3.3. Survival function for increasing time periods, showing the increasing move away from the lognormal distribution towards the steady-state Pareto distribution.

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 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 really necessary as you can avoid modelling the attritional losses altogether: attritional losses are

almost by definition small and plentiful and therefore we can simply resample from the available data. One could also use interpolation but this doesn't add much value in most circumstances.

The theoretical underpinning for the resampling approach to attritional losses is Givenko-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 KS test is obvious and based on that it's possible to put limits to the KS statistic with a certain probability (e.g., we know that $\sup |F_n(x) - F(x)| < 1.358 \sqrt{n}$ with probability of 95% for large enough 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 rarefy when they become large and therefore the distortions arising from interpolation become more important. So 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, assuming that we know that the correct distribution is, e.g., a lognormal, we want to compare these two situations: (a) resampling from the data available; (b) using a lognormal model calibrated *based on the available data*. We can calculate the KS distance between the true underlying model and (a) vs (b). Repeating many times we can have an estimate of the average KS distance for the two situations. Note that the KS distance must be evaluated for the value below a threshold, as simple resampling is not interesting for large losses except in special cases.

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, 1000, 2000, 3000, 5000, 10000). 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 undistinguishably from modelling.

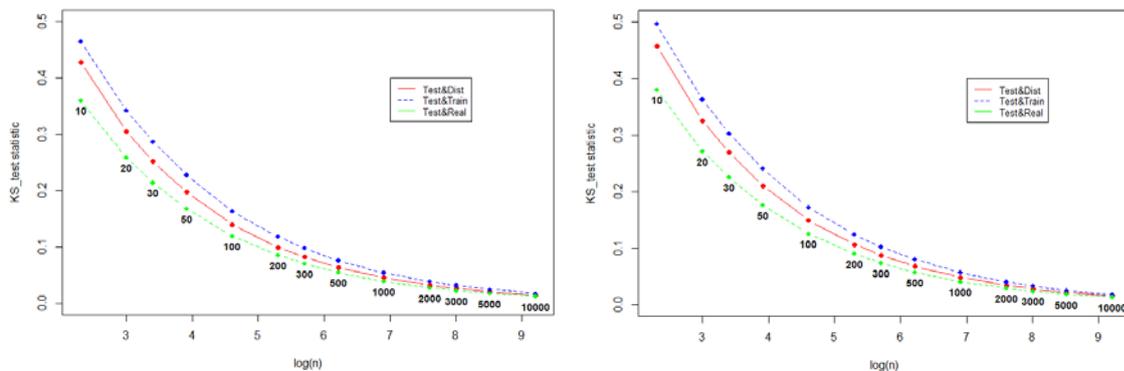


Figure 3.4. KS distance for test set vs (data-calibrated) model, test set vs training set, test set vs true model for different values of the sample size. (Left) No threshold. (Right) 90% threshold. Note that the KS distance (rather than the normalised KS distance, which will remain relatively flat) has been deliberately shown here.

So, 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 case of a lognormal). We have also argued, however, that the lognormal model is normally not a good model, and 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 the empirical distribution (resampling) will always outperform any attritional loss model. The only reason to use a lognormal model or other models is because the model will in general be easier to handle for simulation purposes than a loss data set.

3.4 Severity distributions from first principles in jurisdictions with compensation tables

In some jurisdictions, a compensation table (often called a “Baremo table”) providing a range of compensations for different types of bodily injuries or for different degrees of temporary or permanent disability suffered by the victim of an accident. This significantly constrains the distribution of possible loss amounts. Examples of this can be found in Spain, Italy, and Germany. A simplified version of the Baremo table for the Italian market as produced by the Milan court can be found in Figure 3.5.

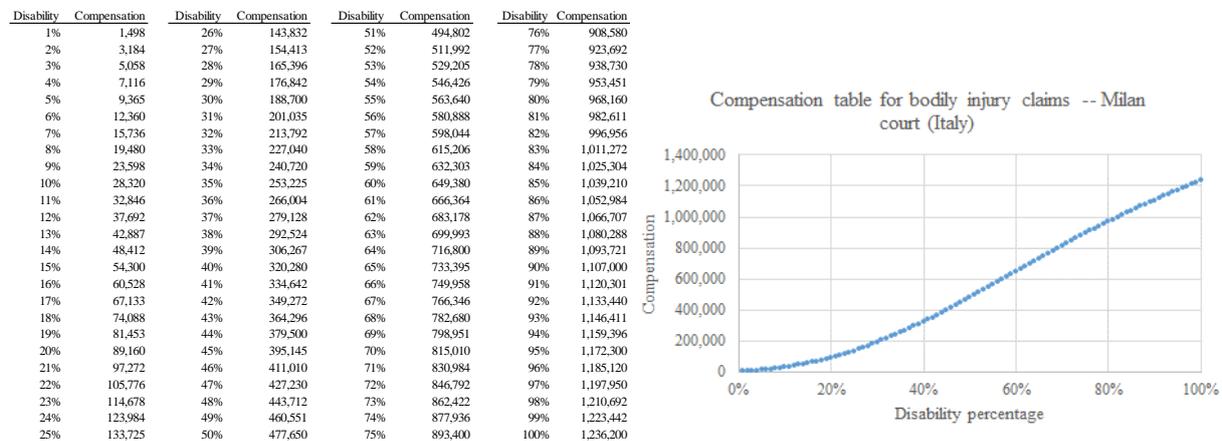


Figure 3.5. (Left) A simplified version of the Baremo table for the Italian market as produced by Milan’s Osservatorio sulla giustizia civile in 2021. The actual table has different compensations by age and allows the court some flexibility around the exact amount. This simplified version only shows the example pay-out for a one-year-old victim, who elicits the largest payment. (Right) A graphical rendition of the Baremo table on the left. All monetary amounts are in EUR.

The table in Figure 3.5 shows the “maximum” pay-out for a given level of disability, although that maximum can be further increased up to 50% under special circumstances. Other complexities include modifications depending on age. Separate tables are available for death and for the compensation of relatives.

In the presence of a Baremo table, the distribution of loss amounts depends on the probability of suffering a certain degree of disability, and (depending on the jurisdiction) on the distribution of outcomes for a specific degree of disability. In case more than one claimant is included for a given loss, the distribution of the number of claimants and the dependency structure between the claims also needs to be taken into consideration.

Let’s start by looking at a simple case – that where the Baremo table has a fixed compensation for a given degree of disability, with no additional flexibility. We will also ignore deaths and compensation to relatives.

3.4.1 Severity distributions arising from a Baremo table with fixed compensations and a single claimant

We assume that there is a compensation table in its simplest form, providing a fixed benefit for a claimant who has suffered a given degree of disability d . Figure 3.5 is an example of such table, but this can be generalised to a generic function $x(d)$ – which we will call the Baremo function – that returns the loss for a given degree of disability (in this case, we can assume that d is a real number with values $0 < d \leq 1$).

The information in the Baremo function is not sufficient to determine the severity distribution. To do that, we also need the probability that, given a loss, the victim suffers a degree of disability D larger than d , $\bar{F}(d) = \Pr(D > d)$. Once one has $\bar{F}(d)$ and $x(d)$, one can derive the survival probability for the loss amount, $\bar{F}(x)$.

Let’s see how this works for some special cases of $\bar{F}(d)$ and $x(d)$.

3.4.1.1 Exponentially decreasing $\bar{F}(d)$, exponentially increasing $x(d)$

Here’s one simple case that we may consider: that in which the probability of suffering a degree of disability falls exponentially, $\bar{F}(d) = \exp(-\lambda d)$ for $\lambda < 1$, $\bar{F}(1) = 0$, while the compensation increases exponentially with the

disability: $x(d) = k \exp(\mu d)$ for $0 < d \leq 1$ (we are here assuming that there is a minimum compensation k that needs to be paid for a claim no matter how small the disability incurred, and that the maximum amount payable is that for 100% disability, $k \exp(\mu)$). By cancelling out d we obtain the relationship:

$$\bar{F}(x) = \begin{cases} \left(\frac{k}{x}\right)^{\frac{\lambda}{\mu}} & k \leq x \leq k \exp(\mu) \\ 0 & x > k \exp(\mu) \end{cases}$$

Which is a Pareto severity distribution for $x \leq k \exp(\mu)$ with exponent $\alpha = \frac{\lambda}{\mu}$ and threshold k equal to the minimum compensation payable. Note that the probability density $f(x)$ has a probability mass equal to $\exp(-\lambda)$ for $x = k \exp(\mu)$.

3.4.1.2 Exponentially decreasing $\bar{F}(d)$, linearly increasing $x(d)$

Another example, which seems to be more in line with the Italian Baremo for bodily injury (Figure 3.5), is where $x(d)$ is linear (any additional point of disability leads to the same additional payment): $x(d) = kd$. In this case, it is immediate to show that the severity distribution remains exponential:

$$\bar{F}(x) = \begin{cases} \exp\left(-\frac{\lambda}{k}x\right) & x \leq k \\ 0 & x > k \end{cases}$$

Slightly more generally, if $x(d)$ increases as a power law $x(d) = (kd)^{\frac{1}{\beta}}$ the severity distribution will be a Weibull distribution $\bar{F}(x) = \exp\left(-\frac{\lambda}{k}x^{\beta}\right)$.

3.4.1.3 General empirical case

More in general, if the $x(d)$ is given by an actual Baremo table/curve like that illustrated in Figure 3.5, and the probability of different levels of disability $\bar{F}(d)$ is derived by actual accident statistics, possibly smoothed and approximated by a known function such as an exponential, a power law or a set of splines, the severity loss distribution can be estimated by Monte Carlo simulation. We'll see an example in Section 3.4.2.3.

3.4.2 Severity distribution from a Baremo table – generalisations

The results in Section 3.4.1 can be generalised in various directions. Five obvious enhancements are:

1. Incorporating the volatility around the compensation granted for a given degree of disability based on age and other factors.
2. Incorporating the possibility of temporary (100%) disability and the volatility around the number of days for which temporary disability is suffered.
3. Incorporating the outcome of death into the calculation and the compensation to the relatives of the victim, by degree of relation.
4. Incorporating the compensation for the relatives of the victim for severe injury.
5. Incorporating the possibility of multiple claimants, accounting for the correlation in the degree of disability/death of the different victims of the same accident.

Incorporating these enhancements requires various degrees of complexity.

Incorporating age-dependency of compensation (Item 1) is straightforward, as it simply requires having a model for the distribution of the age of the victim, and then looking up the relevant compensation from the table (note that the table, in this case, can be smoothed/approximated as a two-variable function).

Incorporating temporary disability (Item 2) in the calculation is relatively straightforward once one has an idea of the distribution of the number of days of disability following an accident. It will also probably be much smaller than the compensation for permanent disability, at least when the number of days is small (in Italy, the daily rate as set by the Milan court is currently up to €150/day).

Allowing for the compensation for the relatives of a victim for death (Item 3) and severe injury (Item 4) is more complex, as it requires a model of how many relatives may be involved (and with what relation degree) for each age range. An additional complication is that some of the relatives may also be the victims (which is especially true when the relation is a spouse/partner or a child/parent) – therefore some drastic simplifications may be needed here. In the case of death, we also need to have an estimate of the conditional probability that a victim dies given that they have had an accident.

Accounting for multiple claimants is also complex, as that requires a model for the number of claimants (which is the easiest part) and a model on how the degree of disability of the different victims is correlated. We will address these issues in the following subsections.

3.4.2.1 Statistics on the number of claimants for a single event.

If more than one claimant is potentially involved, we need to have information on what the likelihood is for a given number of claimants to be involved. Empirical results on the number of claimants for each accident can be used. E.g., Figure 3.6 shows the distribution of the number of claimants for UK municipality (mainly PL) claims over a number of years. In the case of motor claims, we expect the probability of having more than one victim to be much higher than this table suggests.

No. of claimants	No. of accidents
1	178,003
2	1,754
3	268
4	91
5	32
>5	39

Figure 3.6. Statistics around the number of claims for PL claims to UK municipalities.

We can either use the empirical distribution as it is or attempt to model it with some discrete distribution such as a Poisson model. Without loss of generality, we will typically use a zero-truncated distribution, i.e., a distribution for which the probability of no losses is zero (Klugman et al., 2012).

3.4.2.2 Dependency structure of the degree of disability of different claimants.

If more than one claimant is involved in a single loss (and assuming that the claims are filed jointly) the question arises as to the dependency structure among the severity of these claims. Clearly, we expect that there is some correlation between the degree of disability suffered from claimants involved in the same accident.

At one extreme, we could assume that the degree of disability of different claimants is perfectly correlated, and therefore we can pick the same level of disability for all. At the other extreme, we can assume full independence and sample independently from the distribution $\bar{F}(d)$. The truth is likely to lie in between these two extremes. To correlate the losses, two common approaches are:

- to set up a copula that correlates the percentiles of the sampled distribution for the different claimant. This is complex and difficult to calibrate as data is likely to be too sparse;
- to assume a common shock (Meyers, 2007) – i.e., to sample a degree of “severity” of the loss (e.g., mild, average, severe) with a given probability, and use different shapes for $\bar{F}(d)$ for the different cases.

3.4.2.3 Estimation of severity distribution - example

As a relatively easy example of how to estimate the severity distribution from first principles, let us ignore the issue of the correlation between claimants – assuming instead zero correlation – and the issues around compensation for relatives and for temporary disability. We allow for the possibility of death, but in case death occurs, we set the compensation at a fixed value x_{death} . The algorithm for estimating the severity distribution would then go as follows.

ESTIMATING THE SEVERITY DISTRIBUTION FROM A BAREMO TABLE

Inputs: probability of death p_{death} , age distribution $\bar{F}(a)$, disability distribution $\bar{F}(d)$, distribution of the number of claimants with non-zero losses $p(r)$, compensation table $x(a, d)$ by age and disability, fixed compensation for death x_{death} .

For each scenario:

- a. Sample a value r from the zero-truncated distribution of the number of claimants, $p(r)$
- b. For each claimant $j = 1, 2 \dots r$:
 - i. Sample a value from a Bernoulli distribution with parameter p_{death} to determine whether the claimant dies or is permanently injured
 - ii. If claimant j dies:
 - Set the loss to $x_j = x_{\text{death}}$
 Else:
 - Sample a value d_j from the (empirical or modelled) disability distribution, $\bar{F}(d)$
 - Sample a value a_j from the distribution of ages, $\bar{F}(a)$
 - Calculate the loss corresponding to age a_j and disability d_j , based on the compensation table or a model of it: $x_j = x(a_j, d_j)$
- c. Calculate the total loss across all claimants, $x = \sum_{j=1}^r x_j$

Repeat for a large number of scenarios. Sort losses in ascending order to obtain the estimated severity distribution.

The estimated severity distribution for the case where the losses are sampled from the complete Baremo table from the Milan court and with other assumptions specified in the legend is given in Figure 3.

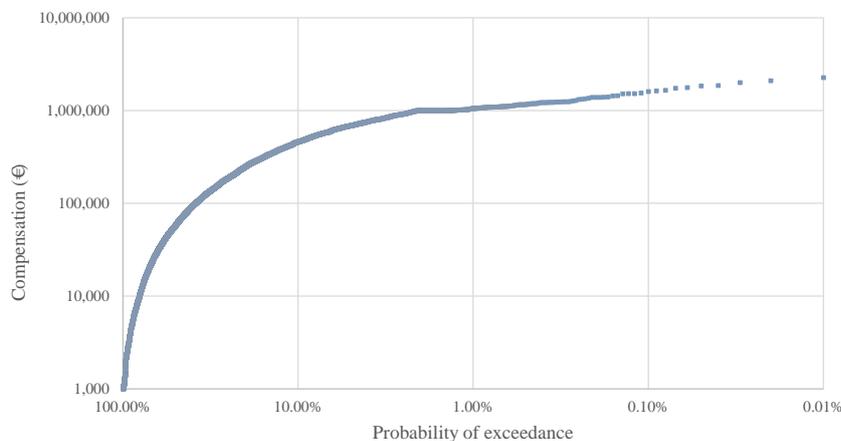


Figure 3.7. The CDF (in log-log scale) of the estimated severity distribution for the complete Milan Baremo table, with exponentially decreasing probability ($\bar{F}(d) = \exp(-\lambda d)$, $\lambda = 0.2$) of a given disability, a zero-truncated Poisson distribution (rate = 0.5 per year) for the number of claimants, and a distribution of ages based on the distribution of ages in the Italian population at large¹² (obviously an approximation as it doesn't necessarily match the distribution of ages of people being drivers or passengers in cars). Zero correlation between disabilities is assumed (which will underestimate the tail of the

¹² <https://www.cia.gov/the-world-factbook/countries/italy/#people-and-society>

distribution). The probability of death given an accident is taken to be $p_{\text{death}} = 1\%$, and the fixed compensation for death to be $x_{\text{death}} = \text{€}1\text{M}$, which is roughly the maximum compensation for three close relations¹³.

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 – i.e., fitting a curve through 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 – approximately, at least – on the MPL but only on the type of construction.

In Europe and other territories, use of Bernegger’s curves (also called MBBEFD curves) are widespread for this purpose.

4.1.1 Bernegger’s curves and their connection to statistical mechanics

Bernegger’s curves are curves borrowed from statistical mechanics that are well-suited for modelling property losses. As Bernegger emphasized from the beginning, their use was purely practical and there was no suggestion of any deeper connection with statistical mechanics.

It is useful to state explicitly this 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 density* for the different curves from statistical physics are given by:

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

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

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

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.

¹³ According to the same document from which the table in Figure 3.11 was taken, the maximum compensation to the relatives for the death of the victim is €336,500 (for a parent, child, or spouse/civil partner) and €146,120 (for a sibling or grandparent).

It should also be mentioned that there is a special (“Swiss Re”) parameterisation dependent on a single parameter c formed from a particular configuration: $b = b(c) = \exp(3.1 - 0.15c(1 + c))$, $g = g(c) = \exp((0.78 + 0.12c)c)$. Residential property fire damage then 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 (for any value of b , and for $g \geq 1$) by $\Pr[X = 1] = \bar{F}(1^-) = \frac{1}{g}$.

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

As already 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) correspond 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 called 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 – i.e., if direct fire propagation between the two corresponding rooms is possible.

In the simplest version, the value of each room/unit is the same, let’s say 1, and each node reached by the fire is considered fully lost (so 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; 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 maximum possible loss (MPL) is taken to be the size of the largest connected component of the graph – the idea being that a fire starting in one connected component cannot propagate to a separate connected component for lack of passageways. This simple framework can of course be refined at will: the value of different nodes may be different, and partial losses at nodes may be possible.

4.2.1.1 Static approach

In this approach, we start from a weighted graph¹⁴ $G = (V, E, W)$, 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 delete each edge e in the graph with a given probability $1 - w_e$. The loss size is given by the size of the connected component of what remains of the graph G which includes the node v^* ; by dividing by the MPL (the largest connected component of the graph *before* edge deletion) we obtain the damage ratio. The process is

¹⁴ A *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.

repeated for a large number of simulations until the distribution of the damage ratios emerge. The algorithm is described in more detail in Parodi & Watson (2019) and a visual representation of it is shown in Figure 4.1. The simulation can be run for both a single property or for a portfolio or similar (but not identical) properties, in which case one also has to pick properties at random.

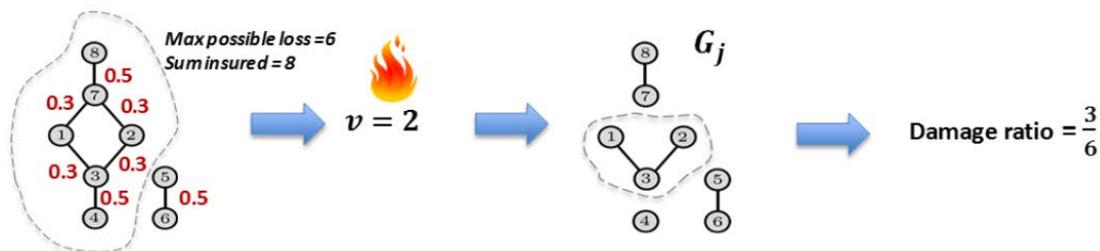


Figure 4.1. In this simulation example, the total insured value (TIV) is 8 (we are assuming that each node represents a unit of worth equal to 1) while the maximum possible loss (MPL) is equal to 6 (the size of the largest connected component). The origin of the fire is selected at random (node 2 in 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.)

4.2.1.2 Dynamic approach

An alternative approach is to view the fire as propagating through a building deterministically – i.e. with the route defined by the graph connectivity – but for a random time.

Weights on the edges are in this approach related to the time it takes for fire to spread through a given passageway and their status (e.g., open vs closed doors).

The fire will spread for a time T (“delay”), which depends on how long it takes for the fire to be put out. 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 starts traversing the graph deterministically. The time to 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 the 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 of similar properties. The results of the simulation with a given set of parameters are shown in Figure 4.2.

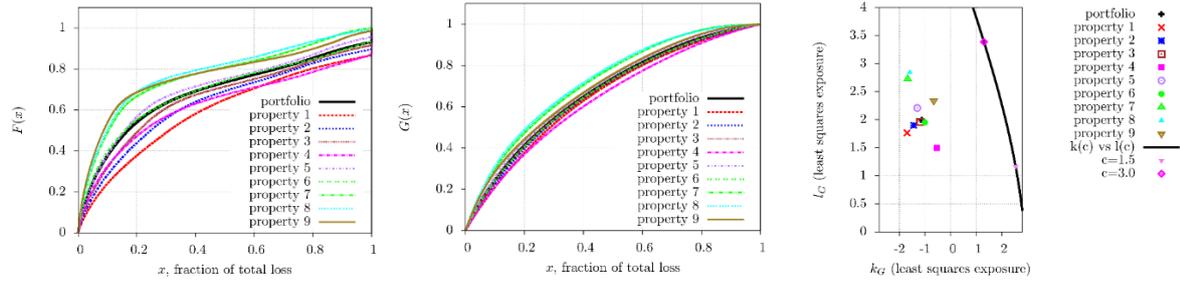


Figure 4.2. The result of the simulation with the random time approach with the following parameters: $\lambda = \frac{1}{2}$, $\Pr(\text{open}) = 0.8$, $n_{\text{open}} = 1$, $n_{\text{closed}} = 10$, $n_{\text{wall}} = 30$. (Left) The severity (top) and exposure (bottom) curves for different property 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).

The Bernegger model gives a good approximation of the various curves coming out of the simulation for both the individual properties and the portfolio, but the values of the parameters are far from those of the Swiss Re c curves. This seems to be true also for a wide range of different choices of the parameters.

So far we have mostly summarised results from Parodi & Watson (2019). We are now going to investigate the relationship between the shape of the graph – or rather, 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 – interpretation of MB vs BE vs 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 will be 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.¹⁵

So, let us assume that the fractional loss is a deterministic function of a random time. Let us further assume that the random time is drawn from the exponential distribution with parameter λ . We can write:

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

The (as-yet unknown) function, $f(t)$, will be referred to as the evolution function and represents the fractional amount of damage done. It should be non-decreasing and has the boundary conditions

$$f(0) = 0, \quad f(t > t_{\text{max}}) = 1$$

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

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

such that

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

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.

¹⁵ Here, 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.

This is a simple, intuitive, and natural interpretation¹⁶. Further, the parameter for the exponential distribution has been set (without loss of generality) to 1.

Now let us derive the evolution function $x = f(t)$ that corresponds to the CDF for the MBBEFD distribution and 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 cases¹⁷.

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

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

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

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:

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

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, i.e., sketching) $f'(t) \sim e^{-f(t)}$ and we see that $f'(t)$ decreases with time – the evolution function is *decelerating*.
3. BE ($\alpha > 0$). Now we must consider the values of b (since $b=1$ is a special case):

¹⁶ Moreover, 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.

¹⁷ The 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, $\alpha + \beta = b(g - 1) \geq 0$.

- 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*.
- 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 $f'(t) = e^{\lambda t} \lambda / \alpha$ - i.e., the evolution function is *accelerating*.
- 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 4.3 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 has in all cases been set to $\lambda = 1$ (without loss of generality, as this is simply a scaling factor).

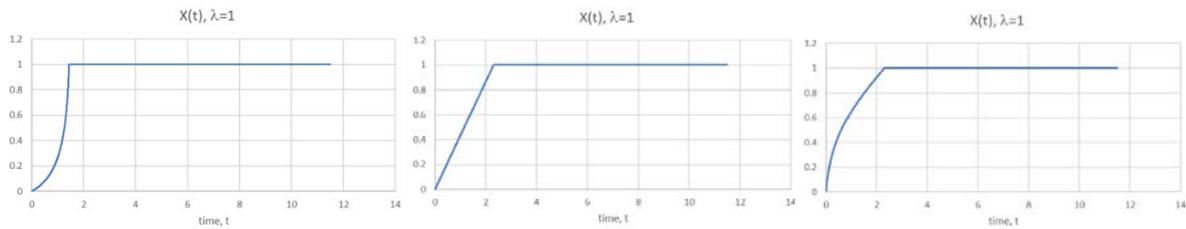


Figure 4.3. (Left) Example of the BE regime: time evolution for a Swiss Re curve with $c = 1.5$ (residential property). Increasing the value of c yields the same type of behaviour but with a lower acceleration/convexity. (Centre) Example of the MB regime: evolution for a Bernegger curve with $b = 0.1, g = 10$. (Right) Example of the FD regime: evolution for a Bernegger curve with $b = 0.01, g = 10$.

4.2.2.3 Connection between the property graph topology and the statistic

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 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 accelerating¹⁸. 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 how a property graph looks like (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 Fermi-Dirac region. However, it is not difficult to imagine graphs that lead to a Bose-Einstein distribution. One such example is a graph where one node – which we can call the *hallway node* for obvious topological reasons – is connected to a very large percentage of other nodes, and each node is separated from the hallway node by two degrees at most. 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 Bose-Einstein 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 similar to considering that alongside fire there are other related phenomena that cause damage, such as smoke – so once a good percentage of the property is out of order the likelihood of salvaging the rest is reduced.

¹⁸ Note that all Swiss Re curves are in the Bose-Einstein region as long as the parameter c is between -0.48 and 21.48 .

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 tree¹⁹ 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).

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 indeed 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. It relates in obvious ways to fractal theory as the structures that emerge during phase transitions typically have fractal dimension.

Indeed, percolation theory had been used before for modelling 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 have mainly looked at 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). So there is an intimation that the Bernegger curves may provide more than a purely formal connection to the losses caused by the spread of fires.

In a nutshell, there appear to be two types of networks: power-law networks (which give rise to a BE statistic) and Cayley trees (which give rise to an FD statistic). According to Bianconi and Barabási (2001), growing networks self-organise into a complex networks in which some nodes have an outsized number of connections to the rest of the network. These networks (which may be evolving) 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 – i.e., 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, this is only a research suggestion at the moment.

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 a Fermi distribution.

In both Bianconi & Barabási (2001) and Bianconi (2015) the key idea is to create a mapping between the graph (network) 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, risks to property traditionally cover insurance against a number of perils traditionally named FLExA (fire, lightning, explosion, aircraft collision) and (in All Risks policies) natural catastrophes. Explosion risk was dealt with in Parodi & Watson (2019) using weighted directed graphs and we refer the interested reader to that paper.

¹⁹ A tree is a graph with no loops; a Cayley tree of degree m has a root node of degree m , leaf nodes (or interface nodes) of degree 1 and intermediate nodes of degree $m + 1$.

4.5 Aviation Hull – Initial steps towards a fire model

Aviation accidents caused by fire account for only 3% of the total number of hull losses²⁰, but pose a significant safety threat. Cabin fires can be categorized as ramp, inflight, and post-crash fires. Post-crash fires, resulting from ground collisions or high-speed landings, are fatal and can cause severe hull damage. The relationship between cabin temperature during a fire and hull damage is essential to establish bespoke fire damage curves. These curves may help derive a severity curve from first principles.

4.5.1 The relationship between burning time and temperature

In 1982, NASA conducted a full-scale fire test²¹ on a B737 aircraft, to understand the rate of fire spread the cabin, over time. NASA's data (Test 24) showed some interesting results. Initially when the fire ignited, during the first ~200 seconds the temperature rises slowly as the fire begins to spread but it then rises rapidly until it reaches flashover at about 514 seconds since ignition.

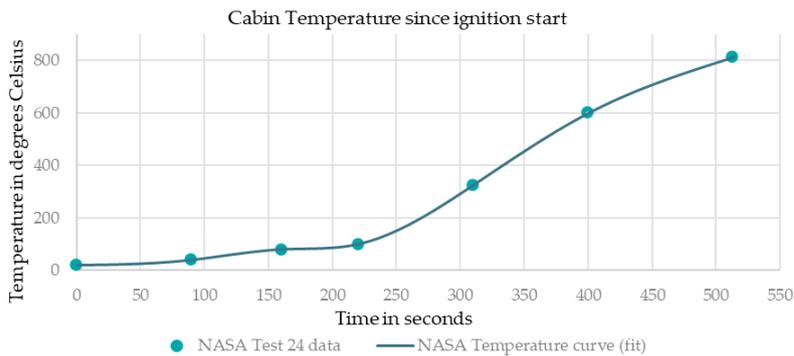


Figure 4.4. Temperature as a function of time since ignition based on NASA's report (p. 675). Only a handful of data points were available in that study, and we fitted these points using an interpolating cubic spline.

Research²² conducted by the FAA on the impact of different cabin materials on the time-to-flashover shows that PVF-phenolic/glass and PVF-phenolic/carbon which are used in composite materials in most modern aircrafts have a time-to-flashover between 4 and 5 minutes.

NIST's fire research²³ related to Fire Dynamics, i.e., how fires start, spread and develop suggests that at ≥ 400 degrees Celsius, temperature of gases at the beginning of room flashover. Based on the NASA-full scale fire test, this temperature is achieved ~325 seconds (~5.5 minutes) since ignition in an aircraft cabin, which ties in approximately with the FAA research for PVF-phenolic/glass and PVF-phenolic/carbon. For our purpose we will assume 325 seconds to be the mean time to flashover for the analysis.

4.5.2 Damage ratio for all fire losses

To estimate the damage ratio (loss as a percentage of hull value), we have used a ratio of the Hull Loss to the Hull value at the time of loss for all fire losses from a third party database. We analysed over 60 fire/explosion losses for all wide body aircrafts between 1974 and 2020. The actual damage ratio curve can be approximated quite well by a Swiss Re curve (Bernegger, 1997) with $c = 1.1$, which gives ~37% probability of a total loss (Figure 4.5).

We also looked at NTSB all aviation accident dataset²⁴ for all accidents classified as 'Boeing' or 'Airbus' and events related to Fire or Explosion and aircraft category classified as 'AIR'. This included 158 losses, of which 114 were classified as minor/substantial/destroyed. The percentage of total losses (destroyed) in the dataset is 34/114~30%, which is in the same ballpark as the 37% we have obtained using the third-party database.

²⁰Airbus – Distribution of Accidents by Accident Category - <https://accidentstats.airbus.com/statistics/accident-categories>

²¹ Full-scale flammability test data for validation of aircraft fire mathematical models <https://ntrs.nasa.gov/citations/19820013292>

²² Background : FAA Fire Safety - <https://www.fire.tc.faa.gov/Research/Background>

²³ Fire Dynamics – NIST - <https://www.nist.gov/el/fire-research-division-73300/firegov-fire-service/fire-dynamics>

²⁴ National Transportation Safety Board (NTSB) data - <https://data.ntsb.gov/avdata/avall.zip>

The probability of a total loss for all widebody aircrafts is ~30% in the NTSB data as compared to ~36.7% in the third party dataset. The probability of a minor loss is about ~50% in the NTSB data and this compares well to the ~45% probability of losses with less than 20% hull damage expected in the third party data.

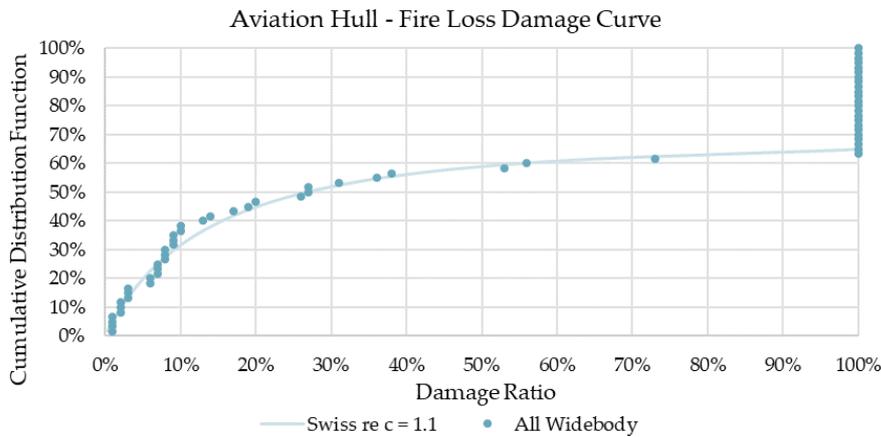


Figure 4.5. The normalised severity curve for aviation hull, based on a data set of 60 fire losses between 1974 and 2020.

4.5.3 The conjecture – damage to the hull is a function of cabin temperature and time to flashover

Based on the damage curve analysis above, we assume that the probability of the total loss of 37% corresponds roughly to the percentage of times that the flashover occurred. In order to analyse the relationship between temperature and damage we first analyse the probability distribution of time to flashover from ignition start. We assume that the distribution of time to flashover is exponential: $F(t) = 1 - \exp(-\lambda t)$, where T is the time to flashover and $\lambda = 1/325s$ (where 325s corresponds to the mean time to flashover based on the FAA data).

Based on the exponential distribution assumption, there is a ~37% chance that a flashover will not occur, matching the probability of a total loss discussed above.

One may therefore put forward the conjecture that there exists a relationship between time to flashover, temperature of the cabin and degree of hull damage. It makes intuitive sense that the longer the fire burns, the higher the temperature in an aircraft cabin and the wider it spreads, leading to higher damage. However, without any experimental evidence that links damage to time/temperature this remains a conjecture.

We have plotted the NASA temperature curve and the damage ratio curve both as a function of time since ignition start. At 325 seconds, the maximum temperature achieved is close to 800 degrees Celsius which is the maximum temperature achieved as per the NASA fire tests and the degree of hull damage is also 100% (total loss).

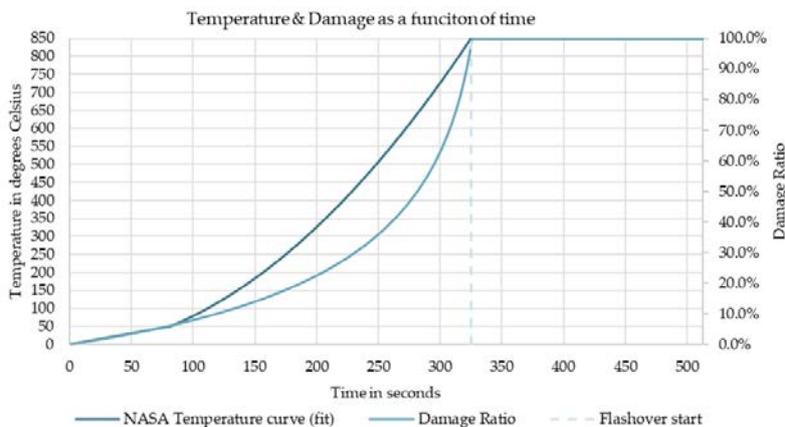


Figure 4.6. Temperature and damage as a function of time: NASA temperature curve vs the damage ratio derived from a Swiss Re curve with $c=1.1$.

The analysis above is based on the conjecture that the damage to an aircraft hull due to fire is proportional to the time since ignition and temperature achieved in the cabin. It provides a framework to estimate the severity of a Hull loss due to fire as a function of cabin material and aircraft type. The conjecture remains to be proven but it provides an example of an approach from first principles in the analysis of the severity of aircraft hull damage losses. There are several limitations and assumptions in this analysis. The assumptions underlying the proposed model are sensitive to the parameters considered and there is significant uncertainty underlying the models.

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 – to predict the distribution of the size of losses. It is a clear example 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 focus on a single example, that of earthquake modelling.

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 (Gutenberg & Richter, 1956), 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 :

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

In other terms, 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-organized criticality**²⁵, 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, analogously 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

²⁵ Self-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.

here is likely to be found for other types of natural catastrophes (Mandelbrot, 1982); similarly with Helmstetter's approach of fractal scaling.

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 – that Zipf's law which we discussed 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 – other things being equal – much more damage than an 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 rate will often be set as a simple multiple of the Property Damage rate. The loss data available will often be expressed in terms of financial amounts as opposed to physical delay times (numbers of days lost) so often combines information on:

- the interruption period;
- the financial impact per day (which will differ for partial losses and total losses²⁶), and
- 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 where a significant amount of oil and gas production may flow through 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);
- any mitigation that may be available: e.g., 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 (e.g., cyber attack; terrorist activity either directly at the location or in the immediate vicinity); pandemic impact upon the workforce etc.

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.

The approach here is not "linear" in the sense that some consideration should be given to the wider "network" of exposures. A more sophisticated analysis would also consider the time element, e.g., to adjust the exposure for insuring conditions such as waiting periods; indemnity periods and maximum daily recovery rates etc.

²⁶ See 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.

As a simple example consider a typical oil and gas exposure where the network can be described as in Figure 6.1. The Direct Business Interruption values at risk and the relative dependency is summarised in Table 6.1.

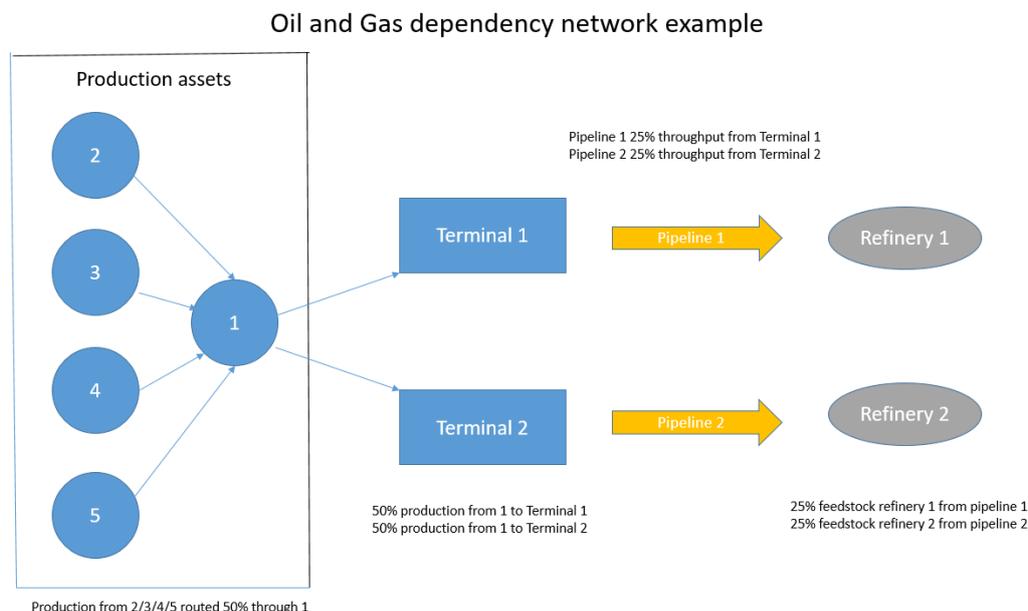


Figure 6.1: Oil and Gas network example.

ID	Name	Type	Direct BI (\$M)	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 6.1: The dependency matrix for Figure 6.1. As mentioned in Section 4.2.1, such matrix can be viewed a representation of a weighted (directed) graph, with an edge from j to k if the value at (j, k) is > 0 .

The critical dependencies can be summarised as follows:

- (1) Assets 2 to 5 are routing production via Asset 1 (a gathering platform for the whole field) so are 50% dependent: e.g., 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%.
- (2) Terminals 1 and 2 are each taking 50% of the production from Asset 1.
- (3) Pipeline 1 gets 25% of its throughput from Terminal 1. Pipeline 2 gets 25% of its throughput from Terminal 2.
- (4) 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 “ripples” 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 i.e. 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 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 6.2.

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

Table 6.2: Business Interruption: lower order effects

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 6.3.

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 6.3: Business Interruption: lower order effects, revised

This can be summarised by considering the relationship between “direct” BI (first order effect) and “indirect” BI (lower order effects) as in Table 6. 5.

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%

Table 6.4: Direct BI vs Indirect BI – original and revised network

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 indeed be much more common to declare a simple “catch all” additional \$100M for contingent business interruption based on Table 6.5 (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 takes into account of these factors and tests the impact of varying the coverage (for example increased waiting periods or sub-limits on indirect losses).

7. Modelling the severity of Cyber Business Interruption losses

As for the case of modelling contingent business interruption through supply-chain disruption, 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 not only interesting from a theoretical point of view but 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, which also translates into shorter contractual indemnity periods.
- Cyber BI is not location-specific: a cyber attack can trigger business interruption losses in different territories.

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 IT service provider (cloud) outage, local IT service outage, 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.

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

1. *Contamination*. Cyber threats are starting to emerge and successfully impacting the operation of the insured by up to $X\%$ depending on the insured's cybersecurity posture²⁷ 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, e.g., 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

²⁷ Cybersecurity posture refers to the cyber defence measures of the insured, e.g., authentication, employee training, deployed security system/solutions and the collective effectiveness of the combination of individual measures.

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 longer time.

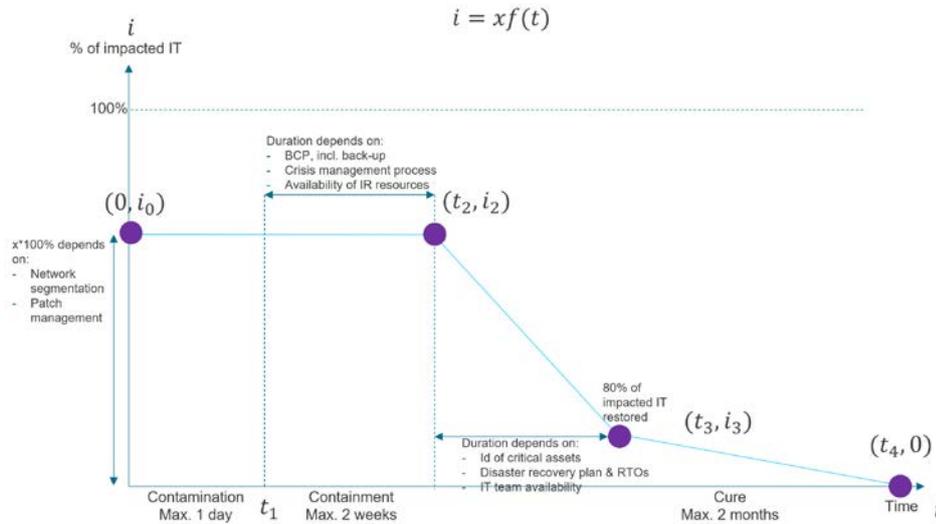


Figure 7.1. A simplified, linearised version of the cyber disruption pattern.

Figure 7.1 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%: e.g., 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 inset of the three different phases.

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. Below is an illustration of such an assessment for a company of a certain size in a given country, for 3 NACIS codes.

	$\max(t_1)$	$\max(t_2 - t_1)$		$\max(t_3 - t_2)$	$\max(t_4 - t_2)$	$\max(x)$
NAICS-based Segmentation	Max Contamination (day)	Max Containment Time (day)	RM Process Disruption (Cyr ius)	Max time to restore 80% of impacted IT (day)	Max time to restore 100% of impacted IT (day)	Max(x)
Agriculture, Forestry, Fishing and Hunting	1	14	1.53	38.45	115.34	0.38
Mining, Quarrying, and Oil and Gas Extraction	1	14	1.84	43.25	129.76	0.49
Utilities	1	14	2.00	45.66	136.97	0.55

Figure 7.2. Example of assessment of the parameters of a cyber disruption pattern.

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.

We denote as \bar{L}_d 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 with the following formula.

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

Monte-Carlo simulations can then be performed to generate one million Cyber BI scenarios for each sector. A few random scenarios generated for sector “Finance and Insurance” are shown in Figure 7.3.

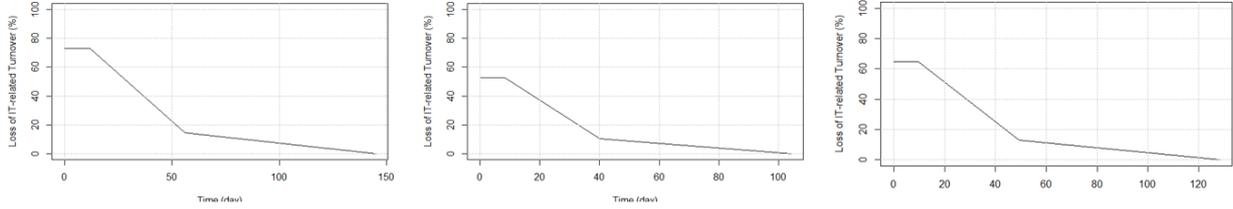


Figure 7.3. Three random scenarios generated by varying the parameters of the cyber disruption period.

Using this simulation, we can then produce a distribution of the possible losses for a particular sector and derive from that the corresponding exposure curve. Figure 7.4 shows both for the sector “Finance and Insurance”.

One advantage of this approach is that the maximum possible loss (MPL) is approximately equal to the maximum simulated loss if the number of simulations is sufficient. For the Finance and Insurance sector, 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 industrial 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 itself can be written as

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

where M is 48.47 according to the calculations above for the Finance and Insurance sector.

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

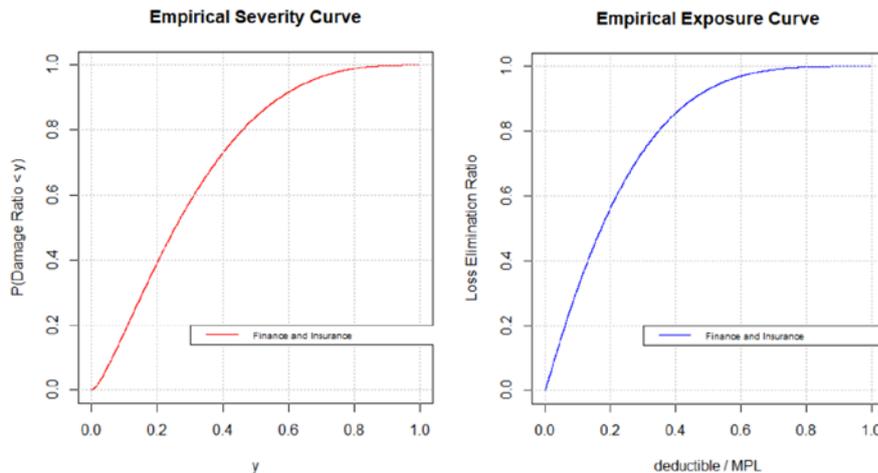


Figure 7.4. The empirically-derived CDF plot for the damage ratio (left) and exposure curve (right) for the Finance and Insurance sector.

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 only 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 in 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), as 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 report is wide enough to deserve 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. As for the derivation of severity curves from compensation tables, moving from the conceptual stage to a well-calibrated approach requires data on the frequency with which different degrees of disability are suffered as a result of accidents, on injury correlation, and on other demographic information.

In Section 4, more research is needed on the relationship between graph structure and exposure curve and to 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 (e.g. extended warranty, machinery breakdown, 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.

9. Acknowledgments

Our thanks to Frank Cuypers, Yuriy Krvavych and the ASTIN section for encouraging us to create this working party; and to Christian Levac and Walter Neuhaus for their help through the process. We also thank our companies: SCOR, Gallagher Specialty, WTW, QBE, Hiscox, Acasta, Marsh for their support in this research initiative. We are indebted to several people for advice on the actual content of this report. 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. 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 Baremo tables for Italy and other countries. Finally, we would like to thank Shane Lenney for reviewing this paper and providing us with helpful advice.

10. References

Anscombe, F. J. (1950), Sampling Theory of the Negative Binomial and Logarithmic Series Distributions, *Biometrika*, Dec. 1950, Vol. 37, No. 3/4, pp. 358-382. Stable URL: <https://www.jstor.org/stable/2332388>

Applebaum, S. (2009). *Lévy processes*. Oxford University Press.

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. In: *Lévy driven volatility models* (Chapter 2). <https://pure.au.dk/ws/files/193874249/levybook.pdf>
- Benckert, L. (1962). The Lognormal Model for the Distribution of one Claim. *ASTIN Bulletin: The Journal of the IAA*, 2(1), 9-23. doi: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>
- Boswell, M. T. (1970). Chance mechanisms generating the negative binomial distribution. *Random counts in scientific work*.
- Champernowne, D. G. (1953). A contribution to the theory of economic growth. *Economic Journal*, 63(250), 318-335.
- Devroye, L. (1986), *Non-Uniform Random Variate Generation*. New York: Springer-Verlag.
- 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.
- Galea, E.R., & Markatos, N.C. (1987). A review of mathematical modelling of aircraft cabin fires. *Applied Mathematical Modelling*, 11(3), 162-176.
- Guggisberg, D. (2004), *Exposure rating*, Zurich: Swiss Re. <http://www.swissre.com>.
- 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, August). 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.
- 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.
- Mitchell-Wallace, P., Wilson, D., & Peacock, N. (Eds.). (2012). *Natural catastrophe risk management and modelling*. John Wiley & Sons.
- Parodi, P., and Watson, P. (2019). Property graphs – A statistical model for fire and explosion losses based on graph theory. *ASTIN Bulletin*, 49(2), 263-297. doi:10.1017/asb.2019.4
- Parodi, P. (2023), *Pricing in General Insurance* (2nd edition), New York: CRC Press, New York.
- Rath, B. and Toth, B. (2009) Erdos-Renyi random graphs + forest fires = self-organized criticality. *Electron. J. Probab.* 14. Paper no. 45, 1290--1327. doi: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.