A graph theoretic approach to the derivation of property loss curves from first principles

Peter Watson
Cass Business School
London
Supervisor: Dr. Pietro Parodi

This dissertation is submitted as part of the requirements for the award of the MSc in Actuarial Science.

August 31, 2017

Abstract

Property fire risk and the associated loss curves for portfolios of properties are often heuristically described using the MBBEFD distribution. However, no known models exist to describe the underlying situation from first principles. In particular, there is no known method for studying the risk associated with specific properties in detail. In this study, two such models are derived utilising graph theory and an exploratory analysis performed. The first model is based on a physical simulation of fire propagation through a graph that represents the property in question, over a randomly distributed time period (the graph vertices correspond to rooms and the graph edges correspond to routes that a fire might take in moving from one room to another). Using an input time distribution fitted to a portfolio of properties, the loss distribution for an individual property may be derived using Monte Carlo techniques. The second model employs random graph theory. In this scenario, each possible edge of the graph for a particular property is included with a probability that represents the possibility of fire transmission. A vertex is selected at random to be the starting point for the fire and the subsequent connected subgraph to which it belongs then represents the loss. Monte Carlo simulation then generates the loss distribution. It is found that the loss distributions generated by both models are well approximated by the MBBEFD distribution. Further, when applied to individual properties, the two models give rise to distinct behaviours that could be used to discriminate between them in practice.

Acknowledgements

I would like to thank my supervisor, Dr. Pietro Parodi, for his enthusiasm and patience. It has been a real pleasure working with him.

Contents

1 Introduction **7**

1.1 Severity curves and exposure rating 7

1.2 Graph theory 11

2 MBBEFD distribution in more detail **13**

2.1 Test Monte Carlo 14

2.2 Parameter estimation 16

2.2.1 Maximum likelihood 16

2.2.2 Least squares exposure 18

2.2.3 Quick way to find g (and c) 19

2.3 Connection between the MBBEFD and other distributions 20

3 Fire simulation model **21**

3.1 A physical illustration 21

3.2 A fictitious scenario 22

3.3 Fire propagation 25

3.4 Time distribution 28

3.5 Results for the residential property portfolio 29

3.6 Other property types: an office block example 34

3.7 Conclusions for the fire simulation model 36

4 A purely graph theoretic approach **37**

4.1 Erdős-Rényi random graphs 38

4.2 Property portfolio graphs 40

4.3 Conclusions for the graph theoretic approach 43

5 Conclusions **46**

A Property portfolio **47**

List of Figures

1.1 An example severity curve. 9

1.2 [left panel] The parametric severity curves for $c = 1.5$ and $c = 8$. [right panel] The corresponding parametric exposure rating curves for $c = 1.5$ and $c = 8$. 11

1.3 Example graph. 11

2.1 [left panel] The parametric severity curves for various values of b and g. The parameter set $b = 11.547$, $g = 5.0$ corresponds to the Swiss Re parameterisation with concavity $c = 1.646$. [right panel] The corresponding exposure rating curves. 14

2.2 [left panel] Monte Carlo simulated severity curves for various sample sizes N alongside the analytic result. The parameter set is $b = 11.547$, $g = 5.0$ (and corresponds to the Swiss Re parameterisation with concavity $c = 1.646$). [right panel] The corresponding exposure rating curves. 15

2.3 A plot of the absolute difference between the analytic and Monte Carlo simulated severity curves, $|F(x) - \tilde{F}(x)|$, for various sample sizes. 16

2.4 A parametric plot of k versus l for the Swiss Re parameterisation as a function of the concavity c. 17

2.5 Scatter plot of log-likelihood fitted values k_F and l_F for 100 runs of Monte Carlo simulations with $N = 1000, 10000, 100000$. [left panel] Parameter set $b = 11.547$, $g = 5.0$, or $k = 2.446$, $l = 1.386$. [right panel] Parameter set $b = 1$, $g = 10$, or $k_F = 0$, $l_F = 2.197$. 18

2.6 Scatter plot of least squares exposure fitted values k_G and l_G for 100 runs of Monte Carlo simulations with $N = 1000, 10000, 100000$. [left panel] Parameter set $b = 11.547$, $g = 5.0$, or $k = 2.446$, $l = 1.386$. [right panel] Parameter set $b = 1$, $g = 10$, or $k_F = 0$, $l_F = 2.197$. . 19

3.1 A (crude rendition of a) house. The upper diagram is the ground floor and the lower diagram is the first floor. Doors are shown in red. The loft (room '12') is not shown. . . . 22

3.2 The graph for the house, Fig. 3.1. Each room (including the loft) is represented by a numbered vertex. For the edges between vertices: thick black lines represent open spaces between rooms, thin red lines represent doors that are open with probability p_d, dashed blue lines represent all other possible routes that a fire may propagate. 26

3.3 Plot of the fractional loss due to a fire in the example house of Fig. 3.1 as a function of time step, for different door probabilities (each with 100 simulations). [top left panel] $p_D = 0.0$ (all doors closed). [top right panel] $p_D = 0.5$ (equal probability for each door to be open/closed). [lower left panel] $p_D = 1.0$ (all doors open). [lower right panel] Compendium of the average fractional loss for the different probabilities. 27

3.4 [left panel] Scatter plot of log-likelihood fitted parameters k_F versus l_F for various sample sizes. [right panel] Scatter plot of least squares exposure fitted parameters k_G versus l_G for various sample sizes. Parameters are: $\tilde{b} = 1.3$, $\tilde{g} = 2.0$, $t_m = 50$, $p_D = 0.8$. 30

3.5 [upper left panel] Comparison of the model Monte Carlo severity curve for the portfolio of properties and the corresponding fitted curves with $p_D = 0.8$. [upper right panel] Comparison of the model Monte Carlo exposure curve for the portfolio of properties and the corresponding fitted curves with $p_D = 0.8$. [lower panels] As for the upper panels but with $p_D = 0$. Parameters are: $\tilde{b} = 1.3$, $\tilde{g} = 2.0$, $t_m = 50$, $N = 100000$. 31

3.6 [upper left panel] Empirical severity curves for the individual properties (plus the portfolio). [upper right panel] Scatter plot of the corresponding log-likelihood fitted parameters k_F versus l_F for the individual properties (plus the portfolio); also shown for reference is the Swiss Re parametric curve of Fig. 2.4. [lower left panel] Empirical exposure curves for the individual properties (plus the portfolio). [lower right panel] Scatter plot of the least squares exposure fitted parameters k_G versus l_G for the individual properties (plus the portfolio). Parameters are: $\tilde{b} = 1.3$, $\tilde{g} = 2.0$, $t_m = 50$, $p_D = 0.8$, $N = 100000$. 33

3.7 Probability density function for the time distribution, $f_{T|T<t_m}(t)$, Eq. (3.6.5), for different parameter sets $\tilde{b}$ and $\tilde{g}$. 35

3.8 [left panel] Empirical severity curves for the 'office' building with different parameter sets, $\tilde{b}$ and $\tilde{g}$, compared to the portfolio of residential properties. [right panel] Scatter plot of the resultant k_F and l_F parameter estimates; also shown for reference is the Swiss Re parametric curve of Fig. 2.4. 35

4.1 [left panel] Scatter plot of log-likelihood fitted parameters k_F versus l_F for various sample sizes using the Erdős-Rényi random graphs. [right panel] Scatter plot of least squares exposure fitted parameters k_G versus l_G, again for various sample sizes. Parameters are: $n = 20$, $p = 0.05$. 38

4.2 [left panels] Plots of the empirical severity distribution alongside the fitted curves using log-likelihood or least squares exposure fitting for the Erdős-Rényi random graph model. [right panels] Plots of the empirical exposure curve alongside the fitted curves using log-likelihood or least squares exposure fitting. Parameters are: $n = 20$, $p = 0.025$ [upper panels], $p = 0.05$ [middle panels], $p = 0.1$ [lower panels]. 39

4.3 [left panel] Plot of log-likelihood estimates k_F versus l_F for different graph orders (n) and probabilities (p) using the Erdős-Rényi random graphs. Shown for reference is the parametric curve for the Swiss Re parametrisation in terms of the concavity c. [right panel] Plot of least squares exposure estimates k_G versus l_G. Monte Carlo sample size is $N = 100000$. 40

4.4 [left panel] Scatter plot of log-likelihood estimates k_F versus l_F for various sample sizes using the residential property portfolio graphs. [right panel] Scatter plot of least squares exposure estimates k_G versus l_G for various sample sizes. Parameters are: $p_D = p_W = 0.1$. 41

4.5 [left panels] Plots of the empirical severity distribution alongside the fitted curves using log-likelihood or least squares exposure fitting for the residential property portfolio graph model. [right panels] Plots of the empirical exposure curve alongside the fitted curves using log-likelihood or least squares exposure fitting. Parameters are: $p_D = p_W = 0$ [upper panels], $p_D = p_W = 0.1$ [middle panels], $p_D = 0.5$, $p_W = 0.1$ [lower panels]. Monte Carlo sample size is $N = 100000$. 42

4.6 [left panel] Plot of log-likelihood estimates k_F versus l_F for different probabilities p_D and p_W using the residential property portfolio graphs. Shown for reference is the parametric curve for the Swiss Re parametrisation in terms of the concavity c. [right panel] Plot of least squares exposure estimates k_G versus l_G. Monte Carlo sample size is $N = 100000$. . 43

4.7 [upper left panel] Empirical severity curves for the individual properties (plus the portfolio). [upper right panel] Scatter plot of the corresponding log-likelihood fitted parameters k_F versus l_F for the individual properties (plus the portfolio); also shown for reference is the Swiss Re parametric curve of Fig. 2.4. [lower left panel] Empirical exposure curves for the individual properties (plus the portfolio). [lower right panel] Scatter plot of the least squares exposure fitted parameters k_G versus l_G for the individual properties (plus the portfolio). Parameters are: $p_D = 0.5$, $p_W = 0.1$, $N = 100000$. 45

A.1 The graphs corresponding to the portfolio of residential properties, 'property 1-9' and the office building 'office'. The different edges follow the notation of Fig. 3.2. 49

List of Tables

3.1 Parameter estimates k_F, l_F and k_G, l_G (with their respective b and g values) for the portfolio and individual properties. For reference, the Swiss Re parameterisation with concavity $c = 1.5$ has $k = 2.54$, $l = 1.17$ ($b = 12.65$, $g = 4.22$). 34

4.1 Parameter estimates k_F, l_F and k_G, l_G (with their respective b and g values) for the portfolio and individual properties. For reference, the Swiss Re parameterisation with concavity $c = 1.5$ has $k = 2.54$, $l = 1.17$ ($b = 12.65$, $g = 4.22$). 44

A.1 Distribution of housing (England and Wales), grouped by number of bedrooms [narchive2016]. 47

A.2 Distribution of properties in the portfolio. 48

Chapter 1

Introduction

As any homeowner will tell you, property insurance is a necessity. For most property owners, residential or commercial, their property represents a significant proportion of their assets and cannot be replaced easily, if at all, without insurance cover. One of the key components of property insurance is fire risk – the risk that some, or all of a building may be damaged in the event of a fire. This study addresses the question of the severity of a loss. The severity of losses due to fire risk is most often heuristically described using the MBBEFD distribution.[1] However, to the best of our knowledge, there is no model for the loss severity based on first principles. In this study, two such models shall be developed and an exploratory analysis performed. The first model separates the propagation of a fire through a physical building and the time during which the fire propagates. By introducing a time distribution fitted to the data, the loss severity distribution for individual buildings may be analysed. The second model uses graph theory to approach the problem from a different angle.

In this first chapter, we introduce severity curves, exposure rating, and the basic graph theory that will be utilised throughout. Next, the MBBEFD distribution will be discussed with reference to various technical and numerical aspects. The third chapter is concerned with the fire simulation model and culminates in a set of results. Likewise, the fourth chapter deals with the purely graph theoretic approach and associated results. The final chapter details the conclusions of the study. The portfolio of properties used in the study (in graph theoretic form) is presented in the appendix.

1.1 Severity curves and exposure rating

Severity curves and exposure rating for property losses are described in detail in many textbooks, see for example, Ref. [parodi2015]. Let us begin by briefly introducing the main concepts.

For an insurer, one of the primary quantities of interest is the expected annual aggregate loss for a portfolio of contracts, $E[S]$. In the collective risk model, the aggregate loss S may be modelled as a

[1] MBBEFD stands for *Maxwell–Boltzmann, Bose–Einstein, Fermi–Dirac* – three sets of physicists that developed the statistical mechanics to describe certain properties of classical, bosonic, and fermionic gases (the latter two being quantum phenomena).

compound distribution:

$$S = \sum_{i=1}^{N} X_i, \qquad E[S] = E[X]\,E[N], \tag{1.1.1}$$

where N is a random number (drawn from a chosen frequency distribution) representing the frequency of claims for the year[2] and the X_i are independent and identically distributed random numbers for the individual claims amounts. Importantly, the frequency of claims N must be independent of the claims amounts X_i and we shall assume this throughout. In the case of fire risk, this is a valid assumption if one discards the possibility of a large fire spreading to neighbouring properties – this could be a concern where, for example, *i*) buildings in cities are constructed from timber and densely packed (as was the case, e.g., for the Great Fire of London in 1666 [gfol1666]), *ii*) there is the possibility of widespread civil unrest or conflict, or *iii*) there is the possibility of a gas explosion affecting several buildings simultaneously.

One of the defining characteristics of property losses is that, depending on the type of property being insured, the loss cannot exceed either the sum insured (SI) or the maximum possible loss (MPL).[3] Thus,

$$X_i \in [0, M], \tag{1.1.2}$$

where M refers to either the SI or the MPL as appropriate. It then makes sense to consider the severity of a loss as a fraction (or percentage) of its corresponding maximum, i.e.,[4]

$$x = \frac{X}{M}. \tag{1.1.3}$$

Moreover, there may exist a finite probability for a loss due to fire to be a total loss (e.g., $P[x = 1] = 0.22$). This may be easily interpreted – when a fire starts, it can spread and if unchecked, will consume the entire building, whereupon it can do no further damage (assuming that it does not spread to another building). The severity curve is the cumulative distribution function for an individual loss, $F(x)$, expressed as a fraction of the maximum loss (x) and a typical example is sketched in Fig. 1.1.

To get the severity curve for a particular building, it would be necessary to have fire loss data from a large number of identical buildings. Given that the risk from fire is (fortunately) low and that buildings come in a myriad of shapes, sizes, and configurations, there is not enough data in an insurer's portfolio to construct an empirical severity curve in this fashion. It is necessary to consider reinsurance data, i.e., the severity curves for the insurance market as a whole, broken down into segments of roughly similar building types. Such data is held by Swiss Re [swiss2004] and is conveniently represented using exposure curves.

To understand exposure curves, consider the following reinsurance scenario: for each loss X, there is

[2] We ignore the complications arising from the time delays between the loss event occurring, the claim being reported, and the claim being settled – in effect assuming that any claims are settled immediately after the corresponding loss occurs.

[3] The MPL for fire risk applies to, for example, the case where an industrial complex is insured as a single entity, but which comprises several buildings – the MPL would then be the worst case scenario where the most valuable building is completely destroyed (assuming that any fire does not spread between buildings).

[4] Note that we break slightly with the convention that random variables be denoted by a capital letter.

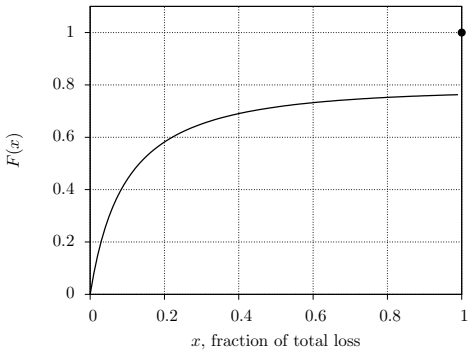

Figure 1.1: An example severity curve.

a deductible D and an excess-of-loss retention level of $D + L$. For the insurer this entails a payout of

$$X^I = \min(X, D + L) - \min(X, D) = \begin{cases} 0, & X < D \\ X - D, & D \leq X \leq D + L \\ L, & X > D + L \end{cases}, \tag{1.1.4}$$

or an expected loss of

$$E\left[X^I\right] = E\left[\min(X, D + L)\right] - E\left[\min(X, D)\right]. \tag{1.1.5}$$

If the deductible (D) and retention levels ($D + L$) for a portfolio are set to common fractions of the sum insured or maximum possible loss, then we can write

$$d = \frac{D}{M}, \quad l = \frac{L}{M}, \tag{1.1.6}$$

such that

$$\frac{E\left[X^I\right]}{E[X]} = \frac{E\left[\min(x, d + l)\right] - E\left[\min(x, d)\right]}{E[x]} \tag{1.1.7}$$

is independent of M. The aggregate loss for the insurer, given a portfolio of roughly similar buildings (i.e., where the severity distribution for each loss is identically distributed), can then be rewritten as

$$\begin{aligned} E\left[S^I\right] = E\left[\sum_{i=0}^{N} X_i^I\right] &= \frac{E\left[X^I\right]}{E[X]} \times E[N]\, E[X] \\ &= \frac{E\left[\min(x, d + l)\right] - E\left[\min(x, d)\right]}{E[x]} E[S]. \end{aligned} \tag{1.1.8}$$

Thus, the reinsurance details can be factorised from the aggregate expected losses. The exposure rating, $G(u)$ is defined as

$$G(u) = \frac{E\left[\min(x, u)\right]}{E[x]}, \quad u \in [0, 1] \tag{1.1.9}$$

(note that G is expressly independent of x due to the expectation value) and represents the ratio of the expected retained loss to the expected loss, where there is no deductible but an excess-of-loss retention fraction u.

There is a one-to-one correspondence between the severity curve and the exposure rating (and its derivative). This follows from the fact that (using integration by parts)

$$E\left[\min(x, u)\right] = \int_0^u dx\ \left[1 - F(x)\right], \tag{1.1.10}$$

so

$$G(u) = \frac{\int_0^u dx\ [1 - F(x)]}{\int_0^1 dx\ [1 - F(x)]}. \tag{1.1.11}$$

Differentiating with respect to u,

$$G'(u) = \frac{1 - F(u)}{E[x]}. \tag{1.1.12}$$

This can be inverted to find $F(u)$ in terms of $G'(u)$ (recalling the discontinuity in F and that $F(0) = 0$):

$$F(u) = \begin{cases} 1 - \frac{G'(u)}{G'(0)}, & 0 \leq u < 1 \\ 1, & u = 1 \end{cases}. \tag{1.1.13}$$

The probability of a total loss may be conveniently expressed as

$$P(x = 1) = 1 - \lim_{x \to 1^-} F(x) = \frac{G'(1)}{G'(0)}. \tag{1.1.14}$$

Notice the special case of the 'gunpowder' scenario where all losses are total losses, i.e., $P[x = 1] = 1$. This means that $F(x < 1) = 0$ such that in Eq. (1.1.11) we have

$$G(u) = \frac{\int_0^u dx}{\int_0^1 dx} = u. \tag{1.1.15}$$

In other words, the gunpowder scenario corresponds to the case where $G(u) = u$ is a straight line. The interpretation of $G(u)$ as the ratio of the expected retained loss to the expected loss is particularly evident in this case: as each loss is a total loss with an expected value of unity (in the scaled units), the payout is equal to the retention level fraction u.

The Swiss Re exposure curves for the different groups of building types can be written in parametric form [bernegger1997] – they are based on the MBBEFD distribution. The result is

$$G(x) = \frac{\ln\left[\frac{(g-1)b + (1-bg)b^x}{(1-b)}\right]}{\ln(bg)}, \qquad F(x) = \begin{cases} \frac{b(g-1)(1-b^x)}{b(g-1)+(1-bg)b^x}, & 0 \leq x < 1 \\ 1, & x = 1 \end{cases}, \tag{1.1.16}$$

where

$$b = \exp\{\alpha + \beta c(1 + c)\}, \qquad g = \exp\{(\gamma + \delta c)c\} \tag{1.1.17}$$

and

$$\alpha = 3.1, \quad \beta = -0.15, \quad \gamma = 0.78, \quad \delta = 0.12. \tag{1.1.18}$$

The remaining parameter c is called the concavity and distinguishes between the groups of buildings: e.g., for residential properties, $c = 1.5$, whereas for major industrial complexes c can go up to 8 [swiss2004] (the gunpowder scenario has $c = 0$). The curves for $c = 1.5$ and $c = 8$ are plotted in Fig. 1.2. Starting with the severity curves on the left, it is clear that for industrial property, the probability of a small loss (relative to the maximum) is high while the probability of a total loss is very small. This is not true for residential property where there is a sizeable probability for a total loss ($P[x = 1] \approx 0.22$). For the exposure rating curves, we note that they are continuous, concave functions. The case $c = 1.5$ more resembles the straight line curve $G(x) = x$ (the gunpowder scenario where all losses are total losses).

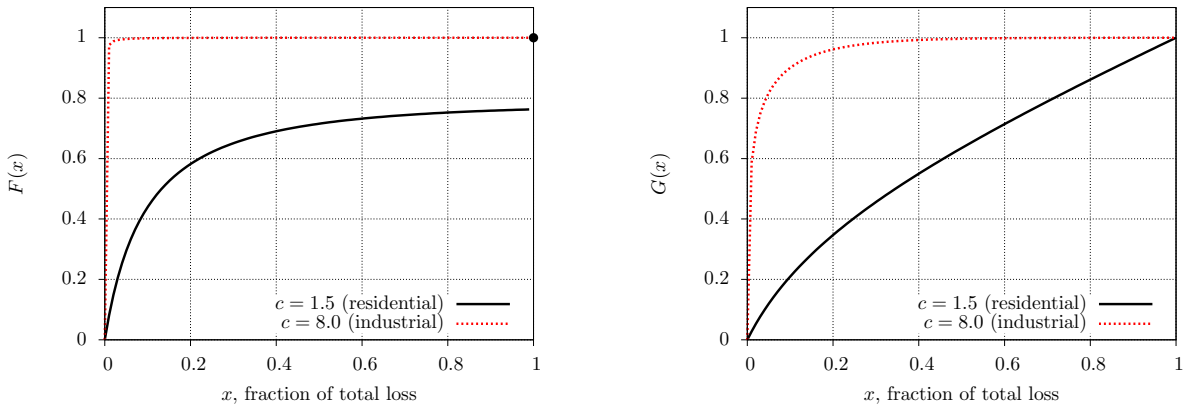

Figure 1.2: [left panel] The parametric severity curves for $c = 1.5$ and $c = 8$. [right panel] The corresponding parametric exposure rating curves for $c = 1.5$ and $c = 8$.

1.2 Graph theory

We shall be using graph theory techniques to model physical buildings in this study, so it is helpful to introduce the nomenclature and some pertinent results (see for example, one of the introductory textbooks: Refs. [bollobas1979, bondy1976]).

A graph consists of a set of vertices (nodes) possibly connected by edges. Denoting the set of vertices V and the set of edges E, the graph G is the ordered pair (V, E) such that E is a subset of the set of unordered pairs of distinct elements of V. We shall denote the vertices $v_1, v_2, \ldots$, and the edges $(v_1, v_2), (v_2, v_3), \ldots$. Note that there is no distinction between the ordering of the vertices for an edge, i.e., $(v_1, v_2) = (v_2, v_1)$.[5] Further, no edge may connect a vertex to itself, i.e., (v_1, v_1) is not allowed. Moreover there can be only one edge between two vertices.

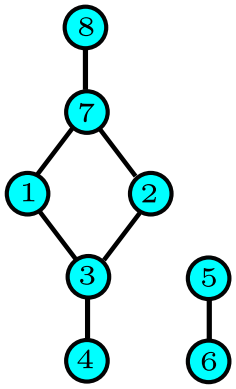

Figure 1.3: Example graph.

An example graph is shown in Fig. 1.3. The set of vertices is

$$V = \{v_1, v_2, \ldots, v_8\} \tag{1.2.19}$$

and the set of edges is

$$E = \{(v_1, v_7), (v_1, v_3), (v_2, v_7), (v_2, v_3), (v_3, v_4), (v_5, v_6), (v_7, v_8)\}\,. \tag{1.2.20}$$

The order of a graph G is the number of vertices and is denoted $|G|$; clearly, $|G| = |V|$. In this study, we are concerned with finite graphs – those with a finite order. The size of the graph G is the number of edges, which we denote $e(G)$. The maximum size for a graph is

$$\text{Max}\,\{e(G)\} = \binom{|G|}{2} = \frac{1}{2}|G|\,(|G| - 1)\,. \tag{1.2.21}$$

[5] Where the ordering of the vertices is relevant, the edge is called an arc. Further, the graph is called a *directed* graph. We shall not be considering such graphs in this study.

The density of a graph, denoted $D(G)$ is defined as the ratio of the number of edges (the size) to the maximum number of edges:

$$D(G) = \frac{2e(G)}{|G|\left(|G|-1\right)}. \tag{1.2.22}$$

In the example graph shown in Fig. 1.3, the order of the graph is $|G| = 8$, the size is $e(G) = 7$, and the density is $D(G) = 0.25$.

Two vertices v_i, v_j are adjacent if the edge (v_i, v_j) exists. Similarly, two edges are adjacent if they share a common vertex. The set of vertices adjacent to v is denoted $\Gamma(v)$ and the degree of v is defined as $d(v) = |\Gamma(v)|$. The so-called "handshaking lemma" relates the total degree to the size of the graph:

$$\sum_{i=1}^{|G|} d(v_i) = 2e(G). \tag{1.2.23}$$

A path is a graph (typically a subgraph of a larger graph) of the form

$$V = \{v_1, v_2, \ldots, v_n\}, \quad E = \{(v_1, v_2), (v_2, v_3), \ldots (v_{n-1}, v_n)\}. \tag{1.2.24}$$

A simple path has no repeated vertices. A graph is connected if, for every distinct pair of vertices, $v_i, v_j \in V$, there is a path (not necessarily simple) from v_i to v_j. The connected components of a graph are subgraphs. In the example, Fig. 1.3, we can clearly identify two connected components. A complete graph is a graph whose vertices all have edges linking to every other vertex (i.e., the degree of each vertex is $|G| - 1$ and the density is unity). Conversely, an empty graph is a graph where there are no edges. Further, a bridge is an edge which, if eliminated from the graph, would result in that graph becoming disconnected (for example, the edge (v_5, v_6) in Fig. 1.3 is a bridge).

Chapter 2

MBBEFD distribution in more detail

Let us consider the severity and exposure curves, $F(x)$ and $G(x)$, respectively, of Eq. (1.1.16) and the MBBEFD distribution in more detail. For our purposes, it will be useful to consider the parameterisation with b and g treated separately rather than the Swiss Re parameterisation with b and g constrained via the concavity c as in Eq. (1.1.17). The parameters must obey the following constraints:

$$b > 0, \qquad g \geq 1, \tag{2.0.1}$$

the latter arising from the interpretation of g as the inverse of the probability of a total loss:

$$P(X = 1) = 1 - \lim_{x \to 1^-} F(x) = \frac{1}{g}. \tag{2.0.2}$$

The curves have special cases for particular values of the parameters that must be implemented numerically in order to avoid spurious division-by-zero type errors (or instabilities). These special cases are derived using asymptotic expansion (they are also highlighted in Ref. [bernegger1997]). The expressions given in Eq. (1.1.16) are modified:

$g = 1$: The first special case is $g = 1$ (the gunpowder scenario). The severity and exposure curves become independent of b and are

$$F(x) = \begin{cases} 0, & 0 \leq x < 1 \\ 1, & x = 1 \end{cases}, \qquad G(x) = x. \tag{2.0.3}$$

$b = 1$: The next case is $b = 1$:

$$F(x) = \begin{cases} \frac{(g-1)x}{1+(g-1)x}, & 0 \leq x < 1 \\ 1, & x = 1 \end{cases}, \qquad G(x) = \frac{\ln\left(1 + (g-1)x\right)}{\ln\left(g\right)}. \tag{2.0.4}$$

An interesting connection between this case and the two parameter Pareto distribution is presented in Section 2.3.

$bg = 1$: The last case is $bg = 1$ (note that under this condition b is further restricted to $b \leq 1$):

$$F(x) = \begin{cases} 1 - b^x, & 0 \leq x < 1 \\ 1, & x = 1 \end{cases}, \qquad G(x) = \frac{1 - b^x}{1 - b}. \tag{2.0.5}$$

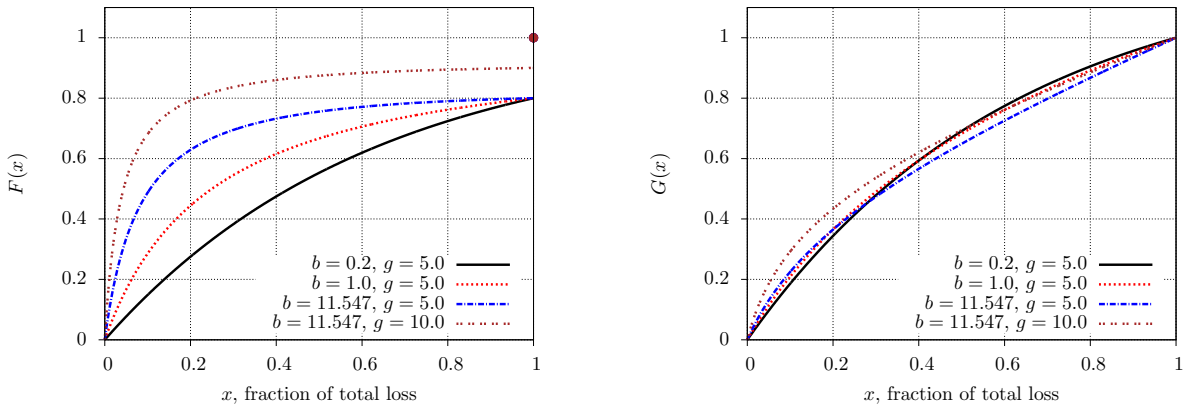

Figure 2.1: [left panel] The parametric severity curves for various values of b and g. The parameter set $b = 11.547$, $g = 5.0$ corresponds to the Swiss Re parameterisation with concavity $c = 1.646$. [right panel] The corresponding exposure rating curves.

There also exists a connection between this and the exponential distribution, discussed in Section 2.3.

The effect of varying the parameters b and g is illustrated in Fig. 2.1 (the parameter set $b = 11.547$, $g = 5.0$ corresponds to the Swiss Re parameterisation with concavity $c = 1.646$). From the severity curve $F(x)$, the roles of the two parameters are particularly evident: g controls which proportion of the losses are total losses (smaller g means fewer partial losses) whereas b affects only the partial losses such that for larger b, there are more small partial losses. The exposure curves show that the variation due to different parameter sets is far less sensitive than for the severity curves – this will become important when considering parameter estimation.

2.1 Test Monte Carlo

It is possible to simulate the severity curve $F(x)$ using Monte Carlo techniques (because F is a cumulative distribution function). Taking a sample, $\underline{u} = (u_1, u_2, \ldots, u_N)$, of N random numbers drawn from the distribution $U \sim \text{Unif}(0, 1)$ we apply the inverse transform method to construct a new sample (allowing for the special discrete case $x = 1$) $\underline{x} = (x_1, x_2, \ldots, x_N)$ drawn from the distribution corresponding to F. The general transform is given by

$$x_i = \begin{cases} \frac{\ln\left[\frac{b(g-1)(1-u_i)}{(1-bg)u_i+b(g-1)}\right]}{\ln(b)}, & 0 \le u_i < 1 - 1/g \\ 1, & u_i \ge 1 - 1/g \end{cases} . \tag{2.1.6}$$

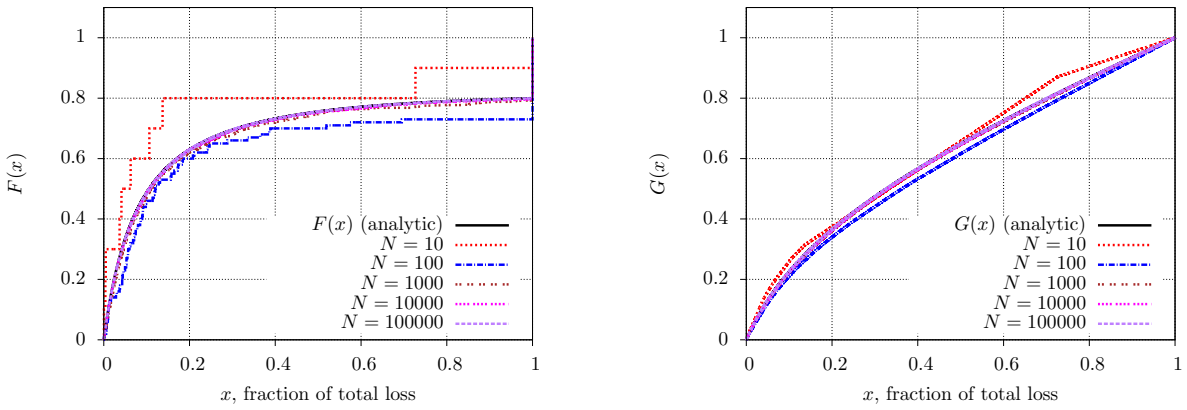

Figure 2.2: [left panel] Monte Carlo simulated severity curves for various sample sizes N alongside the analytic result. The parameter set is $b = 11.547$, $g = 5.0$ (and corresponds to the Swiss Re parameterisation with concavity $c = 1.646$). [right panel] The corresponding exposure rating curves.

Again, we must distinguish the cases for special values of the parameters:

$$\begin{aligned}
\mathbf{g = 1}\text{:} \quad & x_i = 1, \\
\mathbf{b = 1}\text{:} \quad & x_i = \begin{cases} \frac{u_i}{(g-1)(1-u_i)}, & 0 \le u_i < 1 - 1/g \\ 1, & u_i \ge 1 - 1/g \end{cases}, \\
\mathbf{bg = 1}\text{:} \quad & x_i = \begin{cases} \frac{\ln(1-u_i)}{\ln(b)}, & 0 \le u_i < 1 - b \\ 1, & u_i \ge 1 - b \end{cases}.
\end{aligned} \tag{2.1.7}$$

Given the sample $\underline{x}$, the empirical cumulative distribution function is

$$\tilde{F}(x) = \frac{1}{N} \sum_{i=1}^{N} I(x - x_i), \qquad I(y) = \begin{cases} 1, & y \ge 0 \\ 0, & y < 0 \end{cases}. \tag{2.1.8}$$

Similarly, the empirical exposure curve is

$$\tilde{G}(x) = \frac{\sum_{i=1}^{N} \min(x, x_i)}{\sum_{i=1}^{N} x_i}. \tag{2.1.9}$$

The empirical severity and exposure curves with various sample sizes N are compared to the analytic result for the parameters $b = 11.547$ and $g = 5.0$ (corresponding to the Swiss Re parameterisation with concavity $c = 1.646$) in Fig. 2.2. It is seen that for $N \ge 10000$, the curves lie on top of each other. In this respect, it is useful to plot the absolute difference between the analytic and Monte Carlo simulated severity curves, $|F(x) - \tilde{F}(x)|$, for various values of N[1] on a logarithmic scale and this is shown in Fig. 2.3. Clearly, larger sample sizes N come closer to the analytic result. Moreover, the difference scales roughly as $1/\sqrt{N}$, as one would expect from the central limit theorem.

[1]The supremum of $|F(x) - \tilde{F}(x)|$ over the interval $x = [0, 1]$ would then give the Kolmogorov-Smirnov statistic for each value of N.

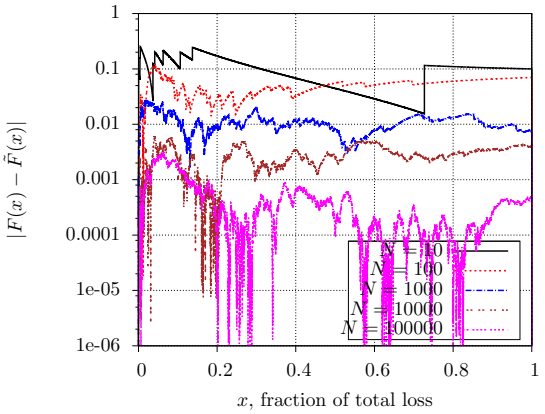

Figure 2.3: A plot of the absolute difference between the analytic and Monte Carlo simulated severity curves, $|F(x) - \tilde{F}(x)|$, for various sample sizes.

2.2 Parameter estimation

We can use the Monte Carlo simulated samples (which are known to belong to the correct distribution) to test various methods of extracting parameter estimates. This also serves to check the numerical implementation of the methods and gives us a rough criterion for convergence of the Monte Carlo simulation.

2.2.1 Maximum likelihood

Let us suppose that we have a sample of size N, $\underline{x} = (x_1, x_2, \ldots, x_N)$, where (if we order the sample) the first m of the x_i are less than one and the remaining n are equal to one (such that $m + n = N$). The likelihood function for the MBBEFD distribution in terms of the parameters b and g is

$$L(b, g|\underline{x}) = \left\{ \prod_{i=1}^{m} f_{X|X<1}(x_i|b, g) \right\} [P(X = 1)]^n , \tag{2.2.10}$$

where

$$P(X = 1) = \frac{1}{g} \tag{2.2.11}$$

and

$$\begin{aligned} f_{X|X<1}(x_i|b, g) &= \frac{d}{dx} F_{X|X<1}(x|b, g)\Big|_{x=x_i} \\ &= \frac{(b-1)(g-1)b^{1-x_i} \ln(b)}{[1 - bg + (g-1)b^{1-x_i}]^2} \end{aligned} \tag{2.2.12}$$

is the probability density function for $x_i < 1$. As before, we must distinguish the special values of the parameters:

$$\mathbf{g = 1}: \quad f_{X|X<1}(x_i|b, g) = 0, \tag{2.2.13}$$

$$\mathbf{b = 1}: \quad f_{X|X<1}(x_i|b, g) = \frac{g-1}{[1 + (g-1)x_i]^2}, \tag{2.2.14}$$

$$\mathbf{bg = 1}: \quad f_{X|X<1}(x_i|b, g) = -b^{x_i} \ln(b) . \tag{2.2.15}$$

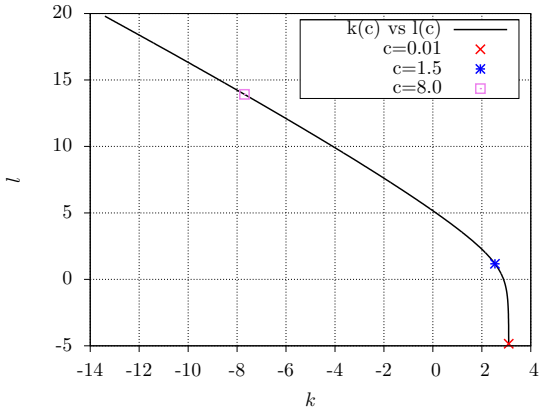

Figure 2.4: A parametric plot of k versus l for the Swiss Re parameterisation as a function of the concavity c.

The log-likelihood function is then

$$l(b, g|\underline{x}) = \sum_{i=1}^{m} \ln \left[f_{X|X<1}(x_i|b, g) \right] - n \ln (g) \,. \tag{2.2.16}$$

In principle, this should be maximised numerically with respect to the parameters b and g to find the maximum likelihood estimators $\hat{b}$ and $\hat{g}$.[2] However, numerically it turns out to be more stable to rewrite the parameters as

$$b = e^k, \qquad g = 1 + e^l \tag{2.2.17}$$

and to maximise the log-likelihood to find the estimates k_F and l_F.[3] This has the effect of enforcing the constraints on b and g, Eq. (2.0.1), but without restricting the numerical search for k and l (constrained optimisation can be delicate to implement, whereas unconstrained optimisation is far more robust). Further, while the parameters b and g may vary over several orders of magnitude, k and l vary more linearly, giving additional stability. Notice that this parameterisation is very similar to the Swiss Re parameterisation of earlier, Eq. (1.1.17). In practise, it is seen that the log-likelihood function in terms of k and l is smooth and has a well-defined global maximum that presents no further numerical difficulties. Figure 2.4 shows a parametric plot of k vs. l for the Swiss Re parameterisation as a function of concavity c. This illustrates the typical values that k and l may take.

To check the veracity of the parameter estimation obtained by maximising the likelihood and the corresponding sampling error associated with Monte Carlo simulation, we perform 100 test Monte Carlo simulations (of the type discussed in Section 2.1) for various sample sizes N and compare the estimates k_F and l_F with the original k and l. A scatter-plot of these resultant estimates is shown in Fig. 2.5 for two different parameter sets: $k = 2.446$, $l = 1.386$ (or $b = 11.547$, $g = 5.0$) and $k = 0$, $l = 2.197$ (or $b = 1$, $g = 10$). It is seen that for increasing sample size N, the parameter estimates do converge to their proper values approximately like $1/\sqrt{N}$, which is to be expected. Moreover, the sampling distributions for the

[2] Actually, $-l(b, g|\underline{x})$ is minimised, using the Python scipy.optimise.minimise routine [scipy2001, python].

[3] The subscript F denotes fitting using the severity curve, as opposed to the next subsection where a fit is made using the exposure curve and denoted by the subscript G.

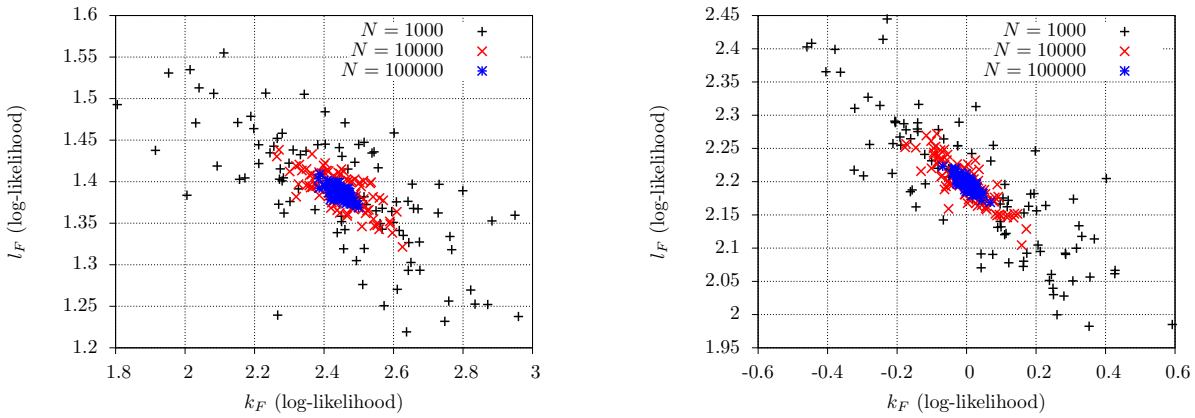

Figure 2.5: Scatter plot of log-likelihood fitted values k_F and l_F for 100 runs of Monte Carlo simulations with $N = 1000, 10000, 100000$. [left panel] Parameter set $b = 11.547$, $g = 5.0$, or $k = 2.446$, $l = 1.386$. [right panel] Parameter set $b = 1$, $g = 10$, or $k_F = 0$, $l_F = 2.197$.

parameter estimates look nicely Gaussian, indicating that everything is in order. We notice that k_F and l_F are clearly negatively correlated for both parameter sets and are thus not independent. We estimate that we can take the error associated with parameter fitting a Monte Carlo simulation with $N = 100000$ sample points to be approximately

$$k_F \approx k \pm 0.1, \qquad l_F \approx l \pm 0.03. \tag{2.2.18}$$

It should be stressed that this is a rough rule of thumb, not an in-depth analysis (if nothing else, the two variables are not independent). In principle, for every simulation performed with a certain parameter set, the (co-) variance of the fitted parameters should be ascertained by performing the Monte Carlo multiple times to build up a sampling distribution, or by using bootstrap techniques.

2.2.2 Least squares exposure

An alternative method for estimating the parameters of the MBBEFD distribution comes from performing a least squares minimisation fit using the exposure curve G. For a sample $\underline{x} = (x_1, x_2, \ldots, x_N)$, this entails minimising the following quantity with respect to b and g:[4]

$$\sum_{i=1}^{N} \left[G\left(u_i | b, g\right) - \frac{\sum_j \min(u_i, x_j)}{\sum_j x_j} \right]^2 \Bigg|_{u_i = x_i} . \tag{2.2.20}$$

[4]There is a numerical subtlety that arises when considering the double sum of this expression for large sample sizes N. For example, when $N = 10^5$, the double sum would involve $\mathcal{O}\left(N^2\right) \sim 10^{10}$ operations which takes an inordinately long time to process. The resolution is to first sort the sample (which for the built-in method employed by Python, 'Timsort' [peters2002, python] involves at worst $\mathcal{O}\left(N \ln(N)\right) \sim 10^6$ operations). The double sum for the sorted sample can then be reduced to a single sum by noting that for index i,

$$\sum_{j=1}^{N} \min(u_i, x_j) \Bigg|_{u_i = x_i} = \sum_{j=1}^{i} x_j + (N - j) x_i. \tag{2.2.19}$$

The sum over j can now be stored cumulatively as one increments the index i, effectively linearising the algorithm.

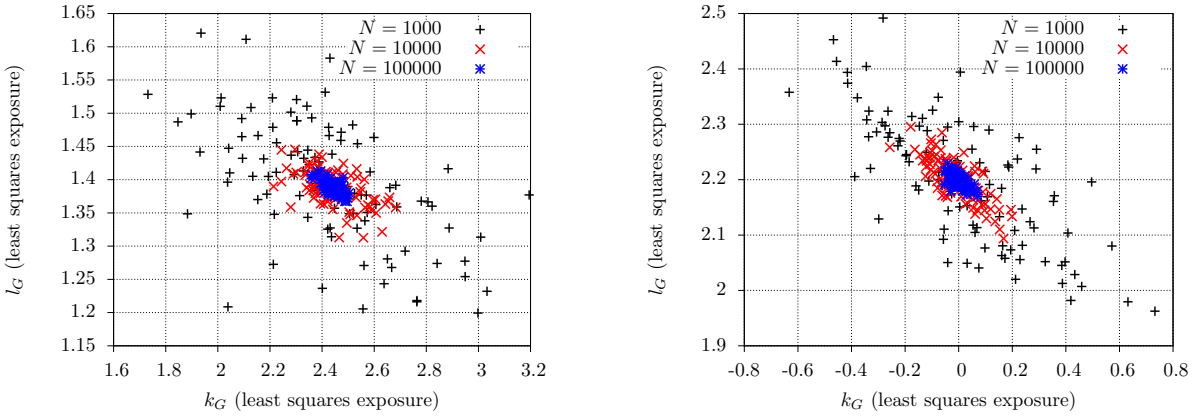

Figure 2.6: Scatter plot of least squares exposure fitted values k_G and l_G for 100 runs of Monte Carlo simulations with $N = 1000, 10000, 100000$. [left panel] Parameter set $b = 11.547$, $g = 5.0$, or $k = 2.446$, $l = 1.386$. [right panel] Parameter set $b = 1$, $g = 10$, or $k_F = 0$, $l_F = 2.197$.

As before, we rewrite the parameters b and g in favour of k and l, Eq. (2.2.17), to ensure numerical stability. At the minimum, we obtain the estimates k_G and l_G.

Just as for the log-likelihood fitting, we can scatter-plot estimates k_G and l_G for multiple Monte Carlo simulations with different sample sizes. This is shown in Fig. 2.6 with the same parameter sets as before. The results are almost identical, showing that the two techniques can complement each other (insofar as in those cases where the sample is known to come from the correct distribution). Again, the parameter estimates converge to their true values as $1/\sqrt{N}$ and there is a clear negative correlation between k_G and l_G.

2.2.3 Quick way to find g (and c)

Given a large sample $\underline{x} = (x_1, x_2, \ldots, x_N)$ where the number of $x_i = 1$, n, is also large, the parameter $g = 1/P(X = 1)$ may be estimated as

$$\hat{g} = \frac{N}{n}. \tag{2.2.21}$$

This intuitive result can be obtained by considering the maximum likelihood for the two cases $X < 1$ and $X = 1$:

$$L(g|\underline{x}) = P(X < 1)^{N-n} P(X = 1)^n. \tag{2.2.22}$$

The corresponding log-likelihood is then

$$l(g) = (N - n) \ln(1 - 1/g) - n \ln(g) \tag{2.2.23}$$

such that at the maximum, $\hat{g} = N/n$. In the case of the Swiss Re parameterisation, the estimated concavity, $\hat{c}$, can then be directly inferred:

$$\frac{N}{n} = \hat{g} = e^{\hat{c}(\gamma + \delta\hat{c})} \qquad \Rightarrow \qquad \hat{c} = \frac{\sqrt{\gamma^2 + 4\delta \ln(N/n)} - \gamma}{2\delta}. \tag{2.2.24}$$

While this result is simple, there is no guarantee that both n and N for the samples that we will be considering will be sufficiently large to warrant such an approach.

2.3 Connection between the MBBEFD and other distributions

Recall that the two-parameter Pareto distribution with $\alpha = 1$, Pareto$(\alpha = 1, \lambda)$, for a random variable Y has the probability density function

$$f_Y(y) = \frac{\lambda}{(\lambda + y)^2}, \quad \lambda > 0, \quad y > 0, \tag{2.3.25}$$

with the corresponding cumulative density function

$$F_Y(y) = \frac{y}{\lambda + y}. \tag{2.3.26}$$

Setting

$$\lambda = \frac{1}{g - 1} \tag{2.3.27}$$

(noting that since $g \geq 1$, $\lambda > 0$ holds true) and manipulating slightly, this becomes

$$Y \sim \text{Pareto}\left(1, 1/(g-1)\right); \qquad f_Y(y) = \frac{(g-1)}{\left[1 + (g-1)y\right]^2}, \quad F_Y(y) = \frac{(g-1)y}{\left[1 + (g-1)y\right]}. \tag{2.3.28}$$

These correspond precisely to the MBBEFD distribution for the case $b = 1$ and where $y < 1$, Eqs. (2.2.14,2.0.4). This means that if we construct a new variable $Z = \min(Y, 1)$, then

$$Z = \min(Y, 1), \qquad F_Z(z) = \begin{cases} \frac{(g-1)z}{[1+(g-1)z]}, & z < 1 \\ 1, & z \geq 1 \end{cases}, \tag{2.3.29}$$

such that $Z \sim \text{MBBEFD}(b = 1, g)$. With the Swiss Re parameterisation, this corresponds to a concavity $c = 4.0735$.

Likewise, recall that the exponential distribution with parameter λ has the cumulative distribution function

$$F_Y(y) = 1 - e^{-\lambda y}, \quad \lambda > 0. \tag{2.3.30}$$

If we set

$$\lambda = -\ln(b) \tag{2.3.31}$$

where $0 < b < 1$, then we have

$$Y \sim \text{Expon}\left(-\ln(b)\right); \qquad F_Y(y) = -\ln(b)\, b^y, \quad F_Y(y) = 1 - b^y. \tag{2.3.32}$$

This now corresponds to the MBBEFD distribution for the case $bg = 1$ (where $b < 1$) and where $y < 1$, Eqs. (2.2.15,2.0.5). Again, we can construct the variable Z such that

$$Z = \min(Y, 1), \qquad F_Z(z) = \begin{cases} 1 - b^z, & z < 1 \\ 1, & z \geq 1 \end{cases}, \tag{2.3.33}$$

and $Z \sim \text{MBBEFD}(b, g = 1/b)$. With the Swiss Re parameterisation, the concavity $c = 25.1145$. This value is admittedly not in the range that one might expect, but makes sense if one considers that the exponential distribution is strongly peaked near the origin, i.e., small times – the large concavity indicates that small (fractional) losses are far more likely.

Treating Z as the minimum of a Pareto or exponential distributed random variable and unity leads to a particular physical interpretation of the MBBEFD distribution, discussed in Section 3.1.

Chapter 3

Fire simulation model

3.1 A physical illustration

In Section 2.3, a connection was made between the MBBEFD distribution and the Pareto distribution, such that for $b = 1$, the random variable X drawn from the MBBEFD distribution was equivalent to $\min(Y, 1)$ where Y is a random variable drawn from the Pareto distribution with parameters $\alpha = 1, \lambda = 1/(g-1)$. Likewise, one could construct an MBBEFD distribution with parameters constrained such that $bg = 1$ from an exponential distribution with parameter $\lambda = -\ln(b)$. This leads to a particularly simple physical picture for the MBBEFD distribution.

Consider filling a bucket with water from a tap. At the beginning, the bucket is empty and the tap is turned on. The water flows with a constant rate of, for example, one bucket per minute (i.e., the rate is in terms of the volume of water per unit time). Now imagine that the tap is turned off after a random time T, where T is Pareto distributed (with $\alpha = 1$) and the time is measured in minutes. The total amount of water that has flowed is $T/(1\text{ min})$–bucketfuls. For T less than one minute, the bucket will be partially full. However, for T longer than one minute, the bucket simply overflows and the volume of water in the bucket remains constant at one bucketful. If we do this many times, the amount of water in the bucket on each attempt will be distributed according to the MBBEFD distribution with parameter $b = 1$.

The same physical idea may be applied to the damage caused by a fire. One direct application of the above example would be to a candle that is lit for a random time T before being extinguished – since the candle burns uniformly, the 'damage' would correspond to the length of the candle that has been burnt and if the time is long enough, the entire candle will burn. One might wonder at how an infamously heavy-tailed distribution such as the Pareto distribution with $\alpha = 1$ can correspond to a real-life situation.[1] However, when discussing the time taken before a fire is put out, consider the following: the longer a fire burns, the larger it can get and consequently the more difficult it can be to bring under control. Also, there are situations where it is simply not possible to put a fire out. Two examples come to

[1] Recall that when $Y \sim \text{Pareto}(\alpha, \lambda)$, the expectation value $E[Y] = \lambda/(\alpha - 1)$ such that when $\alpha = 1$, the expectation value is ill-defined.

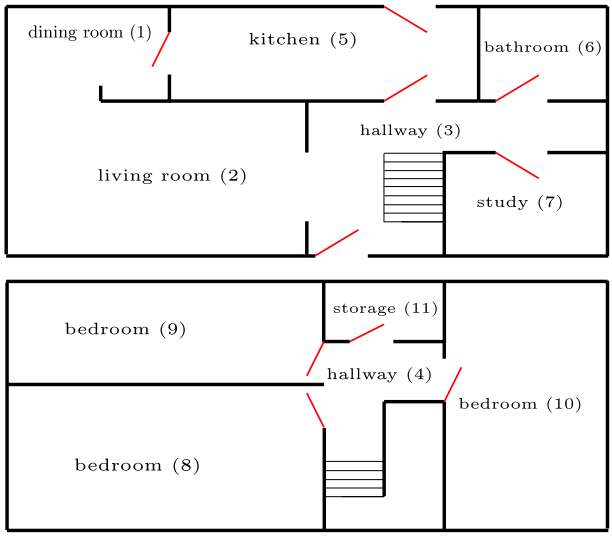

Figure 3.1: A (crude rendition of a) house. The upper diagram is the ground floor and the lower diagram is the first floor. Doors are shown in red. The loft (room '12') is not shown.

mind:[2] the first is an out of control forest fire (frequently in the news) – in this case the firefighters often circumvent their inability to bring the fire under control by creating breaks in the forest such that the spread of the fire is hindered (rather like putting a hole in the side of the bucket in our earlier example); the second is coal mine fires, such as that in Jharia, India which has been reputedly burning for over 100 years [jharia2017].

The case of fire damage to property is complicated by the following factors:

- The fire will not burn uniformly through a property. It is hindered by the presence of closed doors; conversely, it could be aided by the plentiful presence of combustible material.
- The distribution of the total time that the fire burns is not known.

In this chapter, we shall construct a model for the evolution of a fire through a building and introduce an ansatz[3] for the distribution of the time during the fire damages the building.

3.2 A fictitious scenario

Let us start by considering the example residential property shown in Fig. 3.1. There are twelve distinct rooms (we will discuss the loft shortly) and (not counting the front and back doors) eight doors. We are interested in the severity curve for fire risk. I.e., if a fire starts, what is the probability distribution of the claim size, as a fraction of the property value (sum insured)? There are clearly a huge number of variables to the problem, most of which are irrelevant to the modelling process. Let us imagine one particular scenario:

[2] Actually, a third example – the sun (burning for billions of years) might be excluded because technically it is a nuclear fusion reaction. Moreover, there have been no attempts to put it out.

[3] Ansatz is the fancy term for an educated guess, the word coming from German.

> *"After a romantic dinner, an unattended candle on the dining room table falls over, setting the tablecloth alight. In the blink of an eye, the whole table is on fire. Sensing that something is wrong, the occupants investigate and realise to their horror that without a fire extinguisher, the blaze cannot be tackled – they run outside and call the fire brigade. By this time, the fire has spread to the living room but not to the kitchen since both kitchen doors had been kept shut to keep the cat away from the leftovers of the dinner. It does not take long for the fire to reach the lower hallway and move upstairs to the upper hallway. The bathroom door and two of the bedroom doors were open, so the fire spreads. The study door, one bedroom door, and the storage room door were closed, so the fire does not initially spread to these rooms. Since the couple live in the countryside, the fire brigade take half an hour to arrive. By this time, the fire has found its way past one of the kitchen doors and also into the remaining bedroom. The fire brigade then set about their business and put the fire out. Three out of the twelve rooms (the study, the storage room, and the loft) have been left relatively unscathed because the rooms were infrequently used and the doors shut snugly. Fortunately, the cat was outside throughout and was unharmed."*

To be sure, this fictional story contains many irrelevant details. However, it can be broken down and the various elements discussed from the insurance modeller's perspective and the set of model assumptions that will be used.

1. **The owners and their cat** – behind each fire lies a human story and sadly, all too often this story involves fatalities. Here, the model concerns only the property loss and does not consider possible payouts for injury or mortality. Also, in the case of non-residential property, it does not consider loss of business due to the fire.

2. **The cause of fire** – there are a huge number of possible causes of fire and different rooms face different risks. Rather than trying to model this, the simulation of a fire will begin by selecting (with equal probability) a room at random to be the starting point. In comment 5 below, it will be seen how this is consistent.

3. **Fire extinguishers** – in the initial stages of a fire, there is the possibility that the fire can be put out by either the occupants or an installed sprinkler system. We shall assume that this possibility is implicitly included in the time taken to put the fire out (see later).

4. **Fire spreading within a room** – once a fire takes hold in a room, without fire suppression measures (such as a fire extinguisher or sprinkler systems in larger buildings) and assuming that there is a 'normal' amount of combustible material (the room is not stripped bare; equally, it is not filled with explosives) it will engulf the room in a matter of a few minutes. Physically, the exact time will depend on the airflow, the size of the room, and the nature of the contents. For the model, we shall assume that these details are unknowable and even if the insurer had complete data for a portfolio of properties, it would not be feasible to perform a Monte Carlo simulation with such detail. We therefore introduce an arbitrary, but uniform time unit (a few minutes) within which a

fire will completely engulf a room, destroy its contents, and be ready to spread to any neighbouring room(s) that are not already on fire.

5. **The rooms and the value of their contents** – clearly, different rooms have different sizes and contain different objects but a fire will destroy everything without prejudice. Just as above (for modelling how quickly a room will be completely engulfed in flames), keeping track of the location of all items and their respective values is not feasible for the insurer or the model. Once the requirement of knowing 'what is where' is abandoned, it is no longer necessary to keep track of the room designations and they will simply be delineated by a number (and the ordering of these numbers is irrelevant). What is important is to which other rooms this room is connected (this is where the graph theory will come into play). The choice of a random starting point for the fire in comment 2 above also becomes self-evident. There are two possibilities for modelling the value of a room destroyed by fire: *i*) each room is assigned an equal fraction of the property's value, or *ii*) each room is assigned a random value. We shall assume the latter, such that the fractional damage done by the fire is a continuously distributed. Notice that we have not mentioned damage from smoke which will also affect the loss incurred – this model will not include this consideration (which is admittedly a shortcoming).

6. **Fire propagation through open doors** – in the fictitious scenario, some doors were open (allowing the fire to freely propagate) and others were closed. It is a matter of taste whether a household keeps most doors open (e.g., to let air circulate and to let the light in), or not (e.g., to keep particular rooms warm in winter). It is also dependent on the time of day and on whether or not someone is at home. Further, this changes according to the usage of the building: doors may be more frequently open in a residential property, but office doors may be closed. To put it bluntly, this is not something that we know. In the model, we shall treat the possibility of a door being open as a Bernoulli distributed random variable and with the same probability for each door. This probability is an external parameter of the model, to be estimated (and may vary between different classes of buildings).

7. **Fire propagation through closed doors** – once a fire reaches a closed door, it is only a matter of time before it burns through. Here, we assume that all doors have been properly constructed. In general, the current building regulations in England [build2010] stipulate that an internal house door should be able to resist a fire for at least 30 minutes (in order to allow time for evacuation), although this depends on the location and usage of the door. We approximate this to 10 of the arbitrary time steps in our model, after which the fire passes through to the adjacent room.

8. **Fire propagation through walls and ceilings/floors** – this was not mentioned in our fictitious story, but just as for a closed door, the walls and ceilings within a building will eventually succumb in the event of a fire. Again, this is a rather complex issue due to the different types and usages. The regulations [build2010] are correspondingly detailed. For the purposes of modelling, we shall assume that all internal walls, ceilings, and floors can withstand a fire for 30 minutes, or 10 time

steps. Notice that in the example house, Fig. 3.1, the loft was not shown! In principle, fire could spread from one bedroom to another via this loft (this will be included in the model simulation).

9. **Fire propagation via external routes** – in our model, we assume that fire can only spread internally within a building. However, it is possible that flames emanating from windows may cause the fire to spread to the next storey of a building. Further, external cladding (sometimes used for insulation and aesthetic appearance) may ignite and propagate a fire externally.

10. **Total time taken** – in our fictitious scenario, the fire brigade took 30 minutes to arrive (and no mention was made on how long it took to put the fire out). This is unrealistic – for example, the average response time (the time taken between the report being received and the first fire engine being ready to combat the fire) in England for the year 2015/16 was 8 minutes and 47 seconds [response2017]. Clearly, the actual time during which a fire may damage property depends on multiple factors: how long it took before the fire was noticed, the response time for the fire brigade, and how long it takes for the fire brigade to get the fire under control safely. Initially, we shall run each simulation for a definite number of time steps to see how the fire progresses, but we will return to this issue in Section 3.4.

3.3 Fire propagation

In the previous subsection, the example house, Fig. 3.1, had (if we include the loft) twelve distinct rooms. These rooms were connected by either an open space (e.g., between the dining room and living room), a door (e.g., the between the dining room and kitchen), or were separated by a wall (e.g., the bedrooms denoted 8 and 9) or floor/ceiling (e.g., bedroom no. 8 and the living room). The loft is the only room that does not have a door – it has an access, but this is invariably closed, which is why we treat it separately. Fire could propagate via any of the routes between rooms, the only distinction being how long it would take. We thus introduce the following rules for fire propagation in our model:

1. at the beginning of the simulation, each door is open with a probability p_D,

2. the fire is started in a room selected at random,

3. fire propagates from one room to the next room in a single time step if there is an open space or open door connecting them,

4. fire propagates from one room to the next room in ten time steps if there is a closed door, a wall, or a floor/ceiling between them.

Notice that we treat the fire as spreading isotropically and deterministically. The isotropy is a model assumption (i.e., we assume that the fire will try to spread in all directions equally). In principle, one could imagine a scenario where the fire moves in a particular direction due to airflow, but this is beyond the scope here. With the assumptions of isotropy and that the fire propagates with definite speed from one room to another given the nature of the route between them, it follows that the spread of the fire

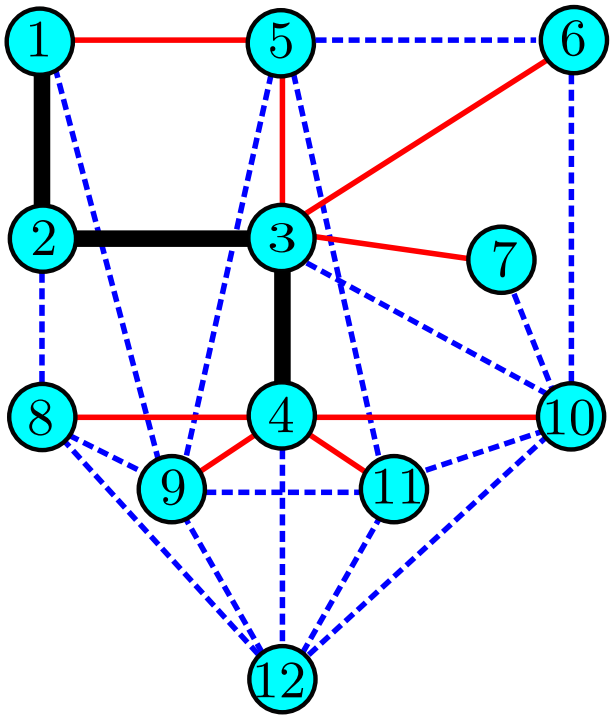

Figure 3.2: The graph for the house, Fig. 3.1. Each room (including the loft) is represented by a numbered vertex. For the edges between vertices: thick black lines represent open spaces between rooms, thin red lines represent doors that are open with probability p_d, dashed blue lines represent all other possible routes that a fire may propagate.

is deterministic – once the fire is started in a given building and the doors have been set to be open or closed, the progress of the fire is known, although complex.

Since the designation of rooms is treated as irrelevant, the example house, Fig. 3.1, can be represented by a graph whose vertices are the rooms and whose edges represent the possible paths of fire propagation. This is shown in Fig. 3.2. The thick black lines represent certain propagation through open spaces, the thin red lines denote doors that may or may not be open, the dashed blue lines represent fire propagation routes through walls or ceilings/floors. Two points become immediately apparent. By including fire propagation through the walls and ceilings/floors, we can see that the house is actually very well-connected: the order of the graph is 12 and the size is 27, giving a density of 0.409. Moreover, if we imagine walking between rooms, we can go between any two rooms by passing through only one or two doors. At first sight, this might seem surprising but intuitively, this is a characteristic of most houses – they are designed such that one can freely move around and when going from one room to another, it would not make sense to have to pass through all the other rooms or lots of doors![4]

In Fig. 3.3, we show plots of the fractional loss as a function of the time step in various simulations for our example house, with different door probabilities p_D. Here, each room is designated with equal value. In the plots, each solid grey lines correspond to the progress of the fire for one simulation (there are

[4]It is also worth pointing out that one difference between residential buildings and public buildings or office spaces is the presence of fire doors that are kept closed at all times.

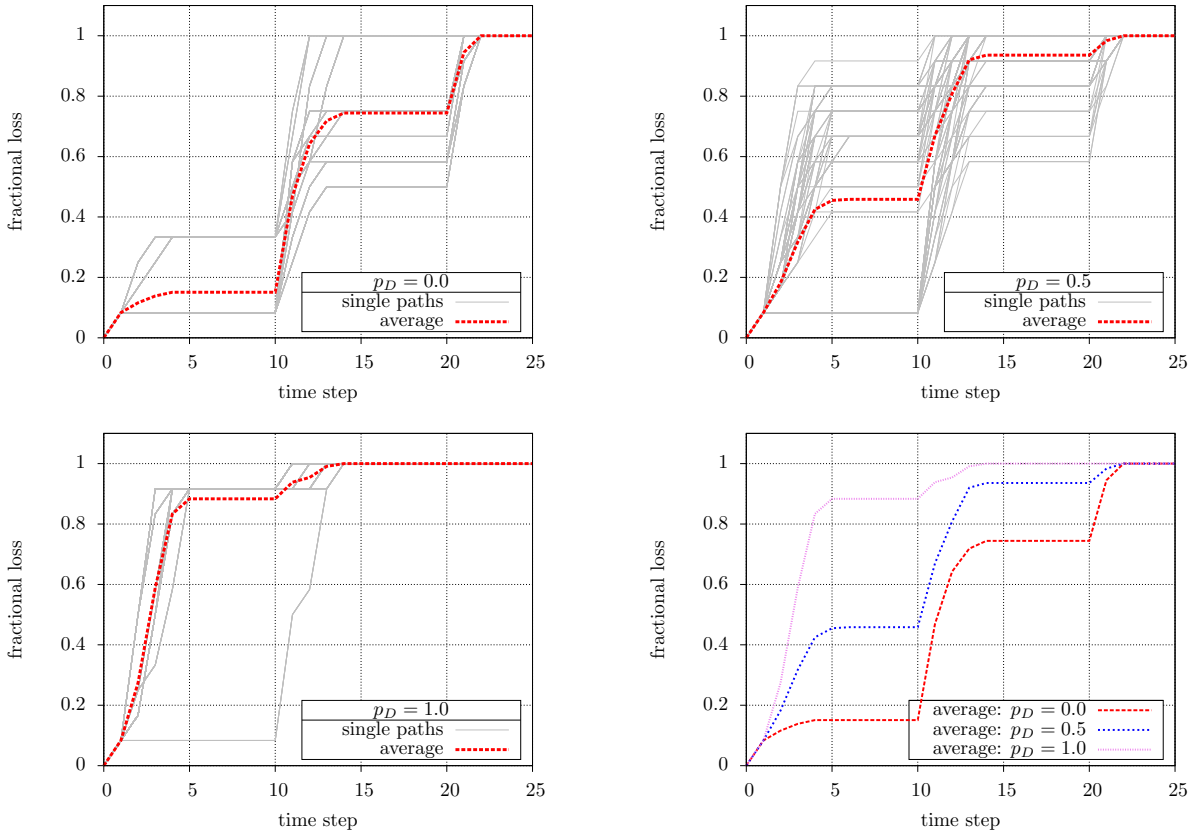

Figure 3.3: Plot of the fractional loss due to a fire in the example house of Fig. 3.1 as a function of time step, for different door probabilities (each with 100 simulations). [top left panel] $p_D = 0.0$ (all doors closed). [top right panel] $p_D = 0.5$ (equal probability for each door to be open/closed). [lower left panel] $p_D = 1.0$ (all doors open). [lower right panel] Compendium of the average fractional loss for the different probabilities.

100 simulations for each plot) and the dotted red lines to the average. This illustrates the deterministic evolution of the damage with time, given that the starting point for the fire and the configuration of the doors is random. We notice that there are characteristic periods of time during which the fire is hemmed in – at these points, the linear spread of the fire is halted until whatever barrier succumbs. What is abundantly evident is that unlike the earlier bucket of water being filled or the candle burning, the more realistic physical simulation here is not uniform.

The lack of uniformity in the propagation of a fire leads to a somewhat subtle point. The MBBEFD curves refer to the insurer's portfolio of property – each individual property burns once. In this study, a small portfolio of residential properties will be 'burnt' many times in our Monte Carlo simulation (for practical reasons – a large portfolio takes time to construct). Because of the possibility of fire propagation pausing within a certain building, this will lead to deviations from the MBBEFD curve. However, even with a small, representative, portfolio several results will emerge, as will be discussed. The portfolio is constructed so as roughly to mimic the distribution of residential properties nationwide and is discussed in the Appendix.

3.4 Time distribution

At the beginning of the chapter, a physical interpretation for the MBBEFD distribution was discussed – fire propagates for a randomly distributed time T. In the context here, this time includes the time taken before the fire is discovered (*if* it is discovered), for the fire brigade to arrive and to put the fire out safely (*if* possible) or, if a sprinkler system is installed, whether or not this is effective. Clearly, there are many factors such that making a reasonable assumption for each possibility would become prohibitively complex. We thus make an ansatz for the distribution of the total time taken. We assume that this ansatz is the same within each class of property (residential, industrial, etc.) – this makes sense, at least for residential property, since the chance of the fire being detected, the response time of the fire brigade, and the possibility of putting the fire out safely will be roughly independent of the property itself.[5] We will initially focus on the case of residential property, only afterwards turning to the more general case.

Our ansatz time distribution is a generalisation of the MBBEFD distribution and the cumulative distribution function is

$$F_T(t) = \begin{cases} \frac{\tilde{b}(\tilde{g}-1)\left(1-\tilde{b}^t\right)}{\tilde{b}(\tilde{g}-1)+\left(1-\tilde{b}\tilde{g}\right)\tilde{b}^t}, & 0 < t < t_m \\ 1, & t \geq t_m \end{cases} . \tag{3.4.1}$$

In the above, $\tilde{b}$ and $\tilde{g}$ are the parameters to be fitted to the portfolio; t_m is interpreted as a maximum time (by which all properties will have burnt completely with certainty). The distribution may be numerically simulated in a manner analogous to the simulation of the MBBEFD distribution described in Section 2.1. Taking a sequence of uniform deviates, $\underline{u} = (u_1, \ldots, u_N)$, the sample $\underline{t} = (t_1, \ldots, t_N)$ is obtained by the transformation

$$t_i = \begin{cases} \frac{\ln\left[\frac{\tilde{b}(\tilde{g}-1)(1-u_i)}{(1-\tilde{b}\tilde{g})u_i+\tilde{b}(\tilde{g}-1)}\right]}{\ln\left(\tilde{b}\right)}, & 0 \leq u_i < u_m \\ 1, & u \geq u_m \end{cases} , \tag{3.4.2}$$

where

$$u_m = \frac{\tilde{b}\,(\tilde{g}-1)\left(1-\tilde{b}^{t_m}\right)}{\tilde{b}\,(\tilde{g}-1)+\left(1-\tilde{b}\tilde{g}\right)\tilde{b}^{t_m}}. \tag{3.4.3}$$

The time is measured in our time step units – by fitting to data, we obviate the need to specify what this time unit is! In what follows, we set $t_m = 50$. The finite probability $P(t = t_m)$ reflects that in some cases, the fire might not be put out at all. It may seem strange that we use the MBBEFD distribution to obtain an MBBEFD distribution but we shall see that the values of $\tilde{b}$ and $\tilde{g}$ are quite different from b and g. The crucial point of the model is that for fixed time distribution (i.e., for a certain class of properties), the different fire propagation characteristics of the different properties relative to the aggregate portfolio will result in different severity and exposure curves. Moreover whereas for the bucket example earlier where the bucket was filled at a unit rate, here the aggregate rate of fire propagation is different from unity (the scale changes).

[5] One might wonder if a fire in a large house is more difficult to put out (and would thus take longer) than for a small house. This is misleading – while the difficulty of putting a fire out is presumably proportional to the size of the fire, this in turn is proportional to the time for which the fire has burnt (and the fire will spread through similar types of building at similar rates). Thus, the information about the relative difficulty of putting the fire out is encoded implicitly in the time distribution already.

Having introduced a continuously distributed random variable for the time, we must specify what happens to the fire propagation between unit time steps. We assume that during each time step, the damage done by a fire that has spread to a room at the start of that time step increases linearly such that the room is completely damaged by the end of that time step (the fire then does no further damage to that room, except to start to burn through any closed doors/walls in order to spread to any further connected rooms). To enforce the continuous nature of the fractional loss x, each room value is assigned a uniformly distributed random variable $U \sim \text{Unif}(0, 1)$. The loss fraction x thus consists of three elements: the room value, whether the room is partially damaged, and which rooms are damaged – all relative to the total value of the rooms combined.

The parameters of the time distribution ansatz, $\tilde{b}$ and $\tilde{g}$, will be fitted using our portfolio of properties such that the original concavity of the Swiss Re parameterisation is reproduced roughly. Individual properties can be analysed then, *relative* to this portfolio. In principle, a rigorous procedure for estimating the parameters of the ansatz for the time distribution could be devised. One way to do this requires using sufficiently large Monte Carlo sample sizes such that the Monte Carlo simulation gives parameter estimates for k and l whose sampling distribution is sufficiently narrow that the underlying parameters of the time distribution ($\tilde{b}$ and $\tilde{g}$) may be fitted. Another way to do this would involve studying the sampling distributions for k and l using multiple Monte Carlo simulations, the mean of this distribution can then be used in the fitting procedure for the time distribution. Unfortunately, neither of these methods are possible here due to limited computer resources (all simulations are performed on a standard laptop and such large scale Monte Carlo studies would take a prohibitive time to run). Instead, we must be content with a trial and error approach such that we get a rough estimate for the parameters.

3.5 Results for the residential property portfolio

Let us assume that for residential property, the probability for a door being open (or more properly, not being tightly shut such that fire cannot easily pass through) is $p_D = 0.8$.[6] Further, let us consider the parameters for the time distribution, Eq. (3.4.1), to be $\tilde{b} = 1.3$, $\tilde{g} = 2.0$, $t_m = 50$ – as discussed in the previous section, these values are chosen so as to roughly reproduce the concavity of the Swiss Re parameterisation (and are found by trial and error) for our portfolio of residential property, namely $c \approx 1.5$ ($b \approx 12.6$, $g \approx 4.2$, or $k \approx 2.5$, $l \approx 1.2$). With a given Monte Carlo sample size N, each individual simulation (as discussed in the previous sections) involves a property being selected from the portfolio of residential properties (see the Appendix), the starting point for the fire being selected at random, the random time during which the fire damages the property being chosen from the distribution, Eq. (3.4.1), and the fire propagating according to definite rules. Each individual simulation gives a fractional loss value x_i such that the overall Monte Carlo simulation generates an empirical sample $\underline{x} = (x_1, \ldots, x_N)$. Under the hypothesis that the distribution is of MBBEFD-type with true parameters b and g, or their surrogates, k and l, related via Eq. (2.2.17), the parameter estimates for the empirical distribution are found using two methods – log-likelihood (giving k_F and l_F) and least squares exposure (giving k_G and

[6] This estimate is based on personal experience, but is emphatically no more than an estimate.

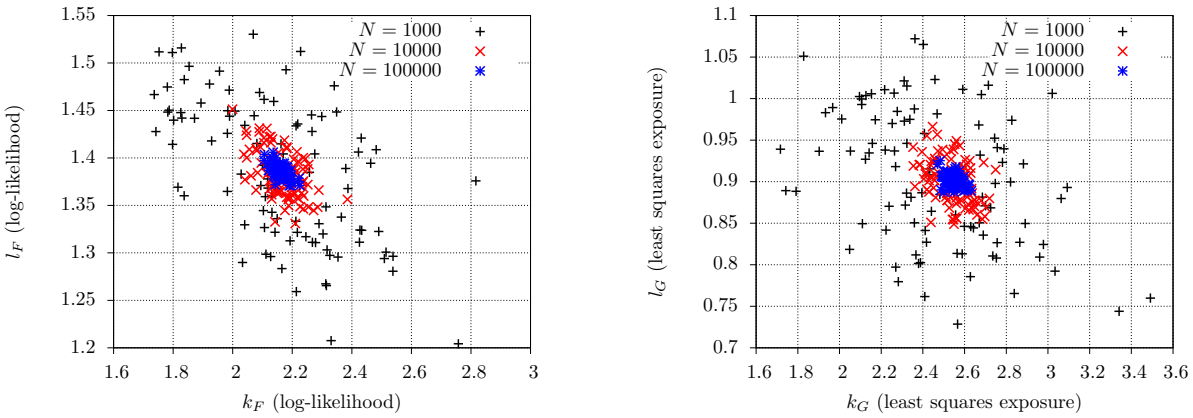

Figure 3.4: [left panel] Scatter plot of log-likelihood fitted parameters k_F versus l_F for various sample sizes. [right panel] Scatter plot of least squares exposure fitted parameters k_G versus l_G for various sample sizes. Parameters are: $\tilde{b} = 1.3$, $\tilde{g} = 2.0$, $t_m = 50$, $p_D = 0.8$.

l_G).

Our first order of business is to check the convergence of the Monte Carlo simulation with respect to the sample size N. This is done in the same way as for Section 2.2 by taking Monte Carlo sample sizes $N \in \{1000, 10000, 100000\}$ and repeating each simulation 100 times to form a scatter plot of both k_F versus l_F and k_G versus l_G. The resultant scatter plots are shown in Fig. 3.4. It is seen that just as in Section 2.2, the distribution is roughly Gaussian, with k and l being negatively correlated. Moreover, for both sets, when $N = 100000$, the scatter plot values are within approximately ± 0.1 for either k_F or k_G and ± 0.03 for either l_F or l_G. These values are commensurate with those of Section 2.2 and we consider the Monte Carlo simulation with $N = 100000$ to have converged sufficiently, at least for our purposes here. Now for the crunch – taking the central values of the parameter estimates from the scatter plot Fig. 3.4, we get

$$k_F \approx 2.15, \quad l_F \approx 1.38, \quad k_G \approx 2.55, \quad l_G \approx 0.9. \tag{3.5.4}$$

Even allowing for the finite convergence tolerance inherent to the Monte Carlo simulation, the two sets of parameter estimates do not agree with each other (although both are in the vicinity of the Swiss Re parameterisation values).

The lack of agreement between the two sets of parameter estimates can be seen in a more direct fashion. The empirical severity and exposure curves from a Monte Carlo simulation with sample size $N = 100000$ are shown as the solid black lines of the upper panels of Fig. 3.5. The resultant fitted severity and exposure curves are plotted alongside. The log-likelihood fitted curves are the dotted red curves; the least squares exposure fitted curves are dot-dashed blue. The first observation from Fig. 3.5 is that the empirical severity curve (upper left panel) does not exactly correspond to an MBBEFD distribution, although the shape of the empirical severity curve does contain all the right features so as to be not entirely unrealistic. Neither the log-likelihood nor the least squares exposure fitted curves give an exact match (and we can now see why the two different fitting methods do not agree with each

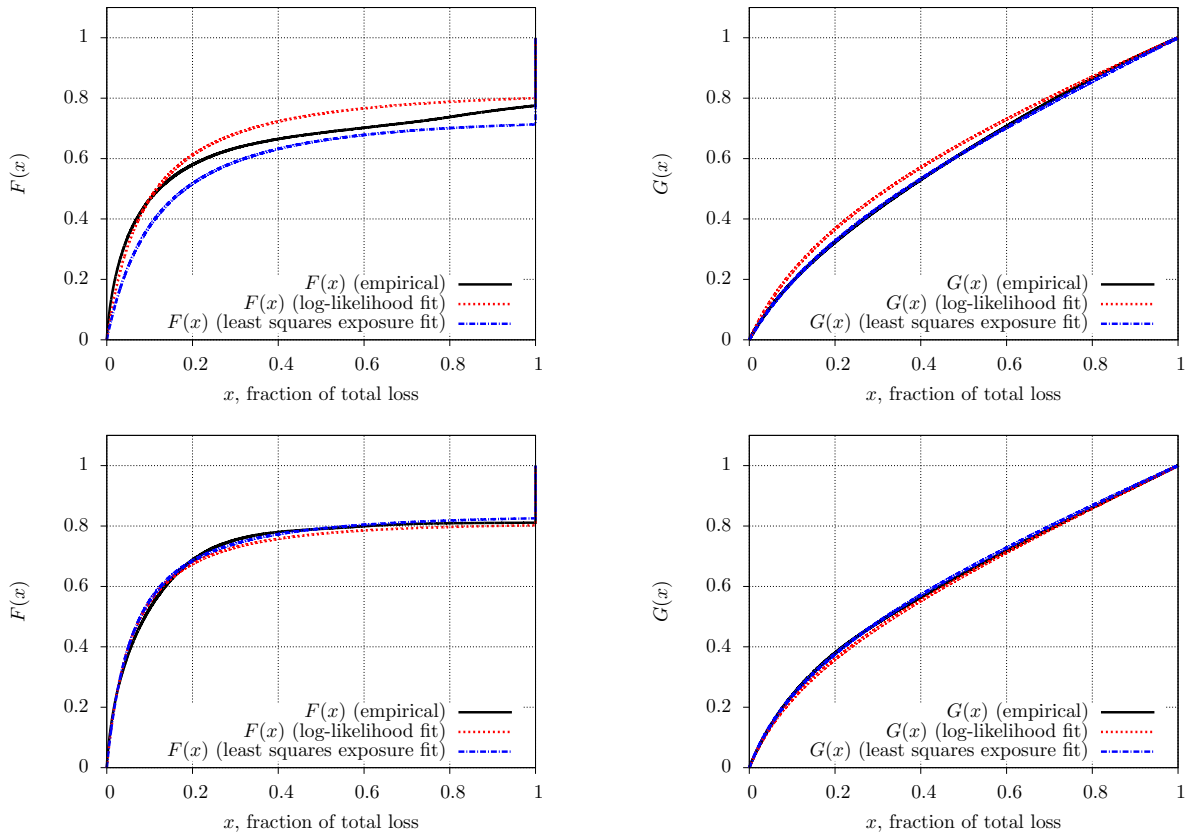

Figure 3.5: [upper left panel] Comparison of the model Monte Carlo severity curve for the portfolio of properties and the corresponding fitted curves with $p_D = 0.8$. [upper right panel] Comparison of the model Monte Carlo exposure curve for the portfolio of properties and the corresponding fitted curves with $p_D = 0.8$. [lower panels] As for the upper panels but with $p_D = 0$. Parameters are: $\tilde{b} = 1.3$, $\tilde{g} = 2.0$, $t_m = 50$, $N = 100000$.

other). If we were to construct the Kolmogorov-Smirnov test statistic, $\sup|F(x) - \tilde{F}(x)|$, where $\tilde{F}$ is the empirical distribution, then from the plot we would see that the result is of the order of 0.1 for both fitting procedures. This can be compared to the test Monte Carlo results shown in Fig. 2.3 – there (where the empirical distribution is simulated directly from the known distribution) such a Kolmogorov-Smirnov test statistic would be associated with a sample size of $N \sim 10 - 100$ not $N = 100000$ as is the case here. If we consider the corresponding empirical exposure curve of Fig. 3.5 (upper right panel), then at first glance things look much better. However, as discussed at the beginning of Chapter 2 with reference to Fig. 2.1, the comparative insensitivity of the exposure curve relative to that of the severity curve masks the true picture. Where the empirical distribution does not exactly match the MBBEFD distribution, the log-likelihood fitting method using the severity curve is more responsive and is the preferred result. Certainly, the least squares exposure fitted severity curve shown in the upper left panel of Fig. 3.5 is consistently below the empirical curve, unlike the log-likelihood fitted curve where approximately half the sample lies above and half below.

The analogous empirical severity and exposure curves for the Monte Carlo simulation of the portfolio

but with all doors firmly shut, $p_D = 0$ are shown in the lower panels of Fig. 3.5. Interestingly, the fitted curves more closely match the empirical curves (with both log-likelihood and least squares exposure methods). The estimated Kolmogorov-Smirnov test statistic in this case is of the order of 0.01 – which from Fig. 2.3 would correspond to a sample size of $N \sim 1000 - 10000$ if the empirical distribution were to be the 'true' one. This result seems far more convincing than for the previous case when $p_D = 0.8$.

The explanation for these results comes from considering two factors: the small size of the portfolio and the uniformity (or lack thereof) of the individual properties. Glancing at the graphs of the properties within the portfolio (see the Appendix), it becomes apparent that when all doors are shut, each graph becomes more or less regular (with a few exceptions for open spaces between rooms). What this means in practise is that the fire propagates uniformly and as discussed in Section 3.1, we know that in this situation, the MBBEFD distribution will be reproduced naturally. This means that when $p_D = 0$, each property will render more faithfully an MBBEFD type curve than when $p_D = 0.8$. Where $p_D = 0.8$, the Kolmogorov-Smirnov test statistic more closely resembles that of a small sample, even though the Monte Carlo sample size (N) is very large. However, the portfolio has only nine properties – each of which is burnt many times. This is not what the MBBEFD distribution represents in real life – in reality, we have a large portfolio where only a small number are burnt exactly once. Since each property is burning in a non-uniform fashion, the deviations from the MBBEFD distribution become more pronounced. There are two possible directions for further study in this regard:[7] *i*) the portfolio of properties should be dramatically expanded – by including more properties of each size, the non-uniformities may cancel out; *ii*) an alternative may be to simplify the model and only consider graphs where the edges are identical (i.e., to impose a uniform time for which the fire propagates between rooms).

Having discussed some of the shortcomings of the model, let us now turn to a more positive result. The central idea behind the model is that the severity and exposure curves for individual properties (or groups of properties) may be analysed relative to the larger portfolio. Knowing the exposure curve for an individual property, rather than simply the concavity for a large class of properties would allow for a more detailed breakdown of the risk and thus the premiums required to cover that risk. Figure 3.6 shows the results for an analysis of the empirical severity and exposure curves for the individual residential properties relative to their portfolio. As before, we use the parameters $\tilde{b} = 1.3$, $\tilde{g} = 2.0$, $t_m = 50$, $N = 100000$, with $p_D = 0.8$, these parameters being 'fitted' (by trial and error) to approximate a concavity of $c \approx 1.5$ for the portfolio using the Swiss Re parameterisation. The empirical exposure curves (lower left panel of Fig. 3.6) are not distinguishable by eye – as before, the lack of sensitivity in the exposure curve hampers such a distinction. However, the empirical severity curves (upper left panel of Fig. 3.6) show a definite result – the individual properties of the portfolio are clearly distinguishable. The smallest property ('property 1') corresponds to the lowest of the curves with a relatively larger number of total losses, whereas the largest property ('property 9') corresponds to one of the upper curves (because each property is non-uniform, the curves are not exactly of the MBBEFD distribution) with more partial losses. The parameter estimates k_F, l_F, and k_G, l_G are shown as scatter plots in the upper and lower right panels of Fig. 3.6, respectively (and tabulated in Table 3.1). It is seen that the individual results are sufficiently widely spaced so as to be

[7] Not possible here, due to the exploratory and time-constrained nature of this study.

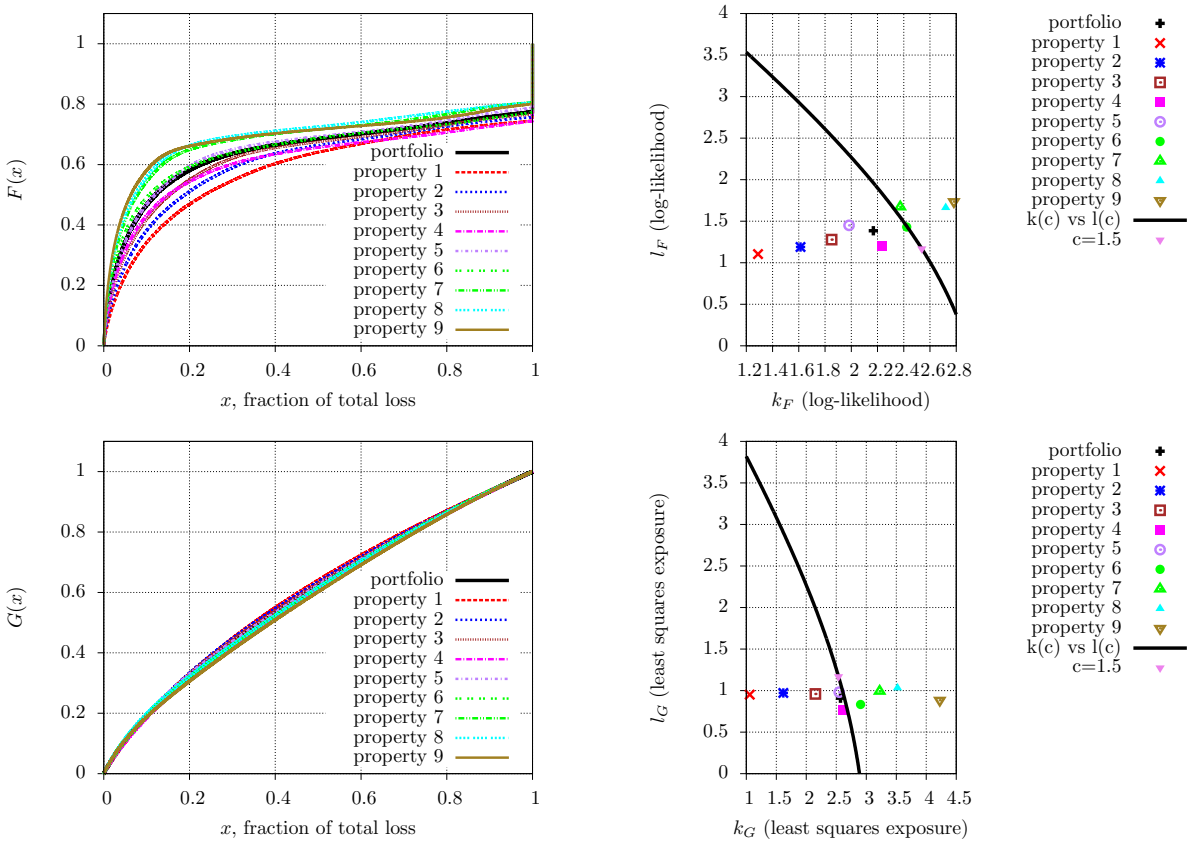

Figure 3.6: [upper left panel] Empirical severity curves for the individual properties (plus the portfolio). [upper right panel] Scatter plot of the corresponding log-likelihood fitted parameters k_F versus l_F for the individual properties (plus the portfolio); also shown for reference is the Swiss Re parametric curve of Fig. 2.4. [lower left panel] Empirical exposure curves for the individual properties (plus the portfolio). [lower right panel] Scatter plot of the least squares exposure fitted parameters k_G versus l_G for the individual properties (plus the portfolio). Parameters are: $\tilde{b} = 1.3$, $\tilde{g} = 2.0$, $t_m = 50$, $p_D = 0.8$, $N = 100000$.

clearly distinguishable from the sampling distribution inherent to the Monte Carlo simulation for sample sizes $N = 100000$, i.e., the distinction between property sizes is not a small effect that can be explained as a statistical fluctuation. For both fitting procedures, l_F and l_G vary slightly, but k_F and k_G increase more or less uniformly with property size (deviations from this trend may be attributable to the lack of uniformity in the properties of our small portfolio). Clearly the increase in k_G is much more pronounced than for k_F – as discussed before, the log-likelihood fitting procedure is favoured because of the better sensitivity. What is very interesting about the scatter plots is the comparison to the parametric curve of the Swiss Re parameterisation (solely dependent on the concavity, c) – the deviations for the different results of the individual properties are nearly perpendicular to the parametric curve. This suggests that the model presented here may be very useful in splitting the Swiss Re classification beyond the concavity.

Table 3.1: Parameter estimates k_F, l_F and k_G, l_G (with their respective b and g values) for the portfolio and individual properties. For reference, the Swiss Re parameterisation with concavity $c = 1.5$ has $k = 2.54$, $l = 1.17$ ($b = 12.65$, $g = 4.22$).

property	log-likelihood				least squares exposure			
	k_F	l_F	b_F	g_F	k_G	l_G	b_G	g_G
portfolio	2.17	1.39	8.76	5.01	2.57	0.91	13.97	3.48
1	1.29	1.10	3.63	4.00	1.06	0.95	2.88	3.59
2	1.61	1.19	5.00	4.29	1.62	0.97	5.05	3.64
3	1.85	1.28	6.36	4.60	2.15	0.96	8.58	3.61
4	2.24	1.20	9.39	4.32	2.60	0.77	13.46	3.16
5	1.98	1.45	7.24	5.26	2.55	0.98	12.81	3.66
6	2.42	1.43	11.24	5.18	2.91	0.83	18.36	3.29
7	2.37	1.67	10.70	6.31	3.22	0.99	25.03	3.69
8	2.72	1.66	15.18	6.26	3.52	1.03	33.78	3.80
9	2.78	1.74	16.12	6.70	4.23	0.88	68.72	3.41

3.6 Other property types: an office block example

So far, we have discussed residential property. We have seen how different properties within a portfolio give rise to different severity curves, given a common time distribution. However, the Swiss Re parameterisation with different concavities tells us that different classes of properties have different behaviour. In the hypothesis of the fire simulation model, this may be attributed to the existence of distinct time distributions. For example, offices and public buildings invariably have fire doors, smoke and fire detectors (sometimes directly linked to the emergency services), security guards, fire extinguishers, and even automatic sprinkler systems. Further, more complex industrial facilities will have bespoke systems for dealing with fires (e.g., airports have their own fire response teams). It is therefore reasonable to hypothesise that the distribution of the random time during which a fire causes damage is more skewed towards small times. Certainly, this is strongly suggested by the increasing concavity. The probability density function for the time distribution (excluding the special cases $\tilde{g} = 1$, $\tilde{b} = 1$, and $\tilde{b}\tilde{g} = 1$) where $T < t_m$ is

$$f_{T|T<t_m}(t) = \frac{(\tilde{b}-1)(\tilde{g}-1)\tilde{b}^{1-t}\ln\left(\tilde{b}\right)}{\left[1-\tilde{b}\tilde{g}+(\tilde{g}-1)\tilde{b}^{1-t}\right]^2}. \tag{3.6.5}$$

$f_{T|T<t_m}(t)$ is plotted in Fig. 4.1 for three illustrative parameter sets – the set with the lowest $\tilde{b}$ and highest $\tilde{g}$ is most strongly skewed towards small times.[8]

[8] We reiterate that the time distribution is an ansatz whose parameters should be obtained from a portfolio of similar properties. That this ansatz is a generalisation of the MBBEFD distribution does not mean that we are simply using an MBBEFD distribution to reproduce another MBBEFD distribution – the point of the model is to identify the behaviour of individual properties *relative* to their portfolio. In this section however, the time distributions used are for illustrative purposes.

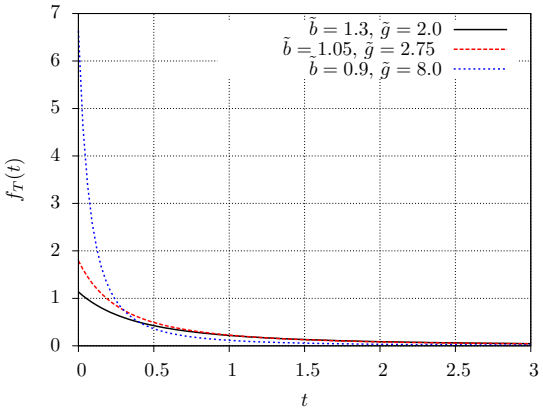

Figure 3.7: Probability density function for the time distribution, $f_{T|T<t_m}(t)$, Eq. (3.6.5), for different parameter sets $\tilde{b}$ and $\tilde{g}$.

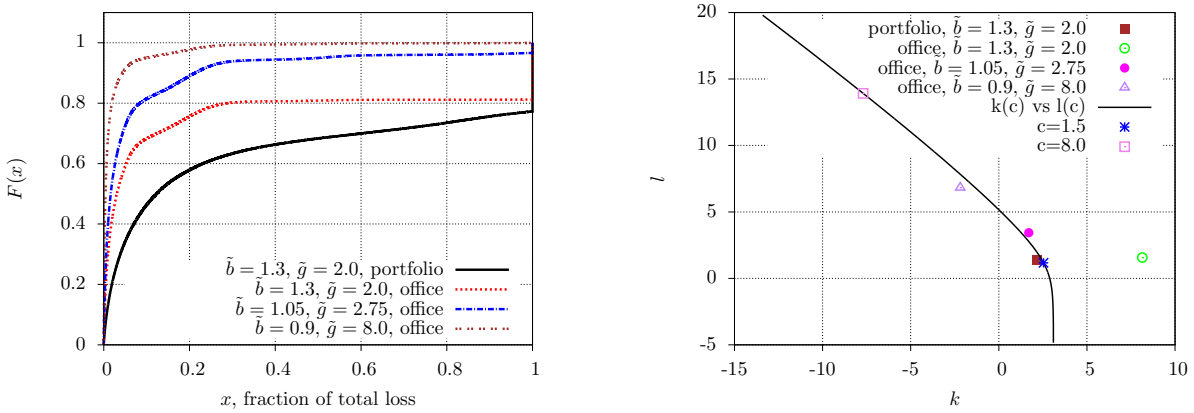

Figure 3.8: [left panel] Empirical severity curves for the 'office' building with different parameter sets, $\tilde{b}$ and $\tilde{g}$, compared to the portfolio of residential properties. [right panel] Scatter plot of the resultant k_F and l_F parameter estimates; also shown for reference is the Swiss Re parametric curve of Fig. 2.4.

Let us now consider the 'office' property with 30 rooms outlined in Fig. A.1 of the Appendix. This has fire doors – if one were to consider only the usual doors (thin red lines) and open spaces (thick black lines) as the edges in the graph, the graph would decompose into five disconnected components. The results for the Monte Carlo simulation (for different parameter sets $\tilde{b}$, $\tilde{g}$, but with $t_m = 50$, $N = 100000$, $p_D = 0.8$ fixed) are shown in Fig. 3.8. In the left panel, the empirical severity curves are shown. The solid black line is the residential portfolio from the previous section, used as a reference. The empirical severity curve for the office using the same time distribution as for the residential property is shown as the dotted red line – this gives a result analogous to the different members of the residential property portfolio and is in effect, just a larger version. Clearly, the fire doors don't have a significant impact in this respect. The dot-dashed blue and dot-dot brown curves show what happens when the time distribution is skewed towards smaller times and the effect is quite evidently that smaller fractional losses are more heavily prevalent. A scatter plot of the parameter estimates k_F, l_F is shown in the right panel of Fig. 3.8.

The residential property portfolio has the time distribution 'fitted' (as discussed in the previous section) such that the concavity $c \approx 1.5$. The office with the same time distribution (the light green circle) has approximately the same l_F but an appreciably larger value of k_F – this is exactly the behaviour exhibited by larger properties within the residential portfolio. The two different time distribution parameter sets (solid pink dot and purple triangle) lie close to the parametric concavity curve but with higher concavity than for the residential portfolio.

This exercise is not intended as an exhaustive study of office properties, but merely an illustration of how the model may be used beyond the case of residential property. In principle, the idea would be to take a portfolio of similar office buildings within the Swiss Re classification and to use the fire simulation model to see if there is a distinction between the various properties, given a common time distribution.

3.7 Conclusions for the fire simulation model

The fire simulation model explored in this chapter is based on the observation that the MBBEFD distribution may be broken down into two components: a deterministic propagation of the fire and a random time during which this fire propagates. The physical picture is akin to filling a bucket with water for a randomly selected time. Although the lack of uniformity inherent to real properties and the small residential portfolio used in this study do lead to deviations from the MBBEFD behaviour, the overall trend is very clear. One of the central conclusions of this model is that larger buildings have larger k (or b) values but similar l (or g) parameter for the distribution of losses, relative to a portfolio of properties with a common time distribution. This is in contrast to the Swiss Re concavity classification where k (b) decreases and l (g) increases along the parametric curve as the concavity increases. An illustration of how different concavities may occur within the fire simulation model was also given. This suggests a definite procedure as to how one might refine the broad concavity classification scheme in order to determine more accurately the losses due to fire risk for different properties and property types.

Chapter 4

A purely graph theoretic approach

In this chapter, we shall consider a more abstract approach to the study of severity and exposure curves involving random graphs [bollobas2001]. The basic observation is that these random graphs exhibit properties not unlike the MBBEFD distribution, as shall be discussed below.

The study of random graphs began with the introduction of two (closely related) models: by Gilbert [gilbert1959] and Erdős-Rényi [erdosrenyi1959]. In the literature, graphs constructed via both models are usually referred to as Erdős-Rényi graphs, but for our purposes we shall be considering the Gilbert model. Recalling the nomenclature from Section 1.2, starting from an empty graph with n vertices such that $V = \{v_1, \ldots, v_n\}$, each of the $\binom{n}{2}$ possible edges $(v_i, v_j)_{j>i}$ is added with probability p. For small p, the graph is nearly empty, with many disconnected components; for large enough p, the graph is connected (and for $p = 1$, the graph will be complete). How the graph 'evolves' from one behaviour to the other for different p in the limit of large n was studied in Ref. [erdosrenyi1960].

Erdős-Rényi random graphs may be considered in another fashion. Starting from a complete graph, each edge is considered in turn and removed with probability $1 - p$. However, in principle this procedure could be applied to any starting graph. Here, we shall investigate complete graphs and the graphs of the residential property portfolio outlined in the Appendix in this fashion. Numerically, the graphs that we deal with in this chapter can be conveniently simulated using the Python 'networkx' package [scipy2008].

Turning to the problem of the severity and exposure curves for fire risk, let us model the loss fraction in the following manner. We start from a complete graph of order n or one of the properties in our portfolio, where (as for the last chapter) the vertices represent the different rooms of a property. The edges are removed as described above. We now select a vertex from the original graph at random and work out the order of the connected subgraph to which it belongs. The loss fraction x is the ratio of the order of the connected subgraph containing the selected vertex to the order of the original graph. This is repeated N times to construct a Monte Carlo sample $\underline{x} = (x_1, \ldots, x_N)$. The underlying idea is that the fire has spread from the selected vertex throughout the connected subgraph such that the distribution of fractional losses is dependent on the probability p and the order of the graph. We know that for small p, there will be many disconnected components such that the loss fraction is likely to be small; conversely, for large p, the graph will almost invariably be connected such that the loss fraction is unity. Will the

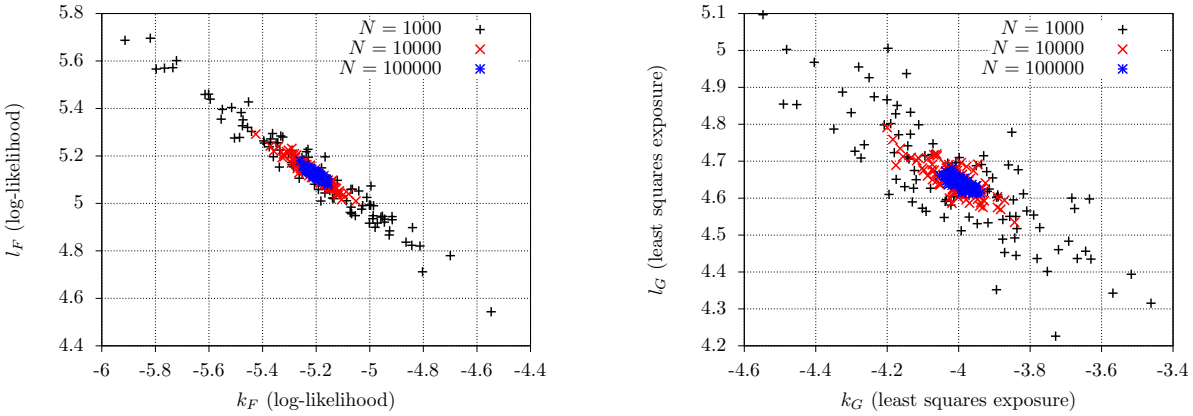

Figure 4.1: [left panel] Scatter plot of log-likelihood fitted parameters k_F versus l_F for various sample sizes using the Erdős-Rényi random graphs. [right panel] Scatter plot of least squares exposure fitted parameters k_G versus l_G, again for various sample sizes. Parameters are: $n = 20$, $p = 0.05$.

distribution of the loss fraction be equivalent to the MBBEFD distribution? No – in this model, the loss fraction will have the discrete values

$$x \in \left\{ \frac{1}{n}, \frac{2}{n}, \ldots, 1 \right\}, \tag{4.0.1}$$

whereas the MBBEFD distribution is for continuous $x \in [0, 1]$. However, as we shall see, the model approximates the MBBEFD distribution rather well and is thus a candidate tool for studying fire risk.

4.1 Erdős-Rényi random graphs

This section concerns Erdős-Rényi random graphs whose starting point is the complete graph of order n. We begin by considering a graph of order $n = 20$, with probability for each edge $p = 0.05$ (such that $pn = 1$ in this case) and check the convergence of the Monte Carlo simulation with respect to the sample size N. This is done, as in the previous chapter, by running the Monte Carlo 100 times for different sample sizes. Under the hypothesis that the distribution of x approximates the MBBEFD distribution (this will be discussed shortly), we extract the parameter estimates k_F, l_F (log-likelihood) and k_G, l_G (least squares exposure), and scatter plot the results to see if a reasonable degree of convergence is achieved. The results are shown in Fig. 4.1. What is seen is very similar to the results of the previous chapter. When $N = 100000$, we see that the tolerance is approximately ± 0.07 for k_F or k_G and ± 0.05 for l_F and l_G. Again, there is a clear negative correlation. The central values are

$$k_F \approx -5.20, \quad l_F \approx 5.25, \quad k_G \approx -4.00, \quad l_G \approx 4.65. \tag{4.1.2}$$

That the two different estimation procedures give different results is an indication that the distribution of losses only approximates the MBBEFD distribution.

Figure 4.2 shows the empirical severity and exposure curves alongside the fitted MBBEFD curves for a graph of order $n = 20$ and with various values of p. For $p = 0.025$, we see that indeed the smaller fractional losses are more prevalent whereas for $p = 0.1$, total losses begin to appear. This is the expected

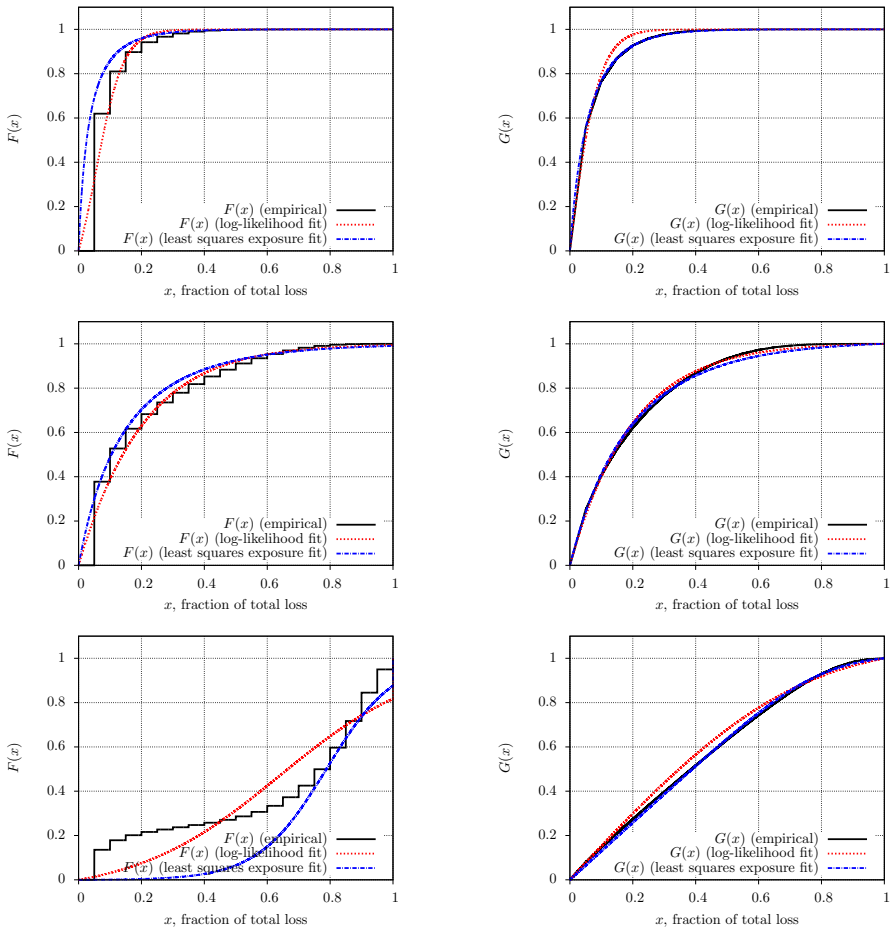

Figure 4.2: [left panels] Plots of the empirical severity distribution alongside the fitted curves using log-likelihood or least squares exposure fitting for the Erdős-Rényi random graph model. [right panels] Plots of the empirical exposure curve alongside the fitted curves using log-likelihood or least squares exposure fitting. Parameters are: $n = 20$, $p = 0.025$ [upper panels], $p = 0.05$ [middle panels], $p = 0.1$ [lower panels].

behaviour. Allowing for the discreteness, the empirical severity curve is a reasonable approximation to the MBBEFD distribution for the two lower values of p with an estimated Kolmogorov-Smirnov test statistic of the order of 0.1. We see that as before, the exposure curves are less sensitive than the severity curves. For the largest probability, $p = 0.1$, the agreement between the empirical curves and the MBBEFD distribution curves is not so good, with an estimated value for the Kolmogorov-Smirnov test statistic of around 0.2. We conclude that the random graph model for complete graphs does not give rise to an MBBEFD distribution for values of p that are larger than $p \sim 1/n$ (this will be seen further in the next paragraph).

Figure 4.3 shows a plot of the parameter estimates k_F, l_F and k_G, l_G for different values of n (the order of the graph) and p (the probability of an edge being present). Comparing the results for the two parameter estimation methods (left and right panels, respectively), it is clear that although the general trend of the results is similar, $k_F \neq k_G$ and $l_F \neq l_G$, especially for p larger than around $1/n$. This is indicative of the fact that while the model distribution resembles the MBBEFD distribution, it is no

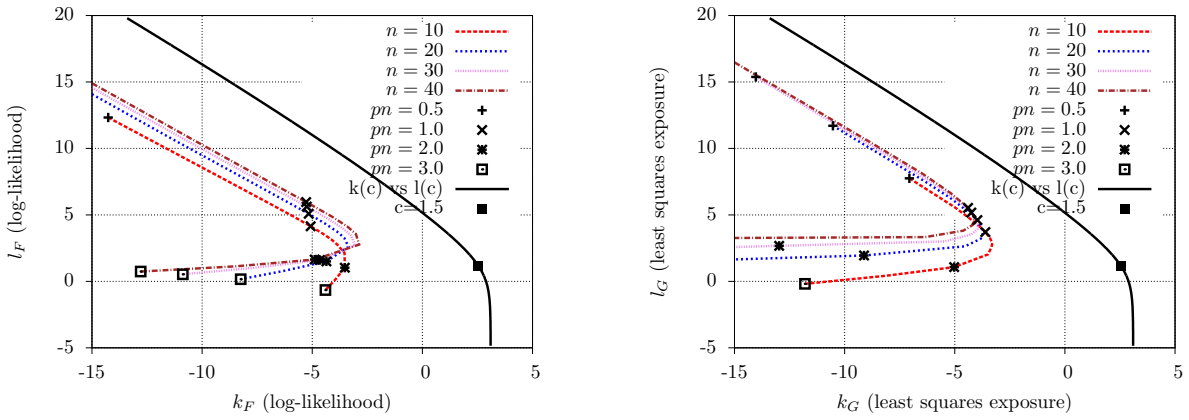

Figure 4.3: [left panel] Plot of log-likelihood estimates k_F versus l_F for different graph orders (n) and probabilities (p) using the Erdős-Rényi random graphs. Shown for reference is the parametric curve for the Swiss Re parametrisation in terms of the concavity c. [right panel] Plot of least squares exposure estimates k_G versus l_G. Monte Carlo sample size is $N = 100000$.

more than an approximation. Also shown in Fig. 4.3 is the parametric curve k versus l for the Swiss Re parameterisation in terms of the concavity. The results for the random graph model are far-removed – for a given value of l, the model values k_F and k_G are typically five less than k for a range of n and p configurations. Recall that $k = \ln(b)$, so this means that the model parameter b_F or b_G would be $\sim e^{-5}b \approx 0.0067b$, i.e., the random graph model with complete graphs gives rise to b-values that are 2–3 orders of magnitude smaller than for the concavities implied by the Swiss Re parameterisation. However, we also note that for lower values of pn, the slope of the curves almost exactly matches that of the parametric curve. We thus conclude that while the Erdős-Rényi random graph model does not match the Swiss Re parameterisation exactly, it does exhibit some interesting similarities.

4.2 Property portfolio graphs

As mentioned above, it is possible to generalise the Erdős-Rényi random graph model and use any graph as a starting point. In this section, we shall consider the graphs of the residential property portfolio presented in the Appendix. Unlike the more abstract complete graphs, these graphs represent physical properties. Recall that in the graphs of the portfolio, the edges represent the open spaces, doors, and walls/ceilings/floors through which a fire may propagate. For the model approach considered here, we consider the fire to propagate through an open space or door with probability p_D; likewise, the fire will propagate through a wall/ceiling/floor with probability p_W. The model for the portfolio then proceeds as follows for each of the N Monte Carlo sample simulations: a property graph is selected from the portfolio using the probability distribution given in the Appendix, each edge (of those that occur in that graph) is removed with a probability $1 - p_D$ or $1 - p_W$ as appropriate, a vertex is selected at random, and the resultant loss fraction is the ratio of the connected subgraph containing the selected vertex to the order of the original graph. The Monte Carlo sample then gives us our empirical distribution from which

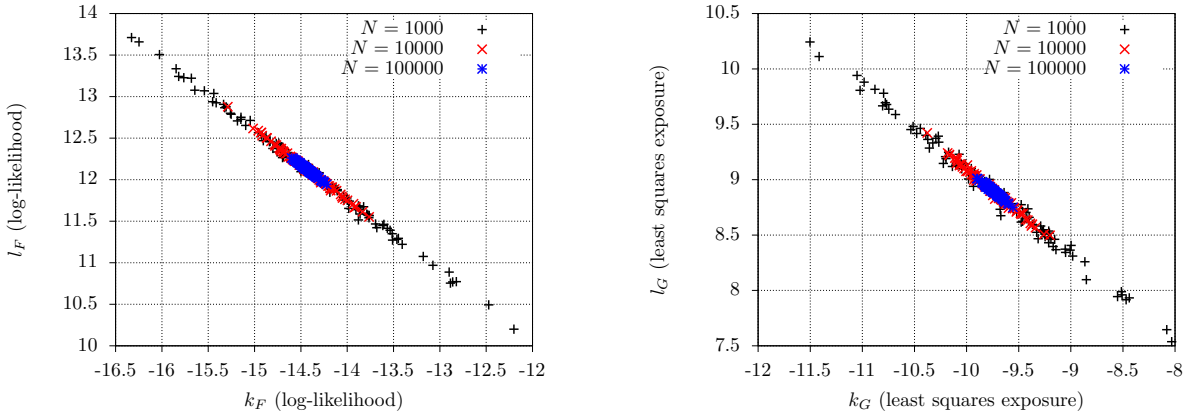

Figure 4.4: [left panel] Scatter plot of log-likelihood estimates k_F versus l_F for various sample sizes using the residential property portfolio graphs. [right panel] Scatter plot of least squares exposure estimates k_G versus l_G for various sample sizes. Parameters are: $p_D = p_W = 0.1$.

the parameters k and l may be estimated under the hypothesis that the distribution is an MBBEFD distribution.

As before, we begin by checking that the Monte Carlo simulation is sufficiently convergent for a given sample size N. The parameter estimates for 100 Monte Carlo samples of different size N using either log-likelihood estimation (k_F, l_F) or least squares exposure (k_G, l_G) are scatter plotted in Fig. 4.4 (the fire propagation probabilities were $p_D = p_W = 0.1$). The pattern is identical to the previous results – a sample size of $N = 100000$ gives a reasonable degree of convergence.

Figure 4.5 shows the empirical severity and exposure curves alongside the corresponding fitted MBBEFD curves for different parameter sets p_D, p_W. The upper panels show the case $p_D = p_W = 0$. This corresponds to the situation where the fire does not spread at all – in effect each loss is of exactly one room from the property. The distribution of losses thus arises directly from the distribution of properties occurring in the portfolio of residential properties for this parameter set and the model effectively has a lower bound for the loss distribution which is clearly an artefact of the discretisation. The agreement between the empirical and fitted MBBEFD severity curves is good, with an estimated Kolmogorov-Smirnov test statistic (allowing for the discretisation) of less than 0.1; the corresponding exposure curves also agree closely, although as discussed previously, this is a less sensitive measure. For the non-trivial parameter sets $p_D = p_W = 0.1$ and $p_D = 0.5$, $p_W = 0.1$, it is seen that the empirical severity curves and log-likelihood fitted MBBEFD curves agree nicely, with an estimated Kolmogorov-Smirnov test statistic of around 0.1. The least squares exposure fitted severity curve also works well for $p_D = 0.5$, $p_W = 0.1$ (in this case, the agreement between the two fitting procedures is striking). The corresponding exposure curves also show good agreement in all cases except for the log-likelihood fitting for the parameter set $p_D = p_W = 0.1$. The loss distribution arising from the random graph model using the portfolio of residential properties thus appears to be a good approximation to the MBBEFD distribution and works better than for the complete graphs of the previous section. This may be due to the distribution of properties within the

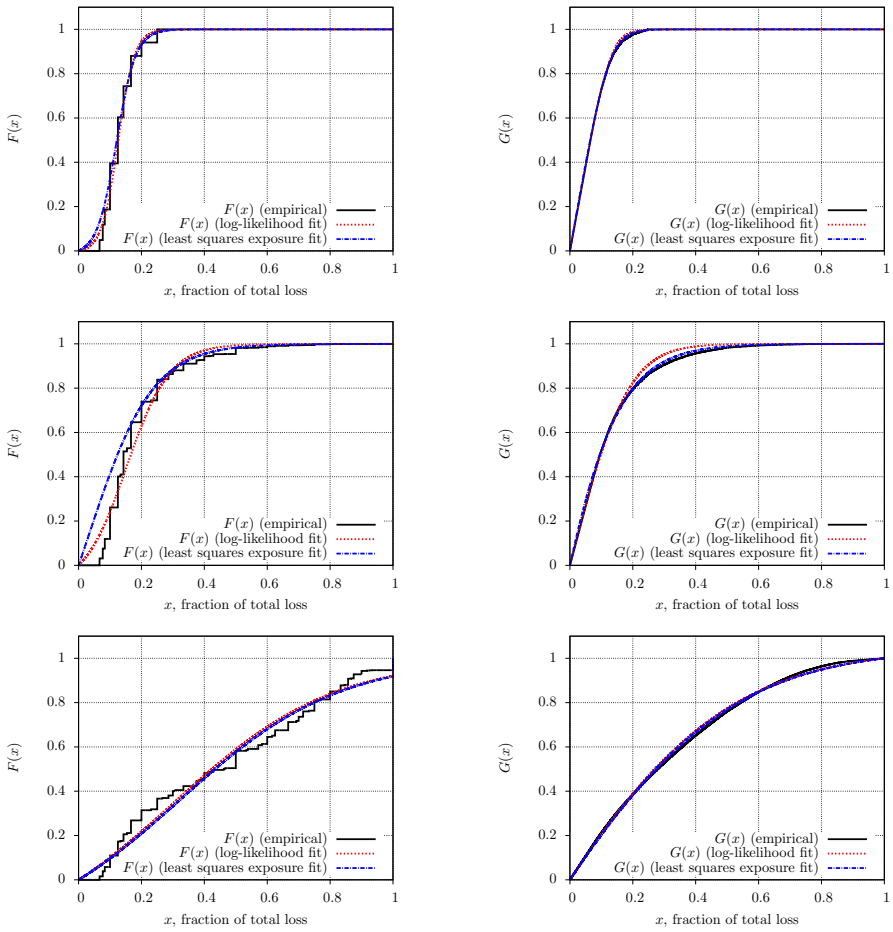

Figure 4.5: [left panels] Plots of the empirical severity distribution alongside the fitted curves using log-likelihood or least squares exposure fitting for the residential property portfolio graph model. [right panels] Plots of the empirical exposure curve alongside the fitted curves using log-likelihood or least squares exposure fitting. Parameters are: $p_D = p_W = 0$ [upper panels], $p_D = p_W = 0.1$ [middle panels], $p_D = 0.5$, $p_W = 0.1$ [lower panels]. Monte Carlo sample size is $N = 100000$.

portfolio (in the sense that more properties are being burnt less often, as was discussed in the previous chapter) or the fact that the starting graph is less connected than for the complete graphs.

Next, we plot the parameter estimates k_F, l_F (log-likelihood estimation) and k_G, l_G (least squares exposure estimation) for various probabilities p_D and p_W in Fig. 4.6. The results are very similar to those of the previous section for complete graphs. Comparing the two estimation methods (left and right panels of the figure), the agreement is better for lower values of p_D and p_W (and is significantly better when compared to Fig. 4.3 for complete graphs). Comparing to the parametric Swiss Re curve, the slope of the various curves for low probabilities almost exactly matches as before. However, the estimates k_F and k_G are again smaller than for the parametric Swiss Re curve.

Finally, let us consider the random graph model for each of the properties of the portfolio individually. The results for the parameter set $p_D = 0.5$, $p_W = 0.1$ and with $N = 100000$ are shown in Fig. 4.7. The empirical severity curves for the individual properties are shown in the upper left panel (the solid black

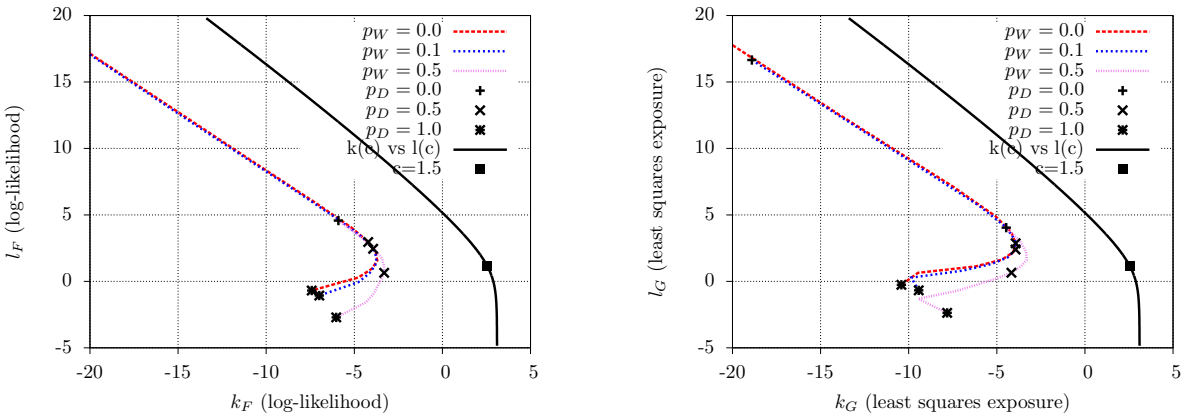

Figure 4.6: [left panel] Plot of log-likelihood estimates k_F versus l_F for different probabilities p_D and p_W using the residential property portfolio graphs. Shown for reference is the parametric curve for the Swiss Re parametrisation in terms of the concavity c. [right panel] Plot of least squares exposure estimates k_G versus l_G. Monte Carlo sample size is $N = 100000$.

line is for the portfolio and corresponds to the empirical severity curve shown in the lower left panel of Fig. 4.5). The effects of the discretisation are prominent. The empirical severity curve for the smallest property (property 1) is shown as the dashed red line and has the largest probability for a total loss. As the property size increases, the probability of a total loss decreases whereas the probability of smaller losses increases. This is what one would expect. The empirical exposure curves are shown in the lower left panel of Fig. 4.7 and also show a clear distinction between the various properties as their size (the order of the graph) increases. Scatter plots of the log-likelihood and least squares exposure parameter estimates k_F, l_F and k_G, l_G are shown in the upper and lower right panels of Fig. 4.7, respectively. The corresponding numerical values are given in Table 4.1. There is good agreement between the two different parameter fitting methods, especially for the values of l. It is seen that as the property size increases, the value of l increases and the value of k decreases slightly. Moreover, the differences are too big to be explained as a statistical fluctuation of the Monte Carlo simulation. Comparing to the Swiss Re parametric curve, we see that the k, l values for the different properties lie on a line almost parallel to the parametric curve, although the values of k are appreciably lower. Given that the larger properties have larger values of l, they would appear to have larger concavity, as one would expect from the Swiss Re parameterisation. Thus, the random graph model for the individual properties seems to share many properties with the Swiss Re parameterisation.

4.3 Conclusions for the graph theoretic approach

In this chapter, a purely graph theoretic approach to the derivation of property loss distributions from first principles has been considered. Considering the rooms of a property as the vertices of a graph and the routes that a fire may propagate as the edges of the graph, each edge is considered in turn and removed with probability $1 - p$, where p is the probability of fire transmission. One vertex is selected

Table 4.1: Parameter estimates k_F, l_F and k_G, l_G (with their respective b and g values) for the portfolio and individual properties. For reference, the Swiss Re parameterisation with concavity $c = 1.5$ has $k = 2.54$, $l = 1.17$ ($b = 12.65$, $g = 4.22$).

property	log-likelihood				least squares exposure			
	k_F	l_F	b_F	g_F	k_G	l_G	b_G	g_G
portfolio	-3.95	2.46	0.019	12.70	-3.96	2.42	0.019	12.25
1	-3.83	1.77	0.022	6.87	-3.45	1.98	0.032	8.24
2	-4.09	2.17	0.017	9.76	-3.60	2.24	0.027	10.39
3	-4.18	2.41	0.015	12.13	-3.78	2.43	0.023	12.36
4	-3.78	1.70	0.023	6.47	-4.18	1.77	0.015	6.87
5	-4.25	2.72	0.014	16.18	-3.82	2.62	0.022	14.74
6	-4.33	2.84	0.013	18.12	-4.09	2.72	0.017	16.18
7	-4.10	2.98	0.017	20.69	-4.16	2.92	0.016	19.54
8	-4.13	3.28	0.016	27.58	-4.34	3.30	0.013	28.11
9	-3.86	3.06	0.021	22.33	-4.51	3.14	0.011	24.10

at random to be the starting point for the fire and the connected subgraph to which it belongs then represents the loss. The loss fraction is then the ratio of the order of this connected subgraph to the order of the original graph. This is repeated many times to generate a Monte Carlo sample.

In the first instance, complete graphs were considered as the starting point for the simulation. These correspond to Erdős-Rényi random graphs. It was seen that for low values of p (lower than $1/n$ where n is the order of the graph), the resultant loss distribution was well approximated by the MBBEFD distribution. Comparing to the Swiss Re parameterisation, while the values of k (or b) were significantly lower for the random graphs, the results evolved parallel to the parametric curves for low p on a scatter plot of k versus l.

The more physical case of the graphs for a portfolio of residential properties was also considered. The loss distributions were again well approximated by the MBBEFD distribution; moreover the agreement between the different parameter estimation methods employed (log-likelihood and least squares exposure) was rather good. It was possible to consider each property individually and it was seen that the resulting parameter estimates lie parallel to the Swiss Re curve on a scatter plot of k versus l. This indicates that the larger properties have larger concavities (insofar as the term applies beyond the the Swiss Re parametric curve) in the random graph model and that the model shares definite characteristics with the Swiss Re parameterisation.

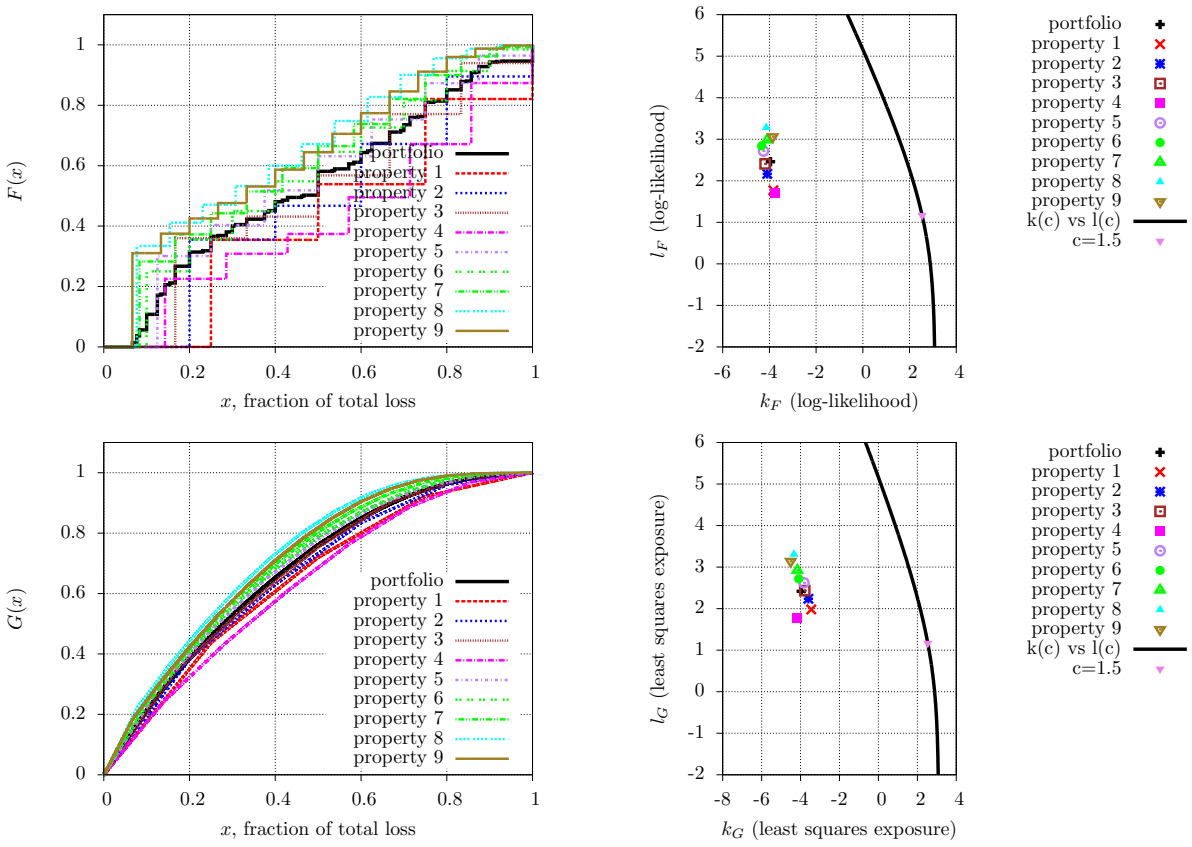

Figure 4.7: [upper left panel] Empirical severity curves for the individual properties (plus the portfolio). [upper right panel] Scatter plot of the corresponding log-likelihood fitted parameters k_F versus l_F for the individual properties (plus the portfolio); also shown for reference is the Swiss Re parametric curve of Fig. 2.4. [lower left panel] Empirical exposure curves for the individual properties (plus the portfolio). [lower right panel] Scatter plot of the least squares exposure fitted parameters k_G versus l_G for the individual properties (plus the portfolio). Parameters are: $p_D = 0.5$, $p_W = 0.1$, $N = 100000$.

Chapter 5

Conclusions

This study has been concerned with modelling property fire risk and exposure curves. After a general introduction, the MBBEFD distribution that is often heuristically used to describe the data was discussed in some detail. Numerical procedures for Monte Carlo simulation and parameter estimation were developed. In particular, two different numerical parameter estimation methods were employed: a maximum likelihood method for the severity curve and a least squares minimisation fit for the exposure curve. Two separate models for the loss severity and exposure curves were developed and results obtained. The first model was based on a physical interpretation of the MBBEFD distribution. The progress of a fire through a graph representing the physical property was simulated over a randomly distributed time period during which the fire caused damage. The time distribution was chosen so as to reproduce the characteristic parameter estimates for a portfolio of properties. Thereafter, the model could be used to study the losses incurred for properties individually, relative to the portfolio. The second (simpler) model was based on random graph theory. Starting from either a complete graph or a graph representing a physical property, the edges were removed according to definite probabilities. A vertex (room) was selected at random to be the starting point of the fire and the loss was given by the order of the connected subgraph to which the vertex belonged. Again, this allowed for a study of individual properties. Both models gave rise to severity and exposure curves that were well approximated by the MBBEFD distribution. The model results were compared to the Swiss Re parameterisation. Of note are the results for individual properties within a portfolio. On a scatter plot of the parameter estimates, the two models gave rise to two distinct behaviours: the first model had parameter estimates that evolved with property size perpendicularly to the Swiss Re parametric curve, the parameter estimates of the second model evolved parallel to the Swiss Re curve. The overall conclusion is that this distinction could be utilised in order to discriminate between the models and determine which (if either) is the proper description in future studies of the data.

Appendix A

Property portfolio

The residential properties that comprise the portfolio used in Chapters 3 and 4 are shown in graph theoretic form in Fig. A.1 ('property 1-9'). They are designed to mimic a realistic sample.[1] Of course in real life, some house sizes are more prevalent than others, so in the Monte Carlo simulations that use the full portfolio, properties are thus chosen from this portfolio with specific probabilities. The UK government National Archives provide census information on the distribution of residential housing, grouped by the number of bedrooms in a household and whether rented or owned [narchive2016] (in this case for England and Wales). Since we are not interested in the ownership of the property, the rented/owned figures are combined using their relative probabilities. The resultant distribution is shown in Table A.1.

No. bedrooms	Probability
1	0.1192
2	0.2768
3	0.4152
4	0.1396
≥ 5	0.0492

Table A.1: Distribution of housing (England and Wales), grouped by number of bedrooms [narchive2016].

Our portfolio consists of nine residential properties and is classified by the number of rooms (not necessarily bedrooms). The distribution is shown in Table A.2. The first two of our properties have been assigned to have one bedroom, with the probability for each of these properties to be chosen from the portfolio to be half that of a single bedroomed household from the National Archive. This is repeated until the last property (number 9) which is assigned to have 5 bedrooms and (given that there is only one five-bedroomed house in the portfolio) takes the full probability as for the National Archive distribution. Notice that the graph density typically decreases with graph order.

[1] Indeed, as the reader might surmise, most of these properties do in fact exist. The graphs are constructed by simply sketching the physical buildings in the same way as for the example property, Fig. 3.1 ('property 7').

Property	No. bedrooms	No. rooms (order)	Graph size	Graph density	Probability
1	1	4	5	0.833	0.0596
2	1	5	7	0.700	0.0596
3	2	6	9	0.600	0.1384
4	2	7	10	0.476	0.1384
5	3	8	15	0.536	0.2076
6	3	10	19	0.422	0.2076
7	4	12	27	0.409	0.1888
8	4	13	28	0.359	0.1888
9	5	15	34	0.324	0.0492

Table A.2: Distribution of properties in the portfolio.

Also shown at the bottom of Fig. A.1 is the graph for an office building. This has 30 vertices and 66 edges, giving a density of 0.152.

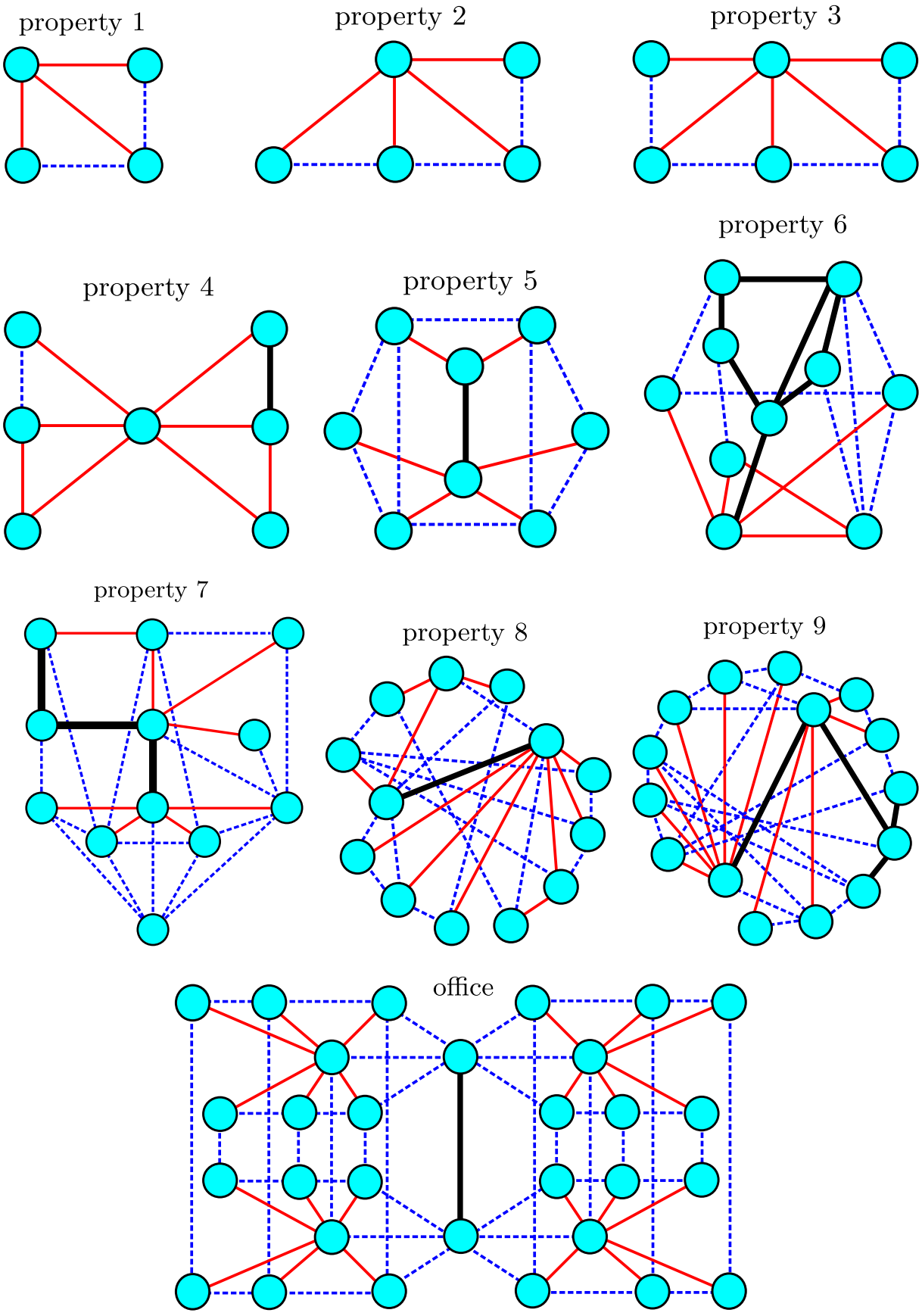

Figure A.1: The graphs corresponding to the portfolio of residential properties, 'property 1-9' and the office building 'office'. The different edges follow the notation of Fig. 3.2.

Bibliography

[bernegger1997] S. Bernegger. The Swiss Re exposure curves and the MBBEFD distribution class. *ASTIN Bulletin*, 27:99–111, 1997.

[bollobas1979] B. Bollobás. *Graph theory: an introductory course*. Springer-Verlag, New York, 1979.

[bollobas2001] B. Bollobás. *Random Graphs*. Cambridge University Press, Cambridge, 2nd edition, 2001.

[bondy1976] J.A. Bondy and U.S.R. Murty. *Graph theory with applications*. MacMillan, London, 1976.

[build2010] UK Government. The Building Regulations 2010, Fire safety, Approved Document B, Volume 1 - Dwelling Houses. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/485420/BR_PDF_AD_B1_2013.pdf, Crown Copyright 2011.

[erdosrenyi1959] P. Erdős and A. Rényi. On random graphs I. *Publicationes Mathematicae*, 6:290–297, 1959.

[erdosrenyi1960] P. Erdős and A. Rényi. On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 5:17–61, 1960.

[gfol1666] Wikipedia. *Great fire of London*. https://en.wikipedia.org/wiki/Great_Fire_of_London, 2017.

[gilbert1959] E. N. Gilbert. Random graphs. *Ann. Math. Statist.*, 30(4):1141–1144, 12 1959.

[jharia2017] Wikipedia. Jharia. https://en.wikipedia.org/wiki/Jharia, 2017.

[narchive2016] The National Archives, Office for National Statistics. Part of 2011 Census, Detailed Characteristics on Housing for Local Authorities in England and Wales Release. Adapted from data from the Office for National Statistics licensed under the Open Government Licence v.3.0. http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/rel/census/2011-census/detailed-characteristics-on-housing-for-local-authorities-in-england-and-wales/sty-households-in-england-and-wales.html, Crown Copyright 2015.

[parodi2015] P. Parodi. *Pricing in general insurance*. CRC Press, Taylor & Francis Group, Boca Raton, 2015.

[peters2002] Tim Peters. Listsort / timsort. In (Python.org), editor, *Python Software Foundation*, 2002.

[python] Guido van Rossum. Python. http://www.python.org.

[response2017] Home Office. Fire Incident Response Times: April 2015 to March 2016, England. Contains public sector information licensed under the Open Government Licence v3.0. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/584351/fire-incident-response-times-1516-hosb0117.pdf, Crown Copyright 2016.

[scipy2001] Eric Jones, Travis Oliphant, Pearu Peterson, et al. SciPy: Open source scientific tools for Python. http://www.scipy.org/, 2001.

[scipy2008] Aric A. Hagberg, Daniel A. Schult, and Pieter J. Swart. Exploring network structure, dynamics, and function using networkx. In Gaël Varoquaux, Travis Vaught, and Jarrod Millman, editors, *Proceedings of the 7th Python in Science Conference*, pages 11 – 15, Pasadena, CA USA, 2008.

[swiss2004] Swiss Re. Exposure rating, 2004.