



Stockholms
universitet

Pricing short-term XL contracts using timeline simulation

Niklas Jungner

Masteruppsats 2010:1
Matematisk statistik
Juni 2010

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm



Mathematical Statistics
Stockholm University
Master Thesis **2010:1**
<http://www.math.su.se>

Pricing short-term XL contracts using timeline simulation

Niklas Jungner*

Juni 2010

Abstract

This thesis looks at a sample problem of pricing a short-term excess of loss contract, valid only during the last three months of the year. We fit (annual) claim frequency and severity distributions to a dataset of historical losses, using the Peaks Over Thresholds (POT) model. The POT model is extended to allow additional distributions. The limited duration of the contract means that we have to take seasonal variation of claim occurrence rates into account. This seasonality is estimated from external data sources, since our own data are too meagre. Finally, we use Monte Carlo timeline simulation to estimate the total loss.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden.
E-mail: niklas.jungner@gmail.com . Supervisor: Dmitrii Silvestrov.

Preface

This report constitutes a Master's thesis, for the degree of Master of Science in Actuarial Mathematics at Stockholm University.

The work was carried out at Sirius International Insurance Corporation, during the spring of 2010.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to Jörgen Olsén, my supervisor at Sirius International, for his invaluable guidance and support.

I am also grateful to the rest of the staff at Sirius, particularly Johan Karlsson and Patrik Dahl, for their advice and assistance, and Dance Zurovac-Jevtic, for providing me with the opportunity to work on this thesis.

Finally, I would also like to thank Professor Dmitrii Silvestrov, my supervisor at Stockholm University, for his input.

Contents

1	Introduction	5
1.1	Types of reinsurance	5
1.1.1	Proportional reinsurance	5
1.1.2	Non-proportional reinsurance	5
1.2	This thesis	6
2	Data	8
2.1	Claim records	8
2.1.1	Trending the data	9
2.2	External data	10
3	Current approach to pricing	12
3.1	The collective risk model	12
3.2	The distribution of S	12
4	The timeline simulation approach to pricing	13
4.1	The advantages of timeline simulation	13
4.2	Our approach	13
4.2.1	Parameter estimation	14
4.2.2	Seasonality estimation	15
4.2.3	Timeline simulation	15
5	Parameter estimation	16
5.1	The Peaks Over Threshold (POT) method	16
5.1.1	Assessing the goodness-of-fit	19
5.2	Number of claims: Poisson	20
5.3	Number of claims: NegBin	20
5.4	Claim severity: (Shifted) Pareto	21
5.5	Claim severity: Burr	22
5.6	Claim severity: Log-Normal	22
5.7	Claim severity: Weibull	23
6	Seasonality estimation	24
6.1	Kernel smoothing: Local linear regression	24
6.1.1	Bandwidth selection	25
6.2	Estimating $\lambda(t)$	26
7	Timeline simulation	29
7.1	No causality or time-dependence	29
7.2	NHPP simulation	29
7.2.1	Mixed NHPP	30

8	Case study: European property insurance	31
8.1	Parameter estimation	31
8.2	Seasonality estimation	34
8.3	Timeline simulation	37
8.3.1	Some remarks	38
9	Conclusions	40
A	Appendix: Properties of the Poisson process	41
A.1	The Poisson distribution	41
A.2	Definition of the (mixed) Poisson process	41
B	Appendix: Claim record	44
C	Appendix: Parameter estimates	46
C.1	$u = 2000000$	46
C.2	$u = 2462963$	46
C.3	$u = 4000000$	47
D	Appendix: ML-estimation of parameters	48
D.1	Number of claims: Poisson	48
D.1.1	Truncated distribution	48
D.1.2	ML parameter estimates	48
D.2	Number of claims: NegBin	49
D.2.1	Truncated distribution	49
D.2.2	ML parameter estimates	49
E	References	51

1 Introduction

Reinsurance can succinctly be described as insurance for insurers. The concept originated in the field of marine insurance, and can be traced back to the Middle Ages. Seafaring was a risky business back in the day, and merchants seeking to protect their income and investments fed a demand for marine insurance. As the stakes became higher, fewer insurers were willing and/or able to take on the greater risks themselves. This created a need for reinsurance, as a way of spreading the risk among many. Today, reinsurance has branched out to cover all types of risks, but the core principle remains the same: to protect regular insurance companies, by taking on part of the risk, in exchange for a slice of the premia.

1.1 Types of reinsurance

Reinsurance cover comes in many different shapes and forms. A distinction is usually made between proportional and non-proportional types of reinsurance.

1.1.1 Proportional reinsurance

Proportional reinsurance is defined in terms of exposure, rather than losses. Typically, the reinsurer takes on a certain percentage of each risk, in return for the same percentage of the premium. The most basic form of such an arrangement is known as quota share reinsurance, where the proportion to be ceded is fixed upon signing the reinsurance contract, and is uniform among all policies covered by the agreement. The second form of proportional insurance is known as surplus insurance. Here, the reinsurance contract specifies a retention level and an upper limit to the cover, rather than a fixed percentage. The cedant will retain all of the risk that fall below the retention, and surrender the remainder (up to the upper limit) to the reinsurance company. This means that the proportion of each risk that is ceded will vary between policies, depending on the sums insured. For example, if the retention is set at 1 000 currency units, an individual policy insuring values up to 2 000 will be reinsured at 50%, whereas one worth 4 000 will be reinsured at 75%. All claims from these policies will be split 50:50 and 25:75 respectively, even if they fail to exceed the retention level.

1.1.2 Non-proportional reinsurance

Proportional reinsurance is less common in the field of non-life insurance and will not be dealt with in the following. The more common alternative, non-proportional reinsurance, is characterised by the fact that such policies only cover losses that exceed a fixed lower limit, known as the retention, or deductible. (Note that this definition of retention differs from the one in

surplus reinsurance, where retention specified the amount of *exposure* that stayed with the cedant.)

Non-proportional reinsurance contracts are thus defined in terms of how the incurred losses are split between cedant and reinsurer, irrespective of the amount of exposure. Such contracts can be divided into two major subclasses: excess of loss reinsurance and stop-loss reinsurance.

In excess of loss reinsurance, often abbreviated XL, the contracts cover the losses of individual events. The cedant is reimbursed for all individual losses that exceed some agreed-upon retention level r , but only up to the value of some limit l . This type of reinsurance policy is often denoted ' l xs r '. If an occurred loss exceeds $r + l$, it will 'spill over' the reinsurance cover and the excess cost will fall upon the cedant. As a consequence, the cedant may sign up for another XL contract with another reinsurer, with retention set at $r + l$, the value at which the cover from the first contract was capped. Large exposures are usually covered by a chain of such successive contracts, with several different reinsurers, each covering their own layer $(r_i, r_i + l_i)$.

The characteristics of an excess of loss contract depend on how we define an 'event'. In so called *Per Risk XL* reinsurance, every single risk (i.e. policy issued by the direct insurer) is treated individually, and we regard claims from different policies as different events, even if they originated from the same incident. In this context, the XL contract is essentially a protection against unusually large claims.

With the other type of XL contract, *Catastrophe XL* or *Cat XL* for short, we treat major catastrophes (e.g. hurricanes, earthquakes, etc.) as single events, and aggregate all the losses that arise from them (i.e. all the claims made on the cedant by its policyholders) into catastrophe losses. This way, the size of a catastrophe loss will reflect both the number of claims received by the cedant and the severity of each claim. In this setting, we can regard the XL contract as a protection against clusters of claims that are moderate in their own right, but substantial when put together.

If we instead aggregate all losses over a full calendar year, we arrive at the stop loss contract. Stop loss reinsurance is basically excess of loss applied to the annual aggregate, and can be used to put a cap on the annual loss.

1.2 This thesis

The backdrop of this thesis is that the commissioning company, Sirius International, would like to be able to simulate claims on a timeline, accounting for seasonality and clustering phenomena, when pricing reinsurance contracts. In one of the current pricing approaches, the annual number of claims for a given cedant and policy is modelled using either a Poisson or a Negative binomial distribution, but this implies that the occurred claims are distributed uniformly over time. In real life, there is likely to be some degree of seasonality, in particular when dealing with natural perils like

windstorms, and the timeline simulation model should be able to reproduce it.

The problem is that data are scarce, since, being a non-life reinsurance company dealing mainly in non-proportional contracts, Sirius are typically only notified of claims that exceed the retention threshold. There might be enough data in a client's loss history to estimate the annual number of claims with reasonable precision, but it will not suffice if our goal is to estimate the seasonal variation.

This thesis looks at a sample problem of pricing a short-term XL contract, valid only during the last three months of the year. The fact that the contract does not span a full year means that we have to take seasonal variation of claim occurrence rates into account. We fit claim frequency and severity distributions to a set of historical claims from a European direct insurer, and combine the fitted distributions with a seasonality function obtained from external data sets. The resulting non-homogeneous point process is used to simulate series of claims on a timeline, and we estimate the premium from the average aggregate loss over repeated realisations of that process.

The thesis is organised in the following way: Section 1 is this introduction; Section 2 describes the available data sets; Section 3 gives a brief description of the current pricing approach; Section 4 presents the timeline simulation approach to pricing; Section 5 outlines the Peaks Over Thresholds method of parameter estimation; Section 6 describes how we estimate the intensity function; Section 7 presents two different approaches to timeline simulation; Section 8 is the actual case study; Section 9 summarises the results.

2 Data

Reinsurance companies typically have quite limited historical data to draw on, especially when dealing with catastrophe excess-of-loss cover. This is due to catastrophic events being few and far between, and the fact that cedants only disclose claims that exceed a certain threshold. The latter also means that our data will be left-truncated, which further add to the difficulty of parameter estimation. Data will also differ greatly in scope, detail and reliability between different cedants. Some provide full information on everything, from exact dates of loss to what types of perils that caused them, whereas others only list major losses and the years they occurred.

Moreover, it is seldom possible to merge data arising from different cedants or policies, as the inherent risks might differ greatly between them. For example, in the case of property insurance, different portfolio compositions could mean that one cedant is more susceptible to wind storm damage than another, even if their insured building stocks are equal in value and found in the same geographic areas.

What little data we have might be enough to make inference on the number of claims in a given year, as well as their severities, but it will most certainly not suffice if we wish to study seasonal variation. Hence, since we wish to incorporate seasonality in our model, we will have to make use of data from external sources.

Below is a short description of the data sets available to us.

2.1 Claim records

By claim records, we mean the reinsurance company's own historical data, or data provided to them as part of the underwriting information. Let's say we have a record of all claims made by a given cedant, on a given policy, over a period of m years. These m years need not be consecutive, as there might have been intermittent years when the policy was not renewed.

We introduce the notation

M_i = lower limit of reported claims in year i

n_i = number of claims in year i with value greater than M_i

x_{ij} = value of claim j in year i , given that $x_{ij} > M_i$

where $i = 1, \dots, m$, $j = 1, \dots, n_i$. When dealing with excess-of-loss policies, $M_i \geq 0$ will typically be the excess point in use that year, or some value below it if such figures were requested as part of the underwriting information. Note that no M_i :s are likely to be the same, since all monetary values have to be adjusted for inflation, as well as other possible trends. For this purpose, we also need some suitable measure of inflation and information on how exposure has varied over time. Inflation measures are easy to come

by (e.g. consumer price index, property price indices, etc.) but measures of exposure will have to be supplied by the cedant. Two examples of such measures are 'total sums insured' or 'earned premium income'. It is assumed that this is all information available to us.

Our claim record can then be presented in a table, akin to Table 1.

<i>Year</i>	<i>Lower limit</i>	<i>No. of claims</i>	<i>Severities</i>		
1	M_1	n_1	x_{11}	\cdots	x_{1n_1}
\vdots	\vdots	\vdots	\vdots	\ddots	\vdots
m	M_m	n_m	x_{m1}	\cdots	x_{mn_m}

Table 1: Typical claim record for some reinsurance policy.

In this thesis, we will be looking at claims resulting from property and loss of production insurance policies, issued by an undisclosed European direct insurer. We are unable to reveal the name of the company, for reasons of confidentiality, but we present the claim record in question in Appendix B. Note that all monetary values have been linearly transformed (to keep from identification) and *trended* (see section 2.1.1 below). About 99% of the sums insured stem from properties found in the direct insurer's country of domicile; the rest come from one of its neighbouring countries.

2.1.1 Trending the data

The procedure of adjusting past loss data for inflation, and other possible changes over time, is known as *trending*. This can be done in various ways, some more complex than others, but a basic model should at the very least compensate for inflation and changes in exposure. We will employ one such very basic model, where the consumer price index (CPI) is used as a measure of inflation, and the earned premium income (EPI) of our cedant is used to adjust for varying exposure. Note that the earned premium income will itself have to be deflated, since it reflects the number of policies as well as the general premium level, and the latter is obviously subject to increases in property prices.

It would be preferable to use some property price index (PPI) as a measure of inflation, instead of the more general CPI, but the 2009 values of the PPI:s from our country under study were not available at the time of writing. Likewise, the Total Sums Insured (TSI) of our cedant is a more accurate measure of exposure, but those figures were not included in the submitted underwriting information, from which our data were obtained.

The trending procedure in the excess of loss case will differ slightly, depending on whether the reported claims are aggregates of several policies (catastrophe XL) or if they have arisen from single policies (per risk XL). In the cat XL case, each event represents a major catastrophe that is likely

to hit every cedant in the exposed area, irrespective of how large their portfolios are. As a consequence, a change in exposure (e.g. number of policies sold) will have little effect on an individual cedant's claim frequency, but it will affect the severity of each claim, since the aggregate loss from each catastrophe ought to be proportional to the number of active policies.

With per risk contracts, we get the opposite situation. Claims are now counted individually, which means that the claim frequency is likely to increase linearly with exposure, but the losses are no longer aggregated and will not be affected. Of course, inflation applies to all monetary values, regardless of the type of contract.

In summary, we have the following two scenarios:

- **Catastrophe XL:** We inflate the losses X_{ij} , reporting limit M_i and number of claims n_i of year i to present values, by

$$\begin{aligned} X_{ij}^{\text{today}} &= X_{ij} \frac{CPI_{\text{today}}}{CPI_{\text{year } i}} \left(\frac{EPI_{\text{today}}}{EPI_{\text{year } i}} \bigg/ \frac{CPI_{\text{today}}}{CPI_{\text{year } i}} \right) = X_{ij} \frac{EPI_{\text{today}}}{EPI_{\text{year } i}} \\ M_i^{\text{today}} &= \dots = M_i \frac{EPI_{\text{today}}}{EPI_{\text{year } i}} \\ n_i^{\text{today}} &= n_i \end{aligned}$$

- **Per risk XL:** We inflate X_{ij} , M_i and n_i by

$$\begin{aligned} X_{ij}^{\text{today}} &= X_{ij} \frac{CPI_{\text{today}}}{CPI_{\text{year } i}} \\ M_i^{\text{today}} &= M_i \frac{CPI_{\text{today}}}{CPI_{\text{year } i}} \\ n_i^{\text{today}} &= n_i \left(\frac{EPI_{\text{today}}}{EPI_{\text{year } i}} \bigg/ \frac{CPI_{\text{today}}}{CPI_{\text{year } i}} \right) = n_i v_i \end{aligned}$$

The data set in our case study consists of individual claims, and hence we use the second approach, 'per risk XL'. The scale factor associated with n_i will be denoted v_i from now on. A table of all values of v_i , relevant to our case study, is presented in Appendix B.

2.2 External data

As hinted earlier, the external data will only be used to capture general seasonality patterns, not to estimate the actual number of claims. That is, we will use it to find the distribution of the occurrence times, conditional on the number of occurrences in a given time interval. For example, if we believe that claims occur according to a non-homogeneous Poisson process, with time-dependent intensity $\lambda(t)$, then the external data will be used to

find the *shape* of $\lambda(t)$; its amplitude will then be determined from our own historical claim records.

One consequence of this approach is that our choice of external data will have no effect on the expected number of annual claims and, by extension, total claim amount, as long as we restrict ourselves to policies that span full calendar years. Hence, the choice of external data set(s) is of lesser importance in such situations, and it can be chosen freely, as long as it specifies the exact loss dates, is large enough to provide reasonably accurate estimates of the daily claim intensities, and exhibits the desired pattern of seasonal variation. A natural choice in many applications would be pooled industry-wide data, if such are available.

But not in ours! We would like to be able to tailor the shape of $\lambda(t)$, to suit the unique risk profile of individual cedants. For example, if one cedant only has exposure against risks in the southern parts of Sweden, then its claim seasonality is likely to differ from the one that appears in pooled, nation- and industry-wide data. (Remember that we are mostly dealing with natural catastrophes, whose rates of occurrence are highly dependent on season, as well as geographic location.) Hence, we are looking for $\lambda(t)$, *conditional* on the exposure of our cedant.

One solution to this problem is offered by risk management software packages, such as the one supplied by Risk Management Solutions, Inc. The software draws on meteorological and geological data from several decades back, and uses advanced (undisclosed) statistical models to predict the frequency and severity of various perils, in various geographic areas. RMS allows us to enter the exact risk profile of our cedant, including information on exposure divided by location, if such is available, and will then generate a long list of simulated events that are likely to affect it. The seasonality of those events can be regarded as representative of the company we are studying, and we can use the simulated data set to estimate the shape of $\lambda(t)$. One could argue about the sensibility in fitting a model to simulated data, rather than the actual observations on which the simulations were based. However, the latter approach would require us to develop our own catastrophe model from scratch, something that the author has neither the time nor the expertise to do. Instead, we utilise the results from the more sophisticated models of RMS.

Because of the nature of our historical claim record (i.e. claims from a European property insurer), we will use simulated data on wind storms and floods in Europe. The simulations spanned 7 136 and 25 908 simulation rounds, and produced 500 000 and 400 000 events, respectively. We will only make use of the simulated claim *dates*.

(Note that we did not have access to information on the exact exposure of this cedant, and used generic simulations instead, covering roughly the same geographic area. This thesis is, after all, just meant for demonstrative purposes — more care might be advisable in a real life situation.)

3 Current approach to pricing

The most commonly used method of modelling the annual claim cost is to use the collective risk model, defined below. This allows us to express the annual cost as a stochastic sum of independent random variables, and determine its probability distribution, e.g. through some recursive formula. This distribution can then be used to predict future annual costs.

3.1 The collective risk model

In the collective risk model, we consider the insurance portfolio as an entity, and pay no attention to the individual policies from which claims arise. It has two main components:

$$\begin{aligned} N &= \# \text{ claims during the contract period (typically one year)} \\ X_j &= \text{loss amount of the } j\text{:th claim, } j = 1, \dots, N \end{aligned}$$

and we can express the aggregate loss S as a stochastic sum

$$S = \sum_{j=1}^N X_j. \quad (1)$$

It is assumed that N, X_1, X_2, \dots are mutually independent and that all $\{X_j\}$ are identically distributed.

Reinsurance companies introduce a payout function $g(\cdot)$, so that their stake in each claim X_j can be expressed as $Y_j = g(X_j)$. Their aggregate loss is then obtained by replacing X_j with Y_j in (1). A typical choice of $g(\cdot)$ is

$$g(X_j) = \min(l, \max(X_j - r, 0)) \quad (2)$$

which corresponds to an l vs r type of contract. Stop-loss contracts, on the other hand, are represented by aggregate payout functions

$$h(S) = \min(L, \max(S - D, 0)) \quad (3)$$

where L and D are the aggregate limit and deductible, respectively.

3.2 The distribution of S

Using historical claim data, we are able to fit distributions to N, X_1, \dots, X_N . These distributions can then be used to obtain the distribution of S , through some recursive method. Two common choices are the recursive formulas of Panjer and Ströter. Other possibilities include normal power or gamma approximations, which allow us to express the cumulative distribution function of S using the standard Gaussian and Gamma cumulative distribution functions, respectively. These methods are presented in Johansson (1997, Chapter 5) and will not be dealt with here.

4 The timeline simulation approach to pricing

4.1 The advantages of timeline simulation

Timeline simulation involves shifting our interest from modelling the number of claims that occur in a specific time-frame, typically one year, to modelling the inter-arrival times between claims. In practice, it means that we estimate the claim arrival intensity $\lambda(\cdot)$ from our data, and use it to simulate series of claims (or other events) on a timeline. With sufficiently many realisations, we are able to estimate the distribution of the aggregate loss S from the observed aggregate claims.

In the most general form, we can denote the probability of an event occurring in $(t, t + \Delta t]$, given that n events have taken place in $(0, t]$ by

$$P(\text{event in } (t, t + \Delta t] | n \text{ events in } (0, t]) = \lambda(t, n, T_1, \dots, T_n, x_1, \dots, x_n) \Delta t$$

that is, we allow the instantaneous claim frequency $\lambda(\cdot)$ to depend on all information available at the time, including severities of previous claims x_j and previous arrival times T_j . Likewise, we allow the claim severities to depend on all previous history.

The timeline simulation method is advocated convincingly by Kreps (2007), who lists some of its advantages. These include

- **Causality:** Current events (may) influence future events.
- **Time dependency:** Allows seasonality and long-term trends. Discounting can be done exactly.
- **Greater realism:** Allows more complex modelling of events, by adding new features (e.g. payments made in instalments at stochastic time intervals, stochastic interest rates, exposure that depends on the previous history.)

One key aspect of the timeline simulation approach is that it can be easily reduced to the collective risk model, by letting λ be constant (equivalent to $N \sim Po(\lambda)$) or by letting $\lambda \sim \Gamma(\alpha, \beta)$ and sampling a new λ for each realisation (equivalent to $N \sim NegBin(\alpha, \beta)$). We will use this fact when fitting our model.

4.2 Our approach

Our approach is a middle ground between the two outlined here. We will adhere to the collective risk model for parameter estimation throughout this thesis, in the sense that we will fit distributions to N and X_1, X_2, \dots , assumed to be independent, and consider insurance portfolios rather than individual policies. But instead of finding the distribution of (1) through some recursive or approximate method, as is commonly done, we will employ

Monte Carlo timeline simulation to estimate the mean value of S directly. That is, we will generate sample paths of the (non-homogeneous) claim process, and estimate $E[S]$ from the annual aggregates of the simulated claims. This procedure allows us to introduce seasonality, delayed payments, etc.

We present this approach in three steps:

1. Parameter estimation
2. Seasonality estimation
3. Timeline simulation

4.2.1 Parameter estimation

We fit distributions to N and X_1, X_2, \dots using the standard Peaks Over Threshold method on our claim record. Maximum Likelihood estimators are used throughout. The following distributions are applied to N

- Poisson
- Negative binomial

and the following are applied to X_1, X_2, \dots

- (Shifted) Pareto
- Burr
- Log-normal
- Weibull

The Poisson distribution for the number of claims in a given time interval arises naturally when the claims can be regarded as independent events, occurring one at a time, and the probability of a claim taking place in an arbitrary time-interval equals the length of that interval, multiplied by some risk parameter (intensity) λ .

The negative binomial distribution can be seen as an extension of the Poisson distribution, where we regard the intensity λ as a Gamma distributed random variable. This reflects an uncertainty in the value of λ , which would be present if the underlying risk fluctuates randomly.

As for the claim severity distributions, the ones listed above are common examples from the literature, known for their heavy tails.

4.2.2 Seasonality estimation

We look for a periodic intensity function $\lambda(t)$, with period 1 year, that describes how claims are distributed over time. The external data set is used as source data, since our own historical data are likely to be too meagre. In principle, one could allow periods longer than one year and let $\lambda(t)$ account for long-term deterministic fluctuations as well, but such phenomena should not be inferred from external data.

$\lambda(t)$ is scaled so that

$$\hat{\lambda} = \int_0^1 \lambda(t) dt, \quad (4)$$

where $\hat{\lambda}$ is the ML estimate of λ , obtained from fitting a $\text{Poisson}(\lambda)$ distribution to N . If $N \sim \text{NegBin}(\alpha, \beta)$, we replace $\hat{\lambda}$ with λ_i in (4), where λ_i are observations from a random variable $\Lambda \sim \Gamma(\alpha, \beta)$, and make sure to repeat the scaling procedure whenever a new value of Λ is drawn.

4.2.3 Timeline simulation

We use a (mixed) non-homogeneous Poisson process to simulate series of claims on a timeline. The simulations are carried out using time transformation. We repeat the simulation procedure for all fitted distributions and estimate the expected total loss $E[S]$ by the average over all simulation rounds.

5 Parameter estimation

We now turn to finding suitable probability distributions for the number of claims in a single year and the severity of each claim.

Reinsurance companies that issue excess-of-loss policies are primarily interested in achieving a good fit for claims that exceed a certain threshold, the excess point. One way of doing this is by employing the Peaks Over Threshold (POT) method.

5.1 The Peaks Over Threshold (POT) method

This section is based on Johansson (1997, Chapter 5), but has been expanded to include more probability distributions.

The idea behind this method is that we restrict ourselves to studying claims that exceed a certain threshold u , and then fit distributions to the excess loss, *conditional* on the claims being greater than u .

We start off with a data set like the one presented in Table 1. In our case, it is the claim record of a property insurance policy, spanning 11 years. It is assumed that all x_{ij} and M_i have been adjusted for inflation, changes in exposure and other factors that could affect the comparability between years (see Section 2.1.1). The observed claim counts n_i should not be adjusted at this stage; we will compensate for changes in exposure during the parameter estimation in Sections 5.2-5.3.

We proceed to transform this data set in the following manner:

1. Select a threshold u .
2. Set

$$\tilde{M}_i = \max(M_i, u)$$

$$\tilde{n}_i = \text{number of claims in year } i \text{ with value greater than } \tilde{M}_i$$

$$\tilde{x}_{ik} = \text{value of claim } k \text{ in year } i, \text{ given that } \tilde{x}_{ik} > \tilde{M}_i,$$

$$\text{where } k = 1, \dots, \tilde{n}_i, \text{ i.e. } \{\tilde{x}_{ik}\} = \{x_{ij} : x_{ij} > \tilde{M}_i\}.$$

3. Subtract \tilde{M}_i from all \tilde{x}_{ik} .

This transformed data set is presented in Table 2.

Choosing a suitable threshold u is something of a balancing act: if we set it too low, we will include too many small claims in the analysis, which could compromise the level of fit in the right tail. On the other hand, too high a threshold will not leave many claim observations for parameter estimation. Mean excess plots could be of use in this regard.

In reality, u will always have to be chosen from values below or equal to the excess point when pricing excess-of-loss contracts, since we want the

<i>Year</i>	<i>Lower limit</i>	<i>No. of claims</i>	<i>Severities - retention</i>		
1	\tilde{M}_1	\tilde{n}_1	$\tilde{x}_{11} - \tilde{M}_1$	\cdots	$\tilde{x}_{1\tilde{n}_1} - \tilde{M}_1$
\vdots	\vdots	\vdots	\vdots	\ddots	\vdots
m	\tilde{M}_m	\tilde{n}_m	$\tilde{x}_{m1} - \tilde{M}_m$	\cdots	$\tilde{x}_{m\tilde{n}_m} - \tilde{M}_m$

Table 2: Transformed claim data set.

excess point to fall within the support of the fitted distributions. Stop-loss contracts can be regarded as a special case, with excess points (of the individual contracts) equal to zero; this leaves $u = 0$ as the only choice.

We now set out to fit some possible distributions of N and $X_k - u$ (defined below) to our data. We assume that all random variables are independent and that X_1, X_2, \dots are identically distributed.

$$N = \# \text{ claims in a year that exceed } u \quad (5)$$

$$X_k = \text{severity of claim } k \text{ in a year, given that } X_k > u, \quad (6)$$

$$k = 1, \dots, N.$$

But our observations $\{\tilde{n}_i\}$ and $\{\tilde{x}_{ik} - \tilde{M}_i\}$ are from the r.v. N_i and $X_{ik} - \tilde{M}_i$, defined by

$$N_i = \# \text{ claims in year } i \text{ that exceed } \tilde{M}_i \quad (7)$$

$$X_{ik} = \text{severity of claim } k \text{ in year } i, \text{ given that } X_{ik} > \tilde{M}_i, \quad (8)$$

$$k = 1, \dots, N_i.$$

and it follows that $N_i \stackrel{d}{=} N$ and $X_{ik} - \tilde{M}_i \stackrel{d}{=} X_k - u$ only for years i where $\tilde{M}_i = u$. However, we still want to incorporate all available observations in our parameter estimations, including those from years where $\tilde{M}_i > u$. Fortunately, as we shall soon see, the distributions of (5)-(6) are related to those of (7)-(8), and the probability/density functions of the former pair will contain the same unknown parameters as those of the latter pair. Thus, if we find the probability/density functions of (7)-(8), we can make use of all available observations.

If we denote the cumulative distribution function of $X_k - u$ by F , we can express the c.d.f. of $X_{ik} - \tilde{M}_i$ in the following way

$$\begin{aligned} G_i(y) &= P(X_{ik} - \tilde{M}_i \leq y) = P(X_k - \tilde{M}_i \leq y | X_k > \tilde{M}_i) \\ &= \frac{P(\tilde{M}_i < X_k \leq \tilde{M}_i + y)}{P(X_k > \tilde{M}_i)} = \frac{P(\tilde{M}_i - u < X_k - u \leq \tilde{M}_i - u + y)}{P(X_k - u > \tilde{M}_i - u)} \\ &= \frac{F(\tilde{M}_i - u + y) - F(\tilde{M}_i - u)}{1 - F(\tilde{M}_i - u)}. \end{aligned} \quad (9)$$

The second equality follows from the fact that all ground-up claims are assumed to be equally distributed; this means that any difference between X_k and X_{ik} is due to them having different truncation points. Derivation then gives us the corresponding density functions

$$g_i(y) = \frac{f(\tilde{M}_i - u + y)}{1 - F(\tilde{M}_i - u)}. \quad (10)$$

If we introduce $y_{ik} = \tilde{x}_{ik} - \tilde{M}_i$, the log-likelihood takes the form

$$l = \sum_{i=1}^m \left[-\tilde{n}_i \log(1 - F(\tilde{M}_i - u)) + \sum_{k=1}^{\tilde{n}_i} \log f(\tilde{M}_i - u + y_{ik}) \right]. \quad (11)$$

The ML equations can be obtained through derivation, but they will invariably have to be solved numerically. It is more convenient to maximise (11) directly; we do it by feeding a custom log-likelihood into the `mle` procedure of Matlab. The solutions are the ML estimates of the parameters from the distribution F , that is the distribution of $X_k - u$. This fitted distribution can then be used to find the distribution of N , see sections 5.2 and 5.3.

It is easy to modify the above expressions so as to allow for right truncation as well. With the conditional c.d.f.

$$\tilde{F}_v(y) = P(X_k - u \leq y | u < X_k \leq u + v) = \frac{P(0 < X_k - u \leq y)}{P(0 < X_k - u \leq v)} = \frac{F(y)}{F(v)}, \quad (12)$$

our observations $\tilde{M}_i < \tilde{x}_{ik} \leq u + v$ can be regarded as observations from the distribution $\tilde{G}_{i,v}$, a modified version of (9), given by

$$\begin{aligned} \tilde{G}_{i,v}(y) &= \frac{\tilde{F}_v(\tilde{M}_i - u + y) - \tilde{F}_v(\tilde{M}_i - u)}{1 - \tilde{F}_v(\tilde{M}_i - u)} \\ &= \frac{F(\tilde{M}_i - u + y) - F(\tilde{M}_i - u)}{F(v) - F(\tilde{M}_i - u)}. \end{aligned} \quad (13)$$

The corresponding density function and log-likelihood becomes

$$\tilde{g}_{i,v}(y) = \frac{f(\tilde{M}_i - u + y)}{F(v) - F(\tilde{M}_i - u)} \quad (14)$$

$$\tilde{l}_v = \sum_{i=1}^m \left[-\tilde{n}_i \log(F(v) - F(\tilde{M}_i - u)) + \sum_{k=1}^{\tilde{n}_i} \log f(\tilde{M}_i - u + y_{ik}) \right] \quad (15)$$

and can be treated in the same way as (10) and (11). The right-truncated case will not be dealt with explicitly in the following.

In the general POT model, as presented in Rootzén and Tajvidi (1995) among others, it is assumed that N and $X_k - u$ have Poisson and generalised

Pareto distributions, respectively. While NegBin could very well be used in lieu of Poisson, the (generalised) Pareto distribution (GPD) is the natural choice for claim severities in the POT framework. This is mainly because $X_k - u$ converge in distribution to generalised Pareto as $u \rightarrow \infty$, for a large class of distributions of X_k . (Among them standard Pareto, Gamma and Log-Normal - see Schirmacher et al. (2005) for more details.) Moreover, the GPD is stable under increasing thresholds u , a unique and desirable property of that class of distributions (see Rootzén and Tajvidi (1995)).

Nevertheless, we will consider the Log-normal, Weibull and Burr distributions as well, since they are other common choices for claim severities, and we cannot assume that u will be large enough for convergence to take place.

5.1.1 Assessing the goodness-of-fit

The most convenient way of comparing the goodness of fit of the different distributions under study is by using the information criteria of Akaike (AIC) and Bayes (BIC). Both are based on the maximised value of the log-likelihood, but add a penalty term, proportional to the number of parameters in the fitted distributions.

With the notation

$m =$ no. of observation years,

$\tilde{n}_{tot} = \sum_{i=1}^m \tilde{n}_i =$ total no. of observed claims,

$k_N =$ no. of parameters in the distribution of N ,

$k_X =$ no. of parameters in the distribution of X ,

$l_{N,max} =$ maximised log-likelihood for the distribution of N ,

$l_{X,max} =$ maximised log-likelihood for the distribution of X ,

the AIC and BIC statistics can be written

$$\begin{aligned} AIC &= AIC_N + AIC_X \\ &= 2(k_N + k_X) - 2(l_{N,max} + l_{X,max}), \end{aligned} \tag{16}$$

$$\begin{aligned} BIC &= BIC_N + BIC_X \\ &= k_N \log m + k_X \log \tilde{n}_{tot} - 2(l_{N,max} + l_{X,max}). \end{aligned} \tag{17}$$

The combinations of distributions for N and X are ranked according to their values of AIC and BIC; smaller values are preferable. From (16) and (17), we see that the difference between AIC and BIC lie solely in the penalty term, with BIC punishing extra parameters more strongly (provided m and \tilde{n}_{tot} are greater than 7).

It should be noted that these criteria are only used to compare competing models, and say nothing about the overall goodness of fit (i.e. the absolute values of AIC and BIC have no meaning).

5.2 Number of claims: Poisson

It holds that $\{X_{ik} : k = 1, \dots, N_i\} \subseteq \{X_k : k = 1, \dots, N\}$ for any i . Since all claims are independent, it follows that the number of variables among $\{X_k : k = 1, \dots, N\}$ that exceed \tilde{M}_i (i.e. N_i), conditional on $N = n$, has a binomial distribution, i.e. $N_i|N = n \sim Bin(n, p_i)$, where $p_i = P(X_k > \tilde{M}_i) = 1 - P(X_k - u \leq \tilde{M}_i - u) = 1 - F(\tilde{M}_i - u)$. Again, F denotes the distribution function of $X_k - u$, which has yet to be determined.

Now, if we assume that $N \sim Po(\lambda)$, then it can be shown (see Appendix D.1) that

$$N_i \sim Po(\lambda p_i),$$

a thinned Poisson distribution with thinning parameter $p_i = 1 - F(\tilde{M}_i - u)$. We see that this reduces to a $Po(\lambda)$ distribution for years where $\tilde{M}_i = u$, as one would expect.

The ML estimate of λ , using all available scaled up observations $\tilde{n}_i v_i$, is also derived in Appendix D.1. It is found to be

$$\hat{\lambda} = \frac{\sum_{i=1}^m \tilde{n}_i v_i}{\sum_{i=1}^m p_i} \quad (18)$$

5.3 Number of claims: NegBin

We now turn to our other choice of claim number distribution, the Negative Binomial one. The setup is the same as in the Poisson case, except that we now have $N \sim NegBin(\alpha, \beta)$, with parameterisation

$$P(N = n) = \frac{\Gamma(\alpha + n)}{\Gamma(\alpha)n!} \left(\frac{\beta}{\beta + 1}\right)^\alpha \left(\frac{1}{\beta + 1}\right)^n. \quad (19)$$

The relation $N_i|N = n \sim Bin(n, p_i)$ is not affected, however. A more common parameterisation is obtained by setting $p = 1/(\beta + 1)$, but we will stick with (19) to emphasise its identity as a mixed Poisson distribution, with a $\Gamma(\alpha, \beta)$ distributed structure variable.

It is shown in Appendix D.2 that, under these conditions,

$$N_i \sim NegBin(\alpha, \beta/p_i).$$

and that the ML estimates of α and β , using the scaled up observations

$\tilde{n}_i v_i$, can be obtained by maximising

$$l(\alpha, \beta) = \sum_{i=1}^m [\log \Gamma(\tilde{n}_i v_i + \alpha) - \log \Gamma(\alpha) + \alpha \log \left(\frac{\beta}{\beta + p_i} \right) - \tilde{n}_i v_i \log(\beta/p_i + 1)] \quad (20)$$

5.4 Claim severity: (Shifted) Pareto

Turning to the claim severities, we start by considering the (shifted) Pareto distribution, also known as the two-parameter Pareto distribution. It is perhaps the most common choice of claim severity distribution, and is defined by the cumulative distribution function

$$F(x) = 1 - \left(\frac{\theta}{\theta + x} \right)^\alpha, \quad x \geq 0, \quad \alpha > 0, \theta > 0. \quad (21)$$

This is essentially a reparameterised version of the two-parameter GPD, studied by Rootzén and Tajvidi (1997), under the constraint $\alpha > 0$. (The GPD permits $\alpha \in (-\infty, \infty)$.) As such, we expect the Pareto distribution to be stable under increasing thresholds, and, indeed, if we assume that $X_k - u \sim \text{Pareto}(\alpha, \theta)$, then the distribution function G_i of $X_{ik} - \tilde{M}_i$, as defined by (9), becomes

$$G_i(y) = 1 - \left(\frac{\theta + \tilde{M}_i - u}{\theta + \tilde{M}_i - u + y} \right)^\alpha. \quad (22)$$

This is also the c.d.f. of a shifted Pareto distribution, and we conclude that $X_{ik} - \tilde{M}_i \sim \text{Pareto}(\alpha, \theta + \tilde{M}_i - u)$.

The log-likelihood (11) has the form

$$l(\alpha, \theta) = \sum_{i=1}^m \sum_{k=1}^{\tilde{n}_i} \left[\log \alpha + \alpha \log(\theta + \tilde{M}_i - u) - (\alpha + 1) \log(\theta + \tilde{M}_i - u + y_{ik}) \right] \quad (23)$$

and can be maximized numerically to obtain the ML estimates.

Sometimes, the standard $SPareto(\alpha, \theta)$ distribution, also known as the one-parameter Pareto distribution, is used instead of the shifted one. Its c.d.f. takes the form

$$H(x) = 1 - \left(\frac{\theta}{x} \right)^\alpha, \quad x \geq \theta, \quad \alpha > 0, \quad (24)$$

and although it appears to have two parameters, θ is simply a lower bound-ary for the support, and must be set in advance.

For our purposes, this means that if we have fitted a $\text{Pareto}(\hat{\alpha}, \hat{\theta})$ distribution to $X_k - u$, then $X_k - u + \hat{\theta} \sim SPareto(\hat{\alpha}, \hat{\theta})$, and if we wish to

simulate a future ground-up claim Z from this distribution, it can be done in either of two ways:

$$Z = u + X = u - \hat{\theta} + Y \quad (25)$$

where $X \sim \text{Pareto}(\hat{\alpha}, \hat{\theta})$, and $Y \sim \text{SPareto}(\hat{\alpha}, \hat{\theta})$.

5.5 Claim severity: Burr

The Burr distribution adds another parameter to the (shifted) Pareto distribution. Its c.d.f. reads

$$F(x) = 1 - \left(\frac{\theta}{\theta + x^\tau} \right)^\alpha, \quad x \geq 0, \quad \alpha > 0, \theta > 0, \tau > 0. \quad (26)$$

As before, if we assume that $X_k - u \sim \text{Burr}(\alpha, \theta, \tau)$, then $G_i(y)$ becomes

$$G_i(y) = 1 - \left(\frac{\theta + (\tilde{M}_i - u)^\tau}{\theta + (\tilde{M}_i - u + y)^\tau} \right)^\alpha. \quad (27)$$

This is not a familiar distribution function, but derivation yields

$$g_i(y) = \frac{\alpha\tau(\tilde{M}_i - u + y)^{\tau-1}[\theta + (\tilde{M}_i - u)^\tau]^\alpha}{[\theta + (\tilde{M}_i - u + y)^\tau]^{\alpha+1}}, \quad (28)$$

which allows us to express the log-likelihood (11) as

$$l(\alpha, \theta, \tau) = \sum_{i=1}^m \sum_{k=1}^{\tilde{n}_i} \left[\log(\alpha\tau) + (\tau - 1) \log(\tilde{M}_i - u + y_{ik}) + \alpha \log[\theta + (\tilde{M}_i - u)^\tau] - (\alpha + 1) \log[\theta + (\tilde{M}_i - u + y_{ik})^\tau] \right]. \quad (29)$$

Numerical maximisation gives us the ML estimates $\hat{\alpha}$, $\hat{\theta}$, and $\hat{\tau}$.

5.6 Claim severity: Log-Normal

A variable $Y = \exp(X)$ is said to have a log-normal distribution if $X \sim N(\mu, \sigma^2)$. Its c.d.f. is written

$$F(x) = \Phi\left(\frac{\log x - \mu}{\sigma}\right), \quad x > 0, \quad -\infty < \mu < \infty, \sigma > 0, \quad (30)$$

where Φ is the standard normal c.d.f. If we assume that $X_k - u \sim \text{LogN}(\mu, \sigma^2)$, then $G_i(y)$, as defined by (9), becomes

$$G_i(y) = \frac{\Phi\left(\frac{\log(\tilde{M}_i - u + y) - \mu}{\sigma}\right) - \Phi\left(\frac{\log(\tilde{M}_i - u) - \mu}{\sigma}\right)}{1 - \Phi\left(\frac{\log(\tilde{M}_i - u) - \mu}{\sigma}\right)} \quad (31)$$

and upon derivation, we get

$$g_i(y) = \frac{\frac{1}{\sqrt{2\pi}\sigma(\tilde{M}_i - u + y)} \exp\left\{-\frac{[\log(\tilde{M}_i - u + y) - \mu]^2}{2\sigma^2}\right\}}{1 - \Phi\left(\frac{\log(\tilde{M}_i - u) - \mu}{\sigma}\right)} \quad (32)$$

which yields the log-likelihood

$$l(\mu, \sigma) = \sum_{i=1}^m \sum_{k=1}^{\tilde{n}_i} \left[-\log\left(1 - \Phi\left(\frac{\log(\tilde{M}_i - u) - \mu}{\sigma}\right)\right) - \log\sigma - \frac{[\log(\tilde{M}_i - u + y_{ik}) - \mu]^2}{2\sigma^2} \right] + \text{constant}. \quad (33)$$

Maximisation gives the ML estimates $\hat{\mu}$ and $\hat{\sigma}$. Note that these are the expectation and standard deviation of $X = \log Y$. The equivalent statistics for Y are given by

$$E[Y] = e^{\mu + \sigma^2/2}, \quad (34)$$

$$\text{Var}(Y) = e^{2(\mu + \sigma^2)} - e^{2(\mu + \sigma^2/2)}. \quad (35)$$

5.7 Claim severity: Weibull

The c.d.f. of the Weibull distribution

$$F(x) = 1 - e^{-(x/c)^\tau}, \quad x \geq 0, \quad c > 0, \quad \tau > 0, \quad (36)$$

resembles that of the exponential distribution, but adds a shape parameter τ . Again, assuming that $X_k - u \sim \text{Weibull}(c, \tau)$ leads to $G_i(y)$, as defined by (9), taking the form

$$G_i(y) = 1 - \exp\left(-\left[\left(\frac{\tilde{M}_i - u + y}{c}\right)^\tau - \left(\frac{\tilde{M}_i - u}{c}\right)^\tau\right]\right). \quad (37)$$

Derivation yields

$$g_i(y) = \frac{\tau}{c} \left(\frac{\tilde{M}_i - u + y}{c}\right)^{\tau-1} \exp\left(-\left[\left(\frac{\tilde{M}_i - u + y}{c}\right)^\tau - \left(\frac{\tilde{M}_i - u}{c}\right)^\tau\right]\right), \quad (38)$$

which results in the log-likelihood

$$l(c, \tau) = \sum_{i=1}^m \sum_{k=1}^{\tilde{n}_i} \left[\log \tau - \tau \log c + (\tau - 1) \log(\tilde{M}_i - u + y_{ik}) - \left[\left(\frac{\tilde{M}_i - u + y_{ik}}{c}\right)^\tau - \left(\frac{\tilde{M}_i - u}{c}\right)^\tau\right] \right], \quad (39)$$

which can be maximised numerically to obtain the ML-estimates \hat{c} and $\hat{\tau}$.

6 Seasonality estimation

Having fitted distributions to N and X_1, X_2, \dots , we now set out to model the seasonal variation of claim occurrence rates. It is assumed in this section that real-life claims arrive according to a non-homogeneous Poisson process $N(t)$, with some unknown intensity function $\lambda(t)$.

Our goal in this section is to estimate $\lambda(t)$. For this purpose, we will use kernel smoothing, in the form of local linear regression, and the next subsection will be devoted to presenting this method in general terms. In the section thereafter, we demonstrate how it can be applied to our data.

6.1 Kernel smoothing: Local linear regression

For smoothing purposes, we employ local linear regression. This involves fitting a unique weighted linear regression model in the near vicinity of every $x \in [\min_i(x_i), \max_i(x_i)]$ that we wish to evaluate, where $\{(x_i, y_i)\}$ are our observations, and using the value of that regression line at x as a smoothed estimate of $y(x)$.

This model can be expressed as

$$\underline{y} = \mathbf{A}_x \underline{\theta}_x + \underline{\epsilon} \quad (40)$$

where

$$\underline{y} = (y_1, y_2, \dots, y_n)^T \quad (41)$$

$$\underline{\theta}_x = (\alpha_x, \beta_x)^T \quad (42)$$

$$\mathbf{A}_x = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 - x & x_2 - x & \cdots & x_n - x \end{pmatrix}^T. \quad (43)$$

We make no assumptions about the error term $\underline{\epsilon}$. Note that the fitted regression line is only evaluated at the centre point x — a new regression model has to be fitted for every new value of x .

As weights, we use a kernel function $K(u)$, i.e. a non-negative, continuous, bounded and symmetric function, such that

$$\int_{-\infty}^{\infty} K(u) du = 1. \quad (44)$$

There are many choices of kernel functions, but the Epanechnikov kernel

$$K(u) = \frac{3}{4} (1 - u^2) \mathbf{1}_{\{|u| \leq 1\}} \quad (45)$$

has been shown to be optimal and will be used henceforth. (Optimal in this case means that it minimises the asymptotic mean integrated squared error, AMISE. See Wand and Jones (1995, Chapter 2) for more details.) From

(45), we see that $K(u) = 0$ for $|u| > 1$, but we want to be able to select our own regression window $(-h, h)$. This can be achieved by rescaling the kernel in the following manner

$$K_h(u) = K\left(\frac{u}{h}\right) \frac{1}{h}. \quad (46)$$

The value h is called the window width, or bandwidth, and specifies the length of (half) the interval around x , from which we include data points in the regression. h is chosen by the user, and determines the level of smoothness of the resulting curve. A large value of h will result in a smooth curve, but comes at the expense of detail; the opposite is true for small values of h . Another way of seeing it is that a large h produces an estimate with large bias, but small variance; again, the opposite holds for small h . Choosing a suitable h is rather tricky, and we will return to the subject shortly.

It should be noted that when x approaches $\min_i(x_i)$ or $\max_i(x_i)$, part of the regression window $(x - h, x + h)$ will overspill the boundaries, which means that the number of observations included when fitting the model is likely to decrease. While it can be shown (Wand and Jones (1995, Chapter 5.5)) that the local linear regression model is free of boundary bias (i.e. the bias near the boundaries are of the same magnitude as elsewhere on the curve), the smaller and more contracted set of included observations will result in greater variance.

With the weights specified by (45) and (46), the least squares problem translates into minimising

$$\sum_{i=1}^n K_h(x_i - x) [y_i - \alpha_x - \beta_x(x_i - x)]^2. \quad (47)$$

General theory on linear regression states that the vector $\hat{\underline{\theta}}_x = (\hat{\alpha}_x, \hat{\beta}_x)^T$ that minimises (47) is given by

$$\hat{\underline{\theta}}_x = (\mathbf{A}_x^T \mathbf{W}_x \mathbf{A}_x)^{-1} \mathbf{A}_x^T \mathbf{W}_x \underline{y} \quad (48)$$

with

$$\mathbf{W}_x = \text{diag} \{K_h(x_1 - x), K_h(x_2 - x), \dots, K_h(x_n - x)\} \quad (49)$$

and the other components given by (41), (42) and (43).

6.1.1 Bandwidth selection

In reality, the choice of kernel function is of secondary importance, a fact discussed in Wand and Jones (1995, Chapter 2.7). What really matters is the bandwidth h , i.e. the length of the interval around each evaluation point x , from which to include data points in the regression. As a result, the subject

of bandwidth selection has been studied quite extensively in the past, and several methods for automatic bandwidth selection have been proposed.

Unfortunately, none of the suggested methods appear to be optimal in all circumstances. One particular class, using so called plug-in estimators, have often been touted in the literature, ahead of the older and less sophisticated methods of cross-validation. The plug-in methods come down to finding an analytic expression for the bandwidth that minimises the asymptotic mean integrated squared error (AMISE), and replacing unknown components with estimates. This method was applied to local regression models by Ruppert et al. (1995). Cross-validation involves subdividing the data into two separate sets; the first one is used to fit the model and the second is used to test its power of prediction.

The criticism of the cross-validation approaches states that they tend to undersmooth the curves, by choosing too small an h . But Loader (1999) showed that the evidence provided in favour of plug-in methods is highly disputable, and that the cross-validation methods often performed better. Similar conclusions were reached by Lee and Solo (1999) in a simulation study.

The cross-validation method was tested whilst writing this thesis, but was not satisfactory. As a consequence, the choice of bandwidth was only assessed graphically. (It made no sense to implement a more complicated method, with only one curve to smooth.)

6.2 Estimating $\lambda(t)$

In our applications, all sets of observations are of the form $\{(i, y_i) : i = 1, \dots, 365\}$ and represent daily values; hence, the data points are equidistant. Furthermore, our one-year periodicity requirement means that we can handle the problem of increased variance near the boundaries 1 and 365, by defining

$$z_{i+n365} = y_i, \quad i = 1, \dots, 365, \quad n \in \mathbb{Z} \quad (50)$$

and applying the local regression model to the extended set of observations $\{(i, z_i) : i \in \mathbb{Z}\}$.

With these modifications, the local regression model, centred around day i , can be written

$$z_{i+\Delta} = \alpha_i + \beta_i \Delta + \epsilon_{i,\Delta}, \quad \Delta = -[h], \dots, [h] \quad (51)$$

and the least-squares estimates $\hat{\theta}_i = (\hat{\alpha}_i, \hat{\beta}_i)^T$ are given by

$$\hat{\theta}_i = (\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{W} z_i \quad (52)$$

where

$$\hat{z}_i = (z_{i-[h]}, z_{i-[h]+1}, \dots, z_{i+[h]})^T \quad (53)$$

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ -[h] & -[h] + 1 & \dots & [h] \end{pmatrix}^T \quad (54)$$

$$\mathbf{W} = \text{diag} \{K_h(-[h]), K_h(-[h] + 1), \dots, K_h([h])\}, \quad (55)$$

that is, the design matrix \mathbf{A} and weight matrix \mathbf{W} remain the same for all i , and they are reduced in size, from 365×2 and 365×365 to $(2h + 1) \times 2$ and $(2h + 1) \times (2h + 1)$, respectively. This should speed up computations somewhat, since the model has to be fitted one time for each $i = 1, \dots, 365$.

Our smoothed estimates \hat{z}_i can also be calculated directly, without performing the matrix operations of (52), by noting that

$$\hat{z}_i = \hat{\alpha}_i + \hat{\beta}_i \cdot 0 = \frac{\sum_{\Delta=-[h]}^{[h]} K_h(\Delta) z_{i+\Delta}}{\sum_{\Delta=-[h]}^{[h]} K_h(\Delta)}, \quad (56)$$

a weighted mean of the observed values $\{z_{i-[h]}, \dots, z_{i+[h]}\}$.

It is also worth noting that this smoothing procedure preserves the sum of the observations, i.e.

$$\begin{aligned} \sum_{i=1}^{365} \hat{z}_i &= \sum_{i=1}^{365} \hat{\alpha}_i = [1 \quad 0] (\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{W} \begin{pmatrix} \sum_{i=1}^{365} z_i \\ \vdots \\ \sum_{i=1}^{365} z_i \end{pmatrix} \\ &= [1 \quad 0] (\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{W} \begin{bmatrix} \sum_{i=1}^{365} z_i \\ \vdots \\ \sum_{i=1}^{365} z_i \end{bmatrix} \\ &= [1 \quad 0] (\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{W} \mathbf{1} \sum_{i=1}^{365} z_i \\ &= [1 \quad 0] (\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{W} \mathbf{A} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \sum_{i=1}^{365} z_i \\ &= \sum_{i=1}^{365} z_i \end{aligned}$$

As a result, we do not have to worry about normalising the data after smoothing it.

We now turn to estimating $\lambda(t)$, which is regarded as a piecewise constant, periodic function, with period length of one year. t is a continuous time parameter, with time-units of one year, but we will temporarily switch over to discrete time, with $k = 1, \dots, 365$ describing the calendar day of events.

We assume that we have obtained at least one external data set of events, with seasonality (believed to be) similar to that of our own claim data. From these external observations, we form an empirical discrete density function π_k , defined by

$$\pi_k = \frac{\# \text{ events on calendar day } k \text{ in the ext. dataset}}{\text{total no. of events in the external dataset}}, \quad k = 1, \dots, 365. \quad (57)$$

Alternatively, we could form linear combinations $\pi_k = \sum_i v_i \pi_{i,k}$ of several external data sets, with $\sum_i v_i = 1$, if, for example, our cedant is exposed to more than one risk and/or market. (RMS simulations are carried out one market and peril at a time.) Regardless of which, we apply local linear regression to π_k , to get a smoothed (but still discrete) density $\tilde{\pi}_k$.

Finally, we return to continuous time, by defining

$$\lambda(t) = \hat{\lambda} \tilde{\pi}_{1+(\lfloor 365 t \rfloor \bmod 365)}, \quad t \geq 0, \quad (58)$$

where $\hat{\lambda}$ is the fitted Poisson parameter from Section 5.

7 Timeline simulation

The timeline simulations can be performed in various different ways, and we will look at two of the options in this section. The first approach assumes that there is no causality between events and that every parameter except the claim intensity is held constant; this allows us to simplify things considerably. The second method involves simulating the trajectories of a non-homogeneous Poisson process (NHPP) using time transformation.

For simplicity, we will assume that each simulation round spans one full calendar year, but both methods can easily be generalised to allow longer, or shorter, time intervals.

7.1 No causality or time-dependence

If we do not intend to have past events influence the future, for example by adjusting parameter values in response to each new claim that occur, and if we regard all parameters except the Poisson intensity as constant, there is no need to simulate the full trajectories of a (mixed) non-homogeneous Poisson process. Instead of simulating the time to the next event, we

1. simulate the number of claims N_j in each simulation round j , from a Poisson or Negative binomial distribution,
2. distribute the N_j claims among days $k = 1, \dots, 365$, by simulating the daily claim counts N_{jk} from a Multinomial($N_j, \tilde{\pi}$) distribution, where $\tilde{\pi} = (\tilde{\pi}_1, \dots, \tilde{\pi}_{365})^T$, the smoothed discrete seasonal density on vector form.

7.2 NHPP simulation

Simulating the sample paths of a non-homogeneous Poisson process (NHPP), however, does allow causality and time-dependent parameters. The easiest way of doing it is via time transformation.

We denote our claim arrival process by $N(t)$, an NHPP with intensity function $\lambda(t)$ given by (58). $\lambda(t)$ is a right continuous step function, and we obtain the mean value function $\mu(t)$ by integrating it

$$\mu(t) = \int_0^t \lambda(s) ds = \hat{\lambda} \int_0^t \tilde{\pi}_{1+(\lfloor 365s \rfloor \bmod 365)} ds, \quad t \geq 0. \quad (59)$$

$\mu(t)$ thus defined is unbounded, piecewise linear, continuous and increasing (and therefore invertible). From Mikosch (2004) (see Appendix A.2), we know that

$$N(\mu^{-1}(t)) \stackrel{d}{=} \tilde{N}(t), \quad t \geq 0, \quad (60)$$

where \tilde{N} is a standard homogeneous Poisson process, and the arrival times of N are therefore given by

$$T_n = \mu^{-1}(\tilde{T}_n), \quad \tilde{T}_n = \tilde{W}_1 + \dots + \tilde{W}_n, \quad n \geq 1, \quad (61)$$

where $\tilde{W}_i \sim \text{Exp}(1)$, i.i.d. Each realisation of the NHPP $N(t)$, up to some time point $t = T_{\text{end}}$ can therefore be generated by the pseudocode

1. Initialise the counter $n = 1$.
2. Initialise $\tilde{T}_n = -\log(U_n)$, where $U_n \sim U(0, 1)$.
3. Loop while $\tilde{T}_n < \mu(T_{\text{end}})$.
 - (a) Set $T_n = \inf \{t : \mu(t) \geq \tilde{T}_n\}$.
 - (b) Set $n = n + 1$.
 - (c) Set $\tilde{T}_n = \tilde{T}_{n-1} - \log(U_n)$, where $U_n \sim U(0, 1)$.

The speed of the algorithm above will be determined by step 3(a), that is, how easy it is to invert μ . We are only concerned with claim *dates*, not what time of day they arrive, and can treat T_n as integers. In that case, we only have to keep track of the values of μ at the end of each day, i.e. μ_k , $k = 1, 2, \dots$. With μ_k stored in a vector `mu`, the arrival times T_n in step 3(a) can quickly and easily be obtained with a single line of Matlab code:

```
Tn = find(mu.*(mu>=Tn_tilde),1,'first');
```

7.2.1 Mixed NHPP

The above procedure can be modified to work for mixed NHPP as well, simply by redefining $\mu(t)$. If we set

$$\mu(t) = \theta \rho(t) = \theta \int_0^t \tilde{\pi}_{1+(\lfloor 365 s \rfloor \bmod 365)} ds, \quad t \geq 0, \quad (62)$$

and sample a new structure variable $\theta \sim \Gamma(\alpha, \beta)$ ahead of each realisation, then the number of claims in each simulation round will be an observation from a negative binomial distribution, with parameters α and $\beta/\rho(t)$. Individual sample paths will not be any different from those of a regular NHPP, but the randomness of $\mu(t)$ will result in different trajectories for each realisation, which combine to form the overdispersion associated with the negative binomial distribution.

8 Case study: European property insurance

Finally, we apply the methods described in the previous sections to some real data, in the form of a claim record from a European property insurer. The data set was described in Section 2, and is available for display in Appendix B (after an anonymising linear transformation).

The contract we have in mind is of the type '10M xs 5M' without reinstatements, and will cover losses that occur between October 1 and December 31. Such short-term contracts are not uncommon during the latter stages of the fiscal year, when cedants might be looking to secure any profits made in the preceding months.

8.1 Parameter estimation

Step one, when fitting a POT-model, is to choose a threshold u . In reality, u will largely be dictated by the type of contract one is about to price. In our case, u must fall below the excess point 5 000 000, but we can select it freely among all such values.

A good starting point is the empirical mean excess plot of our observed losses. The empirical ME-function of a set $\{x_i\}$ is given by

$$e_n(u) = \frac{\sum_{i=1}^n \max(0, x_i - u)}{\sum_{i=1}^n \mathbf{1}_{\{x_i > u\}}}, \quad (63)$$

and is the empirical equivalent of the theoretical ME-function

$$e_F(u) = E[X - u | X > u]. \quad (64)$$

The theoretical ME-functions of the relevant severity distributions are provided in Table 3, where the exponential distribution also has been added for comparison. We see that $e_F(u)$ is constant for the exponential distribution, but increases linearly for the Pareto and Burr distributions. $e_F(u)$ is strictly increasing for Weibull as well (provided $\tau < 1$), but tends to infinity at a lower pace than the ME-functions of Pareto and Burr. It tends to infinity for the Log-Normal distribution too, but has a singularity at $u = e^\mu$. (Generally, $\lim_{u \rightarrow \infty} e_F(u) \rightarrow \infty$ for heavy-tailed distributions, whereas $\lim_{u \rightarrow \infty} e_F(u) \rightarrow 0$ for distributions that are lighter-tailed than the exponential distribution.) A more extensive list of mean excess functions is provided in Mikosch (2004, Chapter 3.2.3).

The point of plotting the empirical ME-functions against u is to see if we can discern any of the functions of Table 3 in our observed data; if, for example, $e_n(u)$ appears to increase linearly past some point $u = u_0$, then u_0 would be a suitable truncation point for fitting a POT-model with Pareto (or Burr) distributed claims.

The empirical ME-function, applied to our observed ground-up claims $\{x_{ij} : i = 1, \dots, 11, j = n_i\}$, is plotted in Figure 1, which, incidentally, also

<i>Distribution</i>	$e_F(u)$
Exponential	λ^{-1}
Pareto	$\frac{\theta+u}{\alpha-1}, \quad \alpha > 1$
Burr	$\frac{u}{\alpha\tau-1} (1 + o(1)), \quad \alpha\tau > 1$
Log-Normal	$\frac{\sigma^2 u}{\log u - \mu} (1 + o(1))$
Weibull	$\frac{u^{1-\tau} c^\tau}{\tau} (1 + o(1))$

Table 3: Mean excess functions of our severity distributions.

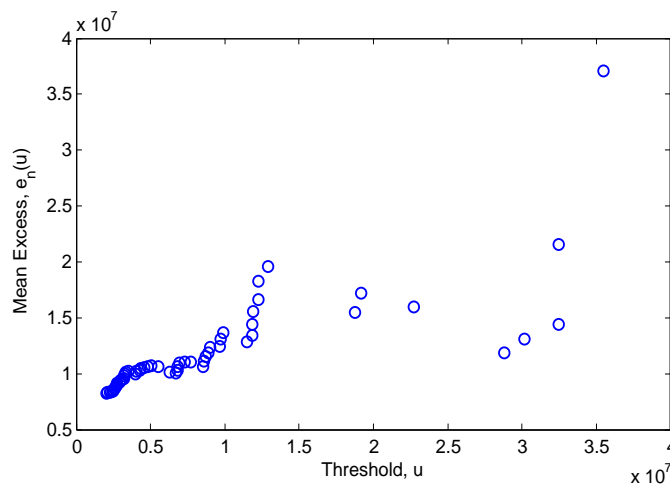


Figure 1: Mean excess plot of our data.

highlights the problems with mean excess plots. Our data were sparse to begin with, and as we increase u , there are few points left from which to form the mean values. The resulting values of $e_n(u)$ are anything but robust — one additional extra large claim could offset the apparent dip between $1.8 \cdot 10^7 \leq u \leq 3.3 \cdot 10^7$ entirely — and Figure 1 alone is no reason to dismiss the distributions of Table 3.

Instead, we zoom in on the more stable points at $u \leq 1.5 \cdot 10^7$. The plot in Figure 2 reveals that $e_n(u)$ appears to change trajectory around $u = 4 \cdot 10^6$, and this could be an interesting truncation point for the POT-model. In addition, we shall consider the cases $u = \min_i(M_i) = 2000000$ and $u = \max_i(M_i) = 2462963$ as well; the first uses all available data points, whereas the second is the lowest truncation point such that the thinning parameters $p_i = 1 - F(\tilde{M}_i - u)$ (see Subsection 5.2) are all equal to 1. (The consequence of having all $p_i = 1$ is that the claim frequency distribution will no longer be dependent on the severity distribution F .)

We see that $e_n(u)$ is hardly linear from $u = \max_i(M_i)$ and up, so we expect some distribution other than Pareto to provide the best fit when using that threshold (or a lower one) in the POT-model. In fact, comparing the

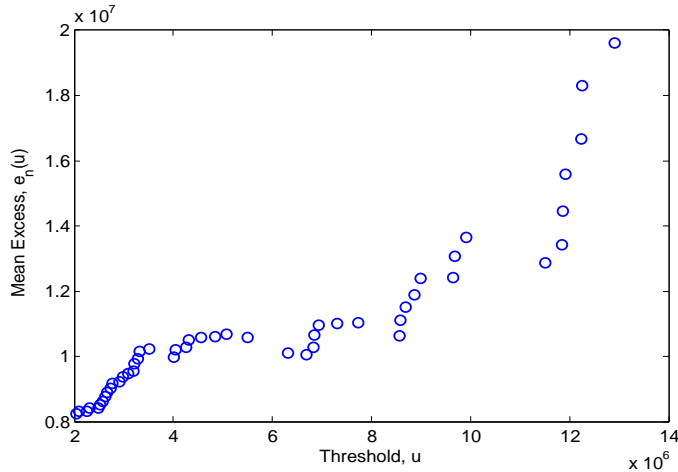


Figure 2: Mean excess plot of our data, zoomed in.

functions of Table 3 with Figure 2 would suggest that a Weibull distribution, with $\tau < 1$, is more appropriate. As for $u = 4 \cdot 10^6$, we cannot rule out a linear relation beyond that point, since the rightmost observations are quite uncertain, and the seeming change of slope around $u = 11 \cdot 10^6$ might only be due to the scarcity of data. Hence, $u = 4 \cdot 10^6$ could very well be a suitable truncation point for fitting a Pareto (or Burr) distribution in the POT framework. Anyway, we shall proceed to fit the POT-model at each of these three truncation points, using all four severity distributions under consideration.

Note that we have overlooked the fact that claims from different years actually have different distributions, since they were sampled from left-truncated distributions, with different truncation points M_i in force each year. However, the differences between the M_i :s are slim when compared to the actual claim values x_{ij} (see Appendix B), and we have chosen to disregard it for now.

But the POT-model, described in Section 5, does take varying truncation points M_i into account. Applying it to our data, with the thresholds u specified above, yields the parameter estimates of Appendix C. Only the Poisson case has been included, since stable estimates of the Negative binomial parameters could not be obtained. (This is because the mean value of $v_i n_i$ exceeds the sample variance, which meant that the ML estimates $\hat{\alpha}$ and $\hat{\beta}$ tended to infinity, and the corresponding negative binomial distribution converged to a Poisson distribution, with intensity $\hat{\alpha}/\hat{\beta}$.)

The tables in Appendix C would suggest that the Weibull distribution gives the best fit for all three truncation points, but it is difficult to say what slightly lower values of AIC and BIC mean in practice. It might therefore be illuminating to plot the fitted cumulative distribution functions against the

empirical ones. Figure 3 depicts the case where $u = 2000000$, and apart from the fitted Log-Normal distribution failing to capture the tail behaviour of the data, none of the other three proposed distributions really stand out. (The fitted Burr curve coincided almost perfectly with the Weibull curve, hence the difficulty of seeing it in the figure. Also, remember that the empirical c.d.f. in Figure 3 is based on observations from distributions with slightly different truncation points M_i , and the empirical probability assigned to the interval $[\min_i(M_i), \max_i(M_i)]$ could be somewhat understated. But the same pattern emerged with $u = 2462963$ and $u = 4000000$, with the only difference being that the relative shortcomings of the Log-Normal distribution were more pronounced in the latter case.)

But one thing that do set the different models apart, in the case when $u = 2000000$, is the estimated Poisson intensity $\hat{\lambda}$; as already mentioned, this is because the Poisson thinning parameters $p_i = 1 - F(\tilde{M}_i - u)$ depend on the fitted severity distributions F . Figure 4 is a zoomed in version of Figure 3, where we have restricted ourselves to $x \in [\min_i(M_i), \max_i(M_i)]$. It is clear that the p_i parameters of the Burr and Weibull distributions are different from those of the other two, but we have no way of telling which ones are more correct. This fact is a source of concern, and we could argue that if the severity distributions F provide a poor fit at that range, it is best not to have them affect the frequency distributions at all. In our case, where the truncation points M_i do not differ much between years, raising the threshold u from $\min_i(M_i) = 2000000$ to $\max_i(M_i) = 2462963$ will only result in a loss of 4 observations, and it is probably wise to do so, in order to achieve independence between the frequency and severity distributions.

Having $u = 4000000$, on the other hand, means losing as many as 19 observations, but we shall consider that case anyway, since the ME plots in Figures 1 and 2 hinted that the characteristics of the severity distribution are altered beyond that point.

All in all, the Burr and Weibull distributions do seem to fit the data about equally well for all three truncation points, but the Weibull distribution has one less parameter and is therefore preferable. However, raising the threshold to $u = 4000000$ means that the Pareto distribution achieves a similar level of fit, just as we predicted whilst looking at the ME plots. The Log-normal distribution, on the other hand, is inadequate irrespective of truncation point.

8.2 Seasonality estimation

As seasonal density π_k , we combine the empirical discrete seasonal densities obtained from the external data sets, covering European wind storms and floods. For simplicity, we use the mixing weights (0.5, 0.5). Again, this is just for illustrative purposes; a real life situation would call for a more informed choice.

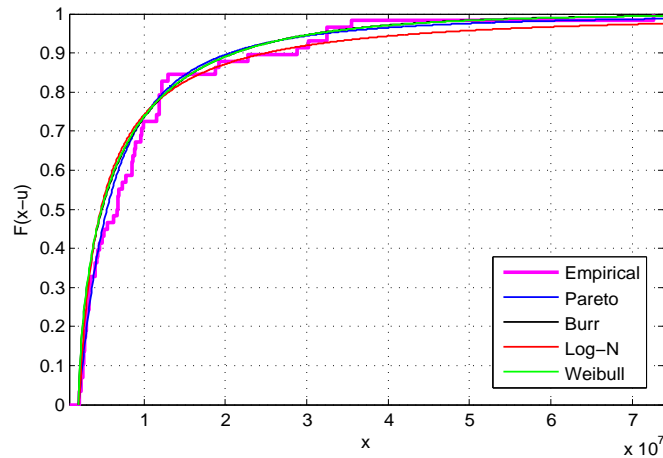


Figure 3: Empirical and fitted c.d.f., $u = 2000000$.

Some trial and error (and subjective judgement) reveal that $h = 15.5$ is a suitable bandwidth for local linear regression in this case. The resulting density $\tilde{\pi}_k$ can be seen in the upper plot of Figure 5, where we have also included the non-smoothed density; the spikes in the lower plot of Figure 5 represent the occurrence dates of the observed claims in our own claim record.

It is hard to say much about the distribution of the observed claims, but they do appear to occur more uniformly than $\tilde{\pi}_k$ would suggest. A great deal of the time spent working on this thesis was devoted to finding a way of adjusting the chosen $\tilde{\pi}_k$, according to the observed claim occurrence dates. We did have some success with multidimensional credibility models (see Bühlmann and Gisler (2005, Chapter 7)), but every attempt at determining the credibility weights remained heuristic, and ultimately it was decided to leave it out of the thesis. Still, the credibility models could be worth considering in practical, real-life situations, where heuristics need not be shunned.

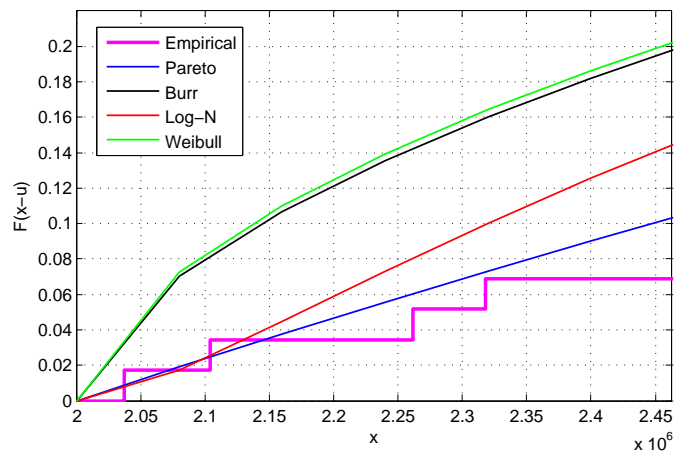


Figure 4: Empirical and fitted c.d.f., zoomed in, $u = 2000000$.

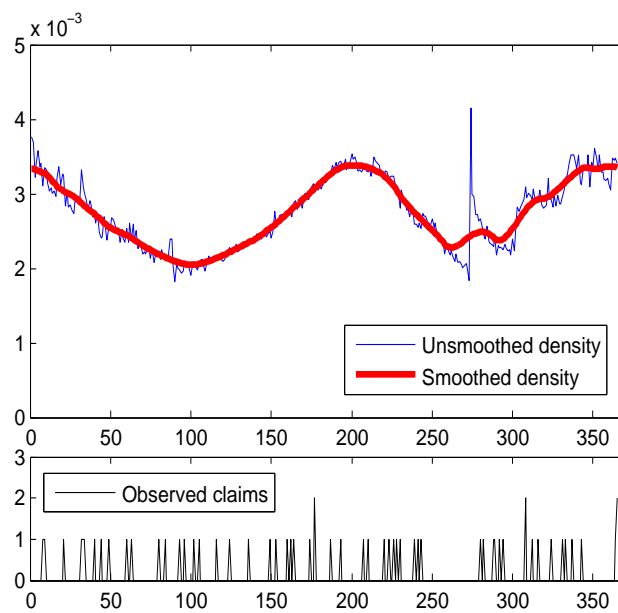


Figure 5: *Upper*: Smoothed and non-smoothed seasonal density functions. *Lower*: Occurrence dates of the observed claims.

8.3 Timeline simulation

Having fitted the frequency and severity distributions, and estimated the seasonality, we now set out to model the losses resulting from our contract. The total annual loss is given by

$$S = \sum_{i=1}^{N^*} Y_i, \quad (65)$$

where N^* is the number of losses that occur during the contractual time-frame, and

$$Y_i = g(X_i) = \min(10000000, \max(X_i - 5000000, 0)), \quad (66)$$

is the payout function of our XL contract.

200 000 simulation rounds, using the NHPP algorithm of Section 7.2 and all the fitted distributions in Appendix C, produced the results presented in Tables 4, 5 and 6. \bar{N}^* , \bar{Y} and \bar{S} are the arithmetic mean values of N^* , Y and S over all simulations, and $s_{N^*}^2$, s_Y^2 and s_S^2 are the corresponding sample variances. It was found that 200 000 simulation rounds were enough to provide \bar{N}^* with two (more or less) stable decimals, but only one for the other mean values; the need for a massive number of iterations is one of the problems with Monte Carlo timeline simulation.

We have plotted the average number of claims per day and simulation round in Figure 6, along with the intensity function $\lambda(t)$; note the matching seasonality.

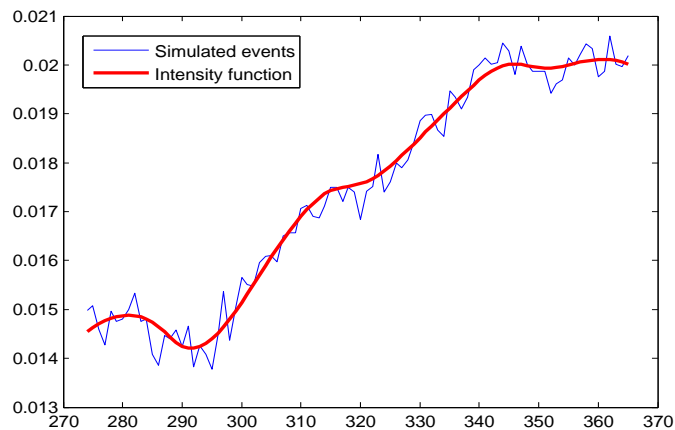


Figure 6: Simulated events from October 1 to December 31.

The premium is generally calculated by scaling up $E[S]$, and we use \bar{S} as an approximation of this value. Hence, \bar{S} is the most important quantity

in the tables. (The scale factor is added to cover administrative costs, add profit and security margins, compensate for the effect this particular contract will have on the reinsurer's capital requirements, etc.)

Starting with the case where $u = 2000000$ (Table 4), we see that the notable differences between the values of \bar{S} are solely due to differences in \bar{N}^* , the average number of claims per simulation round, and the latter are direct consequences of the models having different Poisson intensity estimates $\hat{\lambda}$. We have already established that this is undesirable in our case, and we should be moving on to $u = 2462963$, but not before noting that the best fitting severity distributions (Weibull and Burr) had \bar{S} down to $4.9 \cdot 10^6$.

With $u = 2462963$ and $u = 4000000$ on the other hand (Tables 5 and 6), the best fitting severity distributions, (Burr, Weibull) and (Burr, Pareto, Weibull) respectively, all produced nearly the same value: $\bar{S} = 5.1 \cdot 10^6$. We can regard it as our best approximation of $E[S]$, and the premium should be set accordingly.

8.3.1 Some remarks

The computationally heavy Monte Carlo timeline simulation approach, using the NHPP algorithm, is intended for models far more complex than the one we have considered, and it made no sense to use it here. We could just as well have calculated the exact value of $E[S]$, using the relation

$$E[S] = E[N^*] E[Y], \tag{67}$$

or better yet, obtained the distribution of S through Ströter recursion or some approximate method, see Johansson (1997, Chapter 5). But it was intended merely as an illustrative example, and the same method can be applied even if we expand the model, for example by introducing features such as event causality, exact discounting using stochastic interest rates, and incremental payouts at stochastic time intervals.

It is also important to be aware of the need for exceptionally many iterations that come with simulating heavy-tailed distributions. This is because the large, but rare events at the far end of the tail can have a massive impact, and we need to make sure that the number of such observations in our simulated sample is not disproportionate. But excess of loss contracts with a fixed upper limit are immune to this problem, since the payouts from them are capped anyway. It is probably wise to enforce a fixed upper limit to every payout when performing Monte Carlo simulations, even if the contract itself specifies no such amount; if nothing else, we could set the maximum value at the level of the largest sum insured (minus deductibles, etc.)

<i>Quantity</i>	<i>Pareto</i>	<i>Burr</i>	<i>Log-Normal</i>	<i>Weibull</i>
\bar{N}^*	1.60	1.73	1.63	1.73
$s_{N^*}^2$	1.61	1.72	1.63	1.72
\bar{Y}	$2.9 \cdot 10^6$	$2.8 \cdot 10^6$	$2.8 \cdot 10^6$	$2.8 \cdot 10^6$
s_Y^2	$1.5 \cdot 10^{13}$	$1.5 \cdot 10^{13}$	$1.5 \cdot 10^{13}$	$1.5 \cdot 10^{13}$
\bar{S}	$4.6 \cdot 10^6$	$4.9 \cdot 10^6$	$4.5 \cdot 10^6$	$4.9 \cdot 10^6$
s_S^2	$3.7 \cdot 10^{13}$	$4.0 \cdot 10^{13}$	$3.8 \cdot 10^{13}$	$4.0 \cdot 10^{13}$

Table 4: Results of the simulations when $u = 2000000$.

<i>Quantity</i>	<i>Pareto</i>	<i>Burr</i>	<i>Log-Normal</i>	<i>Weibull</i>
\bar{N}^*	1.42	1.43	1.43	1.43
$s_{N^*}^2$	1.42	1.43	1.43	1.43
\bar{Y}	$3.3 \cdot 10^6$	$3.5 \cdot 10^6$	$3.2 \cdot 10^6$	$3.6 \cdot 10^6$
s_Y^2	$1.5 \cdot 10^{13}$	$1.6 \cdot 10^{13}$	$1.7 \cdot 10^{13}$	$1.7 \cdot 10^{13}$
\bar{S}	$4.7 \cdot 10^6$	$5.0 \cdot 10^6$	$4.5 \cdot 10^6$	$5.1 \cdot 10^6$
s_S^2	$3.8 \cdot 10^{13}$	$4.1 \cdot 10^{13}$	$3.8 \cdot 10^{13}$	$4.2 \cdot 10^{13}$

Table 5: Results of the simulations when $u = 2462963$.

<i>Quantity</i>	<i>Pareto</i>	<i>Burr</i>	<i>Log-Normal</i>	<i>Weibull</i>
\bar{N}^*	1.03	1.03	1.03	1.04
$s_{N^*}^2$	1.04	1.03	1.03	1.03
\bar{Y}	$4.9 \cdot 10^6$	$4.9 \cdot 10^6$	$4.4 \cdot 10^6$	$4.9 \cdot 10^6$
s_Y^2	$1.6 \cdot 10^{13}$	$1.6 \cdot 10^{13}$	$1.7 \cdot 10^{13}$	$1.7 \cdot 10^{13}$
\bar{S}	$5.1 \cdot 10^6$	$5.1 \cdot 10^6$	$4.6 \cdot 10^6$	$5.1 \cdot 10^6$
s_S^2	$4.1 \cdot 10^{13}$	$4.2 \cdot 10^{13}$	$3.8 \cdot 10^{13}$	$4.2 \cdot 10^{13}$

Table 6: Results of the simulations when $u = 4000000$.

9 Conclusions

We have looked at a general way of pricing reinsurance contracts using Monte Carlo timeline simulation, and we applied it to a sample problem of pricing a short term excess of loss contract. In most insurance applications, we focus on modelling the annual number of claims, but with a short term contract, we also have to take seasonal variation into account.

External data sets were used to estimate the seasonality, since our own data set on historical losses was insufficient for the purpose. This meant that we had to put some effort into finding external data sets that exhibited a seasonal variation (believed to be) similar to that of our own data. A logical continuation to this approach would be to find a way of modifying this estimated seasonality, according to the occurrence dates of our own observed claims, however few they are. We did experiment with multidimensional credibility models in an attempt to do this, but we were unable to determine the required credibility weights in a satisfactory way.

The Peaks Over Threshold model was used for parameter estimation, and we extended it to allow more distributions than the usual Poisson-Pareto combination. This proved to be a good idea, since both the Burr and Weibull distribution provided a better fit than the Pareto distribution did. Parameter estimation is considerably more complicated for the new severity distributions, since neither of them are stable under increasing thresholds, but we were able to sidestep this problem by feeding a custom log-likelihood into Matlab's `mle` procedure, and supplying some carefully chosen start values.

The timeline simulation approach to pricing was found to be painfully slow, mostly because it required hundreds of thousands of iterations to give us reasonably stable estimates of $E[S]$, the expected total loss during the contract period. In the case of very simplistic models, like the one we considered in our case study, there is no benefit to be had from using the method, and it ought to be our last choice. But in more complex models, where we incorporate features such as causality between events and exact discounting with stochastic interest rates, there is no other way of estimating the expected total loss, save for pure guesswork.

We did model the claim arrivals with a standard non-homogeneous Poisson process (NHPP), rather than a mixed one, simply because the variance-to-mean ratio of the annual claim counts in our dataset was lower than one, thus making the Gamma-mixed NHPP and its negative binomial distributed claim counts unfit for the task. But the low variance could simply be a result of us having so few observation years (and having had the luck to avoid years with unusually many claims till now). A useful extension would be to replace the standard NHPP with a mixed NHPP, and let the user specify a variance-to-mean ratio of his/her own liking.

A Appendix: Properties of the Poisson process

Poisson and mixed Poisson processes are used in this thesis, and we state here some relevant definitions, theorems and general properties that apply to them. The following section has been adapted mainly from Grandell (1997) and Mikosch (2004, Chapter 2), and we refer to those books for proofs and a much more exhaustive treatment of the subject.

A.1 The Poisson distribution

A random variable N that is Poisson distributed has the probability function

$$P(N = n) = \frac{\lambda^n}{n!} e^{-\lambda}, \quad n = 0, 1, \dots \quad (68)$$

The intensity λ is a positive, real-valued constant, and it turns out that $E[N] = Var(N) = \lambda$. If we replace λ by a random variable Λ , with distribution U , we get the probability function

$$P(N = n) = \int_{0-}^{\infty} \frac{\lambda^n}{n!} e^{-\lambda} dU(\lambda), \quad n = 0, 1, \dots \quad (69)$$

This is known as a mixed Poisson distribution, with structure distribution U . Its mean and variance are

$$E[N] = E[\Lambda], \quad Var(N) = E[\Lambda] + Var(\Lambda) \geq E[N]$$

i.e. the variance is greater than in the regular Poisson case. The structure distribution adds extra uncertainty, which results in overdispersion, or variance greater than the mean. A popular choice of U in the world of insurance is the Gamma distribution, with density function

$$u(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}, \quad \lambda \geq 0. \quad (70)$$

The resulting mixed Poisson distribution is the familiar Negative Binomial distribution, with probability function

$$P(N = n) = \frac{\Gamma(\alpha + n)}{\Gamma(\alpha)n!} \left(\frac{\beta}{\beta + 1}\right)^\alpha \left(\frac{1}{\beta + 1}\right)^n. \quad (71)$$

A.2 Definition of the (mixed) Poisson process

The Poisson process (cf. Def. 2.1.1 in Mikosch(2004))

A stochastic process $\{N(t), t \geq 0\}$ is called a (non-homogeneous) Poisson process if

1. $N(0) = 0$ a.s.

2. The process has independent increments.
3. There exists a non-decreasing right-continuous function $\mu : [0, \infty) \rightarrow [0, \infty)$, with $\mu(0) = 0$, such that the increments $N(s, t]$ for $0 \leq s < t < \infty$ have a $\text{Poisson}(\mu(s, t])$ distribution. We call μ the mean value function of N .
4. With probability 1, the sample paths $\{N(t, \omega), t \geq 0\}$ of N are right-continuous for $t \geq 0$ and have limits from the left for $t > 0$. We say that N has càdlàg sample paths.

We say that N has an intensity function $\lambda(\cdot)$ if, for any $s < t$, $\mu(s, t]$ has the representation

$$\mu(s, t] = \int_s^t \lambda(y) dy, \quad s < t, \quad (72)$$

for some non-negative measurable $\lambda(\cdot)$. We write $\mu(0, t] = \mu(t)$.

If $\mu(s, t] = \lambda(t - s)$, then N is said to be a homogeneous Poisson process, with intensity λ . The interarrival times of a homogeneous Poisson process are $\text{Exp}(\lambda)$ distributed.

Time-transformed Poisson processes (cf. Prop. 2.1.5 in Mikosch (2004))

Let μ be the mean value function of a (non-homogeneous) Poisson process N , and let \tilde{N} be a homogeneous Poisson process with $\lambda = 1$. Then

1. The process $\{\tilde{N}(\mu(t)), t \geq 0\}$ is a Poisson process, with mean value function μ .
2. If μ is continuous, increasing, and $\lim_{t \rightarrow \infty} \mu(t) = \infty$, then $\{N(\mu^{-1}(t)), t \geq 0\}$ is a homogeneous Poisson process with $\lambda = 1$,

The arrival times of N can be written

$$T_n = \mu^{-1}(\tilde{T}_n), \quad \tilde{T}_n = \tilde{W}_1 + \dots + \tilde{W}_n, \quad n \geq 1, \quad (73)$$

where $\tilde{W}_i \sim \text{Exp}(1)$, i.i.d.

The mixed time-transformed Poisson process (cf. Def. 2.3.1 in Mikosch (2004))

Let \tilde{N} be a homogeneous Poisson process with $\lambda = 1$, and let μ be the mean value function of a (non-homogeneous) Poisson process on $[0, \infty)$. Let $\theta > 0$ a.s. be a (non-degenerate) random variable, independent of \tilde{N} . Then the process

$$N(t) = \tilde{N}(\theta \mu(t)), \quad t \geq 0, \quad (74)$$

is said to be a mixed Poisson process, with structure variable θ and time-transformation μ .

The Pólya process

If the Gamma distribution (70) is chosen as structure distribution, i.e. $\theta \sim \Gamma(\alpha, \beta)$, and we set $\mu(t) = t$, then N , defined by (74), turns into the well-known Pólya process. It follows that $N(t)$ has a negative binomial distribution with parameters α and β/t , i.e.

$$P(N(t) = n) = \frac{\Gamma(\alpha + n)}{\Gamma(\alpha)n!} \left(\frac{\beta/t}{\beta/t + 1} \right)^\alpha \left(\frac{1}{\beta/t + 1} \right)^n, \quad (75)$$

which means that

$$E[N(t)] = \frac{\alpha}{\beta}t, \quad \text{Var}(N(t)) = \frac{\alpha}{\beta}t \left(\frac{\beta + t}{\beta} \right).$$

The *state-dependent* intensity function of the Pólya process is written

$$\lambda_n(t) = \frac{\alpha + n}{\beta + t}. \quad (76)$$

B Appendix: Claim record

<i>Year</i>	<i>Month</i>	<i>Day</i>	<i>Threshold</i>	<i>Total claim amount</i>
1999	05	29	2 462 963	12 249 719
1999	08	16	2 462 963	2 975 570
1999	10	16	2 462 963	2 664 310
1999	10	19	2 462 963	4 843 643
1999	10	21	2 462 963	28 803 160
1999	12	31	2 462 963	2 581 505
2000	01	08	2 394 000	8 998 041
2000	02	29	2 394 000	19 192 770
2000	04	12	2 394 000	2 525 048
2000	11	12	2 394 000	3 506 784
2001	02	18	2 337 891	6 690 285
2001	03	21	2 337 891	6 840 724
2001	04	15	2 337 891	7 735 949
2001	07	06	2 337 891	22 735 453
2001	11	20	2 337 891	5 502 992
2002	01	21	2 284 351	12 231 508
2002	02	02	2 284 351	4 319 936
2002	06	23	2 284 351	11 920 201
2002	08	27	2 284 351	4 047 870
2002	10	15	2 284 351	18 747 798
2003	01	09	2 237 383	11 863 474
2003	02	01	2 237 383	3 315 314
2003	02	09	2 237 383	8 583 331
2003	02	13	2 237 383	72 600 645
2003	04	03	2 237 383	3 288 922
2003	06	11	2 237 383	12 915 567
2004	06	02	2 210 526	7 310 449
2004	08	14	2 210 526	4 556 231
2004	11	04	2 210 526	2 909 459

Table 7: Past claims on the cedant in our case study. The column 'threshold' lists the minimum reported claim value for each year. All values have been adjusted for inflation.

<i>Year</i>	<i>Month</i>	<i>Day</i>	<i>Threshold</i>	<i>Total claim amount</i>
2005	04	26	2 172 414	35 481 413
2005	06	09	2 172 414	2 622 755
2005	06	26	2 172 414	2 774 559
2005	07	29	2 172 414	4 013 452
2005	08	11	2 172 414	32 467 085
2005	12	30	2 172 414	6 318 109
2006	02	03	2 131 790	8 562 321
2006	03	25	2 131 790	4 261 930
2006	04	06	2 131 790	2 261 505
2006	05	16	2 131 790	8 693 243
2006	07	12	2 131 790	3 213 754
2006	08	18	2 131 790	32 453 439
2006	10	09	2 131 790	8 872 902
2006	11	04	2 131 790	11 847 422
2006	12	03	2 131 790	3 083 331
2007	05	04	2 096 322	11 513 232
2007	06	13	2 096 322	2 103 970
2007	11	08	2 096 322	6 934 739
2007	11	27	2 096 322	5 070 308
2007	11	29	2 096 322	3 204 505
2008	03	04	2 027 096	2 037 138
2008	06	26	2 027 096	9 915 595
2008	07	26	2 027 096	2 736 579
2008	10	07	2 027 096	30 172 131
2008	12	09	2 027 096	2 318 316
2008	12	31	2 027 096	9 692 615
2009	08	08	2 000 000	2 480 056
2009	08	29	2 000 000	6 822 232
2009	08	31	2 000 000	9 640 000

Table 8: Past claims, continued...

<i>Year:</i>	1999	2000	2001	2002	2003	2004
v_i	1.208	1.220	1.278	1.229	1.045	0.973
<i>Year:</i>	2005	2006	2007	2008	2009	
v_i	0.957	0.971	0.992	1.033	1.000	

Table 9: Scale factor, v_i .

C Appendix: Parameter estimates

C.1 $u = 2000000$

Pareto	<i>Est.</i>	<i>ll</i>	<i>ul</i>	Burr	<i>Est.</i>	<i>ll</i>	<i>ul</i>
$\hat{\alpha}$	1.89	0.24	3.53	$\hat{\alpha}$	2.58E6	1.55E6	3.59E6
$\hat{\theta}$	7.77E6	0	1.79E7	$\hat{\theta}$	4.35E10	3.10E10	5.60E10
				$\hat{\tau}$	0.63	0.60	0.66
<i>Poisson</i>				<i>Poisson</i>			
$\hat{\lambda}$	5.97	4.49	7.46	$\hat{\lambda}$	6.43	4.83	8.02
AIC	1855.1			AIC	1843.6		
BIC	1859.6			BIC	1850.2		
Log-N	<i>Est.</i>	<i>ll</i>	<i>ul</i>	Weibull	<i>Est.</i>	<i>ll</i>	<i>ul</i>
$\hat{\mu}$	14.81	14.29	15.33	\hat{c}	5.03E6	2.32E6	7.75E6
$\hat{\sigma}$	1.67	1.28	2.06	$\hat{\tau}$	0.62	0.44	0.80
<i>Poisson</i>				<i>Poisson</i>			
$\hat{\lambda}$	6.07	4.56	7.58	$\hat{\lambda}$	6.45	4.85	8.05
AIC	1854.7			AIC	1841.2		
BIC	1859.2			BIC	1845.7		

Table 10: Parameter estimates for the POT-model, with $u = 2000000$.

C.2 $u = 2462963$

Pareto	<i>Est.</i>	<i>ll</i>	<i>ul</i>	Burr	<i>Est.</i>	<i>ll</i>	<i>ul</i>
$\hat{\alpha}$	2.08	0.11	4.06	$\hat{\alpha}$	21.39	0	239.48
$\hat{\theta}$	9.79E6	0	2.30E7	$\hat{\theta}$	2.19E6	0	1.92E7
				$\hat{\tau}$	0.74	0.48	0.99
<i>Poisson</i>				<i>Poisson</i>			
$\hat{\lambda}$	5.31	3.95	6.68	$\hat{\lambda}$	5.31	3.95	6.68
AIC	1746.8			AIC	1745.8		
BIC	1751.2			BIC	1752.1		
Log-N	<i>Est.</i>	<i>ll</i>	<i>ul</i>	Weibull	<i>Est.</i>	<i>ll</i>	<i>ul</i>
$\hat{\mu}$	14.91	14.46	15.37	\hat{c}	6.64E6	4.03E6	9.25E6
$\hat{\sigma}$	1.72	1.39	2.04	$\hat{\tau}$	0.72	0.57	0.86
<i>Poisson</i>				<i>Poisson</i>			
$\hat{\lambda}$	5.31	3.95	6.68	$\hat{\lambda}$	5.31	3.95	6.68
AIC	1749.8			AIC	1743.8		
BIC	1754.2			BIC	1748.2		

Table 11: Parameter estimates for the POT-model, with $u = 2462963$.

C.3 $u = 4000000$

Pareto	<i>Est.</i>	<i>ll</i>	<i>ul</i>	Burr	<i>Est.</i>	<i>ll</i>	<i>ul</i>
$\hat{\alpha}$	3.17	0	7.53	$\hat{\alpha}$	25.61	0	385.78
$\hat{\theta}$	2.16E7	0	5.95E7	$\hat{\theta}$	8.85E6	0	1.03E8
				$\hat{\tau}$	0.80	0.47	1.13
<i>Poisson</i>				<i>Poisson</i>			
$\hat{\lambda}$	3.86	2.70	5.02	$\hat{\lambda}$	3.86	2.70	5.02
AIC	1305.9			AIC	1306.8		
BIC	1309.6			BIC	1312.2		
Log-N	<i>Est.</i>	<i>ll</i>	<i>ul</i>	Weibull	<i>Est.</i>	<i>ll</i>	<i>ul</i>
$\hat{\mu}$	15.21	14.67	15.74	\hat{c}	8.44E6	0.49E7	1.20E7
$\hat{\sigma}$	1.70	1.32	2.07	$\hat{\tau}$	0.78	0.59	0.97
<i>Poisson</i>				<i>Poisson</i>			
$\hat{\lambda}$	3.86	2.70	5.02	$\hat{\lambda}$	3.86	2.70	5.02
AIC	1314.3			AIC	1304.8		
BIC	1318.0			BIC	1308.5		

Table 12: Parameter estimates for the POT-model, with $u = 4000000$.

D Appendix: ML-estimation of parameters

D.1 Number of claims: Poisson

D.1.1 Truncated distribution

If

- $N_i|N = n \sim \text{Bin}(n, p_i)$
- $N \sim \text{Po}(\lambda)$

then

$$\begin{aligned}
 P(N_i = \tilde{n}_i) &= \sum_{n=\tilde{n}_i}^{\infty} P(N_i = \tilde{n}_i|N = n) P(N = n) \\
 &= \sum_{n=\tilde{n}_i}^{\infty} \binom{n}{\tilde{n}_i} p_i^{\tilde{n}_i} (1-p_i)^{n-\tilde{n}_i} \frac{\lambda^n}{n!} e^{-\lambda} \\
 &= \frac{(\lambda p_i)^{\tilde{n}_i}}{\tilde{n}_i!} e^{-\lambda} \sum_{n=\tilde{n}_i}^{\infty} \frac{[\lambda(1-p_i)]^{n-\tilde{n}_i}}{(n-\tilde{n}_i)!} \\
 &= \frac{(\lambda p_i)^{\tilde{n}_i}}{\tilde{n}_i!} e^{-\lambda} e^{\lambda(1-p_i)} = \frac{(\lambda p_i)^{\tilde{n}_i}}{\tilde{n}_i!} e^{-\lambda p_i},
 \end{aligned}$$

which means that $N_i \sim \text{Po}(\lambda p_i)$, a thinned Poisson distribution with thinning parameter p_i .

D.1.2 ML parameter estimates

$$N_i \sim \text{Po}(\lambda p_i) \quad \Leftrightarrow \quad P(N_i = \tilde{n}_i) = \frac{(\lambda p_i)^{\tilde{n}_i}}{\tilde{n}_i!} e^{-\lambda p_i}$$

$$L(\lambda) = \prod_{i=1}^m \frac{(\lambda p_i)^{\tilde{n}_i}}{\tilde{n}_i!} e^{-\lambda p_i}$$

We take the logarithm and replace \tilde{n}_i by the scaled up values $\tilde{n}_i v_i$ (see Section 2.1.1). This gives us the log-likelihood

$$l(\lambda) = \sum_{i=1}^m [\tilde{n}_i v_i \log \lambda - \lambda p_i] + \text{constant}$$

$$\frac{dl(\lambda)}{d\lambda} = \sum_{i=1}^m \left[\frac{\tilde{n}_i v_i}{\lambda} - p_i \right]$$

$$\Rightarrow \quad \hat{\lambda} = \frac{\sum_{i=1}^m \tilde{n}_i v_i}{\sum_{i=1}^m p_i} \quad (77)$$

D.2 Number of claims: NegBin

D.2.1 Truncated distribution

If

- $N_i|N = n \sim \text{Bin}(n, p_i)$
- $N \sim \text{NegBin}(\alpha, \beta)$
- $\Lambda \sim \Gamma(\alpha, \beta)$

then

$$\begin{aligned}
P(N_i = \tilde{n}_i) &= \sum_{n=\tilde{n}_i}^{\infty} P(N_i = \tilde{n}_i|N = n) P(N = n) \\
&= \int_0^{\infty} \sum_{n=\tilde{n}_i}^{\infty} P(N_i = \tilde{n}_i|N = n) P(N = n|\Lambda = \lambda) f_{\Lambda}(\lambda) d\lambda \\
&= \int_0^{\infty} \sum_{n=\tilde{n}_i}^{\infty} \binom{n}{\tilde{n}_i} p_i^{\tilde{n}_i} (1-p_i)^{n-\tilde{n}_i} \frac{\lambda^n}{n!} e^{-\lambda} \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} d\lambda \\
&= \int_0^{\infty} \frac{\beta^{\alpha}}{\Gamma(\alpha)} \frac{(\lambda p_i)^{\tilde{n}_i}}{\tilde{n}_i!} e^{-(\beta+1)\lambda} \lambda^{\alpha-1} \sum_{n=\tilde{n}_i}^{\infty} \frac{[\lambda(1-p_i)]^{n-\tilde{n}_i}}{(n-\tilde{n}_i)!} d\lambda \\
&= \int_0^{\infty} \frac{\beta^{\alpha}}{\Gamma(\alpha)} \frac{(\lambda p_i)^{\tilde{n}_i}}{\tilde{n}_i!} e^{-(\beta+1)\lambda} \lambda^{\alpha-1} e^{\lambda(1-p_i)} d\lambda \\
&= \frac{\beta^{\alpha}}{\Gamma(\alpha)} \frac{p_i^{\tilde{n}_i}}{\tilde{n}_i!} \int_0^{\infty} \lambda^{\tilde{n}_i+\alpha-1} e^{-\lambda(\beta+p_i)} d\lambda = \left\{ \lambda = \frac{x}{\beta+p_i} \right\} \\
&= \frac{\beta^{\alpha}}{\Gamma(\alpha)} \frac{p_i^{\tilde{n}_i}}{\tilde{n}_i!} \left(\frac{1}{\beta+p_i} \right)^{\tilde{n}_i+\alpha} \int_0^{\infty} x^{\tilde{n}_i+\alpha-1} e^{-x} dx \\
&= \frac{\beta^{\alpha}}{\Gamma(\alpha)} \frac{p_i^{\tilde{n}_i}}{\tilde{n}_i!} \left(\frac{1}{\beta+p_i} \right)^{\tilde{n}_i+\alpha} \Gamma(\tilde{n}_i+\alpha) \\
&= \frac{\Gamma(\tilde{n}_i+\alpha)}{\Gamma(\alpha)\tilde{n}_i!} \left(\frac{\beta/p_i}{\beta/p_i+1} \right)^{\alpha} \left(\frac{1}{\beta/p_i+1} \right)^{\tilde{n}_i}
\end{aligned}$$

We recognize the last expression as the probability function of a thinned negative binomial distribution, and conclude that $N_i \sim \text{NegBin}(\alpha, \beta/p_i)$.

D.2.2 ML parameter estimates

$$\begin{aligned}
N_i \sim \text{NegBin}(\alpha, \beta/p_i) &\Leftrightarrow \\
P(N_i = \tilde{n}_i) &= \frac{\Gamma(\tilde{n}_i+\alpha)}{\Gamma(\alpha)\tilde{n}_i!} \left(\frac{\beta/p_i}{\beta/p_i+1} \right)^{\alpha} \left(\frac{1}{\beta/p_i+1} \right)^{\tilde{n}_i}
\end{aligned}$$

$$\begin{aligned}
L(\alpha, \beta) &= \prod_{i=1}^m \frac{\Gamma(\tilde{n}_i + \alpha)}{\Gamma(\alpha)\tilde{n}_i!} \left(\frac{\beta/p_i}{\beta/p_i + 1}\right)^\alpha \left(\frac{1}{\beta/p_i + 1}\right)^{\tilde{n}_i} \\
&= \prod_{i=1}^m \frac{\alpha(\alpha + 1) \cdots (\alpha + \tilde{n}_i - 1)}{\tilde{n}_i!} \left(\frac{\beta/p_i}{\beta/p_i + 1}\right)^\alpha \left(\frac{1}{\beta/p_i + 1}\right)^{\tilde{n}_i}
\end{aligned}$$

Again, taking the logarithm, replacing \tilde{n}_i by $\tilde{n}_i v_i$, and placing all terms that do not depend on α or β in a constant term C , yields

$$\begin{aligned}
l(\alpha, \beta) &= \sum_{i=1}^m [\log \Gamma(\tilde{n}_i v_i + \alpha) - \log \Gamma(\alpha) + \\
&\quad + \alpha \log \left(\frac{\beta}{\beta + p_i}\right) - \tilde{n}_i v_i \log(\beta/p_i + 1)] + C. \quad (78)
\end{aligned}$$

The ML-estimates $\hat{\alpha}$ and $\hat{\beta}$ are obtained by maximising (78), which has to be done numerically.

E References

- Bühlmann, H. and Gisler, A. (2005): *A Course in Credibility Theory and its Applications*. Berlin: Springer-Verlag.
- Grandell, J. (1997): *Mixed Poisson Processes*. London: Chapman & Hall.
- Johansson, B. (1997, reprint 2008): *Matematiska modeller inom sakförsäkring*. Stockholm: Division of mathematical statistics, Department of Mathematics, Stockholm University.
- Kreps, R. (2007): Timeline Simulation. In: Brehm, P. et al. *Enterprise Risk Analysis for Property & Liability Insurance Companies*. New York: Guy Carpenter, Chapter 3.4.
- Lee, T. and Solo, V. (1999): Bandwidth Selection for Local Linear Regression: A Simulation Study. *Computational Statistics*, **14**, 515-532.
- Loader, C. (1999): Bandwidth selection: classical or plug-in? *The Annals of Statistics*, **27**(2), 415-438.
- Mikosch, T. (2004): *Non-Life Insurance Mathematics*. Berlin: Springer-Verlag.
- Rootzén, H. and Tajvidi, N. (1995): Extreme value statistics and wind storm losses: a case study. *Technical report 1995:5*. Gothenburg: Division of mathematical statistics, Chalmers University of Technology.
- Ruppert, D. et al. (1995): An Effective Bandwidth Selector for Local Least Squares Regression. *Journal of the American Statistical Association*, **90**, 1257-1270.
- Schirmacher, D. et al. (2005): Stochastic Excess-of-Loss Pricing within a Financial Framework. *Casualty Actuarial Society Forum*, Spring 2005, 297-351.
- Wand, M.P. and Jones M.C. (1995): *Kernel Smoothing*. London: Chapman & Hall.