Non-life insurance risk models under inflation

Abstract
In the first part of the paper we present some classical actuarial models (the collective and individual risk model) and the probability theory behind. A discussion of pros and cons of each approach leads to an alternative approach where the losses on each policy is modelled by an individual compound Poisson process. We estimate this model using generalized linear models (GLM). In the second part we introduce a framework for incorporating empirical claim severity inflation in the severity models. This gives a method for automatic update of the insurance tariff. The framework is a generalization a commonly used method of discounting, modelling and inflating (which we denote the DMI framework). A possible modification to the DMI framework is proposed, which makes it applicable to frequency models too. We suggest some methods to compare risk models, especially with respect to their performance over time, are suggested. Finally the methods are applied on a real life motor insurance dataset and we find that the models under the DMI framework are superior to traditional models without inflation adjustments. The reader is expected to have a background in probability theory and have experience with GLM modelling.

Keywords: Insurance, severity, inflation, risk model, actuarial model, generalized linear models, generalized additive models, Lorentz curve, DMI framework, discounting, inflating, automated.

JEL Classification: G22, C52, C13, C53, C14.

August 2010
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.7</td>
<td>Bibliographic notes</td>
<td>52</td>
</tr>
<tr>
<td>6</td>
<td>Model comparison</td>
<td>53</td>
</tr>
<tr>
<td>6.1</td>
<td>Correct ordering of risk</td>
<td>53</td>
</tr>
<tr>
<td>6.2</td>
<td>Overall correct level</td>
<td>54</td>
</tr>
<tr>
<td>6.3</td>
<td>Absence of time variation</td>
<td>55</td>
</tr>
<tr>
<td>6.4</td>
<td>Dispersion in estimates</td>
<td>56</td>
</tr>
<tr>
<td>6.5</td>
<td>Bibliographic notes</td>
<td>57</td>
</tr>
<tr>
<td>7</td>
<td>Empirical study</td>
<td>58</td>
</tr>
<tr>
<td>7.1</td>
<td>Data</td>
<td>58</td>
</tr>
<tr>
<td>7.2</td>
<td>Methods</td>
<td>58</td>
</tr>
<tr>
<td>7.3</td>
<td>Explorative analysis</td>
<td>59</td>
</tr>
<tr>
<td>7.4</td>
<td>Results</td>
<td>63</td>
</tr>
<tr>
<td>7.4.1</td>
<td>GLM analysis and inflation smoothing</td>
<td>63</td>
</tr>
<tr>
<td>7.4.2</td>
<td>Model comparison</td>
<td>64</td>
</tr>
<tr>
<td>8</td>
<td>Conclusions</td>
<td>68</td>
</tr>
<tr>
<td>A</td>
<td>Description of variables</td>
<td>70</td>
</tr>
<tr>
<td>B</td>
<td>Explorative analysis for GLM</td>
<td>72</td>
</tr>
<tr>
<td>C</td>
<td>Comparison graphs</td>
<td>74</td>
</tr>
<tr>
<td>References</td>
<td>78</td>
<td></td>
</tr>
</tbody>
</table>


1 Introduction

1.1 Acknowledgements

I would like to thank my colleagues Fredrik Thuring, Lars Max Jensen and Sarah Bentzen which have all contributed with valuable comments during the writing process. Also a special thank to Kenneth Wolstrup for giving access to data for the empirical chapter and my supervisor Dorthe Kronborg for her supportive and flexible supervision.

1.2 Insurance in general

In this introductory chapter some general notions from insurance will be presented. Most of the topics will be known to readers familiar with insurance mathematics or employees in the insurance industry. A first step is to get a definition of the word risk as this is a central part of insurance. By risk we shall mean any uncertain future loss. What separates a risk from an ordinary expense is the uncertainty in both the occurrence and the magnitude of it. Usually the occurrence time of a risk is unknown and so is the magnitude of the associated loss. This creates the foundation of the entire risk industry which among others includes insurance companies, pension funds and the banking and asset trading sector. Common to most of these industries are that they make a living out of trading risk with their counter parties. When many small independent risks are pooled together they become more predictable and hence easier to manage. This makes it possible for companies to manage large pools of risks that are difficult to manage on their own.

Many types of risk exist and many of them are generic in the sense that they apply to a wide range of businesses. Examples are financial risk, operational risk and market risk. The type of risk dealt with in this paper is insurance obligation risk (insurance risk in short). When a customer buys insurance the risk if transferred from the insured to the insurer. This means that the insurer will have to pay the losses on some specified risk over a certain exposure period. Clearly companies do not offer such a service for free and the insured has to pay a premium to the insurance to take the risk.

Related to every premium for transferring risk to an insurance company is the expected loss of the risk. This quantity is often referred to as the risk premium, office premium or net premium of the risk. We will use the first term throughout the paper. Clearly the premium paid for a theft insurance on an Audi A8 should differ from a similar insurance on a Toyota Yaris as the expected loss of the Audi is larger than the expected loss of the Toyota. This brings us to our next topic. Insurance risk is usually divided into two components, namely the frequency and the severity of the losses. The (expected\(^1\)) risk premium will depend on both (of these two) quantities. If a certain type of risk occurs very often the expected loss becomes high due to the frequency. Analogously, if the losses are very large when they occur this will bring up the expected loss due to the severity of the claims. The two quantities

\(^1\)When mentioning the risk premium, frequency or severity we will mostly mean the expectation to these quantities unless otherwise is clear from the context.
will play a central role in the determination of the risk premium in Chapter 3 and 4.

Determination of the risk premium allows the insurance company to manage the risk better. For a given insurance coverage (theft, fire, death, etc.) we will talk about good or bad risk, or equivalently worse or better risk. When a customer is said to be bad (good) risk we mean that the customer has a higher (lower) risk premium than other customers with the same coverage. Be aware that this is not equivalent to saying that the customer is a bad (good) customer. A bad risk customer who pays a premium high enough to cover his expected losses (plus some costs) will be a better customer than a good risk customer which only pays half of the losses he is expected to cause.

The risk premium presented above is a technical measure which is used to manage the risk. The actual premium paid by the insured (also known as market price, gross premium or street premium) will usually be higher than the risk premium since the insurance company besides paying the losses also has some costs of running the business (wages, buildings rent, cost of capital, etc.). The relationship between the risk premium and the market price can take any form, but mostly a high risk premium will result in a high market price. General principles for calculating the market price from the risk premium exists, but in practice the calculations may be influenced by many other factors. These factors may include the cost structure of the company, market profile, price elasticity and general marketing considerations. Hence this part of the pricing process may be both more qualitative and specific (to the company) than the determination of the risk premium. In the present paper we limit our scope to only discuss how to determine the correct risk premium based on quantitative methods.

1.3 Risk selection

Risk premiums represent the expected loss on the underlying risk. Since different customers represents different risk, this explains why insurance premiums often vary among customers. Most people will agree that young drivers are bad risk in a motor insurance portfolio. In the same way it is natural that a carpenter is worse risk than an office worker when considering a full time personal accident insurance. One of the most important tasks of the pricing actuary is finding a good model to describe these risk differences and let the market price reflect the findings.

There is no single method for determination of the risk premium. In the presence of many customers with homogeneous coverages (as is usually the case in personal insurance) one often relies of quantitative statistical methods. The level of sophistication of these methods may vary much between companies depending on the available data, IT systems, skills of employees, legal restrictions and even internal politics. However there is a lower bound of sophistication often referred to as the flat rate risk premium.
Definition 1.1. (Flat rate risk premium)
The flat rate risk premium is the average of the losses on a portfolio over a given period.
Let \( \{X_i, i = 1, \ldots, n\} \) be the observed losses over this period. Then the flat rate risk
premium is calculated by,

\[
\bar{\mu} = \frac{1}{n} \sum_{i=1}^{n} X_i .
\]  

(1.1)

The flat rate model assumes that all customers represent homogeneous risk and conse-
quently the risk premium is equal for all customers. Given that the portfolio is static over
time this method may actually give a good prediction of the total loss of a portfolio. By
static we mean than the mix of different customers in the portfolio stays constant over time
which is often approximately satisfied. Since the total loss along with the premiums are
the main concerns in relation to solvency and budgets, one could be tempted to use this
simple method only. There is at least one good reason not to do so which is explained below.

In a competitive market several insurance companies exists but will usually quote different
prices. Assuming that the products are homogeneous and price is the only parameter influ-
cencing the choice of the rational customer, then a flat rate will attract bad risk. To illustrate
this, assume that two companies A and B exists. Company A uses a flat rate premium
while company B uses more sophisticated tools to determine two prices, one for good risk
customers and one for bad risk customers. Because company A pools good and bad risk they
have a higher price for the low risk customers than company B which distinguishes between
the two types of customers. Meanwhile the bad risk customers of company A pays a lower
price than those of company B (in company A the good risk customers suffers from being
pooled with the bad risk customers). When customers realize this, the low risk customers
from company A will move to company B and the high risk customer from company B will
move to company A. In the next period company A will have more high risk customers which
will force them to raise their flat rate premium (as there are fewer good risk customers left to
pay for the bad risk). Company B on the other hand keeps its price levels, and in particular
do not change the price for the low risk customers. Company A will continue to attract
worse risk and loose good risk (if any left) which will constantly force its flat rate premium
upward. Company B is able to keep their price levels due to intelligent pricing. The essence
here is that in company B customers pay for the risk they represent while in company A
the good risk customers pays for the bad risk customers risk. This principle is denoted
anti-selection, see (Bühlmann and Gisler, 2005, p. 11). The principle is demonstrated in
Figure 1.1.

The anti-selection effect is also relevant when no companies use a flat rate. As long as some
company offers a more fair premium than the competitor, anti-selection may occur. The
term fair is here to be understood in the sense that it reflects the risk the customer actually
represents. A fair price is achieved by developing sophisticated models which differentiate
customers. These models are often referred to as micro models (opposed to macro models).
Since large data amounts are often needed to reveal the complex nature of the risk patterns,
Figure 1.1: Illustration of the anti-selection principle in a competitive market. Company A uses flat rate premium while company B quote different prices for different customers. Customers are attracted by the lowest price.

larger companies will usually have an advantage over smaller competitors. Furthermore the resources required to develop micro models are almost equally time consuming for large and small datasets. But since the impact of the models is somehow proportional to the portfolio volume this gives the larger companies a greater effect per resource spend on micro pricing.

It may seem impossible for an insurance company to exist unless it is market leader on fair price. Yet reality shows that many companies exist and that prices do differ in between companies. There are several reasons why this is possible without anti-selection forcing the companies out of the market. We give a few examples below.

• Non-homogeneous products
  Though the main coverage is the same, the companies usually have differentiated policy terms. Heterogeneous products cannot be compared by price alone.

• Multiple selection parameters
  Price is not the only parameter on which the companies differentiate and by which customers choose. Reliability and customer service of the company are very important factors when buying an insurance product. Marketing, communication, company profile and distribution channels may also affect the choice.

• The price of micro pricing
  Sophisticated models are expensive to maintain. The costs saved by using a simple model may allow the company to lower its overall price level

• Customer attraction
  Setting a flat rate price on a product may appeal to customers due to its simple nature. This strategy can be used to attract new customers on one product and afterwards cross sell to other lines of business with sophisticated prices.

The above discussion hopefully reflects that micro pricing is an important task, but deviations from the fair price do not necessarily ruin the company. During this paper we will
aim for a price as fair as possible, but at the same time strive for a simple structure which is related to the flat rate risk premium.

1.4 Types of risk change

In order to explain the motivation of the models given in this paper, we first need to discuss risk changes over time. Traditional insurance models are for risk what photographs are to the world; A picture of the situation at a given time. But both the world and the risk patterns of it changes dynamically over time. We now define two concepts that can be used to characterize how a risk patterns change over time. These two concepts are not well established actuarial notions, but we believe that anyone involved in insurance may recognise the substance of them. In any case they must necessarily be accepted in order to understand the models and their motivation. The concepts are changes in risk profile and changes in risk level.

Definition 1.2. (Change in risk profile)

A change in risk profile (over time) occurs when the relative ordering of the risks changes. The change is said to be heterogeneous.

Definition 1.3. (Change in risk level)

A change in risk level (over time) occurs when all risks changes in the same direction, with no changes in the relative ordering of the risks. The change is said to be homogeneous.

To demonstrate the difference between profile and level changes consider the constructed claim frequencies in Table 1.1. The portfolio is divided into four segments based on some demographical properties of the insured. First consider the change in risk level. Note that all segments represented in the table experience an increase in the frequency. Both before and after the change young males a generally worse risk and drivers older than 25 years are generally better risk. These qualitative interpretations remain unaffected over the change because the increase in some sense is homogeneous over the segments. Now consider the table of changes in risk profile. When a portfolio experience profile changes it means that different groups experience different changes. They change heterogeneously. From the table it is clear that only young males has a change in risk while the three other groups are unaffected.

The above table and discussion is very simplified for illustrative purposes. In real life applications both types of changes will usually be present simultaneously. Furthermore our use of the term “relative ordering” in Definition 1.2 and 1.3 suggest that both sub tables of Table 1.1 exhibit change in risk profile since e.g. \( \frac{5\%}{2\%} \neq \frac{6\%}{3\%} \) in the risk level table. These interpretational problems suggest that we should not see a change as being exclusively of one of the two types. Rather we should say that a change is more one of the types than the other. Hereby we can characterize the main qualitative nature of the change. We now explain how this separation serves a purpose.
1.5 Motivation and outline of the paper

We consider the types of risk change as being a scale with the two types in each direction. Any change can be placed on this scale and thereby give a qualitative characterization of it. The motivation for doing so is that each type of change calls for a special kind of attention. Since the risk models are designed to capture the risk pattern as fair as possible a change in risk profile will demand for the models to be updated. When the relative ordering of the real world risk changes, we need to change our models to mirror this. Otherwise we may be exposed to anti-selection. On the other hand, if the change is only in risk level our attention should be focused differently. Clearly we may still be subject to anti-selection, not because the relative ordering of the risks is wrong, but because the level is. But when the level is wrong another problem arises too. If the expected loss on average underestimates the actual loss, the company will lose money since the premium charged does not match with the payouts (including costs). Note however, that the level problem can be managed by scaling up the risk premium by some factor until it is at the right level. Since only the level of the risk changes, there is no need to update the underlying risk model as the relative ordering of the risks remain the same (no change in risk profile).

The aim of the present paper is to give some methods for updating risk models according to changes in risk level on an automated basis. The intention of doing so is twofold. 1) We believe (and hope) that frequent automated updates of simple level adjustments can be performed better than if done manually on a more infrequent basis. 2) The automation liberate human resources which can be focused on model updates related to changes in risk profile, which by nature are more complex and demanding.

Related to the changes if risk level is the the term inflation. Our approach throughout the paper will, briefly explained, be to adjust the risk premiums according to the observed inflation as if it was only due to changes in risk level. We will naturally discuss the problems related to this approach later. For now we would only like to emphasize that our treatment of inflation here is substantially different from what is normally seen in actuarial literature. Inflation plays a very important role for insurances with long termed payout patterns (e.g. life insurance, workers compensation). Also the reserves are affected by inflation because it changes the future payout patterns. Hence inflation is often mentioned in relation to these

<table>
<thead>
<tr>
<th>Year</th>
<th>Gender/Age group</th>
<th>Change in risk profile</th>
<th>Change in risk level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>18-25</td>
<td>25+</td>
</tr>
<tr>
<td>n</td>
<td>Male</td>
<td>5%</td>
<td>2%</td>
</tr>
<tr>
<td>n</td>
<td>Female</td>
<td>4%</td>
<td>2%</td>
</tr>
<tr>
<td>n + m</td>
<td>Male</td>
<td>10%</td>
<td>2%</td>
</tr>
<tr>
<td>n + m</td>
<td>Female</td>
<td>4%</td>
<td>2%</td>
</tr>
</tbody>
</table>

Table 1.1: Constructed annual claim frequencies of a portfolio. The table shows the qualitative difference between changes in risk profile and risk level.
topics in literature. Our treatment concerns pricing of short term business only. We are not interested in predicting inflation 30 years out in the future and we are not interested in the inflations effect on the reserves. The goal of our contribution is to make sure that the insurance risk premiums are at the correct level today, i.e. follows the observed inflation in the risk patterns. Hence the intended audience for this presentation is mainly practitioners in the pricing department of an insurance company.

The outline of the rest of the paper is as follows. In Chapter 2 we give some results from probability theory that are essential to the risk models. In Chapter 3 we present the traditional risk models of insurance and discuss an alternative model that is basically an improvement of the traditional models. The core in the improvement is to model the loss on each policy by a compound Poisson process and we demonstrate how this approach naturally relates to the traditional models. Next, in Chapter 4, we present methods for estimating the risk models using generalized linear models (GLM). Chapter 5 presents the main models of this paper. These models are designed to automatically adjust the level of the severity estimates proportionally, according to the observed inflation in the risk pattern. Especially the DMI framework (the main contribution of the paper) is treated in details because it is very flexible and yet quite simple. We discuss the theoretical justification of our positive expectation to this framework and highlight some possible problems related to it. In Chapter 6 we give some methods to evaluate the models, especially regarding their performance over time. Finally in Chapter 7 we apply the methods to a real life dataset from a Scandinavian motor insurance portfolio. We find that the DMI framework is superior to traditional models with respect to the comparison measures. In particular we find the that DMI framework is better to keep the correct level in periods of (absolutely) high inflation. In Chapter 8 we combine the results of the paper and discuss possible development issues of the models.

1.6 Bibliographic notes

We recommend (Ohlsson and Johansson, 2010, chapter 1) for a general introduction to the terminology of insurance pricing. The reader may also find the discussion in (Bühlmann and Gisler, 2005, chapter 1) helpful, especially in relation to anti-selection. For a discussion of some theoretical methods to derive the market premium from the risk premium see (Kaas, Goovaerts, Dhaene, and Denuit, 2001, chapter 5).
2 Probabilistic prerequisites

We now present some results from probability theory that are important to the subsequent models. Our focus in this chapter is only on the probability theory while the relation to insurance risk is treated in Chapter 3. Most of the results are well known from other branches of probability theory but are included here for completeness.

2.1 Random variables, distributions, expectations and variance

Random variables (r.v.) play an important role in insurance mathematics. We denote a random variable by a capital letter, e.g. $X$. In general we will not distinguish between a r.v. variable $X$ and its distribution. Hence expressions like "let $X$ have a gamma distribution" and "let $X$ be a gamma distribution" are equivalent. When the outcome of a r.v. is no longer uncertain we denote the realized value by the corresponding non-capital letter, e.g. $x$. When talking about a sequence of independent identically distributed (iid.) r.v. $\{Y_i, i = 1, 2, \ldots\}$ we will sometimes just refer to $Y$ without subscript when the index is superfluous.

Let $X$ be a r.v. defined on the set $S$ with distribution (function) $F$, shortly $X \sim F$. This means that

$$F(x) = \mathbb{P}(X \leq x) = \int_{y \in S : y \leq x} dF(y) = \int_{y \in S : y \leq x} f(y) dy$$

(2.1)

where $\mathbb{P}$ is used for "probability" and

$$f(y) = \frac{d}{dy} F(y) .$$

(2.2)

The expected value, symbolised by $\mathbb{E}$, of $X$ is defined by,

$$\mu_X = \mathbb{E}(X) = \int_S x dF(x) .$$

(2.3)

The expected value is also referred to as the mean or the 1st central moment of $X$. The variance, symbolised by $\mathbb{V}$, of $X$ is defined by,

$$\sigma_X^2 = \mathbb{V}(X) = \int_S (x - \mathbb{E}(X))^2 dF(x) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 .$$

(2.4)

The variance is also referred to as the second central moment of $X$. When it is clear from the context to which r.v. the expectations is taken the subscripts may be omitted, i.e. $\mu_X = \mu$ and $\sigma_X^2 = \sigma^2$. In general the $\alpha$’th (non-centralized) moment is defined as,

$$m_{\alpha} = \int_S x^\alpha dF(x) .$$

(2.5)

2.2 Moment generating function

The moment generating function (m.g.f.) of a random variable $X$ (equivalently the distribution of $X$) is a unique function characterizing the r.v. just like the probability density function or the distribution function.
Definition 2.1. (Moment generating function)

The moment generating function \( M_X \) of a random variable \( X \) is defined by,

\[
M_X(r) = \mathbb{E}(e^{rX}), \quad r \in \mathbb{R}.
\] (2.6)

The uniqueness property is of great importance since it makes it possible to determine the distribution of a r.v. from its m.g.f. This can be useful in determination of the distribution of a sum of independent r.v.’s.

Proposition 2.2. (Sums of independent random variables)

Let \( \{Y_1, ..., Y_N\} \) be independent random variables with m.g.f. \( M_{Y_1}, ..., M_{Y_N} \). The moment generating function of the sum \( S = Y_1 + ... + Y_N \) is then,

\[
M_S(r) = M_{Y_1}(r) \cdots M_{Y_N}(r).
\] (2.7)

Proof. Simply calculate the expectation to the sum and use the independence of the random variables.

\[
M_S(r) = M_{Y_1 + ... + Y_N}(r) = \mathbb{E}(e^{r(Y_1 + ... + Y_N)}) = \mathbb{E}(e^{rY_1}) \cdots \mathbb{E}(e^{rY_N}) = M_{Y_1}(r) \cdots M_{Y_N}(r).
\]

By use of Proposition 2.2 one can determine the m.g.f. of a sum of random variables, \( S \). By uniqueness of the m.g.f. it may be possible to identify the m.g.f. as one from a known distribution.

As the name indicates, the m.g.f. can be used to determine the moments of random variables. The \( \alpha \)’th moment of a r.v. can be found by simply calculating the \( \alpha \)’th derivative at zero (provided that it exists).

\[
m_\alpha = \mathbb{E}(X^\alpha) = \left. \frac{\partial^\alpha M_X(r)}{\partial r^\alpha} \right|_{r=0}.
\] (2.8)

To see why this is true just apply a traditional Taylor series expansion of \( e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} \) to the m.g.f.,

\[
M_X(r) = \sum_{k=0}^{\infty} \frac{\mathbb{E}(X^k)r^k}{k!}.
\] (2.9)

When differentiating, only the term corresponding to the \( \alpha \)’th moment will be non-zero \((r^k = 0 \text{ for } r = 0, k \neq 0)\).
2.3 Poisson process

The Poisson process plays a prominent role in classical insurance mathematics. We start by a formal definition.

**Definition 2.3. (Poisson process)**

A continuous time stochastic process \( N_t, t \geq 0 \) is said to be a homogeneous Poisson process with intensity \( \lambda > 0 \) if

1. \( N_0 = 0 \).
2. \( \{N_t\} \) has stationary and independent increments.
3. \( \mathbb{P}(N_{t+h} - N_t = 0) = 1 - \lambda h + o(h) \) as \( h \to 0 \).
4. \( \mathbb{P}(N_{t+h} - N_t = 1) = \lambda h + o(h) \) as \( h \to 0 \).
5. \( \mathbb{P}(N_{t+h} - N_t \geq 2) = o(h) \) as \( h \to 0 \).

The expression \( o(h) \) represents a function that is much smaller than \( h \) for small \( h \) such that \( \lim_{h \to 0} \frac{o(h)}{h} = 0 \).

Note that there exist many equivalent definitions of this process. The Poisson process is a counting process which means that it takes positive integer values, \( N_t \in \mathbb{N} \). The following property shall be used extensively later on.

**Proposition 2.4. (Poisson distributed increments)**

The number of jumps of a Poisson process \( N_t \) with intensity \( \lambda \) in any interval of length \( s \) is Poisson distributed with intensity \( \lambda s \).

\[
\mathbb{P}(N_{t+s} - N_t = k) = e^{-\lambda s} \frac{(\lambda s)^k}{k!}.
\]

(2.10)

**Proof.** Divide the interval \( (t,t+s] \) into \( n \) subintervals of length \( h = s/n \) where \( n \) is chosen large. Now the probability of two or more jumps in any subinterval follows from Definition 2.3.v.,

\[
\mathbb{P} \left( \bigcup_{k=1}^n (N_{t+kh} - N_{t+(k-1)h} \geq 2) \right) \leq \sum_{k=1}^n \mathbb{P}(N_{t+kh} - N_{t+(k-1)h} \geq 2)
\]

\[
= \frac{n o(h)}{s/n} \quad \frac{o(h)}{h} \to 0 \quad \text{as} \quad n \to \infty.
\]

(2.11)
tion 2.3.ii) \( N_t \) has a binomial distribution with parameters \((n, \lambda h + o(h))\), from Definition 2.3.iv. Now we need the following well know result.

**Lemma 2.5. (Poisson approximation)**

Let \( X_n \sim Bin(n, p) \), \( Y \sim Poi(\lambda) \) and \( np \to \lambda \) as \( n \to \infty \) for fixed \( \lambda \in \mathbb{R}_+ \). Then

\[
\Pr(X_n < x) \approx \Pr(Y < x) \quad \text{as } n \to \infty.
\] (2.12)

**Proof.** See (Shiryaev, 1995, p. 64).

Since \( N_t \sim Bin(n, \lambda(s/n)) \) as \( n \to \infty \), from Lemma 2.5 we have that \( N_t \sim Poi(\lambda s) \) as \( n \to \infty \). This ends the proof.

The moment generating function of a Poisson distributed r.v. is simple to derive.

**Corollary 2.6. (Moment generating function of the Poisson distribution)**

The moment generating function of a Poisson distributed random variable \( X \sim Poi(\lambda) \) is,

\[
M_X(r) = e^{\lambda(e^r - 1)}.
\] (2.13)

**Proof.** Simply calculate the moment generating function (Definition 2.1) using the probability density function of Proposition 2.4 (Poisson p.d.f).

\[
M_X(r) = E(e^{rX}) = \sum_{k=0}^{\infty} e^{rk} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(e^r \lambda)^k}{k!} = e^{-\lambda} e^{e^r \lambda} = e^{\lambda(e^r - 1)}
\] (2.14)

Here, in the fourth line, we use the Taylor series expansion for the exponential function \( e^x \) with \( x \) replaced by \( e^r \lambda \).

2.4 Compound Poisson process

The compound Poisson process is a generalization of the Poisson process. Instead of jumps of size one the compound Poisson process can have jumps of arbitrary size.

**Definition 2.7. (Compound Poisson process)**

\( S_t \) is said to be a compound Poisson process if,

\[
S_t = \sum_{i=1}^{N_t} X_i,
\] (2.15)

where \( N_t, t \geq 0 \) is a homogeneous Poisson process independent of \( \{X_i, i \geq 1\} \) which is a series of iid. random variables with common distribution.
In insurance context $N_t$ counts the number of claims over a certain period while $X_i$ are the claim severities. Therefore the distribution of $X$ in Definition 2.7 will be referred to as the *severity distribution*. To easy notation we will sometimes leave out the time subscript $t$ of $S_t$ and then just implicitly assume that the intensity $\tilde{\lambda}$ of the counting process is adjusted according to the time period under consideration ($\tilde{\lambda} = \lambda t$). We will then consider the process over one period, which could be a month, a year or any other time interval. Similarly the notation of a Poisson processes $N_t$ will sometimes just be referred to as $N$.

The expectation to a compound Poisson process is easily derived using the law of iterative expectations.

\[
E(S) = E(E(S|N)) = E(N E(X)) = E(N) E(X) = \lambda E(X). \tag{2.16}
\]

Here the third equation follows from independence according to Definition 2.7 and the last equation follows directly from Proposition 2.4 and the fact that $E(N) = \lambda$ for $N \sim \text{Poi}(\lambda)$.

**Corollary 2.8. (Moment generating function of compound Poisson distribution)**

The moment generating function of a random variable $S$ having a compound Poisson distribution with Poisson intensity $\lambda$ and severity distribution $X$ is,

\[
M_S(r) = e^{\lambda (M_X(r) - 1)}. \tag{2.17}
\]

**Proof.** Simply calculate the m.g.f. as defined in Definition 2.1 and use the m.g.f. of the Poisson distribution (Corrolary 2.6).

\[
M_S(r) = E(e^{r S}) = E\left(e^{r (X_1 + \ldots + X_N)}\right) = E\left(E\left(e^{r (X_1 + \ldots + X_N)} \mid N = n\right)\right)
= \sum_{n=0}^{\infty} \left(M_X(r)^n \mathbb{P}(N = n) = \mathbb{E}\left((M_X(r))^N\right) = \mathbb{E}\left(e^{N \log M_X(r)}\right) = M_N(\log M_X(r)) = e^{\lambda (M_X(r) - 1)}.
\]

This corollary immediately leads to the following very important theorem.
Theorem 2.9. (Sum of compound Poisson distributions)

Let \( \{S_i, i = 1, ..., k\} \) be random variables, each having a compound Poisson distribution with Poisson intensity \( \lambda_i \) and severity distribution \( F_i \). Further let the \( S_i \)'s be independent. Then the sum \( S = S_1 + ... + S_k \) has a compound Poisson distribution with Poisson intensity \( \lambda = \sum_{i=1}^{k} \lambda_i \) and severity distribution \( F = \frac{1}{k} \sum_{i=1}^{k} \lambda_i F_i \).

Proof. Let \( M_i(r) \) be the moment generating function of \( S_i \). Since the \( S_i \)'s are independent we can use Proposition 2.2 combined with the compound Poisson m.g.f. from Corollary 2.8.

\[
M_S(r) = \mathbb{E}(e^{rS}) = \prod_{i=1}^{k} M_i(r) = \prod_{i=1}^{k} e^{\lambda_i M_i(r) - 1} = e^{\sum_{i=1}^{k} \lambda_i (M_i(r) - 1)} = e^{\sum_{i=1}^{k} \lambda_i M_i(r) - \lambda} = e^{\lambda \left( \frac{1}{k} \sum_{i=1}^{k} \lambda_i M_i(r) - 1 \right)} .
\]

(2.19)

By uniqueness of m.g.f.'s, this is the m.g.f. of a compound Poisson distribution with Poisson intensity \( \lambda = \sum_{i=1}^{k} \lambda_i \) and a severity distribution with m.g.f. \( \frac{1}{k} \sum_{i=1}^{k} \lambda_i F_i(r) \). To identify the severity distribution notice that a r.v. \( X \) having a mixture distribution \( F = \sum w_i F_i \) with weights \( w_i \) on component distribution \( F_i \) has a moment generating function,

\[
M_F(r) = \mathbb{E}(e^{rX}) = \sum w_i \mathbb{E}_{F_i}(e^{rX}) = \sum w_i M_{F_i}(r) .
\]

(2.20)

By uniqueness of m.g.f. the severity distribution of the compound Poisson distribution \( S \) is therefore a mixture of the underlying severity distributions \( F_i \) with weights proportional to the intensity \( \lambda_i \) by which they occur. Hence, \( F = \frac{1}{k} \sum_{i=1}^{k} \lambda_i F_i \). \( \square \)

2.5 Bibliographic notes

The theory presented in this chapter is available in most textbook on actuarial theory such as (Boland, 2007) and (Kaas, Goovaerts, Dhaene, and Denuit, 2001). Also the technical volume by (Rolski, Schmidli, Schmidt, and Teugels, 1998) contains the results as well as their relation to a wide range of topics in actuarial theory. For a general book on stochastic processes Ross (1996) is good choice. (Shiryaev, 1995) presents some general probability theory.
3 Insurance risk: Theory

Insurance risk is a diverse field of science. Several interrelated sub branches exists, such as pricing, credibility theory, reserve setting and calculation of ruin probabilities. A cornerstone in most of these branches is an objective understanding of the risk in a portfolio, line of business or single policy. The purpose of this chapter is to present several models used to describe these risk (losses). First we present two of the classic insurance risk models, the \textit{individual risk model} and the \textit{collective risk model}. Next we discuss a model which in some sense is a mixture of the two classic models. We will discuss why this model is problematic for several reasons and propose an alternative model that seems to solve these problems. Though this alternative model is not new (but sparse) in literature, the relation to the classical risk models seems absent. We show here how this alternative model naturally arises from both the individual and collective risk model and how it inherits the desirable properties of each.

3.1 Collective risk model

The collective risk model is a methods to estimate the aggregate loss of an insurance portfolio. Since the number of claims in the portfolio during the time interval \((0, t]\) is a random variable \(N_t\), we may assume a certain distributional assumptions on it. It could e.g. be that \(N_t\) has a Poisson, binomial or negative binomial distribution. Each of the \(N_t\) claims will represent a random loss \(\{X_i, i = 1, ..., N_t\}\) and we assume these to be iid. Hence the aggregate loss on the portfolio in the time interval \((0, t]\) is,

\[
S_t = \sum_{i=1}^{N_t} X_i.
\]

(3.1)

We will assume that \(N_t\) is independent of the \(X_i\’s\), i.e. that the number of claims does not influence the severity of the claims. Though this is not always true it is crucial for the simplicity of the method. The expression in (3.1) looks quite familiar to the definition of the compound Poisson process (Definition 2.7). Note however, that the distribution of \(N_t\) in (3.1) may differ from the Poisson distribution. We say that \(S_t\) has a \textit{compound distribution} and call it compound Poisson, compound binomial, compound negative binomial, etc. according to the distribution of \(N_t\). We will only focus on the compound Poisson distribution here, but should emphasize the definition of the collective risk model is more general than the Poisson case. Below we give a formal definition of the Poisson case.

\begin{definition}
(Collective risk model, Poisson)
Let \(N_t\) be a homogeneous Poisson process with intensity \(\lambda\) counting the number of claims in a portfolio in the time interval \([0, t]\). Let \(\{X_i, i \geq 1\}\) be a sequence of iid. severities independent of \(N_t\). The (Poisson) collective risk model is now defined as the aggregate loss \(S_t\) of the portfolio during the time interval \([0, t]\),

\[
S_{0, t} = S_t = \sum_{i=1}^{N_t} X_i.
\]

(3.2)
\end{definition}
3.2 Individual risk model

The individual risk model is an alternative to the collective risk model. It also models the aggregate loss, but takes a slightly different view on the risk. Instead of summing losses over claims the individual risk model sums losses over policies. It is assumed that the losses on the policies, \( \{\tilde{Y}_j, j = 1, ..., n\} \), are independent but not necessarily identically distributed. Further the model assumes that each policy can have at most one loss during the period under consideration. Using this design gives at most \( n \) losses for a portfolio with \( n \) policies. Usually not all policies in the portfolio will experience a loss during a period. Therefore a severity variable \( \tilde{X}_j \) is introduced which is conditional on the occurrence of a loss on policy \( j \). At the same time a Bernoulli r.v. is introduced, \( I_j = 1_{\{\text{Loss on policy } j \text{ in period } (0,t]\}} \), which indicates whether the policy has a loss, such than \( I_j = 0 \) if there is no loss on policy \( j \) and \( I_j = 1 \) otherwise. The policy loss variable \( \tilde{Y}_j \) is then defined as the product \( \tilde{Y}_j = I_j \tilde{X}_j \).

Before giving a formal definition of the individual risk model a notational convention is needed. A portfolio of \( n \) policies \( j = 1, ..., n \) will be denoted by the set \( \mathcal{P} = \{1, ..., n\} \). Hence the size of a portfolio is \(|\mathcal{P}| = n \). We summarize the individual risk model in a definition.

**Definition 3.2. (Individual risk model)**

Let \( I_j \) be an indicator random variable which is one if there is a loss on policy \( j \) during the time interval \((0,t]\) and zero otherwise. Let \( \tilde{X}_j \) be the random variable representing the severity given a loss occur on policy \( j \). Assume that \( \tilde{Y}_j = \tilde{X}_j I_j \) is independent of \( \tilde{Y}_i = \tilde{X}_i I_i \) for all \( j \neq i \) and that \( I_j \) is independent of \( \tilde{X}_j \). The individual risk model is now defined as the aggregate loss of the portfolio \( \mathcal{P} \) during the time interval \((0,t]\),

\[
\tilde{S}_{0,t} = \tilde{S}_t = \sum_{j \in \mathcal{P}} \tilde{Y}_j = \sum_{j \in \mathcal{P}} I_j \tilde{X}_j .
\]

(3.3)

Note that \( \mathbb{E}(I_j) = \mathbb{P}(\text{Loss on policy } j \text{ in period } (0,t]) = q_j \). Hence one may interpret the policy loss variable \( \tilde{Y}_j \) as having a compound Bernoulli distribution with the severity distribution \( \tilde{X}_j \) as its second distribution,

\[
\tilde{Y}_j = \sum_{i=1}^{I_j} \tilde{X}_j .
\]

(3.4)

Due to the independence of \( I_j \) and \( \tilde{X}_j \) the compound Bernoulli distribution share the following property of the mean with the compound Poisson distribution, see (2.16).

\[
\mathbb{E}(\tilde{Y}_j) = \mathbb{E}(I_j)\mathbb{E}(\tilde{X}_j) .
\]

(3.5)

3.3 Comparison and interpretation of the risk models

Both the individual and collective risk model are quite intuitive. The collective risk model sums over a stochastic number of claims while the individual risk model sums over losses on a deterministic number of policies. Unfortunately both models have drawbacks. We start
the discussion by those of the individual risk model.

The individual risk model assumes only one claim per policy per time period. We can interpret this assumption in two ways. 1) One interpretation is that there can actually only occur one claim. However most ordinary insurance contracts for the non-commercial segment do allow for several claims. In this case the natural interpretation of the assumption may be that 2) if several claims occur on a policy they are collapsed into one loss. Hence the Bernoulli indicator $I_j$ indicates whether there is at least one claim on the policy. Analogously the severity $\tilde{X}_j$ may express the severity of the sum over several claims. This way of collapsing data has the consequence that we throw away information. After the collapse it is not possible to distinguish losses occurred from one claim from losses occurred from several claims. The motivation for collapsing data is of course simplicity, since it gives a simple Bernoulli structure of our risk model. Regarding the first interpretation, the assumption may actually be acceptable in a few cases: If the contract terms specify that the insurance can only be used once, if the probability of several claims in a time interval is small or if data are only available on collapsed level (which could be the case for a reinsurance company).

Another unpleasant feature of the individual risk model is that the compound Bernoulli distribution is not closed under convolution as the compound Poisson distribution. The problems related to the individual risk model are compensated by one very important property. That is the fact that the claim probability as well as the severity distribution do not have to be identically distributed for different policies. This makes it possible to estimate risk on policy level. We will denote this property as heterogeneity in risk. The collective risk model on the other hand pools all risk and assume iid severity distribution (homogeneous risk) for all policies. This is a problem, because if we cannot identify the customers with the highest severities (on average) we may attract bad risk (anti-selection, see Section 1.3 for details).

In the next sections two related risk models are presented. Both of these aim to capture the desirable properties of each of the above models.

### 3.4 Approximation to the individual risk model

**The approximation**

The main problem with the individual risk model is that it only allows for one claim per policy. In this section we will show how the compound Poisson distribution can be used to approximate the individual risk model and thereby solve the problem with multiple claims. The presentation shown here is a standard approach for approximating the individual risk, see e.g. (Boland, 2007), (Kaas, Goovaerts, Dhaene, and Denuit, 2001) or (Tse, 2009). However the method has a problematic interpretation which we will return to at the end of the section.

Remember that the policy loss variables $\tilde{Y}_j$ from (3.4) are compound Bernoulli r.v.'s. The

---

2 Convolution is needed when summing r.v. The property of closure under convolution (summation) means that the sum of $n$ r.v. from same distributional family also belong to this distributional family.
sum $\tilde{S}$ over the $n$ policies in the individual risk model is therefore a sum of $n$ Bernoulli r.v. which makes the distribution of $\tilde{S}$ complex to determine (due to the non-closure under convolution). However the expected number of claims in the model is easy to calculate,

$$E\left(\sum_{j \in \mathcal{P}} \mathbb{1}_j\right) = \sum_{j \in \mathcal{P}} E(\mathbb{1}_j) = \sum_{j \in \mathcal{P}} q_j := \lambda .$$ \hspace{1cm} (3.6)

Now introduce the notational convention $X_{ji}$ for the $i$’th claim on policy $j$ to allow for multiple claims on a single policy. When we do not refer to any specific claim or the distribution of any of the claims on policy $j$ we shall use the notation $X_j \cdot$. To avoid complex convolutions under the individual risk model we now approximate the compound Bernoulli r.v. $\tilde{Y}_j$ by a compound Poisson r.v. with the same expected number of claims. Let $Y_j$ be a compound Poisson distribution with $\lambda_j = q_j$ as the Poisson intensity and $X_j \cdot$ as the severity distribution. By Theorem 2.9 the aggregate loss $S$ in the approximation is itself a compound Poisson distribution with intensity $\lambda$ and the \{ $\frac{\lambda_1}{\lambda}, \ldots, \frac{\lambda_n}{\lambda}$ \} mixture of the $X_j \cdot$’s as severity distribution $W$. Since the policy loss probabilities $q_j$ are based on one time period of arbitrary length, the intensity of the Poisson process $N$ will also be measured against this time scale. Hence the one period expected number of claims of the aggregate compound Poisson process $S$ is $E(N) = \lambda \cdot 1 = \lambda$.

Comparison of mean and variance

We now compare the mean and variance of the aggregate loss in the individual risk model and the above approximation. The presentation is partly replicated from (Boland, 2007, section 3.2.2.1). We first need some auxiliary results.

Lemma 3.3. \textit{(Variance of compound distributions)}

The variance of a compound distribution $S$ with claim count variable $N$ and severity variable $X$ is,

$$\mathbb{V}(S) = E(N)\mathbb{V}(X) + E(X)^2\mathbb{V}(N) .$$ \hspace{1cm} (3.7)

\textbf{Proof.} By the law of total variance\footnote{See (Sydsæter, Strøm, and Berck, 2005, Formula 33.44), $\mathbb{V}(Y) = E(\mathbb{V}(Y|X)) + \mathbb{V}(E(Y|X))$.} we have,

$$\begin{align*}
\mathbb{V}(S) &= E(\mathbb{V}(S|N)) + \mathbb{V}(E(S|N)) \\
&= E(\mathbb{V}(X_1 + \ldots + X_N)) + \mathbb{V}(E(X_1 + \ldots + X_N)) \\
&= E(N\mathbb{V}(X)) + \mathbb{V}(N\mathbb{E}(X)) \\
&= E(N)\mathbb{E}(\mathbb{V}(X)) + \mathbb{E}(X)^2\mathbb{V}(N) \\
&= E(N)\mathbb{V}(X) + \mathbb{E}(X)^2\mathbb{V}(N) .
\end{align*}$$ \hspace{1cm} (3.8)

Here we use independence between the $X_i$’s, the expectation rule $E(X+Y) = E(X) + E(Y)$ and the two variance rules $\mathbb{V}(X+Y) = \mathbb{V}(X) + \mathbb{V}(Y) + 2\text{COV}(X,Y)$ and $\mathbb{V}(aX) = a^2\mathbb{V}(X)$. Note that $\text{COV}(X_i, X_j) = 0, i \neq j$ (due to independence) which is the reason why the covariance is absent from the derivation. \hfill $\square$
From Lemma 3.3 we can calculate the variance for two particular compound distributions.

**Corollary 3.4. (Variance of compound Poisson distribution)**

The variance of a compound Poisson distribution $S$ with Poisson intensity $\lambda$ and severity distribution $X$ is,

$$\text{Var}(S) = \lambda \mathbb{E}(X^2). \quad (3.9)$$

**Proof.** Follows by simple calculations using Lemma 3.3.

$$\text{Var}(S) = \mathbb{E}(N)\text{Var}(X) + \mathbb{E}(X)^2\text{Var}(N)$$

$$= \lambda \mathbb{E}(X) + \mathbb{E}(X)^2\lambda$$

$$= \lambda(\mathbb{E}(X) + \mathbb{E}(X)^2)$$

$$= \lambda\mathbb{E}(X^2), \quad (3.10)$$

using (2.4).

**Corollary 3.5. (Variance of compound Bernoulli distribution)**

The variance of a compound Bernoulli distribution $S$ with success parameter $q$ and severity distribution $X$ is,

$$\text{Var}(S) = q\mathbb{E}(X^2) + q(1-q)\mathbb{E}(X)^2. \quad (3.11)$$

**Proof.** A Bernoulli r.v. $I$ with success probability $q$ has mean $\mathbb{E}(I) = q$ and variance $\text{Var}(I) = q(1-q)$. Just insert these two quantities in the formula of Lemma 3.3.

Now back to the problem of calculating the mean and variance under the individual risk model $\tilde{S}$ and the approximation $S$. Let $\mu_j = \mathbb{E}(X_j) = \mathbb{E}(\tilde{X}_j)$ and $\sigma_j^2 = \mathbb{V}(X_j) = \mathbb{V}(\tilde{X}_j)$. We first calculate the mean by use of (2.16) and (3.5).

$$\mathbb{E}(S) = \lambda\mathbb{E}(W)$$

$$= \lambda \sum_{j \in P} \frac{\lambda_j}{\lambda} \mathbb{E}(X_j)$$

$$= \sum_{j \in P} \lambda_j \mathbb{E}(X_j)$$

$$= \sum_{j \in P} q_j\mu_j = \sum_{j \in P} \mathbb{E}(I_j)\mathbb{E}(\tilde{X}_j) = \mathbb{E}(\tilde{S}). \quad (3.12)$$

From this it is clear that the expected number of claims is the same for the individual risk model and its approximation. This is not surprising since by construction the expected number of claims on policy level are equal in the two methods. We proceed with a calculation
of the variance using Corollary 3.4 and Corollary 3.5.

\[ \text{Var}(S) = \lambda \mathbb{E}(W^2) = \lambda \sum_{j \in P} \frac{\lambda_j}{\lambda} \mathbb{E}(X_j^2) \]

\[ = \frac{\lambda}{\lambda} \sum_{j \in P} \lambda_j (\mathbb{V}(X_j) + \mathbb{E}(X_j)^2) \]

\[ = \sum_{j \in P} \lambda_j (\sigma_j^2 + \mu_j^2) \]

\[ = \sum_{j \in P} q_j (\sigma_j^2 + \mu_j^2) \]

\[ = \sum_{j \in P} q_j \sigma_j^2 + q_j \mu_j^2 \]

\[ \geq \sum_{j \in P} q_j \sigma_j^2 + q_j (1 - q_j) \mu_j^2 = \text{Var}(\tilde{S}). \]  (3.13)

From this derivation it is clear that the variance is greater in the approximation than in the individual risk model. This is because multiple claims per policy are possible in the approximation. However the difference should be relatively small in most cases and is easily calculated by,

\[ \text{Var}(S) - \text{Var}(\tilde{S}) = \sum_{j \in P} q_j \sigma_j^2 + q_j \mu_j^2 - \sum_{j \in P} q_j \sigma_j^2 + q_j (1 - q_j) \mu_j^2 \]

\[ = \sum_{j \in P} q_j \mu_j^2. \]  (3.14)

**Points of criticism**

The approximation presented in this section is constructed to inherit the best properties of the traditional risk models (collective and individual). The desirable property of closure under convolution from the compound Poisson distribution (equivalently the collective risk model) and the heterogeneity in risk from the individual risk model are both preserved in this approximation. It has been shown that a small increase in variance of the risk model is the trade off. As argued in (Kaas, Goovaerts, Dhaene, and Denuit, 2001) this will tend a risk averse decision maker (insurer) to make more conservative decisions.

It is the opinion of the author that there are some problems related to the interpretation of the approximation presented here. Below we shall try to clarify these.

As mentioned in Section 3.3 we can interpret the "one claim" assumption of the individual risk models in two ways. If the present risk actually only allows for one claim there is no problem. Then the above approximation with the condition \( X_j = \tilde{X}_j \) allows us to calculate an approximate distribution of the aggregate loss \( S \) only with a small increase in variance (depending on \( \mu_j \) and \( q_j \) of course).

If the risk on the other hand allows for several claims we must interpret the individual risk model in the sense of collapsing data. By introducing the approximation we allow the counting distribution to have multiple claims per policy. Meanwhile the severity distribution is the same as in the individual risk model, \( X_j = \tilde{X}_j \). But remember that under this
interpretation of the "one claim" assumption the severity may express the loss on multiple claims. This is problematic, since then the approximation accounts for multiple claims in both the counting distribution and the severity distribution. Though the calculations are quite unaffected by this, it becomes very difficult to interpret the components of this model.

Intuitively one would expect such a model with double account for multiple claims to overestimate the loss. As shown in (3.12) this is not the case. This is due to differences in the probability of no claims. The approximation has an intensity of $\lambda = q$ (subscript $j$ left out). Hence the Poisson probability of no claims is $P(N = 0) = e^{-q}$, which is to be compared to the Bernoulli probability of no claims $(1 - q)$. We can show that $e^{-q} \geq 1 - q$, $q \in [0, 1]$, i.e. that the probability of no claims is greater in the approximation than in the individual risk model. We will use a geometrical argument. First notice that

$$\frac{d}{dq} e^{-q} = -e^{-q} \Rightarrow \frac{d}{dq} (1 - q) = -1.$$  

Further $e^{-q}|_{q=0} = (1 - q)|_{q=0} = 1$. This means that the line $y = 1 - q$ is exactly tangent to the curve $y = e^{-q}$ at $q = 0$. Since the second derivative of the exponential function, $\frac{d^2}{dq^2} e^{-q} = e^{-q}$, is positive, it is a convex function. Hence the curve $y = e^{-y}$ stays above $y = 1 - q$ for all $q \in \mathbb{R}/\{0\}$ and only touches the line at $q = 0$. Figure 3.1 illustrates the arguments. This explains why the expected number of claims of the two methods are equal despite the possibility of several claims in the approximation. In principle there is nothing wrong with the approximation. However one should be careful in interpreting both the severity distribution and the probability of no claims.

In the next section we will present an alternative method which also have the properties of closure under convolution and heterogeneity risk but at the same time has a natural
interpretation. The method is neither new practice nor in literature, see (Ohlsson and Johansson, 2010). But to the author’s knowledge the presentation of the close connection to the classical risk models is not previously described as below.

3.5 Compound Poisson risk model

The model presented in this section is in many ways a combination of the two classical risk models just as the approximation above. The classical risk models have roots back to the beginning of the 1900’th century. Though the model of this section theoretically could be applied back in the beginning of the 1900’th century, the true benefits of the model first became achievable with the invention of the modern computer and in particular Generalized Linear Models (GLM) which developed intensively during the 1980’s. The methods used here is rarely discussed in literature though the industry has used it for years. The reason for this may be the models similarity to the collective risk model that is found in almost every non-life actuarial textbook.

We start our presentation by a general assumption about the arrivals of the claims on policy level.

Assumption 3.6. (Compound Poisson arrivals of claims)

Let \( Y_j \) be a compound Poisson distribution with counting distribution \( N_j \) having intensity \( \lambda_j \) and severities \( \{X_{ji}, i = 1, 2, \ldots\} \). Assume that the loss on policy \( j \) has the distribution of \( Y_j \) which is independent of \( Y_i, i \neq j \).

Now the definition of the compound Poisson risk model is straight forward.

Definition 3.7. (Compound Poisson risk model)

Let \( P \) be an insurance portfolio of \( n \) policies. Under Assumption 3.6 the compound Poisson risk models is defined as the aggregate loss of \( P \),

\[
S = \sum_{j \in P} Y_j.
\]  

Since the loss on each policy is given by a compound Poisson process, multiple claims on a single policy are possible. In addition the heterogeneity in risk is preserved as each policy has a unique loss distribution \( Y_j \). Further the aggregate loss of the portfolio has a compound Poisson distribution according to Theorem 2.9. Hence the model has all the desirable properties of both the collective and individual risk model.

Note that other compounding distributions could be used as well, but with the loss of the property of closure under convolution. Essentially we are just replacing the compound Bernoulli distribution of (3.4) in the individual risk model by another compound distribution, in this case the compound Poisson. Actually we could see individual risk models as a general class of risk models that are sums over compound distributions with the compound Poisson and compound Bernoulli (standard) as special cases. However in our aim to align
definitions with existing literature, we shall mean the Bernoulli case when referring to the individual risk model.

We can derive inference on portfolio level by use of formula (2.16) and Corollary 3.4 since the losses have a compound Poisson distribution. By Theorem 2.9, the severity distribution of the aggregate compound Poisson distribution is a mixture of \( n \) independent variables \( X_j \). Hence the portfolio inference is just a sum over the underlying expectations and variances, respectively.

\[
\begin{align*}
\mathbb{E}(S) &= \lambda \sum_{j \in P} \frac{\lambda_j}{\lambda} \mathbb{E}(X_j) = \sum_{j \in P} \lambda_j \mathbb{E}(X_j) \quad \text{(3.18)} \\
\mathbb{V}(S) &= \lambda \sum_{j \in P} \frac{\lambda_j}{\lambda} \mathbb{E}(X_j^2) = \sum_{j \in P} \lambda_j \mathbb{E}(X_j^2) \quad , \quad \text{(3.19)}
\end{align*}
\]

where \( \lambda = \sum_{j \in P} \lambda_j \).

### 3.5.1 Relation to the individual risk model

Besides having a natural interpretation, the compound Poisson risk model also has a connection to the individual risk model. Clearly it can be interpreted as a standard individual risk model with \( \mathbb{E}(I_j) = 1 \) and a compound Poisson distribution as the severity distribution. However this contradicts with the normal interpretation of \( I_j \) as we implicitly assume that all policies will experience a loss. This is compensated for by a severity distribution (the compound Poisson distribution) having a point mass at zero, but this is indeed not a desirable way to interpret an individual risk model. Instead we will see how the assumption of the compound Poisson risk model leads to a more natural interpretation of the components of the it.

Consider an individual risk model under Assumption 3.6. That is that claims actually occur according to a compound Poisson process on each policy, but we model the risk using the individual risk model. Note that this assumption excludes the interpretation of the individual risk models as having only one claim and hence we interpret it in the sense of collapsing data as described in Section 3.3. Throughout this section (3.5.1) we will mean the individual risk model under this interpretation when referring to the individual risk model.

Under this model the Bernoulli claim count for policy \( j \) has the interpretation as the probability of at least one claim, \( q_j = \mathbb{E}(I_j) = P(N_j > 0) = 1 - e^{-\lambda_j} \). The severity r.v. \( \tilde{X}_j \) is a little more complex to understand. As argued previously we should let this variable express losses on policies rather than on single claims. Remember that the severity in the individual risk model is conditional on the occurrence of a loss. Because several claims can occur on a single policy (Assumption 3.6), \( \tilde{X}_j \) is a sum of \( N_j | N_j > 0 \) claims of size \( X_{ji} \), hence

\[
\tilde{X}_j = \sum_{i=1}^{N_j} X_{ji} | N_j > 0 = Y_j | N_j > 0.
\]

That is the claim severity variable is a compound Poisson distribution \textit{conditional} on the occurrence of at least one claim. The occurrence is
determined by the Bernoulli indicator $I_j$. We can show that the distribution of a loss, using $\tilde{X}_j$ and $I_j$, is in fact equal to the distribution of the underlying compound Poisson process $Y_j$.

\[
P(\tilde{Y}_j \leq y) = \mathbb{P}(I_j = 0) + \mathbb{P}(\tilde{X}_j \leq y) \mathbb{P}(I_j = 1)
= \mathbb{P}(N_j = 0) + \mathbb{P}(Y_j \leq y|N_j > 0) \mathbb{P}(N_j > 0)
= \mathbb{P}(Y_j \leq y|N_j = 0) \mathbb{P}(N_j = 0) + \mathbb{P}(Y_j \leq y|N_j > 0) \mathbb{P}(N_j > 0)
= \mathbb{P}(Y_j \leq y) \iff Y_j \overset{d}{=} \bar{Y}_j = I_j \cdot \tilde{X}_j.
\]

Here we make use of the law of total probability in the first equation where there is a point mass $\mathbb{P}(I_j = 0)$ at zero. Next, the individual risk model quantities $I_j$ and $\tilde{X}_j$ are substituted with their compound Poisson origins and finally we use the law of total probability inversely to arrive at the result. Note that $\mathbb{P}(Y_j < y|N_j = 0) = 1$, $\forall y \geq 0$ since $Y_j = 0$ when no claims occur.

**Comparison of mean and variance**

Though the above argument illustrates the relationship between the individual risk model and the compound Poisson risk model sufficiently, we would like to demonstrate that the mean and variance are equal in the two models. First we need a useful lemma.

**Lemma 3.8.** (Conditional expectation to Compound distribution)

Let $S$ be a compound distribution with counting distribution $N$ and severity distribution $X$. Then,

\[
\mathbb{E}(S) = \mathbb{E}(S|N > 0) \mathbb{P}(N > 0).
\]  

**Proof.** Follows by the law of iterated expectations.

\[
\mathbb{E}(S) = \mathbb{E}(\mathbb{E}(S|N \geq 0)) = \mathbb{E}(S|N > 0) \mathbb{P}(N > 0) + \mathbb{E}(S|N = 0) \mathbb{P}(N = 0)
= \mathbb{E}(S|N > 0) \mathbb{P}(N > 0) + 0 \cdot \mathbb{P}(N = 0)
= \mathbb{E}(S|N > 0) \mathbb{P}(N > 0).
\]

Using the expectation formula (3.5) for the individual risk model and Lemma 3.8 it is easy to see that the expectations to the individual risk model $\tilde{S}$ coincide with the expectation to the compound Poisson risk model $S$.

\[
\mathbb{E}(\tilde{S}) = \sum_{j \in P} \mathbb{E}(\tilde{Y}_j) = \sum_{j \in P} \mathbb{E}(I_j) \mathbb{E}(\tilde{X}_j) = \sum_{j \in P} \mathbb{E}(\tilde{X}_j) \mathbb{P}(N_j > 0)
= \sum_{j \in P} \mathbb{E}(Y_j|N_j > 0) \mathbb{P}(N_j > 0) = \sum_{j \in P} \mathbb{E}(Y_j) = \mathbb{E}(S).
\]
The variance requires a little more computation but the principles are the same.

\[
V(\bar{S}) = \sum_{j \in P} V(\bar{Y}_j) = \sum_{j \in P} q_j V(\bar{X}_j) + q_j (1 - q_j) E(\bar{X}_j)^2
\]

\[
= \sum_{j \in P} q_j V(Y_j | N_j > 0) + q_j (1 - q_j) E(Y_j | N_j > 0)^2
\]

\[
= \sum_{j \in P} q_j \left( E(Y_j^2 | N_j > 0) - E(Y_j | N_j > 0)^2 \right) + q_j E(Y_j | N_j > 0)^2 - q_j^2 E(Y_j | N_j > 0)^2
\]

\[
= \sum_{j \in P} E(Y_j^2 | N_j > 0) P(N_j > 0) - E(Y_j | N_j > 0)^2 P(N_j > 0)
\]

\[
= \sum_{j \in P} E(Y_j^2) - E(Y_j)^2
\]

\[
= \sum_{j \in P} V(Y_j) = V(S).
\]

(3.25)

Here we use Corollary 3.5 and Lemma 3.8 combined with equation (2.4). The rest of the derivation is simple manipulation and substitution.

### 3.5.2 Relation to the collective risk model

**General decomposition**

We shall now see how the compound Poisson risk model relates to the collective risk model. Most of the discussion is in fact general for all the risk models, and we will emphasize when it only concerns the collective risk model. We will refer to the principle presented in this section as decomposition of risk. The principle is illustrated by an example.

**Example 3.9. (Decomposition of risk)**

Consider an insurance company offering building and motor insurance. Since there is no reason to believe that the severity distributions of these two portfolios are identical (though they might be) it seems reasonable to model the buildings risk and motor risk separately by applying any risk model to two separate dataset containing each of the portfolios. In doing so we realize that the coverages on the motor policies are very different in nature. One coverage has to replace the car in the case of a car fire which may easily cost £10,000. Another coverage insures damages to the windows from small stones thrown up from the road. Since the replacement cost of a front window may be around £1,000 it is clear that the severity distributions for these two coverages are quite different. This motivates that we once again decompose the risk into smaller datasets that each includes data only from a separate coverage. Now consider the glass coverage alone. One may realize that there are segments of different risk in the portfolio. Policies signed by customers who drive more than 18,000 miles per year may have a higher frequency than customers who drives 3,000 miles per year (since more hours on the road results in higher exposure to glass damages). Further we may observe that the window replacement costs (severity) are different for Audi and Mazda.
When to stop this decomposition of risk is a delicate question. The smaller the subset of data becomes the higher is the variability in the distribution estimates. Fortunately methods exist which can estimate distributions on very small dataset while still preserving the large scale robustness of the estimates. We will return to these in Chapter 4. Note that the decomposition is a generic principle which applies to all risk models.

When risk estimates are calculated on the bottom level of a risk decomposition (say on coverage level) we can sum up the result to achieve estimates on policy level, line of business (LoB) level or company level. We demonstrate the principle for a portfolio with $|L|$ lines of business (LoB), $|C|$ types of coverages for LoB $l$ and $|P_{lc}|$ policies for coverage $c$ in LoB $l$. Formally we have for $P_{lc}, C_l, L$ with $j = 1,..., |P_{lc}|$, $c = 1,..., |C_l|$, $l = 1,..., |L|$, 

$$S = \sum_{l \in L} \sum_{c \in C_l} \sum_{j \in P_{lc}} S_j(l, c)$$  \hspace{1cm} (3.26)

where $S_j(l, c)$ is the risk premium of the $j$'th policy on the $c$'th coverage in the $l$'th LoB. Note that an implicit assumption behind such a summation is independence between the risk premiums $S_j(l, c)$. Though the notation here makes the principle look complex it is not. For example if a motor insurance premium has to be estimated, the claims are divided into two components that represent the coverages of the portfolio: Comprehensive (casco) and third part liability (TPL). Then the risk premium is estimates for each separate dataset yielding a price of £300 for the TPL and £500 for the comprehensive coverage. The risk premium of the policy is then simply $S = \£300 + \£500 = \£800$.

Throughout this paper we will always assume that we are dealing with a model on the lowest decomposition level. Hence we can exclude the cumbersome notation $S_j(l, c)$ for the simpler $S_j$ since only one coverage is dealt with. Note that other decompositions than mentioned here may be used instead. The task of aggregating the separate parts in a particular application are easily handled by following the principle above.

**DECOMPOSITION INTO POLICY SEGMENTS (ONLY COLLECTIVE RISK MODEL)**

We now return to a discussion regarding the collective risk model only. We will demonstrate how the decomposition principle naturally leads to the compound Poisson risk model. For the discussion we will assume that data is already decomposed into a given coverage. Example 3.9 illustrates there may be differences in the repair cost for different car brands (and thereby policies). This is problematic when applying the the collective risk model which assume iid. severities. However we can use the principle of decomposition to split the data, beyond coverage level, into policy segments in which policies share some common characteristics (e.g. they are all Audis, Citroëns, etc.). We can proceed this process until we will comfortable with the assumption of iid. severities. Then the collective risk model can in fact be applied on each segment without compromising its assumptions of iid. severities. As the number of the segments increases (a more granulated decomposition) we will in the extreme case have a segment for each policy, hence $n$ segment in a portfolio of $n$ policies. Since the collective risk model has a compound Poisson process for each segment we have in fact a compound Poisson process for each policy. This is exactly the approach taken in
the compound Poisson risk model from Definition 3.7. Note here, that the general discussion on decomposition had no special relation to the collective risk model. But by applying the principle to a deeper level the compound Poisson risk model naturally arises from the collective risk model.

The above discussions show that the compound Poisson risk model has a natural and intuitive relationship to the collective risk model and, if we apply Assumption 3.6, it can be interpreted under the individual risk model. In the next section we will briefly discuss the pros and cons of each model and select a candidate for the rest of the paper.

### 3.6 Summary of risk models

We now summarize the risk models presented so far. The collective risk model is a simple compound Poisson process over a stochastic number of claims. The drawback of the model is that it assumes all claims to have the same severity distribution (homogeneity). The individual risk model has unique severity distribution for each policy (heterogeneity) but relies on an assumption of maximum one claim per policy per time interval. Since most insurance contract allows for multiple claims this assumption makes the model inappropriate (though justified by simplicity). The approximation to the individual risk model captured the desirable features of each of the classical risk models, but its interpretation was very difficult when several claims on a single policy was possible. Finally we presented the compound Poisson risk model that also inherited the nice properties of the classical risk models and further had a natural interpretation. The model simply assumed a compound Poisson process per policy. We saw how this approach naturally arose from the collective risk model due to the principle of decomposition. Further, it was illustrated how the model related to the individual risk model. We summarize the models of this chapter in Table 3.1.

<table>
<thead>
<tr>
<th>Model</th>
<th>Closed under convolution</th>
<th>Heterogeneous risk</th>
<th>Natural interpretation of variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual risk model</td>
<td>No</td>
<td>Yes</td>
<td>Yes *</td>
</tr>
<tr>
<td>Collective risk model</td>
<td>Yes</td>
<td>No</td>
<td>Yes **</td>
</tr>
<tr>
<td>Approximation to IRM</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Compound Poisson risk model</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes **</td>
</tr>
</tbody>
</table>

* Claim variable as the loss on the policy given there is at least one claim on the policy.

** Counting variable as an indicator of whether at least one claim has occurred on the policy.

Table 3.1: Properties of risk models.

In the subsequent chapters we will use the compound Poisson risk model due to its desirable properties. In the following chapter we shall see how the estimation is easily performed.
3.7 Bibliographic notes

Mathematical risk theory and the models presented in this chapter has long historical merits. It origins back to the work of Filip Lundberg, a celebrated Swedish actuary, who wrote the first works on what should become mathematical risk theory in the start of 20\textsuperscript{th} century. Two of the early textbooks (in modern time) which has become very popular are Gerber (1979) and (Grandell, 1991). More present book are (Boland, 2007), (Kaas, Goovaerts, Dhaene, and Denuit, 2009) and (Tse, 2009) which all are good choices.
4 Insurance risk: Application

4.1 Conditional risk measures

The risk models of Chapter 3 give a framework to determine risk at both aggregate and individual level. The current chapter proceeds along this line and explains how to estimate the parameters of these models. The methods presented are quite general and can be used for all the risk models, but for the discussion we will have the compound Poisson risk model in mind. We will assume that the estimation is carried out on the bottom level of the risk decomposition, i.e. on a dataset only containing observations from a segment of a larger dataset (we have coverage in mind here).

Individual policies and policy segments have been an important part of the discussion of the risk models. It should be quite obvious how to separate policies as they are distinct by nature. Meanwhile the definition of a segment is less clear. We assume that every policy is supplied with some customer and policy specific information. For a motor insurance this may be motor power, weight of car, age of driver or annual mileages. On a building insurance it may the value of the building, number of habitants or the size of the building in square meters. On a contents insurance it may be the maximum value of the content insured, age of the policy holder or whether a burglary alarm has been installed or not. On personal accident insurance occupation and age may be important factors to ask the customer for. We will assume that all this information on policy $j$ is available in a column vector $Z_j = \{z_{j1}, z_{j2}, ..., z_{jm}\}'$ which we will denote as the covariates of policy $j$. Here $'$ denotes the transpose to a matrix or a vector. By a segment we understand a group of policies which share characteristics in their covariates e.g. all customers older than 50 year, driving a car with less than 100 horsepower. The information required to classify policies into segments is available in the matrix of all covariates, $Z = \{Z_1, ..., Z_n\}'$. We say that two policies are in the same segments if they are indistinguishable in the covariates, i.e. $Z_i = Z_j$. If we want to segment by only a subset of the covariates we can just redefine $Z_j$ to only include these.

A portfolio $\mathcal{P}$ can in principle consist of both single policies and policy segments. The formulas of aggregate losses given so far, e.g. (3.26), are not affected by this since the counting process of a segments will just have a correspondingly smaller Poisson intensity (and thereby greater mean) than a single policy. Related to this intensity is a time exposure $\tau_j$. A segment with $m$ policies each having an intensity of $\lambda$ will, in a probabilistic sense, act a single policy with intensity $m\lambda$ due to independence. While without influence on the expectation formulas, the exposure plays a prominent role in the estimation of the counting distribution. For a detailed description of the role of the exposure we refer to (Ohlsson and Johansson, 2010, chapter 1). We will in general let a portfolio $\mathcal{P}$ consist of all the individual policies for simplicity (i.e. no segments). Below is given a formal definition of claim frequency, claim severity and risk premium of the individual customers based on the compound Poisson risk model.
Definition 4.1. (Claim severity)
Under the compound Poisson risk model of Definition 3.17, the claims severity $\mu_j$ is the conditional expectation to the severity distribution of $X_j$ of policy $j$ given $Z_j$,

$$\mu_j = \mathbb{E}(X_j | Z_j) . \quad (4.1)$$

The unconditional expectation is called the flat rate claim severity,

$$\bar{\mu} = \bar{\mu}_j = \mathbb{E}(X_j) . \quad (4.2)$$

Note that the expectation $\mu_j$ to the severity distribution $X_j$ is conditional on the covariates. Hence all policies in the same segment will have the same (expected) claim severity, $\mu_j = \mu_i$ for $Z_j = Z_i$. A similar conditional structure applies to the claim frequency.

Definition 4.2. (Claim frequency)
Under the compound Poisson risk model of Definition 3.17, the claims frequency $\lambda_j$ is the conditional expectation to the number of claims $N_j$ of policy $j$ given $Z_j$,

$$\lambda_j = \mathbb{E}(N_j | Z_j) . \quad (4.3)$$

The unconditional expectation is called the flat rate claim frequency,

$$\bar{\lambda} = \bar{\lambda}_j = \mathbb{E}(N_j) . \quad (4.4)$$

Note that the definition of both the claim severity and frequency have an unconditional flat rate equivalent. For both cases we do not use the information related to the policy segment, $Z_j$. Hence the expectation is the same for all policies in the portfolio.

The risk premium is a measure for the total monetary loss on the policy. Since the compound Poisson risk model calculates this by a compound Poisson distribution, we use (2.16) to see that the risk premium is just the product of the expectations to frequency and severity.

Definition 4.3. (Risk premium)
Under the compound Poisson risk model of Definition 3.17, the risk premium $S_j = Y_j$ is the conditional expectation to the aggregate loss of policy $j$ given $Z_j$,

$$\mathbb{E}(S_j) = \mathbb{E} \left( \sum_{i=1}^{N_j} X_{ji} | Z_j \right) = \mathbb{E}(N_j | Z_j) \mathbb{E}(X_j | Z_j) = \mu_j \lambda_j . \quad (4.5)$$

The unconditional expectation is called the flat rate risk premium,

$$\mathbb{E}(\bar{S}) = \mathbb{E}(\bar{S}_j) = \mathbb{E} \left( \sum_{i=1}^{N_j} X_{ji} \right) = \mathbb{E}(N_j) \mathbb{E}(X_j) = \bar{\mu}_j \bar{\lambda}_j = \bar{\mu} \bar{\lambda} . \quad (4.6)$$

The topic of the following section is how to estimate the quantities defined above.
4.2 Estimation of conditional risk measures (GLM)

Generalized linear models are an extension of classic linear models, also known as ordinary least squares. The classical method has origins back to 1794 when Carl Friedrich Gauss first described the method (though it was first published years later). However the classical method only applies under a normal distribution assumption. (Nelder and Wedderburn, 1972) presented the method of generalized linear models (GLM) which extended the classical method to include other distributions as well. In 1983 the textbook (McCullagh and Nelder, 1983) appeared and became (and still is) a standard reference on the topic. Below we will briefly describe the method, but the reader is assumed to have some experience with GLM in advance. We describe the method in relation to insurance but should emphasize that it is of general character and is used in such diverse fields as medicine, marketing and engineering.

Generalized linear models is a method for linking the mean of a response \( Y_j \) to a set of covariates \( Z_j = \{z_{j1}, ..., z_{jm}\}' \) through a linear function. GLM differs from classical regression analysis by two important aspects. First of all the response distribution do not need to be normal, but can be any distribution in exponential family\(^4\). This is desirable in relation to insurance since claims are mostly positive (\( \mathbb{R}_+ \)) which makes the normal distribution inappropriate (\( \mathbb{R} \)). Further, the variance of the normal distribution is independent of the mean (homoskedasticity) resulting in the same variability in estimates for high and low risk policies. This may not be satisfactory since we would usually expect a higher variability for larger claims.

The second difference is that the linearity of the means can be on a different scale under GLM. Hence if we model the mean of the log response, the model will be additive on the log scale, but multiplicative on the original scale. This allows for relative risk differences in premiums rather than additive differences, which may often be preferable.

Following the discussion in Chapter 4.1 the covariates could be the age of the policy holder, value of a car or geographical location. Hence the covariates can be continuous (such as age) or discrete (such as geographical location). We will refer to continuous variables as *variates* and discrete variables as *factors*. The linearity in factors is achieved by introducing a set of dummy variables for each factor. Assume that geographic location is coded with three levels: "Central city", "Urban" and "Other". We then replace the original variable with three new indicator variables \( z_{j1}, z_{j2} \) and \( z_{j3} \) such that \( z_{j1} = 1 \) if the original variable is "Central city", \( z_{j2} = 1 \) if the original variable is "Urban", \( z_{j3} = 1 \) if the original variable is "Other", while \( z_{ji} = 0, i = 1, 2, 3 \) otherwise. Note that these three dummy variables hold the same information as the original geographic coding but are now represented only by binary numbers. When referring to covariates in technical discussions we will mean the set of dummy variables together with the variates. For qualitative discussion the distinction between the dummy variables and the corresponding factor is superfluous. The original factors will never be used directly in calculations.

\(^4\)The family of exponential distributions include among others the normal, binomial, Poisson and gamma distributions.
For the normal distribution case the linearity of a GLM model could be,
\[ Y_j = \beta_0 + \beta_1 z_{j1} + \ldots + \beta_m z_{jm} + \epsilon_j = \mu_j + \epsilon_j, \] (4.7)
where
\[ \mathbb{E}(Y_j|z_{j1}, \ldots, z_{jm}) = \mu_j = \beta_0 + \beta_1 z_{j1} + \ldots + \beta_m z_{jm} \] (4.8)
and \( \epsilon_j \) is a normally distributed error term. This can be written in vector notation as,
\[ \mu_j = Z_j' \beta \] (4.9)
with \( Z_j = \{1, z_{j1}, \ldots, z_{jm}\}' \) and \( \beta = \{\beta_0, \beta_1, \ldots, \beta_m\}' \). Since such an equation exists for each policy we have a whole system of equations,
\[ \mu = Z \beta, \] (4.10)
where \( \mu = \{\mu_1, \ldots, \mu_n\}' \) and \( Z = \{Z_1, \ldots, Z_n\}' \). In the terminology of GLM the matrix \( Z \) is denoted the design matrix, \( \beta \) is the parameter vector, \( \mu \) is the mean vector and \( \epsilon = \{\epsilon_1, \ldots, \epsilon_n\}' \) is the error vector.

One of the great advantages of GLM over classical regression is the introduction of a link function \( g \). The link function makes it possible to model on a different scale. Hence if we let \( g(y) = \log(y) \) we model the means of the log responses. To calculate the estimates on the ordinary scale we have to take the exponential function (inverse link) to a linear expression which gives a multiplicative expression. The system of equations in (4.10) is slightly changed to include the link function,
\[ g(\mu) = Z \beta. \] (4.11)
For the normal distribution the link function is just the identity, i.e. \( g(y) = y \). For the Poisson distribution the log link is usually used and hence the positive responses are mapped from \( \mathbb{R}_+ \to \mathbb{R} \). In this way no restrictions are to be made on the linear predictors as the inverse mapping (exponential) will send the estimates back on \( \mathbb{R}_+ \) no matter how negative they might be.

The overall goal of setting up an GLM model is to estimate the mean vector \( \mu \) (our expectation to the response). In doing so we have to estimate \( \beta \). This is done by maximum likelihood estimation and often requires numerical optimization such as the Newton-Raphson method. The likelihood function of \( n \) independent observations is,
\[ L(y_1, \ldots, y_n|\beta, Z) = \prod_{j=1}^{n} f(y_j|\beta, Z_j), \] (4.12)
where \( f \) is the probability density function of the response distribution (observations are assumed iid.). The values of \( \beta \) that maximizes \( L \) is called the maximum likelihood estimates. To simplify optimization one usually takes the logarithm to \( L \) to get the log likelihood,
\[ l(y_1, \ldots, y_n|\beta, Z) = \log L(y_1, \ldots, y_n|\beta, Z) = \sum_{j=1}^{n} \log f(y_j|\beta, Z_j). \] (4.13)
As the logarithm is a monotonously increasing function the values of $\beta$ that maximizes $L$ are identical to those which maximize $l$. Hence maximization is often performed by differentiating the simpler linear expression of the log likelihood $l$. Methods for optimizing the log likelihood on an automated basis are implemented in most analytical software such as SAS, R or SPSS. In general we will use the hat symbol $\hat{}$ to indicate if a quantity is estimated. Thus the output from the GLM analysis is $\hat{\beta}$.

GLM’s relation to the insurance models should be quite clear from the notation. The severities and the frequencies will be the responses of two separate GLM estimations. Since the policy information $Z_j$ is available for each policy these will serve as covariates in the analysis.

One assumption of GLM is that the observations are independent. Note that this is fulfilled for the severity case because the claims are iid. within policies and independent between policies. For the frequency the Poisson process is the key to independence. Even when the same policy is represented in multiple observations the observations will be iid. due to the independent increments of the Poisson process (as long as the time periods of the observations are disjunct).

The output from the GLM analysis can be used to calculate the risk premium (expected loss) on each policy according to Definition 4.3. Note that that risk premium will rarely predict the losses perfectly but only on average. As long as the premiums are segment-wise fair in the sense that each group of similar looking risk has a correct risk estimate there should be no problems with anti-selection.

As mentioned in the discussion of the risk models we would present a method to deal with the small scale variability in the risk estimates. GLM is this tool! By use of GLM we estimate the relationship between the mean and the covariates through $\beta$ based on the entire dataset. This gives very few parameters compared to individual policy estimates. The combination of more data and fewer parameters gives much more reasonable estimates (less influence of random fluctuation). Hence the idea of decomposing the portfolio of a giving coverage into segments (as described in Section 3.5.2) is indeed possible to apply in practice.

4.3 Bibliographic notes

Generalized linear models is a widely used and described subject. The models first appeared in (Nelder and Wedderburn, 1972). Later textbook treatments include (McCullagh and Nelder, 1983) and the more present (Dobson and Barnett, 2008). The first actuarial textbook including GLM was apparently (Kaas, Goovaerts, Dhaene, and Denuit, 2001, chapter 8) which afterwards has been extended with an extra chapter on the topic. For a short, but still comprehensive presentation of GLM with emphasis on insurance we highly recommend (Boland, 2007, chapter 7). Finally (de Jong and Heller, 2008) and (Ohlsson and Johansson, 2010) must be mentioned as they are totally devoted to GLM on insurance data. Especially the later book is of interest in relation to the present paper as it takes the same approach as the compound Poisson risk model.
5 Inflation models

The risk models of Chapter 3 are the foundation for our next developments. A key problem with these models is that they do not take into account the risk development over time. In our everyday life we experience how prices change from year to year, typically upwards. This is due to inflation, more specifically price inflation. Hence we can ask ourselves whether it is appropriate to pool severities from year $n$ with severities from year $n - 5$ and then estimate the means based on such a dataset. Most people will agree that it is not. The topic of the current chapter is to present some simple methods to deal with inflation. Especially the discussion of the DMI framework (discount, model, inflate) is interesting due to its simple structure. Further, the notion of inflation will be generalized beyond ordinary price inflation in order to deal with other types of inflation as well. Though the principles are not new, since they are often used in practice in various disguises, we here aim to formalize and generalize the structure of these methods.

To illustrate by how different patterns risk can evolve over time we present three real life examples in Figure 5.1.

(a) Decrease in severity for a coverage on a personal accident insurance.
(b) No change in severity for a coverage on a household insurance.
(c) Increase in severity for a coverage on a motor insurance.

Figure 5.1: Smoothed indexes of scaled mean severities (base month is one) for different coverages from a real life insurance portfolio. The actual coverage names are left out for competitive reasons.

The figure shows that various inflation patterns can occur. Further, the magnitude of panel (c) illustrates that the inflation may be a non-negligible factor (over 60% increase over 55 months).

5.1 Structure and motivation

Before presenting the inflation models we discuss how they relate to the compound Poisson risk model. A feature of this model is that it has a separation in claim frequency and claim severity. In principle this division is not strictly necessary as the risk premium could just be modelled directly, which is in fact the approach taken by an interesting class of models
called Tweedie models. However, the division has proven useful since a frequency model and severity model often capture different aspects of the risk. In a completely similar manner, we now extend the severity model by splitting it into an inflation part and base severity model which is cleaned for the inflation effect. Hopefully, an improvement of the risk model is achieved by adding this extra component to the risk models. In Chapter 6, some tools are given to determine whether this is the case or not. The details of the mathematical relationship are given later, but for now, just think of it as $\mu = \mu_b f(I_k)$. Here $\mu_b$ is the base severity estimate, and $f(I_k)$ is a monotone function of the inflation. The multiplication of these two quantities gives the actual severity estimate. The principle is illustrated in Figure 5.2.

![Figure 5.2: Components of the risk premium](image)

The motivation for introducing this additional division of the severity model has roots back to the discussion of risk changes in Section 1.4. As argued, we believe that a change in the risk pattern is characterized by changes in risk profile or risk level (or both). If we cannot identify changes in risk profile correctly we may be exposed to anti-selection without our notice. On the other hand, a lack in the monitoring of risk level will (typically) result in insufficient premiums causing the profit of the company to fall. Both of these aspects of risk is indeed important to understand and mirror in the risk models. However, the changes in risk level are far simpler in their nature than changes in risk profile. Hence, the motivation for splitting the severity into a base model and an inflation part is that the inflation part can be automated to a very high degree on a modern computer. This has three likely advantages. 1) It will liberate actuarial resources which can be focused on the changes in risk profile. 2) The level adjustments of the risk premium may improve as they change more frequently than when performed manually. 3) The statistical work of estimating changes in risk profile just seems more interesting to execute than simple level adjustments. Hence, the work of the actuary becomes more interesting and complex while the computer handles the statistically uninteresting level changes. Hopefully, this justifies the apparently extra work involved in the implementation of inflation models.

To convince the reader that level changes are simpler than full-scale reestimation of the parameters $\hat{\beta}$, we need to consider the tasks involved in a successful implementing a new tariff. These tasks may or may not be executed by an actuary, but in every case, they will demand resources from some department in the company (if they are relevant).
Data generation

Covariates, claim history, customer information and policy terms need to be gathered to a large dataset ready for estimation. Usually these data will come from many different sources and the task of gathering and consistency-test data can be time consuming.

Estimation

The decomposition of a premium into coverages gives a lot of models to estimate.

Model control/evaluation

All the models has to be verified as valid according to the model assumptions and perhaps internal political considerations. As an example it is not desirable that the premium increases with the size of the deductible or duration of the customer, even when empirical findings shows that this is the case.

Market analysis

The new tariffs relation to competitor prices is very important in order to stay competitive.

Revision of market adjustments

There may be adjustments to the market price in order to meet certain political requirements or the price sensitivity of the customers. These will need revision.

IT implementation

Rarely statistical models from the analysis software can be executed directly. Usually some kind of IT implementation is needed.

IT tests

The IT implementation gives an extra possibility of errors. The tariff in the final IT systems has to be verified against the analytical software.

Internal communication

The employees outside the actuarial department have to be informed about the changes made and their effect in different customer segments.

External communication

Perhaps the new premium has to be communicated to some external organisation. This could be some public authority, trade association or price aggregator.

Pruning of existing customers

The existing customers have to be informed about the price change so they can accept or leave the contract.

Forecasting

Change of a tariff will usually call for forecasting of many of the of the key performance indicators of the budget.

The main difference between an update of a traditional risk model and an inflation risk model is that under the later structure only an inflation index is changed while the base risk
estimates remains unchanged. For a traditional risk model all the parameter estimates \( \hat{\beta} \) are updated. A risk model (base or not) is intended to differentiate risk estimates between customers. Assume that two policies \( A \) and \( B \) are priced \( P_A > P_B \) under an old tariff. Now if the risk model is completely re-estimated, we have for the new price (indicated by *) that \( P_A^* \gtrsim P_B^* \) because the relative risk between customers may change in any direction due to new parameter estimates. When updating only the inflation part of a risk model, the prices will be changed proportionally, \( P_A^* = f(I_k) \cdot P_A > f(I_k) \cdot P_B = P_B^* \) for some scaling constant \( f(I_k) > 0 \). In this way the relative risk between the customers remains the same. This affects most of the listed elements of a tariff update. Communication (internal, external, pruning) becomes much easier as it can be expressed by a percentage increase. A simple term understood by customers, marketing colleagues, leaders, etc. The data needed for an inflation index update only requires claim history, but no covariates, customer information, etc. Hence the data generation is much simplified. IT implementation also becomes easier as only one number has to be changed, instead of a full GLM structure (\( \hat{\beta} \)). Clearly this is reflected in the IT test too. Market analysis, market adjustments and forecasting are still affected, but since all segments experience the same relative increase it is usually a matter of scaling of the results of previous analysis. Most important of all is perhaps that the update of an inflation index is likely to be a familiar job since annual price inflation adjustments are standard procedure in many companies. The only change is the actual magnitude of the inflation. We will clarify this in the following sections.

The notion of inflation used here will be general. Normally inflation is related to the increase of monetary units (prices, wage, etc.) over time. Especially price inflation seems relevant to insurance as this affects the severities. But, as Figure 5.1.c shows, inflation in severities can far exceed ordinary price inflation, which in the same period only accounted for a 10% increase. This indicates the existence of other types of inflation which for example could be changes in crime pattern, fraud pattern, legal changes, technology improvements or changes in contract terms. The list is long and it seems an impossible task to measure the different kinds of inflation in details for each coverage. This leads us to a very important principle of the models to be presented. We will work with only one inflation, namely the observed inflation in the severities. The origin of the inflation is indeed interesting but even more important is it to adjust the premiums properly when changes in the risk level occur. By treating all inflations at once we ensure that no import inflations are forgotten. If the inflation is strong enough to manifest itself in data we will use it and adjust models accordingly. Note also that this principle allows for inflation in the claim frequencies even though these are not related to any price inflation (since they are counts). This is convenient since many of the other types of inflation may affect the frequency to change and hence an inflation discussion is just as relevant here as for the severities. Unfortunately the main model to be presented are not directly applicable to frequency models for reasons that will be clear later. Hence we will proceed our discussion with severities in mind.
5.2 Measurement of inflation

5.2.1 Simple measures of inflation

We now introduce some simple measures of inflation that will be used in the models. There should be made a distinction between an inflation and an inflation index. The former denotes the relative changes in percent whereas the later just measures some development under time (usually monetary). Usually inflation indexes are scaled such that the value is 1 or 100 in some base period. Since we will use inflation indexes almost exclusively here we will mostly refer to these as just "inflation" for convenience. The inflation will be measured over months and in some special cases years, but note that the methods work on any approximately equally spaced time intervals.

Price inflation

Incorporation of price inflation is common in practice. Since this paper suggests an improvement to these models we shall use price inflation as a benchmark. Clearly price inflation may differ for repairs on households and replacement cost for new cars. Hence the detail level of a price inflation should be judged from the situation. In the empirical chapter we will just use an overall consumer price index publish by Statistics Denmark (table PRIS6). Since this index is monthly we generate an annual version by simply selecting the first value each year.

Raw inflation

The raw inflation estimates are the core in our improvement of the traditional models. It measures the empirical inflation in the severities over \( n \) disjunct time periods of arbitrary length (here months). For a given coverage let \( \bar{\mu}_n, n \in \mathbb{Z} \) be the mean severity (flat rate) of all claims reported in period \( n \) on the coverage. The inflation (not the index!) from period \( n - 1 \) to \( n \) is then simply calculated by,

\[
i_n = \frac{\bar{\mu}_n}{\bar{\mu}_{n-1}} - 1, \quad n \in \mathbb{N}_+ .
\]

We let the starting point for the period enumeration be such that the first month in the estimation data is \( n = 1 \). Periods \( n < 1 \) will be ignored though they are well defined. Now with the convention that \( \bar{\mu}_0 = \bar{\mu}_1 \) (first month is the base month), we define the inflation index,

\[
I_n = \frac{\bar{\mu}_n}{\bar{\mu}_0} .
\]

Note the relationship,

\[
I_n = \frac{\bar{\mu}_1 \bar{\mu}_2 \cdots \bar{\mu}_{n-1} \bar{\mu}_n}{\bar{\mu}_0 \bar{\mu}_1 \cdots \bar{\mu}_{n-2} \bar{\mu}_{n-1}} = (1 + i_1)(1 + i_2) \cdots (1 + i_{n-1})(1 + i_n) = \prod_{j=1}^{n} (1 + i_j) .
\]
this in Section 5.3. We will use the convention of always moving the severities back in time to period one when discounting.

**No inflation**

Some of the subsequent models do not incorporate inflation at all. Instead of omitting inflation from the models we introduce an inflation (index) which is just 1 in every period. In this way we can keep a generic modelling approach and still leave out inflation by just using this constant index.

### 5.2.2 Smoothed and filtered measures of inflation

Insurance is related to accidents. Fortunately these are relatively rare but for the same reason the data amount on severities is small. In addition, variance in the losses makes the observations quite heterogeneous. Consequently the raw estimates of the inflation can be quite volatile from month to month. This is problematic since a high (or low) observed inflation in a given month not necessarily captures the general inflation trend. We solve this problem by introducing a smoother that aims to extract the signal (the "true" inflation) from a noisy measure (the raw inflation). Assume we have an raw inflation time series \( \{ I_n, n = 1, \ldots, R \} \). A smoother is then designed to give the best estimate of the signal behind \( I_k \) based on all the observations including \( I_m, m > k \).

In practice we cannot use future observations when we want to estimate the true inflation today. This introduces the concept of filters. Filters also aim to extract the signal but are only allowed to use historical observations, \( I_m, m \leq k \). Note that all the simple inflation estimates of Section 5.2.1 are filters, though they do not make use of the past. A simple example of a filter that actually uses past observation is a simple moving average. In this section we present some filters and smoothers which are more advanced than those few examples mentioned above. To limit the scope of the paper we will not go into details with the statistical theory underlying these methods, but instead gives some guidance to relevant literature in bibliographic section.

Note that a filter may be used as a smoother (though often poor), but the opposite is not the case since the smoother uses information from the future. Even though the smoothers and filters aim to remove the noise and extract the true inflation there may be situations where the noise is too dominating. In such situations we suggest to measure inflation on a higher level, e.g., by collapsing three types of coverages to gain more data and thereby less volatility in the means. Risk models may still be estimated on coverage level.

To dampen the large claims effect on the smoother we model the log inflation and afterwards transform the smoothed log inflation back to the original axis. Note that, for simplicity, we will ignore exposure weights in our calculation of inflation and simply take equal weight on each observed raw inflation estimate. This can be problematic in some cases where the number of claims in different months varies much. Take as an example a yacht insurance which is mainly exposed during the summer. Then the few claims during the winter will
have the same impact on the smoothed inflation as the majority of the claims during the summer. Clearly there is room for improvement of the methods in this respect but we will ignore this problem here.

We will in general denote the inflation filter value of period \( n \) using filtering method \( M \) as \( \bar{I}_M^n \). Similarly we let \( \hat{I}_S^n \) denote the value of the inflation smoother in period \( n \) using methods \( S \). When superscripts \( M \) or \( S \) are omitted the estimator can be an arbitrary smoother or filter.

**Exponential moving average**

Below we present the formulas on the normal scale. To calculate on the logarithmic scale instead, simply take the logarithm to every \( I_k \) in the calculations and at the end transform the filtered value back to the normal scale using the exponential function.

As mentioned, the moving average (MA) is a simple filter. It simply takes the average over the \( p \) most present observations. The value of the MA with lag \( p \) in period \( n \) is

\[
\bar{I}_{MA,p}^n = \frac{1}{p} \sum_{j=n-p+1}^{n} I_j .
\]  

(5.5)

Though extremely simple and very useful in many situations, the MA method has a main drawback as it places equal weight on every observation in the mean. In many situations it would be more natural to put more weight on the most recent observations. The exponentially weighted moving average (EWMA) solves this problem. We introduce a weighting \( w_j \) of the moving average and choose this weighting to be exponentially declining.

\[
\bar{I}_{E,\lambda}^n = \sum_{j=0}^{n-1} w_j I_{n-j} = \sum_{j=0}^{n-1} \lambda(1-\lambda)^j I_{n-j} , \quad 0 < \lambda \leq 1 .
\]  

(5.6)

In this way we put most weight on the most recent observations. Note that the weights form a series of geometrically decaying numbers which sums approximately to one, \( \lim_{n \to \infty} \sum_{j=0}^{n} w_j = 1 \), even for relatively short time series. We can rewrite the formula to a recursion formula by extracting the first term.

\[
\bar{I}_{E,\lambda}^n = \lambda I_n + \sum_{j=1}^{n-1} \lambda(1-\lambda)^j I_{n-j} = \lambda I_n + \sum_{j=0}^{(n-1)-1} \lambda(1-\lambda)^{j+1} I_{(n-1)-j} \\
= \lambda I_n + (1-\lambda) \sum_{j=0}^{(n-1)-1} \lambda(1-\lambda)^j I_{(n-1)-j} = \lambda I_n + (1-\lambda) \bar{I}_{E,\lambda}^{n-1} .
\]  

(5.7)

We initiate the smoother by \( \bar{I}_{E,\lambda}^1 = I_1 \). From the recursion it is evident that the EWMA of the \( n \)'th observation is an weighted average of the EWMA of the \( (n-1) \)'th observation and the new observation \( I_n \). Due to the computational efficiency of this recursion the EWMA has been widely used in various statistical branches, especially in the period 1950-1960 where powerful computers were not yet available. Note that when \( \lambda \to 0 \) all weight is placed on the new observation which correspond to the raw inflation estimates. When \( \lambda \to 1 \) the filter is not updated with new information at all, hence \( \bar{I}_{E,\lambda}^n = I_1 \). Since the EWMA method only uses historical information it is a filter. An extension of the methods to include linear
trends exists, called the Holt-Winter EWMA. Since the computational effort of implementing EWMA is so small we will apply this and leave out the standard moving average.

**Generalized additive models**

Generalized additive models (GAM) is a generalization of generalized linear models. It is therefore natural to present the method using notation similar to that used for the GLM presentation in Section 4.2. Though GLM allows the model covariates to be functions of the original covariates (e.g. \( z^2 \) or \( \log(z) \)), there are situations where we have no reason to assume any specific functional structure. The GAM idea is to let data talk and estimate the dependence structure with as few assumptions as possible about the functional form of the relationship.

Let \((z_j, y_j), j = 1, ..., n, z_1 < ... < z_n\) be some pair of observations for which we want to determine a function \( f(z) \) that satisfies \( f(z_j) = y_j \forall j \). Splines are such functions that are piecewise polynomials joined in the \( z_j \)'s (called knots) such that their derivatives match to a certain order (smoothness) and the function match the observation, \( f(z_j) = y_j \). We want this function to be as smooth as possible while still fitting the data. The smoothness is measured by the movements in the second derivative of \( f \), \( J(f) = \int_{z_1}^{z_n} (f''(z))^2 \, dz \). It can be proved that the cubic spline is the smoothest function connecting the points \((z_j, y_j)\) and hence this is a popular choice in the estimation of the \( f_i \)'s. This means that the function \( f \) is produced by joining \( n - 1 \) third degree polynomials at the knots.

In practice the requirement \( z_1 < ... < z_n \) may not be met since ties may exists \((z_i = z_j)\) with different responses, \( Y_i \neq Y_j \). Or perhaps it is not desirable that the curve travels all the points \((z_i, y_i)\). This introduces smoothing splines which are also piecewise polynomials which, instead of fitting all the knots, creates a smooth function through these (resulting in a fitting error).

The representation of a GAM looks very much like the GLM case and from the formula the additivity in the covariates \( z_{j1}, ..., z_{jm} \) is clear.

\[
g(\mu_j) = \beta_0 + \sum_{i=1}^{m} f_i(z_{ji}) \quad , \tag{5.8}
\]

where

\[
\mu_j = E(Y_j|z_{j1}, ..., z_{jm}) \quad . \tag{5.9}
\]

Here the \( f_i \)'s are twice differentiable smooth functions of the covariates. Since we shall only work with the univariate model \( g(\mu_j) = \beta_0 + f(z_{j1}) \) we will leave out the index \( i \) in subsequent discussions.

A non-smooth function will clearly fit data best (minimize the error) so there should be a trade off between the smoothness of \( f \) and the models fit to data. In GAM this is controlled

\[5\] Of all continuous functions on \([z_1, z_n]\) interpolating the \((z_j, y_j)\)'s and which have an absolutely continuous first derivative.
through a smoothness penalization $\lambda$ of the model deviance.

$$D(y, \mu) + \lambda \int_{z_n}^{z_{n+1}} (f''(z))^2 dz,$$

where the traditional deviance $D$ depends on $f$ though the model formula (5.8). A high $\lambda$ will result in a smoother function $f$ since the penalty of variability is large. The optimal choice of $\lambda$ is not obvious, but general computational techniques for its determination exists. In the empirical chapter we will take a more pragmatic approach. Though we will not go into details here, we should mention that the estimation of the model can be formulated using linear terms and hence estimated by use of linear algebra and numerical optimization.

For our inflation estimation we use the claim month $k$ as a covariate to estimate the true inflation of month $k$. Using the log link to dampen the effect of large claims, the model for the inflation smoother (with raw inflation as the response) is,

$$\log(I_k) = \beta_0 + f(k) + \epsilon_k$$

where $\epsilon_k \sim N(0, \sigma^2)$ iid. The estimation outputs $\hat{\beta}_0$ and $\hat{f}$ which can be used to calculate a smoothed inflation on the log scale. This is easily transformed back to the normal scale, which gives the GAM smoothed inflation $I_{G,\lambda}^k = \exp(\hat{\beta}_0 + \hat{f}(k))$.

The GAM method is an example of a smoother since it uses observations on both sides of the observation we want to predict (since the function is smooth at both $z_n$ and $z_{n+1}$). Hence we cannot use the GAM estimates directly as a filter. However we can construct a filter $I_{G,\lambda}^k$ if we for the estimation of $I_{G,\lambda}^k$ only use the observations $I_j, j \leq k$ as input to the GAM estimation. Doing this for each of the $k$’s and storing only $I_{G,\lambda}^k = I_{G,\lambda}^k$ for the $k$’th estimation gives a time series $I_{G,\lambda}^k$ which is more volatile than the smoothed analogue. We will denote this as the GAM filter. The extra volatility naturally arises since the boundary values are harder to estimate for the GAM procedure when only historical data is available.

The flexibility of GAM makes it in many respects superior to GLM. The reader may therefore wonder why not use it for the main risk models and not just for the inflation smoothing. The answer is mainly a matter of tradition but new research suggests that the use of these models is emerging in the actuarial society (see bibliographic section). Two important reasons for not using GAM are the computational aspects as well as how to communicate and interpret these models. Another problem is that the widely used SAS software package do not allow for weights in the GAM procedure which makes it inappropriate for frequency modelling.

Other smoothers and filters

When talking of smoothers and filters we feel obliged to mention the celebrated Kalman filter. The Kalman filter is a very useful and flexible tool for extracting signals from noisy time series. Essentially it assumes that a signal is received through a measure with noise (usually normally distributed). The maximum likelihood estimates of the signal can then be derived. The method both filters and smooths the observation and hence makes it ideal...
for the current purpose. However we omit the Kalman filter from this presentation as the underlying theory is quite comprehensive, but we encourage the reader to experiment with the method. Another useful tool for smoothing data is locally weighted scatterplot smoothing (LOWESS) also called just local regression. This method is not far from the GAM models presented above. The idea is to fit a low degree polynomial at each of the knots. This is done by weighting the closest observation highest using weighted least squares. Note that GAM gets smoothness from the knot conditions \( f''(z_k) = f''(z_{k+1}) \) whereas in LOWESS the smoothness is achieved by the weighting functions.

5.3 The DMI framework: Discount, model, inflate

5.3.1 The framework

We are now ready to present the main model(s) of the entire paper. Actually it is more of a framework than a model since no statistical assumptions are related the framework but only to its components. The purpose of introducing the framework is to incorporate inflation into the risk estimates. As we shall see, the DMI framework offers a very generic approach to this task which can be used together with various risk models and inflation smoothers. The framework requires three components to be defined. 1) A smoothed (or filtered) inflation time series covering the same period as the estimation data. 2) A risk model for severities and methods for estimating the relevant parameters. 3) A method for filtering the inflation along with sufficient inflation data to calculate this filter. With these three components at hand the framework is very simple and is explained by the below three step procedure.

<table>
<thead>
<tr>
<th>The DMI framework</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Discount</strong></td>
</tr>
<tr>
<td><strong>Model</strong></td>
</tr>
<tr>
<td><strong>Inflate</strong></td>
</tr>
</tbody>
</table>

Clearly we need a few more comments on these steps which are given below. For the discussion we assume the severities are estimated using GLM.

**Discount**

We use the notion of discounting for the process of moving severities back to period one. By this we mean discounting in the classical financial sense when comparing monetary amounts originated in different periods. The only difference is that we discount by an empirical inflation rather than the traditional interest rate. Note that we stick to the definition of discounting even when severities are increased by the process (this would be the case when the inflation is decreasing).
The discounting is performed as follows. Let $\bar{X}_{hk}, h = 1, \ldots, n_k, k = 1, \ldots, R$ be observed losses during an estimation period ranging from month 1 to $R$ where $n_k$ is the number of claims in the $k$’th period. We discount the observed severities by the (arbitrary) smoothed inflation $\hat{I}_k$,

$$X_{hk} = \frac{\bar{X}_{hk}}{\hat{I}_k}.$$  \hfill (5.12)

In this way all severities are comparable and denoted in the monetary index of period one. The motivation for performing this discounting prior to estimation is that GLM requires data to be iid. given the covariates. A GLM estimation on non-discounted data under a period with non-zero inflation will result in a systematic time-variation in the residuals of the model (see Chapter 7 for an example). This variation will vary in the opposite direction of the inflation since the models overestimates (underestimates) in periods of low (high) inflation. By discounting prior to estimation we approximately\(^6\) ensures that the mean severities are equal over time which is a good start for modelling iid. data.

**Model**

The model of step 2 in the DMI framework correspond to the base severity model of Figure 5.2.b. Since the data is discounted prior to the estimation, all the severities are in the index of period one. To obtain the base severity model simply join the discounted severities into a response vector $X = \{X_{11}, \ldots, X_{n1}, \ldots, X_{1R}, \ldots, X_{nR} \}^\prime$ and keep the respective covariates in a design matrix $Z$. The estimation is then easily performed using GLM on the model,

$$g(\mu_b) = \mathbb{E}(X|Z) = Z\beta.$$ \hfill (5.13)

We have here used GLM as the risk model, but the same principle applies to other methods. When models are estimated on discounted input their predictions will naturally also be denoted in discounted units.

**Inflate**

The last step of the DMI framework concerns predicting severities. Similar to the notion of discounting we use the notion of inflating for the process of moving risk estimates forth in time. Since the base severity estimates are in period one index, the output from the GLM prediction has to be inflated up to the period of the prediction. For this purpose we need a filtered inflation as we cannot make use of future inflation. Note that the (arbitrary) filtered inflation $\bar{I}_n$ of period $n$ is used as predictor for the inflation in period $n + 1$. This is a principle we will adopt for all the filters used. This correspond to straight line (zero slope) prediction of the inflation index. Often this is an erroneous assumption but on the other hand it may be almost impossible to give a good prediction of the inflation. Meanwhile the intention with the inflation models is to update the inflation index frequently and hence the error made is small. If the filtering method gives theoretical estimates for the inflation in

\(^6\)Exactly if we use the raw inflation (just the mean of the period). The smoothed (or filtered) inflations will naturally only capture a part of the actual variation.
the next period (this is the case for the Kalman filter) these can of course be using instead. The inflated severity estimate \( \tilde{\mu}_{hk} \) for the \( h \)'th claim in the period \( k \) is,

\[
\tilde{\mu}_{hk} = \hat{\mu}_{hk} \tilde{I}_{k-1},
\]

where \( \hat{\mu}_{hk} \) is the base severity estimate for the \( h \)'th claim in the \( k \)'th period. That is the \( (n_1 + n_2 + \ldots + n_{k-1} + h) \)'th element of the GLM prediction vector \( \hat{\mu}_h \).

### 5.3.2 Additional comments

Below we give some additional comments on the DMI framework which are not directly related to any of the steps, but to the framework as a whole. We start by a discussion of some problems related to the framework. Afterwards we state some expectations to DMI models. Whether or not these turn out to be true will depend on the portfolio under consideration. In any case the discussions serves good as a motivation for the DMI framework. Finally we end the section by a discussion of the relation to traditional risk models.

**Problems related to changes in risk profile**

The DMI framework is designed for changes in risk level (from Section 1.4). For changes in risk profile the framework is a little more problematic. Note that the framework will adjust to the change, but it will punish the wrong customers. To see this consider a portfolio consisting of equally many of two types of risk, A and B, which are equally risky (same severities) in period one. A doubling of the severities for type A in period two will give a 50% increase in the severity of the total portfolio (change in risk profile, because there is no change for B). Here we use formula (5.2),

\[
I_1 = \frac{\hat{\mu}_1}{\mu_0} = 1
\]

\[
I_2 = \frac{\hat{\mu}_2}{\mu_0} = \frac{1}{2} \mu_A + \frac{1}{2} \mu_B = \frac{1.5\mu}{\mu} = 1.5,
\]

where \( \mu = \mu_A = \mu_B \) is the mean severity of the two types in period one. The DMI framework will now distribute the increase in the severities equally out on both types, despite type B actually has the same risk as in period one. This will make the premiums *unfair* and give room for anti-selection. Note however that the DMI framework does not forbid updates of the base severity model (which for the above situation is needed both for the DMI framework and the traditional approach). The key difference is the following. In the traditional approach the premium level is too low but there is an exposure to anti-selection in only one direction (bad risk comes in). In the DMI framework the overall level of the premiums is correct but instead there is an exposure to anti-selection in both directions (bad risk comes in, good risk goes out).
Problems related to changes in risk mix

Risk mix changes means than the distribution of customers with respect to covariates changes over time. Consider a portfolio consisting of two observable types of risk, called high \((H)\) and low \((L)\) with respective mean severities \(\mu_H, \mu_L\), \(\mu_H > \mu_L\). This could be males and females. We say that a portfolio has experienced a risk mix change e.g. if there is equally many customers of the two types in period one, but 75% og type \(H\) in the second period. In that case the overall inflation of the portfolio is,

\[
I_2 = \frac{\bar{\mu}_2}{\bar{\mu}_0} = \frac{3\mu_H + \frac{1}{4}\mu_L}{\frac{7}{4}\mu_H + \frac{2}{4}\mu_L} > 1, \tag{5.17}
\]

since \(\mu_H > \mu_L\). The DMI framework will now distribute the overall increase equally out on both type \(H\) and \(L\), despite none of the types has experienced changes in the severities. The problem here is that the raw inflation estimates are not cleaned for the effect of risk mix changes. Hence both the inflation index and base severity model will adjust for the overall increase in premiums resulting in an overestimation of the true severities. The solution to the problem (if present) is not easy but must be found in the calculation of the raw inflation estimates. In some way these has to be cleaned for the effect of risk mix changes. For situations where no serious risk mix changes occur we believe that this problem is of minor influence but this will be clear during a proper test and evaluation phase of the model estimation (through poor predictive power). In any case an insurer should be alert if significant changes in risk profile occur since this can be an indication of anti-selection.

Better parameter estimates

As previously explained the DMI framework should remove most of the time variation in the residuals of the risk models. Related to this is a theoretical hypothesis concerning the estimates of these models. Consider a model estimated directly on non-discounted data which exhibit, say declining, inflation. A possible covariate to include in the estimation is the insurance limit of the policy (maximum payout). It is common practice to increase these limits every year to catch up with ordinary price inflation. Hence the distribution of the insurance limit across policies is slowly sliding upwards each year. Now when performing an estimation, the parameter estimate \(\hat{\beta}_i\) for insurance limit may get undesirably high (absolutely) since there is a negative correlation between the severities (which are decreasing for some reason) and the insurance limit (which are increasing due to business practice). This despite, that there actually may be no real dependence between the severities and the insurance limit. To understand why this is problem, consider the above example in the time after the estimation. Assume that the inflation turns and starts to increase. The insurance limit will continue its path upward unaffected, because price inflation is still going up. Since the dependency between severity and insurance limit was initially estimated negative, the model will now lower its severity estimates every year due to the increased insurance limits. This is indeed troublesome since the actual inflation is going up. The same principle applies to other covariates for which the distribution across policies changes systematically over time.

Our hypothesis is that by discounting the severities prior to the GLM analysis the parame-
Better fit to overall level
The DMI framework is designed to automatically maintain the correct level of a risk model. Naturally we will expect it to perform better than the traditional models in this respect. To justify this expectation consider a portfolio with inflation over time. For the discussion we will assume that the inflation is increasing over time but similar arguments apply for the decreasing case. A traditional risk model estimated with GLM will always have an overall level which is too low in such an environment. This is because one of the assumptions of the GLM models is that the residuals have zero mean, \( \mathbb{E}(\epsilon) = 0 \). Hence the models will overall fit to the level of the estimation period which will be lower than the level of future periods (because of increasing inflation). The DMI framework on the other hand adjusts its estimates according to actual inflations every month and hence the level of the estimates are higher than the level of the original estimation data. Only in the case where the inflation is completely due to risk mix changes shall we expect the standard methods to perform as good as (or better than) the DMI framework with respect to overall fit. In the case with no inflation the DMI framework makes (approximately) no correction to the base severity estimates and hence should perform identical to a traditional model.

Generic structure
The above description of the DMI framework is an attempt to generalize a simple method often used in practice where the severity estimates are adjusted according to the price inflation every year. This model easily fits into the DMI framework since the inflation estimates then just are equal for groups of 12 successive months in a row, \( I_1 = \ldots, I_{12} = \ldots, I_{13} = \ldots, I_{24} = \ldots \). The DMI framework is indeed pragmatic in nature though sound statistical methods underpins each contributing component. A simultaneous estimation (inflation and risk) with some kind of time series specification of the inflation factor would properly be more interesting from a theoretical point of view. Yet the DMI framework is interesting due to it generality and easy implementation. Notice that no restrictions has been made on the statistical methods used for smoothing and filtering of the inflation. Neither was anything assumed about the risk model or estimation methods. This means that the methods should be applicable to a wide range of models. In fact every insurance company that are able to perform level adjustments of the premiums should be able to implement the DMI framework without change of existing structure. Fortunately most insurers are capable of doing so since annual price increases are common practice. Further, the flexibility of the inflation smoothers and filters gives the opportunity to tailor the method to the particular coverage or line of business. For heavy tailed risk other methods than those mentioned here may be preferable. Or perhaps one line of business should be maintained by empirical inflation while another is maintained by annual price inflation. In all cases the DMI framework gives a general framework which are easy to implement and gives a high degree of flexibility.
**Application to frequency models**

The DMI framework as described above is not applicable on frequency models. This is because a discounting of the positive integer counts would result in responses on the real line, \( \mathbb{N} \rightarrow \mathbb{R} \). Since the Poisson probability density function only takes non-negative integer values it is impossible to calculate the likelihood and thereby the model estimates. Still, it should be possible to apply the framework by a little modification. Clearly it is possible to measure the development in claim frequencies and estimate a smoothed inflation \( \hat{I}_k \). But instead of discounting the claim count it should be possible to modify the time exposure of the observation to achieve the same effect. By artificially increasing (decreasing) the exposure \( \tau_j \) in periods with high (low) inflation we can ensure that the mean intensity \( \bar{\lambda} \) stays approximately equal over time. Hopefully an estimation on such a dataset will remove (potential) time residuals. Exactly how to adjust the exposure times is a challenging question since an exposure period will often include several months. A naive idea could be the following. For an observation \( j \) with exposure from month \( m_1 \) to \( m_2, m_1 < m_2 \), calculate the average inflation during the period, \( h_j = \frac{1}{m_2 - m_1 + 1} \sum_{k=m_1}^{m_2} \hat{I}_k \). Next, adjust the exposure length, \( \tilde{\tau}_j = \tau_j h_j \), and use \( \tilde{\tau}_j \) during the estimation instead of \( \tau_j \). Note that this technique of adjusting the exposure can also be applied to the severity models where the standard exposure is just one for all observations (then the discounting of severities should of course be excluded). The above discussion suggests that there is potential for a generalization of the DMI framework to include frequency models. We will however leave this topic to future research and concentrate on the methods with discounting of severities (rather than exposures).

**Epilogue on the DMI framework**

Hopefully the above discussion highlights the motivations and problems related to the DMI framework. An implicit assumption of the DMI framework is that all the observed inflation is due to changes in risk level. This might seem naive at first sight, but remember that the traditional assumption is that no changes at all will occur (the future is like the past). Despite the fact that most changes probably are *not* due changes in risk level alone, we expect it to be better than assuming that the risk does not change at all.

One point of critique of the DMI framework relates to situations where the base severity model needs an update because of changes in risk profile. Note however, that in this situation similar problems will be present in a traditional modelling approach. But changes in risk level may also occur and in this case the DMI framework will hopefully give a good estimate where a standard model may fail. Hence the time needed to take care of changes in risk level in a standard approach will be liberated under the DMI framework. These liberated resources may then be used to handle changes in risk profile by more frequent updates of the base severity model. Since this will reduce the likelihood of anti-selection, we should in fact also prefer the DMI framework with respect to changes in risk profile. The reader may find it helpful to consider the DMI framework as an "autopilot" of the risk models. The framework automatically steers in the direction previously instructed (base severity model) without interruption from the actuary. When a new base severity model is available, the
direction of the autopilot is changed to this new model.

## 5.4 The regression framework

We will now give an alternative method for dealing with inflation in the risk models. The DMI framework handles inflation completely outside the risk estimation. Herein lays the roots to both its flexibility and rigidity. It makes it possible to use arbitrary risk models under the same framework, but at the same time it leaves no room for weighting the importance of the inflation index. This motivates another approach where the inflation is incorporated directly in the model estimation. A natural way forward is to use the inflation as a model covariate. However there is a theoretical problem related to this since the response is a (minor) part of the inflation estimate, see (5.1) and (4.2). Hence if a severity is sufficiently large the inflation will be high too and the dependence between the response and the inflation may be overestimated. To deal with this we suggest three alternative methods.

1) One could use the inflation from the previous month as a covariate instead. However, for smoothers this inflation will still be influenced by the (future) response and hence one should use a filtered inflation if this approach is taken. 2) For each response calculate the inflation on the entire dataset with the response observation excluded (a technique known as jackknife or leave-one-out). Then the response will not affect the inflation covariate. However the interpretation, use and computation of this method is troublesome. 3) Ignore the problem. For light-tailed data the influence of a single observation on the inflation estimate is very little. Further, the smoothers and filters are designed exactly to dampen these effects. Hence we believe that the problem is negligible in practice for most dataset. However the reader should be cautious while working on small and (or) heavy tailed datasets.

An important question is how to incorporate the inflation as a covariate. It would be natural to use \((I_k - I_1)\) such that the inflation has no influence in the first month. But for multiplicative models (log link) this is just a matter of parametrization of the intercept as demonstrated below where we for illustrative purposes has separated the intercept \(\beta_0\) and the inflation parameter \(\gamma\) from the rest of the parameter estimates \(\beta\).

\[
\hat{\mu}_k = e^{\beta_0} e^{Z \beta} e^{\gamma I_k} = e^{\beta_0 + Z \beta} e^{\gamma (I_k + (I_k - I_1))} = e^{\beta_0 + \gamma I_1} e^{Z \beta} e^{\gamma (I_k - I_1)}. \tag{5.18}
\]

Hence by using \(I_k - I_1\), instead of \(I_k\) directly, as a covariate the intercept on the transformed scale changes from \(\beta_0\) to \(\beta_0 + \gamma I_1\). To simplify notation we therefore use \(I_k\) directly as a covariate.

A final note should be made on the relation to frequency models. Note that no discounting of the response is performed in the regression framework. Hence the framework can also be applied directly to frequency models with no problems.

## 5.5 Updates of base risk models

The inflation models (DMI and regression framework) are designed to automatically adjust the level as new information becomes available. Note however that this does not exclude
full re-estimations of the model. These are still needed in cases with changes in risk profile. In practice the time liberated by automatic level updates should optimally be used to refine the base risk models. We will however take a different approach in the empirical section. To demonstrate the benefits of the inflations models we will allow the base models to be updated only one time at the beginning of an evaluation period. Afterwards only the inflation is updated (on a monthly basis). Note that this framework is fully automated after the initial estimation.

The benchmark models presented below has more flexibility. These are intended to replicate the work of an actuary and hence allows for full re-estimations at some specified frequency. When time goes we can update these models so they reflect the new risk patterns. Note that since parameter estimates are changes the relative ordering of the risk changes at each re-estimation.

The purpose of forcing the base base models to be fixed for the inflation frameworks is to ease the comparison. If the base risk models changes over time it is difficult to say whether any difference is due to the inclusion of inflation or changes in the base model. In practice, when the base models of the inflation framework are updated, we would naturally just expect the models to perform even better.

5.6 Benchmark models

For illustrative purposes we will compare the inflations models with some simple benchmark models. As mentioned, the purpose of these models are (to some degree) to replicate the work of an actuary who performs re-estimations of the model by a certain frequency. We will take an approach where price inflation is used to update the models on an annual basis since we know this practice is common in the Scandinavian market. Technically this method is just a scaling of the existing risk premium \( S \) such that the new premium \( S^* = S(1 + p) \) where \( p \) is the price inflation (say 4%). We recognise that this method fits into the DMI framework work with annual price inflation used instead of empirical inflation. However the models differ slightly since the parameter estimates \( \hat{\beta} \) are updated frequently in some of the benchmark models (according to the above discussion).

We will choose the covariates at all the re-estimations to be identical to those of the inflations models for fairness in comparison. An objection to this choice is absolutely fair. As the benchmark models are intended to replicate the actuarial work, a full risk analysis with possibility for incorporation of new covariates would seem more fair. However, we choose this "fixed parameter" approach for simplicity and easy of calculations (we return to this problem in Chapter 7). Still, the parameter estimates for the fixed covariates may change by some frequency and hence gives the benchmark models an advantage over the inflation models which are only estimated initially.

**Benchmark model 1**

The first benchmark model represents a quite passive strategy. The base severity model is estimated only once on severities discounted by annual price inflation. Then the annual price index is used to update the model to the price inflation of the prediction month. We use
annual price inflation (instead of monthly), since this is a commonly used approach which probably has its heritage in the annual frequency of policy renewals. Note that once estimated, the parameters of this model are never updated (same as for the inflation models). This is not representative for a real life model in the long run, but over short time periods (<3 years) it may not be far from reality.

**Benchmark model 2**
The second benchmark model is identical to the first except that we allow the parameter estimates to be updated every 10th month. The data used for each re-estimation is chosen such that the number of (successive) months included is the same as for the initial estimation. Hence we leave out the first 10 months when re-estimating for the first time (since 10 newer months has become available since the initial estimation). For the second reestimation we leave out the first 20 months and so on. When predicting severities the most present model estimated before the time of prediction will be used.

**Benchmark model 3**
The third benchmark model similar to benchmark model 2 except that the model updates are performed every fifth month.

**Benchmark model 4 (flat rate)**
As our last benchmark model we will take a flat rate premium. We make no discounting of the severities but simply take the their average over the estimation period which is the flat rate risk premium \( \bar{\mu} \). The model is only estimated one time initially. Note that also this model falls into the DMI framework with no inflation \( (I_n = 1 \forall n \in \mathbb{N}) \) and a simple average as the risk model. The intention of this benchmark model is to evaluate the performance of the lower boundary of sophistication and use it in comparisons.

### 5.7 Bibliographic notes

The treatment of inflation here it different from many ordinary discussions of inflation, because it makes no parametric assumptions about the development over time and because it is based on empirical claims inflation. Hence we have not been able to find any literature directly related to these model. However (Daykin, Pertikäinen, and Pesonen, 1994, chapter 7) gives a good review of the different types of inflation faced buy an insurance company. Especially the treatment of the Wilkie model is interesting as inspiration for a stochastic extension of the inflation models of this paper.

The first textbook on GAM models were Hastie and Tibshirani (1990). A more present and applied approach is (Wood, 2003). In the recent years the actuarial use of GAM has evolved. We suggest (Ohlsson and Johansson, 2010, chapter 5) and (Grgić, 2008) for the particular case of GAM insurance risk models. For textbooks on the Kalman filter we recommend (Shumway and Stoffer, 2006) and especially (Durbin and Koopman, 2001). For an introduction to local regression see (Loader, 1999).
6 Model comparison

An essential part of introducing a new model is comparison with previous models. In Chapter 5 we defined some benchmark models which will serve as the basis for such comparisons. In this chapter we will give a few methods that are helpful when comparing models. In order to perform any comparison, we must determine with respect to what. We state the following three objectives as a basis for the comparison and motivate them below.

1. Correct ordering of risk.
2. Overall prediction is correct (also over time).
3. Dispersion in estimates.

The first objective is quite natural. A risk model should give the highest estimate to the customers with the largest losses (on average). Furthermore, we want the prediction to fit the correct level since otherwise the business is unprofitable or overpriced. We include the time term in the objective to ensure that the correct level is stable over time and not a result of (lucky) chance over an evaluation period. Finally, the estimates should be as dispersed as possible. To understand this, note that a model which has a perfect prediction of the ordering of the risks will be approximately equal to the flat rate model if the difference in premiums between the worst and the best customer is very small, say £1. Hence we need dispersion in the estimates to ensure that the ordering has an effect. We will return to this topic later.

The focus of the following comparison methods will be on the expected severities versus the observed severities. No attention will be given to the variance of the estimates. The motivation for doing so is that the mean severity plays an essential role in pricing of non-life insurance product through the relation between the risk premium and the market price. The variance on the other hand is of less interest to the pricing actuary though it affects the price indirectly through the cost of capital of the reserves (which should be higher for volatile businesses).

Some of the methods below are graphical in their nature. We will try to capture the qualitative interpretation of the graphs in a single statistics to ease our presentation later. However, the graphs should always be consulted for verification before a final model is chosen. Throughout the chapter we denote the observed claim severities by \( \{X_i, i = 1, \ldots, n\} \) and their estimated counterparts by \( \hat{X}_i \).

6.1 Correct ordering of risk

In economics, Lorentz curves are typically used to explain the distributions of wealth among the members of a society. Typically, there is a skew, such that the wealthiest \( x \) percent of the society owns more than \( x \) percent of the total wealth. This is often illustrated in the graph of the Lorentz curve by plotting the relative number of society members against the share their wealth contribute to the total wealth. We will use a similar approach to reveal how good a model is to distinguish good and bad risk (objective one above). Let the model
estimates $\hat{X}_i$ for the severities be sorted in descending order (take the occurrence of the claims as given) such that,

$$\hat{X}_1 \geq \ldots \geq \hat{X}_n.$$  \hfill (6.1)

If the model has a good predictive power the observed severities will be higher for the first claims in the ordered list. Let $X_{[i]}$ denote the observed severity corresponding to the $i$'th claim in the list $\hat{X}_1, \ldots, \hat{X}_n$. Further let $S$ be the aggregate loss of all claims, $S = \sum_{i=1}^n X_i$. Now the points $(i/n, \text{L}(i))$ where,

$$\text{L}(i) = \frac{\sum_{j=1}^i X_{[j]}}{S},$$  \hfill (6.2)

form a graph. A typical example of such a graph can be found in Figure 7.3.b which we will return to later. For a good model the curve tends up towards the top left corner. Contrary to the traditional Lorentz curve however, there is an upper limit for the curvature, given by the observed severities. No model can make a better ordering than one which exactly replicates the ordering of the observed severities, $X_{[i]} = X_{(i)}$. Hence we find the upper limit by simply ordering the observed severities directly (perfect foresight) to form the points $(i/n, \text{U}(i))$ where,

$$\text{U}(i) = \frac{\sum_{j=1}^i X_{(j)}}{S},$$  \hfill (6.3)

with $X_{(1)} \geq X_{(2)} \geq \ldots \geq X_{(n)}$. At the other extreme the points $(i/n, i/n)$, $i = 1, \ldots, n$ represents a completely random sorting of the claims (no predictive power at all). Typically the plot for a severity model will graphically lie somewhere in between the graphs of perfect foresight and no foresight, respectively. The closer the plot is to perfect foresight the better the model is. We summarize this by taking the ratio of the integrals under the two plots down to the line of no foresight,

$$\Psi = \frac{\sum_{i=1}^n (\text{L}(i) - \frac{i}{n})}{\sum_{i=1}^n (\text{U}(i) - \frac{i}{n})}.$$  \hfill (6.4)

This ratio will never exceed one as the denominator correspond to perfect foresight (maximum integral). However there is a theoretical possibility that that $\Psi$ turns negative since the model may be worse that random selection. In practice this should not be a problem since the models generally are better than that. A higher value of $\Psi$ should always be preferred as long as the underlying graphs of $\text{U}(i)$ and $\text{L}(i)$ does not suggest any problems with the model.

### 6.2 Overall correct level

Our next measure is the most simple and yet, perhaps, the most important. As we have previously mentioned that the level of a risk model is essential since if set to low the premiums charged will not be sufficient to cover the losses. On the other hand if the level is too high...
the premiums will be uncompetitive. To measure a model’s performance with respect to the *overall* level we simply calculate the ratio of the expected loss over the observed loss.

\[ \Omega = \frac{\sum_{i=1}^{n} \hat{X}_i}{\sum_{i=1}^{n} X_i}. \] (6.5)

\( \Omega \in [0, \infty) \) since both expectations and losses are positive by assumption. Further the measure has the pleasant property that \( \Omega - 1 \) expresses the relative under or over estimation. Hence an \( \Omega = 0.85 \) corresponds to an under estimation of 15%. Clearly \( \Omega = 1 \) is the optimal value to obtain. In general we will expect inflation models to outperform traditional models with respect to this measure since the former are designed in particular to keep the correct level.

### 6.3 Absence of time variation

A model exhibiting time variation in the residual (overestimates in some period and underestimates in other periods) may still have an \( \Omega \approx 1 \). Though seasonal fluctuations can often safely be ignored for annual premiums other trends should be taken into consideration as soon as possible. To evaluate the *predictive performance over time* we introduce yet another graph. We now sort the claims in order of occurrence in time such that if we let the accident time of the \( i \)’th claim in the ordered dataset be denoted by \( \tau_i \) we have,

\[ \tau_1 \leq \ldots \leq \tau_n. \] (6.6)

Now denote the observed cumulative loss up to claim \( m \) as \( C(m) = \sum_{i=1}^{m} X_i \) with an estimated counterpart \( \hat{C}(m) = \sum_{i=1}^{m} \hat{X}_i \). Plotting these two sums against each other in a \((C(m), \hat{C}(m))\), \( m = 1, \ldots, n \) graph should approximately give a straight line. If the overall level of the model is correct (\( \Omega \approx 1 \)) the line will have a slope of one while for general underestimation (overestimation) the slope will be less (greater) than one. The purpose of the measure here is to reveal systematic time variation and not to investigate the level of the model. Hence we adjust the graph and instead plot \((C(m), \Omega^{-1} \hat{C}(m))\) such that any model should lie as close as possible to a line with slope one. Systematic time variation should manifest itself as deviation from this line. An example of such a graph is found in Figure 7.3.c. Since deviations in both directions are harmful to the business we shall punish both in our evaluation. Once again we formulate a single measure which captures the conclusions from the graph by taking the absolute integral of the difference between the graph and a straight line (slope one).

\[ \Phi = \frac{1}{n} \sum_{i=1}^{n} \left| \Omega^{-1} \hat{C}(i) - C(i) \right|. \] (6.7)

Here we have introduced a scaling by \( \frac{1}{n} \) to reduce the magnitude of the measure, but this does not change the qualitative results of a comparison. The method described here relates very closely to the mean of the residuals per months. We have chosen this cumulative approach as it punishes hard if a positive mean residual one period is not followed by a negative mean in near future (so the curve can get back to the line). This is because such
a situation is exactly the definition of time variation. Note that the special case of perfect prediction of the severities has $\Phi = 0$ while no upper limit exists, hence $\Phi \in [0, \infty)$. We will naturally prefer a $\Phi$ as low as possible. We shall again emphasize that these summary measures mainly are introduced for presentational reasons and that the graph always should be consulted before conclusions are drawn.

### 6.4 Dispersion in estimates

As argued in the introduction to the comparison methods, it is important that risk estimates are dispersed since if they are approximately equal the ordering will have no power in practice. The above methods consider only the level and ordering of the risk and hence we need additional measure for the dispersion between customers risk estimates. We should emphasize that this is different from considering the dispersion of the estimates which is a measure for the uncertainty in the prediction. Though this is also important to consider (as small as possible is preferable) we have limited the scope to not include measures of this kind.

A straightforward measure for the dispersion is the standard deviation between the predicted severities. But if a model in general underestimates the true severities ($\Omega < 1$) we should expect the standard deviation to be smaller than for a model with the correct level ($\Omega \approx 1$) since the severities are positive, $X \in \mathbb{R}_+$. To adjust for this effect we measure the standard deviation relative to the mean of the r.v. This measure is known as the coefficient of variation,

$$CoVa(X) = \frac{\sigma_X}{\mu_X} = \frac{\sqrt{\text{Var}(X)}}{\mathbb{E}(X)}.$$  \hspace{1cm} (6.8)

Note that the coefficient of variation usually only are useful for positive variables measured on a ratio scale. Consider for example the market prices for an insurance that are subject to a new tax of £50 (additive). This shifts the premiums up £50 causing the mean to increase while the standard deviation is unchanged. Consequently the coefficient of variation decreases but the dispersion in the estimates is the same as before the tax. When multiplicative models are used for the severities there should be no problem in this respect as they are on a ratio scale (relative measures).

In general a coefficient of variation as large as possible is preferable. However this measure is of less importance than the others. This is because we can artificially scale up and down the dispersion of the market price as long as we have a good ordering of the risks (measured by $\Psi$). To see this consider a model with a good ordering of risk but with a low dispersion in the estimates. Now assume that the lack of dispersion in the severity model is also reflected in the risk premium and that the market premiums $\{P_i, i = 1, ..., n\}$ are proportional to the risk premium (both are often approximately satisfied). We can now change the dispersion of the market premiums $P_i$ by applying the following formula to get the modified premium $\tilde{P}_i$.

$$\tilde{P}_i = (P_i - \bar{P})(1 + c) + \bar{P},$$ \hspace{1cm} (6.9)
where $\bar{P} = \frac{1}{n} \sum_{i=1}^{n} P_i$ and $c \in \mathbb{R}$ is some real constant which is greater (smaller) than zero for more (less) dispersion. Clearly we need a method to avoid $\bar{P}_i \leq 0$ but this should not be a big problem to handle. Hence, though we generally prefer high coefficient of variation, dispersion is not of great importance as we can increase it artificially.

6.5 Bibliographic notes

The methods of this chapter are rather ad hoc and the references are consequently sparse. However for the first two graphical methods the theory of ROC curves (receiver operating characteristic) has been a great inspiration. A contemporary volume on this topic is (Krzanowski and Hand, 2009). For an introduction and supplementary references on the coefficient of variation we suggest (Martin and Gray, 1971).
7 Empirical study

We now turn our attention towards the practical use of the methods described so far. For the demonstration we use a real life dataset from a Scandinavian motor portfolio. Naturally the dataset is chosen because it exhibit heavy inflation over the period under consideration but the methods should be useful for dataset with less inflation too.

7.1 Data

The data used in this chapter originates from a motor portfolio in a large Scandinavian insurance company. They were collected during a period of 55 months starting January 2005 and ending July 2009, both inclusive. The dataset contains only claims on a single coverage of the comprehensive part of the motor insurance (this is line with Section 3.5.2). The actual name of the coverage is left out for competitive reasons. The dataset contains 1598 claims causing an aggregate loss of 2.244 million £. Note that the severities are scaled by a random factor, again, for competitive reasons. A short list of variables in the dataset is given in Table 7.1. For a detailed description of the variables Appendix A should be consulted.

<table>
<thead>
<tr>
<th>Group</th>
<th>Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>The car</td>
<td>CarAge, CarGroup, NewValue, TradeValue, Fuel, Weight, SecondCar.</td>
</tr>
<tr>
<td>The owner and user</td>
<td>OwnerSex, UserSex, OwnerAge, UserAge, UserOwnsCar, PostalDistrict, HouseholdInt100m, HouseholdInt1km, HouseholdInt10km, Duration.</td>
</tr>
<tr>
<td>The insurance</td>
<td>InsuranceLimit, Deductible, Coverages, PolicyID, SpecialCoverage, YearSinceClaim.</td>
</tr>
<tr>
<td>The claims</td>
<td>Claims, Severity, ExpoStartDate, ExpoEndDate, Exposure, ClaimDate, ClaimMonth, Year.</td>
</tr>
</tbody>
</table>

Table 7.1: Short list of variables in motor insurance dataset.

7.2 Methods

The methods applied are those described previously in this paper. Hence we will not go into detail with the technicalities here but refer to the relevant sections. Yet a few comments on the application are necessary.

We will use GLM method for estimation of our risk models under the compound Poisson risk model. This choice is mostly a matter of tradition and other methods may actually perform better. Since the focus of this paper is to investigate the effect of inflation in risk models we naturally have to limit ourselves regarding other aspects of the estimation. Hence we will use a Poisson distribution for claim counts and a gamma distribution for the severities without questioning the appropriateness of this (popular) choice. However, we will only concentrate on the severity part of the estimation. We use the log link \( g(y) = \log(y) \) to transform the
strictly positive severities over to entire real line, \( g : \mathbb{R}_+ \to \mathbb{R} \). A linear model on this scale correspond to a multiplicative model on the original scale.

As mentioned in Section 5.6 we will, for comparative reasons, let the covariates used in different models will be fixed. Hence the covariates used in the benchmark models are identical to those of the of the DMI and regression framework. Unfortunately this has the unpleasant property that some covariates may become insignificant in some of the models. Therefore we will restrict the estimation to include only the most significant variables and leave out interactions. In this way we hope to capture the main effects while still having significant parameters when applied to different models. Again, the correctness of the mean (expected loss) is our main concern when estimating insurance prices and hence no focus will be given to the variance of the estimates.

For the smoothing of inflation the methods of Chapter 5 will be applied. We will use various combinations of smoothers and filters to find the best combination. When referring to the inflation used prior to the estimation we will use the term \textit{discounting inflation} (used under the estimation for the regression framework). Analogously for the inflation used for prediction we will use the term \textit{prediction inflation}. In Table 7.4 a list of all the combinations used is given. Further comments to the table are found in Section 7.4.2.

Finally we use the methods of Chapter 6 to compare the models. Instead of evaluating the models directly on the estimation data we use an \textit{out of sample} evaluation. We split the 55 months of a data into two subset consisting of month \{1, ..., 30\} and \{31, ..., 55\} which we will refer to as the \textit{training} and \textit{test} data, respectively. The estimation is then performed only on the training data and the resulting model is then used to predict the severities from month 31 and forth. For the inflation frameworks the months \{1, ..., 30\} are used for estimation of the base severity model and afterwards only the inflation index is updated (on a monthly basis). The benchmark models on the other hand are re-estimated frequently as described in Section 5.6. Hence for each re-estimation we use the most presents 30 months. Now when predicting severities of month \( n \) we use the model of the newest reestimation performed before month \( n \). That is, if re-estimations are performed in months 30, 40 and 50 we use the model of month 30 to predict in month 39 and 40 and the model of month 40 to predict in month 41.

### 7.3 Explorative analysis

We now proceed to an explorative analysis of the data. The purpose of this is to obtain an understanding of data which can be helpful in the estimation phase. We divide our discussion into three small parts starting with some general observations on the inflation.

**General observations on inflation**

Since we are dealing with inflation models it is of particular interest to investigate whether inflation is present in the data or not. We find the raw inflation by averaging the severities for each month, according to formula (5.2). Figure 7.1.a shows the raw inflation along with
several smoothed or filtered versions of it. From the graph it is clear that for this particular dataset the raw inflation is quite volatile. However there still seems to be an upward trend indicating that the severities are getting larger over the 55 months. The GAM smoother (which we will take as the best estimate for the true inflation) indicates that this increase is approximately in order of 60% which is far beyond the price inflation which for the period only accounts for a 10% increase. This suggests that the inflations models with empirical inflation may by useful.

![Graph of Inflation Measures](image)

(a) Grey solid line is the raw inflation, thin black solid line is the GAM filter, dotted black line is EWMA and thick solid line is the GAM smoother.

(b) Top is estimated on severities discounted by the GAM smoother. Bottom is estimated on raw severities without any discounting.

Figure 7.1: Inflation measures (left). Mean residuals against time for a flat rate model (right).

To demonstrate the problems caused by inflation we estimate a flat rate model on the entire dataset (ignoring test and training for now). The monthly mean of the studentized residual for this model is depicted in Figure 7.1.b bottom panel. From the figure it is clear that a systematic time variation in the residual is present which violates our model assumption (zero covariance between errors and covariates, $\text{Cov}(\epsilon_j, \tau_j) = 0$). Since the means are based on studentized residual (for comparative reasons) the graph does not reveal anything about the magnitude of this problem. A similar graph of the raw residuals shows that at the most problematic month (month 55) there is a systematic under estimation of £488 on average! If we alternatively estimate the flat rate model on discounted severities (DMI framework) the time variation becomes negligible (Figure 7.1.b top panel). Though not exactly surprising this finding just supports our previous hypotheses and motivation.

**The type of inflation**

After observing the presence of inflation it is of interest to find out whether it is caused by changes in risk profile or risk level (according to Section 1.4). Several ways to determine this exists, but here we take a quite simple approach which are easily extended. We divide the observation into groups based on covariates and time and then observe whether the infla-
tion is homogeneous or heterogeneous. Since data are sparse (approximately 30 claims each month) we cannot investigate to small segments of data. Hence for each continuous covariate we form a binary indicator which is one if the observation is in the upper 50% empirical quantile of the range of the variable and zero otherwise. Similarly we divide the observation in time such that months $T_1 = \{ 1, ..., 30 \}$ represents the first period and $T_2 = \{31, ..., 55 \}$ the second. Note that the division of the observation in time should be such that one period represent months with low inflation while the other represents month with high inflation. Hence if the inflation graph of Figure 7.1.a were "bell-shaped" with a mode in the intermediate months we would probably choose $T_1 = \{1, ..., 15, 40, ..., 55 \}$ and $T_2 = \{16, ..., 40 \}$.

Before averaging we truncate the severities by taking minimum of the actual severity and the 95% empirical quantile, since otherwise the estimated mean fluctuates wildly. We can now produce a $2 \times 2$ table showing the mean severity of each group formed by interactions of the periods and the binary variable. The $2 \times 2$ tables may reveal the nature of the inflation. If the top 50% (where binary indicator is one) of the observations have a substantially different inflation than the bottom 50% we may interpret this a change in risk profile for the variable. If, on the other hand, the inflation is close for the two groups this indicates changes in risk level. We have applied this method for all the variables in the dataset (discrete variables just keeps their natural grouping) and find indications that the inflation is mainly caused by changes in risk level. In Table 7.2 we reproduce two of the output tables. For the variable NewValue we find almost no difference in the inflation whereas for HouseholdInt10km we observe one of the largest differences. The point at which we say that the difference is big is quite fuzzy just as the distinction between changes in risk profile and level. In general we recommend experiments with truncation values and group generation in order to get a convincing output from the method. Also the inclusion of medians in the tables is valuable.

<table>
<thead>
<tr>
<th></th>
<th>NewValue</th>
<th></th>
<th></th>
<th>HouseholdInt10km</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean severity $T_1$</td>
<td>Mean severity $T_2$</td>
<td>Increase factor</td>
<td>Mean severity $T_1$</td>
<td>Mean severity $T_2$</td>
</tr>
<tr>
<td>Bottom 50%</td>
<td>£865</td>
<td>£1042</td>
<td>1.20</td>
<td>£906</td>
<td>£1144</td>
</tr>
<tr>
<td>Top 50%</td>
<td>£992</td>
<td>£1201</td>
<td>1.21</td>
<td>£953</td>
<td>£1100</td>
</tr>
</tbody>
</table>

Table 7.2: Table for identification of the type of risk change. For NewValue (left) the inflation is apparently identical for the two groups. For the HouseholdInt10km (right) there is a tendency that rural areas experience a larger inflation.

**GLM analysis**

We now proceed to a few comments on the risk models and their GLM estimation. We perform our explorative analysis on the dataset with severities discounted by the GAM smoothed inflation. Central to GLM is that the mean should be linear in the parameters $\beta$. This includes situations where (continuous) covariates are included as $z, z^2, \log(z), \exp(z)$, etc. as long as the $\beta$ terms are linear. To find an appropriate transformation of the covariates
one can use explorative graphs of the observed mean severities as a function of the covariates. However care should be taken when transformations are used. Often these transformations behave wildly when extrapolated just slightly outside the range of the observed data (especially polynomials of $z$). Another import consideration is that the final tariff should be intelligible, to some degree, to the customers. Though a tariff might be perfectly fair in the sense of correct risk, it may be perceived unfair by the customers if it behaves against their intuition. In these days of on-line insurance sales and offering the effect of the pricing components are easily tested and may discourage customers to buy if perceived unfair. We find that inclusion of continuous covariates as linear terms (no transformation) often gives a good compromise between predictability and reasonable structure in the tariff. In Figure 7.2 we illustrate empirical means for a few of the most promising covariates. Fortunately a (close to) linear relationship is present in all three graphs, at least where the majority of the exposure is. We use plots of this kind as inspiration to the GLM analysis. All og the plots are gathered in Appendix B.

Note that the "flattening" of the line in panel (c) is due to very little exposure. Though we could include max(deductible, £850) as a covariate we would prefer to extrapolate the linear effect over the entire scale of the covariate. However the deductible plays special role in the discussion of perceived fairness. Note that the slope is positive meaning that the higher deductible the customer choose the greater do we expect the losses to be. This is natural since all the small claims are never reported for the high deductible policies. However we cannot justify this structure in the customer premium and hence we leave out the deductible from the models\(^8\).

Figure 7.2: Plot of empirical severities against grouped continuous variables. The thin line is the observed mean severity for the group interval. The bold line is a GAM smoothing of the observed severities over these groups. Bars show the exposure in percent of total exposure.

\(^8\)Note that the frequency is almost always decreasing in the size of the deductible. Hence the total effect of lower frequency and higher severity may actually result in a smaller risk premium but we cannot be certain of that!
7.4 Results

The results of the empirical work are dichotomous. The risk models and the impact of each covariate on the final risk estimate is usually of primary importance in applied insurance pricing. But since these results are very specific to the company, portfolio and market under consideration they may be of less interest to the readers of this paper. Hence we will naturally scope the presentation to mainly concern the impact of the inflation models. For completeness we will however give a short presentation of the GLM analysis carried out during the estimation.

7.4.1 GLM analysis and inflation smoothing

Our GLM analysis takes its starting point in the explorative analysis, specifically the plot in Figure 7.2 (and those for the rest of the covariates). These plots give an idea about the marginal dependence structure in the dataset. Two important properties to look for in this respect are linearity and magnitude of the effects. By magnitude should be understood by how much the mean vary with the variable. Note however that this information cannot stand alone as the variance of the mean will also affect the significance of the effect. The property of linearity is only important because we here are working with GLM. If highly non-linear dependency structures are observed over several variables one should consider using other statistical methods such as GAM or make appropriate transformations and grouping of the covariates. The covariates chosen as the base for all the estimated models are listed in Table 7.3 together with the parameters estimates for a model under the DMI framework using the GAM smoother for discounting.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>P-value</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>6.7604</td>
<td>&lt; 0.001</td>
<td></td>
</tr>
<tr>
<td>HouseholdInt10km</td>
<td>0.4207 × 10⁻³</td>
<td>0.0436</td>
<td>Household intensity (per hectare).</td>
</tr>
<tr>
<td>SpecialCoverageGpd</td>
<td>-4.8090 × 10⁻¹</td>
<td>&lt; 0.001</td>
<td>Used if special coverage is &quot;No&quot;.</td>
</tr>
<tr>
<td>TradeT</td>
<td>-7.8430 × 10⁻³</td>
<td>0.0247</td>
<td>Trade value of car in £1,000.</td>
</tr>
<tr>
<td>InsuranceLimitT</td>
<td>1.5514 × 10⁻²</td>
<td>&lt; 0.001</td>
<td>Insurance limit in £1,000.</td>
</tr>
</tbody>
</table>

Table 7.3: Parameters included in all the models. The regression models have one additional parameter for the inflation. Model estimates in the table are for a discounting model with the GAM smoother as discounting inflation.

We will not go into details with the estimates since they represents only one of many models. Noticeable is however the effect of SpecialCoverageGpd (a grouped version of SpecialCoverage where "Yes" and "Unknown" are collapsed) is remarkable since a special coverage gives lower expected severity. This is kind of counter intuitive but may rely on the customers choices and behaviour. More careful customers may both buy an extra coverage and act in a claim reducing way.
For the inflation models we use a pragmatic approach for selecting the smoothing parameter \( \lambda \) for the GAM and EWMA methods. Simple trial and error attempts were performed until the resulting smoothed or filtered graph captured the tendency in the raw inflation in the (subjectively judged) best possible way. The result is seen in Figure 7.1.a. For the EWMA we found \( \lambda = 0.15 \) appropriate. For the GAM estimates the SAS system used requires a specification of the degrees of freedom (d.f.). We found that a choice of 10 d.f were good for the 55 months of data.

### 7.4.2 Model comparison

We now present the main results of the applied inflation models. We compare the predictive power of the models with respect to the objectives described in Chapter 6 to enlighten the differences between the them. The objectives are replicated here for convenience: Correct ordering of risk, dispersion in estimates and correct overall prediction (also in time). The models to be considered arise by combining the DMI and regression framework, respectively, with different combinations of smoothers and filters from Chapter 5. Further we have a benchmark structure which will not be far from what many companies do in practice. Table 7.4 presents the different models along with the comparison measures. We will discuss these in details below.

<table>
<thead>
<tr>
<th>Model name</th>
<th>Method</th>
<th>Discounting</th>
<th>Prediction</th>
<th>( \uparrow )</th>
<th>( \downarrow )</th>
<th>( = 1 )</th>
<th>( \uparrow )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>inflation</td>
<td>inflation</td>
<td>( \Psi )</td>
<td>( \Phi )</td>
<td>( \Omega )</td>
<td>( CoVa(\hat{X}) )</td>
</tr>
<tr>
<td>D1 *</td>
<td>DMI</td>
<td>GAM</td>
<td>EWMA</td>
<td>0.399</td>
<td>22.2</td>
<td>0.999</td>
<td>0.444</td>
</tr>
<tr>
<td>D3 *</td>
<td>DMI</td>
<td>EWMA</td>
<td>EWMA</td>
<td>0.406</td>
<td>21.9</td>
<td>1.057</td>
<td>0.470</td>
</tr>
<tr>
<td>R1 *</td>
<td>Regression</td>
<td>GAM</td>
<td>EWMA</td>
<td>0.397</td>
<td>21.6</td>
<td>0.976</td>
<td>0.455</td>
</tr>
<tr>
<td>R4 *</td>
<td>Regression</td>
<td>Raw</td>
<td>Raw</td>
<td>0.331</td>
<td>15.4</td>
<td>1.011</td>
<td>0.594</td>
</tr>
<tr>
<td>R4 *</td>
<td>DMI</td>
<td>Raw</td>
<td>Raw</td>
<td>0.318</td>
<td>14.6</td>
<td>1.055</td>
<td>0.533</td>
</tr>
<tr>
<td>D2</td>
<td>DMI</td>
<td>GAM</td>
<td>GAM filter</td>
<td>0.387</td>
<td>22.5</td>
<td>1.162</td>
<td>0.450</td>
</tr>
<tr>
<td>D5</td>
<td>DMI</td>
<td>Price</td>
<td>Price</td>
<td>0.382</td>
<td>53.2</td>
<td>0.811</td>
<td>0.494</td>
</tr>
<tr>
<td>R2</td>
<td>Regression</td>
<td>GAM</td>
<td>GAM filter</td>
<td>0.380</td>
<td>22.2</td>
<td>1.173</td>
<td>0.474</td>
</tr>
<tr>
<td>R3</td>
<td>Regression</td>
<td>EWMA</td>
<td>EWMA</td>
<td>0.402</td>
<td>22.6</td>
<td>1.178</td>
<td>0.487</td>
</tr>
<tr>
<td>R5</td>
<td>Regression</td>
<td>Price</td>
<td>Price</td>
<td>0.399</td>
<td>35.1</td>
<td>0.937</td>
<td>0.477</td>
</tr>
<tr>
<td>B1</td>
<td>Benchmark 1</td>
<td>Price, annual</td>
<td>Price, annual</td>
<td>0.383</td>
<td>53.2</td>
<td>0.810</td>
<td>0.495</td>
</tr>
<tr>
<td>B2</td>
<td>Benchmark 2</td>
<td>Price, annual</td>
<td>Price, annual</td>
<td>0.378</td>
<td>33.8</td>
<td>0.885</td>
<td>0.547</td>
</tr>
<tr>
<td>B3</td>
<td>Benchmark 3</td>
<td>Price, annual</td>
<td>Price, annual</td>
<td>0.369</td>
<td>30.2</td>
<td>0.859</td>
<td>0.553</td>
</tr>
<tr>
<td>B4</td>
<td>Flat rate</td>
<td>None</td>
<td>None</td>
<td>0.014</td>
<td>58.3</td>
<td>0.686</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Table 7.4: Comparison of results. \( \Psi \) is the Lorentz plot, \( \Phi \) is the time plot, \( \Omega \) is the overall correctness and \( CoVa(\hat{X}) \) is the coefficient of variation of the predicted severities. The symbols above the measures indicates in which direction we optimally would prefer the measures to go. Models marked with * are not dominated by any other model w.r.t. \( \Psi, \Phi \) and \( \Omega \) (see explanation below).
The models
From the Table 7.4 it is clear that we have included many (but not all possible) combinations of inflation and model methods. Starting from the bottom we have the benchmark models which are updated with varying frequency. B4 is included as a lower bound of the comparison measures (we cannot perform unintentionally worse than this). The three other benchmark models are intended to emulate the the work of an actuary. As mentioned, we fix the parameters used in these models even though this does not mirror reality in which we are free to include or exclude parameters at model updates. However the parameters stay significant (or close to) in most of the estimations. A critique of this method seems fair, since it does not replicate reality, where an actuary may include new variables. Hence one may be tempted to devalue the importance of the results. But remember from Section 5.3.2 that standard models (the benchmark) will always be “one step behind” the correct risk level if inflation is present and are not entirely due to risk mix changes (which is not the case here). Inclusion of additional covariates may therefore improve Ψ and perhaps Φ, but not Ω! Hence our method with fixed covariates should not change the value of Ω noticeably for the benchmark models. Hence the inflation models (here) are superior at least with respect to Ω.

Note that these benchmark models actually are fully automated for the purpose of this applications (because the covariates are fixed). In practice however, we cannot undertake a complete reestimation of the parameters without some kind of supervision of the result. Further the aspects mentioned in Section 5.1 will require resources to be carried out for each coverage. Hence the benchmark models (at least B2 and B3) are way more labour intensive to maintain than the inflations models which require only one complete estimation and afterwards just index updates. We should keep this in mind when comparing the results.

For both the DMI and regression framework five models exists. We include price inflation and raw inflations mainly for comparative reason and expect no positive results from these. The final three models uses empirical inflation. We prefer the GAM approach for smoothing but include EWMA due to its simplicity. Model D3 is solely based on EWMA and should be a good starting point for practitioners uncomfortable with GAM.

The measures
Before we discuss the results a demonstration of the graphical methods is appropriate. Figure 7.3 presents the graphs of the Lorentz plot and time plot of Section 6.1 and 6.3, respectively. The plots hopefully underpin the motivation for the measures Ψ and Φ which are intended to capture the qualitative properties of these graphs. Φ measures the (absolute) integral between straight line and the time plot in panel (a). Note that before the time plots are constructed the predicted severities are adjusted proportionally by $\Omega^{-1}$ such that their total sum equals the same amount as the actual losses. From panel (c) it is evident that the model overestimates slightly for the first half of the losses and then under estimates. Be aware that the over and under estimation is after the correction. Before the correction model B1 overall underestimates by $\Omega = -19\%$.

$\Psi$ is the ratio of the integrals from the Lorentz curve down to the $45^\circ$-line for the perfect model and our model. In the case show in panel (b) the curves of our model covers 39.9%
of the area up to the curve for the perfect model.

(a) Time plot, model D1. The solid line is perfect prediction in time, the dotted line is our models prediction in time.

(b) Lorentz plot, model D1. Thin solid line is random selection, bold solid line is perfect selection, dotted line is our selection.

(c) Time plot, model B1. The solid line is perfect prediction in time, the dotted line is our models prediction in time.

Figure 7.3: Graphical methods of comparison. The measures \( \Psi \) and \( \Phi \) are designed to capture the respective properties of the graphs in a single measure.

The results

Table 7.4 shows the different models with their respective measures for comparison. The table may be cumbersome to interpret at first sight. But in Section 6.4 it was argued that the coefficient of variation could safely be ignored (at least for small differences) as we could artificially create dispersion in the market premiums. With this in mind we can return to the interpretation of Table 7.4 and conclude that the models marked with * are are not dominated by any other of the models with respect to the three most important measures \( \Psi \), \( \Phi \) and \( \Omega \). Considering only these three measures we will for example never choose D2 over D1 as D1 is superior in all measures. A list of domination is given in Table 7.5. Unless we have other reasons to prefer a dominated model (e.g. simplicity) we should focus on these superior models. For this reason we will not consider D2, D5, R2, R3 and R5 as desirable models. Further the models D4 and R4 (raw inflation) are mainly included for comparison and will not be used as they performs poorly on \( \Psi \). This is because the raw inflation of month \( n \) is a very poor estimate of the inflation in month \( n + 1 \) (see Figure 7.1.a). We should emphasize that the graphs behind the measures \( \Psi \) and \( \Phi \) do all look acceptable and very alike. These graphs can be found in Appendix C.

This leaves us with models D1, D3 and R3 to compare with the benchmark models. Note that all these three models dominates each of the benchmark models with respect to \( \Psi \), \( \Phi \) and \( \Omega \)! Most unexpected is perhaps the superiority with respect to \( \Psi \). At first glance we should expect a benchmark model such as B3 to outperform the inflations models since it has five full re-estimations during the evaluation period compared to just one initial for the inflations models. But as argued in Section 5.3.2, one hypothesis is that the data gets more "cleaned" and hence reveal the true nature of the risk patterns better. This may be the
Table 7.5: Domination between models w.r.t. to comparison measures.

<table>
<thead>
<tr>
<th>Model</th>
<th>Dominated by ... w.r.t {\Psi, \Phi, \Omega}</th>
<th>Dominated by ... w.r.t {\Psi, \Phi, \Omega, CoVa(\hat{X})}</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>D2</td>
<td>D1, D3, R1</td>
<td>D3, R1</td>
</tr>
<tr>
<td>D3</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>D4</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>D5</td>
<td>D1, D2, D3, R1, R3, R5</td>
<td>None</td>
</tr>
<tr>
<td>R1</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>R2</td>
<td>D1, D3, R1</td>
<td>None</td>
</tr>
<tr>
<td>R3</td>
<td>D3</td>
<td>None</td>
</tr>
<tr>
<td>R4</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>R5</td>
<td>D3</td>
<td>None</td>
</tr>
<tr>
<td>B1</td>
<td>D1, D2, D3, R1, R3, R5</td>
<td>None</td>
</tr>
<tr>
<td>B2</td>
<td>D1, D3, R1</td>
<td>None</td>
</tr>
<tr>
<td>B3</td>
<td>D1, D3, R1</td>
<td>None</td>
</tr>
<tr>
<td>B4</td>
<td>All</td>
<td>All</td>
</tr>
</tbody>
</table>

case for the data used here. However the argument only expresses a hypothesis and should not be seen as an general argument for superiority of the method with respect to \(\Psi\). On the other hand we will usually expect \(\Phi\) and \(\Omega\) to be better for inflations models as both these measures concerns the correctness of level which is exactly what the inflation models are tailored to keep correct.

Which of the three models D1, D3 and R1 to prefer is partly a matter of taste. However the regression models are dangerous for several reasons. First of all we have our previous argument, that the response is included partly as a covariate through the inflation (see Section 5.4 for details). This can be problematic for heavy tailed data or small sample sizes. The second problem is the that the model is free to choose any parameter estimate \(\gamma\) for the inflation. Hence if there is no inflation in the estimation data there is a high probability that \(\gamma \approx 0\), which will result in a model with no inflation adjustments even in future periods where inflation may be present. Regarding this problem the DMI framework seems more robust since the adjustments are performed manually prior and post estimation. Furthermore the DMI framework allows for arbitrary kinds of base severity models, which is a clear advantage too.

As the reader might have sensed we prefer the DMI models D1 and D3 over the regression model R1. Consequently our choice is then reduced to choosing between the discounting inflations which are the only difference between these two models. When looking at Figure 7.1.a most people will agree that the GAM smoother captures the trend better than EWMA. For this reason we prefer the GAM smoother and hence model D1. However the model D3 is still a good choice.

67
8 Conclusions

In the paper at hand we have discussed insurance risk in general and suggested a method to combine traditional models with inflation estimates. In Chapter 3 we presented the individual risk model and the collective risk model. The problematic assumptions of both of these models were addressed and an extended method was suggested. The extended method (the compound Poisson risk model) had all the pleasant properties of the classical risk models without an noticeable increase in complexity. This model was the foundation of the subsequent chapters which gave some methods to incorporate inflation in the risk estimates. The main framework for doing so was the DMI framework where claim severities were discounted prior to estimation and afterwards the predicted severities were inflated up according to empirical inflation. The performance of this framework compared to the traditional approach was measured by several comparison methods which was tailored to examine the performance with respect to some overall objectives (correct ordering of risk, time performance, overall prediction level and dispersion in estimates). Finally in the empirical chapter the models were applied on a real life dataset from a Scandinavian motor portfolio. The results from the application showed that the inflation models performed better than traditional models despite their simpler structure (no re-estimations were allowed).

The DMI framework for handling inflation in severities was a generalization of a method often used in practice. Usually discounting of data follows ordinary price inflation. We showed that the DMI framework was capable of handling this standard approach as a special case. The generalization of the methods was to extend the notion of inflation from being only price inflation to use the empirical inflation instead. Though the method proved useful for data with severe inflation in the empirical chapter they should work as well in zero inflation environments. This is because in this case the framework approximately corresponds to a normal GLM analysis with no discounting. Further, theoretical considerations imply that the methods should in general be better since the time variation in the residual should be reduced.

We therefore conclude that the ideas that originally motivated the framework indeed has their justification. A general superiority of the inflation models has not been proved but since the traditional methods are a special case, we should expect the DMI framework to be as least as good as this. Due to the generic nature of the DMI framework the methods is applicable to a wide range of existing risk models. This allows the framework to be tailored to the existing structure of most insurance companies without having to change the statistical methods of the risk models. This is an important property since changes of that kind can be costly and long termed. The easy implementation allows companies to experiment with the methods with a very little effort and always the possibility to go back to existing structures by simply setting the inflation index to one in every month.

Though the DMI framework is useful, many questions are left open. One is whether the method is superior in general. Though we have our hypothesises that it is, it hard to say
anything with confidence. Since the framework contains no parametric assumptions the answer to the question is most probably found by establishment of a realistic simulation study. Secondly, inference about the comparison measures is not available at this point. This makes it impossible to say whether any superiority is significant. The methods of bootstrapping may be a possible way forward in the search for such inference. In both cases simulation techniques will require us to make some distributional assumptions (except non-parametric bootstrapping) which will make the conclusions dependent on these.

The DMI framework explicitly make an error when the portfolio exhibit changes in risk mix. Development of inflation estimates which are cleaned for the this effect is of great importance for the implementation of the models. Also a modification of the framework to include frequency models as well would be a very interesting next step in research.

In the evaluation of the models some comparison measures are needed. We have suggested a few and motivated these, but the development of new or existing measures to enlighten different aspects of the models will indeed serve good for the comparisons and conclusions.

Finally the implementation in the market premium has not been given any attention during this paper. Most likely, it is not desirable that prices changes every month. Hence principle must be developed regarding the dependence between the market premium and risk premiums. A natural method could be to let the market premium have a slower update frequency such that it is only updated say every quarter.

We end the paper by concluding that inclusion of inflation in risk models is possible and beneficial and that there seems to be potential for a lot of future research regarding the DMI framework and inflation models in general. We only hope that the present paper has inspired the reader to start considering the importance of inflation in risk estimates.
### Appendix A  Description of variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CarAge</td>
<td>Continuous</td>
<td>The age of the car in years at the date of the accident.</td>
</tr>
<tr>
<td>CarGroup</td>
<td>Discrete</td>
<td>A manual grouping of cars by characteristics. Higher numbers represent more expensive cars.</td>
</tr>
<tr>
<td>NewValue</td>
<td>Continuous</td>
<td>The value of the car as brand new, indexed up to the accident date by price index. Measured in £.</td>
</tr>
<tr>
<td>TradeValue</td>
<td>Continuous</td>
<td>The actual value of the car (follows some depreciation tables). Measured in £.</td>
</tr>
<tr>
<td>Fuel</td>
<td>Discrete</td>
<td>&quot;G&quot; for gasoline, &quot;D&quot; for diesel.</td>
</tr>
<tr>
<td>Weight</td>
<td>Continuous</td>
<td>Weight of car with no goods or fuel, measured in kilograms.</td>
</tr>
<tr>
<td>SecondCar</td>
<td>Discrete</td>
<td>&quot;Yes&quot; or &quot;No&quot;. If the car is the secondary car for the user the variable is &quot;Yes&quot;.</td>
</tr>
<tr>
<td>OwnerSex</td>
<td>Discrete</td>
<td>Sex of the owner of the car and insurance.</td>
</tr>
<tr>
<td>UserSex</td>
<td>Discrete</td>
<td>Sex of the user of the car. Mostly identical to OwnerSex.</td>
</tr>
<tr>
<td>OwnerAge</td>
<td>Continuous</td>
<td>Age of the owner of the car and insurance.</td>
</tr>
<tr>
<td>UserAge</td>
<td>Continuous</td>
<td>Age of the user of the car. Mostly identical to OwnerAge.</td>
</tr>
<tr>
<td>UserOwnsCar</td>
<td>Discrete</td>
<td>&quot;Yes&quot; or &quot;No&quot;. A &quot;Yes&quot; incates that the user is the owner of the car.</td>
</tr>
<tr>
<td>PostalDistrict</td>
<td>Discrete</td>
<td>Four digit number indicating the postal district of the customers. Larger cities are centred around round numbers such as 2000 and 8000.</td>
</tr>
<tr>
<td>HouseholdInt100m</td>
<td>Continuous</td>
<td>The household intensity per hectare, measured by the hectare around the customers household.</td>
</tr>
<tr>
<td>HouseholdInt1km</td>
<td>Continuous</td>
<td>The household intensity per hectare, measured by the nearest square kilometre around the customers household.</td>
</tr>
<tr>
<td>HouseholdInt10km</td>
<td>Continuous</td>
<td>The household intensity per hectare, measured by the nearest 10×10 kilometre cell around the customers household.</td>
</tr>
<tr>
<td>Duration</td>
<td>Continuous</td>
<td>The number of years the customers has been active in the company.</td>
</tr>
<tr>
<td>InsuranceLimit</td>
<td>Continuous</td>
<td>The maximum payout on the coverage in £.</td>
</tr>
<tr>
<td>Deductible</td>
<td>Continuous</td>
<td>The amount of a loss, which the customers has to pay himself, measured in £.</td>
</tr>
<tr>
<td>Coverages</td>
<td>Discrete</td>
<td>Six cipher binary indicator from the IT-systems. If forth digit is one the insurance has comprehensive coverage. Not interesting for modelling purposes.</td>
</tr>
<tr>
<td>Covariate</td>
<td>Type</td>
<td>Description</td>
</tr>
<tr>
<td>-----------------</td>
<td>----------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>PolicyID</td>
<td>Discrete</td>
<td>A unique identification number to identify the row.</td>
</tr>
<tr>
<td>SpecialCoverage</td>
<td>Discrete</td>
<td>&quot;Yes&quot;, &quot;No&quot; or &quot;Unknown&quot;. If &quot;Yes&quot; some kind of extra coverage is written on the policy. &quot;Unknown&quot; occurs due to some old policies from a IT-system with a different data structure.</td>
</tr>
<tr>
<td>YearSinceClaim</td>
<td>Continuous</td>
<td>Number of years since the last claim on the policy.</td>
</tr>
<tr>
<td>Claims</td>
<td>Continuous</td>
<td>The number of claims. Mostly equals one, but sometimes two due some technicalities.</td>
</tr>
<tr>
<td>Severity</td>
<td>Continuous</td>
<td>The sum of the payout plus RBNS, IBNR and IBNER, measured in £.</td>
</tr>
<tr>
<td>ExpoStartDate</td>
<td>Date</td>
<td>The date of the exposure start (not relevant for severities).</td>
</tr>
<tr>
<td>ExpoEndDate</td>
<td>Date</td>
<td>The date of the exposure end (not relevant for severities).</td>
</tr>
<tr>
<td>Exposure</td>
<td>Continuous</td>
<td>The length of the exposure measured in years (not relevant for severities).</td>
</tr>
<tr>
<td>ClaimDate</td>
<td>Date</td>
<td>The date of the accident.</td>
</tr>
<tr>
<td>ClaimMonth</td>
<td>Continuous</td>
<td>The month the accident measured since January 2005, this being month number one.</td>
</tr>
<tr>
<td>Year</td>
<td>Discrete</td>
<td>The year of the accident.</td>
</tr>
</tbody>
</table>

Table A.1: List of covariates in dataset from empirical chapter.
Appendix B  Explorative analysis for GLM

Figure B.1: Explorative GLM analysis. Empirical means of severities against grouped continuous covariates. The thin line shows the actual means and the bold line shows a GAM smoothed version. The bars pictures the share of the total exposure.
Figure B.1: Explorative GLM analysis. (a)-(c): Empirical means of severities against grouped continuous covariates. The thin line shows the actual means and the bold line shows a GAM smoothed version. (d)-(i): Bold line shows the actual means over the values of the discrete variable. (All): The bars pictures the share of the total exposure.
Appendix C  Comparison graphs

Figure C.1: Lorentz plots for model comparison, Section 6.1. Bold solid line corresponds to perfect ordering, thin solid line corresponds to random selection and dotted line shows the relevant models ordering. The closer the dotted line is to the solid bold line, the better the model describes data.
Figure C.1: Lorentz plots for model comparison, Section 6.1. Bold solid line corresponds to perfect ordering, thin solid line corresponds to random selection and dotted line shows the relevant models ordering. The closer the dotted line is to the solid bold line, the better the model describes data.
Figure C.2: Time plot, Section 6.3. Solid line corresponds to perfect prediction, dotted line shows the relevant models predictive power over time. The closer the dotted line is to the solid line the better is the model.
Figure C.2: Time plot, Section 6.3. Solid line corresponds to perfect prediction, dotted line shows the relevant models predictive power over time. The closer the dotted line is to the solid line the better is the model.
References


