Heterogeneity of Hazard Rates in Insurance.
Spreeuw, J.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (http://dare.uva.nl)
Chapter 6

Prediction of Claim Numbers Based on Hazard Rates

6.1 Introduction

As demonstrated in Chapter 2, a life insurance portfolio is a typical example of an approach by means of hazard rates, since once a life dies, it leaves the portfolio for ever. As a consequence, statistical models based on hazard functions can be used in life business. Moreover, such models have found many applications concerning duration data, e.g. biometrics, labor economics and reliability theory.

Keiding et al. (1998) showed that they can be applied to non-life problems as well. The main topic of the paper mentioned involves the estimation of the hazard of occurrence of claims due to casualty insurance portfolios of several types, where as time dimension either calendar time or time elapsed since the last claim of each type is taken into account. The aim of this chapter is to demonstrate that the hazard rate approach can be used in IBNR (Incurred But Not Reported) and RBNS (Reported But Not Settled) problems as well. We will consider the prediction of future claim numbers, by which we mean claims which have already been incurred by a contract but either have not been reported yet (IBNR) or have already been reported but not yet settled (RBNS). In the remainder of this chapter, the terminology used will be in accordance with the IBNR case.

The set-up of this chapter is as follows. In Section 6.2, the problem is introduced and the run-off triangle of claim numbers is displayed. After that, in Section 6.3, an overview of papers concerning the modelling of claim numbers will be given. It will be clarified that the modelling of claim numbers is usually done as part of setting up an IBNR micro model where assumptions concerning claim numbers and claim sizes are made separately. In this respect, they are to be distinguished from macro models, where only aggregate claim amounts are considered. It will be explained that models concerning claim numbers, as far as they appeared in literature, are either simplified or fully parametric. Section 6.4 will, after imposing an additional assumption concerning the reporting of claims, translate the problem considered in terms of hazard functions. After this, some statistical methods based on a proportional structure of hazard functions, at least in a continuous time dimension, will be proposed. In Section 6.5, we will consider one of the most important models of this class, namely the Cox model, which has the property of
being semiparametric rather than parametric. Besides, if the number of claims is small compared to the number of contracts under consideration, the hazard function can be approximately regarded as the product of two effects. Hence it includes the chain ladder method and separation method as special cases, which will be discussed in Subsections 6.5.1 and 6.5.2. A numerical example is considered in Section 6.6. Section 6.7 gives conclusions as well as recommendations for further research.

6.2 The run-off triangle of claim numbers

A certain portfolio has been observed during \( J \) years. Let \( n_i \) be the number of contracts which pay premiums during the \( i \)-th year (or the period \([i-1,i)\)). These contracts are exposed to claims from the \( i \)-th year on. So we have the following, well known, run-off triangle for claim numbers:

<table>
<thead>
<tr>
<th># of contracts</th>
<th>Year of origin</th>
<th>Development year</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n_1 )</td>
<td>1</td>
<td>( n_{11} )</td>
</tr>
<tr>
<td>( n_1 )</td>
<td>2</td>
<td>( n_{12} )</td>
</tr>
<tr>
<td>( n_1 )</td>
<td>( J )</td>
<td>( n_{1J} )</td>
</tr>
<tr>
<td>( n_2 )</td>
<td>1</td>
<td>( n_{11} )</td>
</tr>
<tr>
<td>( n_2 )</td>
<td>2</td>
<td>( n_{22} )</td>
</tr>
<tr>
<td>( \vdots )</td>
<td>( \vdots )</td>
<td>( \vdots )</td>
</tr>
<tr>
<td>( n_J )</td>
<td>( J )</td>
<td>( n_{JJ} )</td>
</tr>
</tbody>
</table>

Figure 5.1 IBNR-triangle of claim numbers.

In this triangle, \( n_{ij} \ (i = 1, \ldots, J, i + j \leq J + 1) \) denotes the observed number of claims that have occurred in year \( i \) and are reported \( j - 1 \) years thereafter. The IBNR-problem concerns predicting the future claim numbers \( n_{ij} \) with \( i = 2, \ldots, J \) and \( i + j \geq J + 2 \).

6.3 Overview of comparable literature on prediction of claim numbers

One of the first articles on forecasting future IBNR claim numbers is Verbeek (1972). Let \( N_{ij} \) denote the random variable representing the number of claims occurring in year of origin \( i \) which are reported in year \( i + j - 1 \). Verbeek (1972) assumes that these numbers are Poisson distributed, with expectation equal to the product of two parameters to be estimated ("Po" is short for Poisson):

\[
N_{ij} \sim \text{Po} (\beta_{i+j-1} p_j) \tag{6.1}
\]

In the specification above, \( \beta_{i+j-1} \) and \( p_j \) represent the effects of calendar year of report and year of development, respectively. Hence Verbeek's model is to be interpreted as the separation method applied to claim numbers. In order to avoid overparametrization of
the model one usually imposes a restriction on the p-values, for instance \( \sum_{j=1}^{J} p_j = 1 \), as Verbeek did. In that case, \( \sum_{j=1}^{J} p_j \) can be considered as the proportion of claims settled upon the end of the \( K \)-th development year, assuming that all claims are settled after \( J \) years. This method results in the same outcomes of the estimates as the arithmetic separation method, which is based on mechanical smoothing and uses a restriction similar to the one above.

Verbeek's paper is one of the few considering only claim numbers. In most papers, such specifications do not stand on their own, but instead are part of a micro model on aggregate claim amounts. Micro models have, contrary to macro models which are based on aggregate claims data, the property that they focus on insured individuals. As a consequence, claim numbers and individual claim amounts are usually specified simultaneously. In this section, we will restrict ourselves to contributions where the modeling of claim numbers is based on the estimation of fixed quantities, because the main model in this chapter has the same property. Contributions where this applies and which will be discussed below, are Bühlmann et al. (1980), Witting (1987) and the special cases in Norberg (1986).

Norberg (1986), first of all, assumes that to each year \( i \), a latent general risk condition corresponds, represented by a random variable, say \( B_i \), in such a way that, conditionally given \( B_i = \beta_i \):

- the number of claims in year \( i \), denoted by \( N_i \), is Poisson distributed with parameter \( \nu_i \beta_i \), where \( \nu_i \) represents the amount of risk exposed, e.g. the number of contracts in force at the beginning of year \( i \), and
- the number of claims in year \( i \) reported in year \( i + j - 1 \), denoted by \( N_{ij} \), has the following distribution:

\[
N_{ij} \sim \text{Po}(\nu_i \beta_i p_j), \tag{6.2}
\]

with \( p_j \) denoting the probability that a claim is reported \( j \) years after occurrence. The domain of \( j \) is \( \{1, \ldots, M\} \) with \( M + 1 \) denoting the maximal time that can elapse between occurrence and notification of a claim.

Finally, quantities related to different occurrence years are assumed to be independent.

Note that the expected value of each random variable \( N_{ij} \) is described as the product of an effect of the year of origin and an effect concerning the development year. So we have that, in case \( \beta_i \) is fixed, instead of being the outcome of a random variable, as considered in Section 4 of Norberg (1986), Norberg’s specification of reported claim numbers reduces to the chain ladder method as \( \nu_i \) is known.

**Definition 1 (Chain ladder method)** When we write about "chain ladder method", we mean that the relevant quantities (e.g. the parameter in a Poisson distribution) can be expressed as a product of an effect of year of origin and an effect of development year. For an overview of such models, the reader is referred to Mack (1994). They are not to be confused with the original chain ladder method based on mechanical smoothing. In that case we will use the terminology "original chain ladder method". In Remark 3, estimates of \( n_{ij} \) corresponding to that approach will be given.
Concerning \( N_t \) as defined above, Bühlmann et al. (1980) and Witting (1987) use the same specification as Norberg (1986), except that \( \beta_i \) is supposed to be fixed and identical for all \( i \). Apart from that, however, the approach in both Bühlmann et al. (1980) and Witting (1987) is entirely different. They first concentrate on the time elapsed between the occurrence and first reporting of the claim, which is assumed to be identically distributed irrespective of the year of origin, and then assume each claim, once it has been reported for the first time, to follow a certain stochastic process. We will not consider this model any further. It should, however, be noted that, assuming

- a claim can only be reported once (this will also be assumed in the next section), and

- once a claim is reported, its size is known instantaneously,

the specification in both papers has much in common with the original chain ladder method.

So, only the papers by Norberg (1986) and Verbeek (1972) remain to be considered. Note that both involve a distributional assumption. In fact, both models, and related ones, are generalized linear models in the sense of Nelder & Wedderburn (1972).

The conclusion of the above survey is that the problem of the prediction of future claim numbers is either simplified by assuming that there are no significant factors except development time (Bühlmann et al., 1980 and Witting, 1987) or tackled by using a fully parametric approach (Norberg, 1986 and Verbeek, 1972).

In this chapter it will be shown that a method exists which not only allows for including more effects than development time alone, but, in addition, is semiparametric, provided that the available data can be considered as failure data or duration data. This requires the assumption that a claim occurred can only be reported once, which in many lines of non-life insurance is a not too severe restriction. In the next section we will translate the given prediction problem in terms of hazard functions and then introduce some available prediction methods.

### 6.4 The hazard rate approach

In the remainder of this chapter it is assumed that each contract can report at most one claim per year of occurrence. If we instead are dealing with an RBNS (Reported But Not Settled) problem, the assumption is that there will be at most one claim payment for each claim reported in some year of origin.

So we have that at the beginning of year \( j \) there are still \( n_t - \sum_{j_0=1}^{j-1} n_{t,j_0} \) contracts exposed to having had a claim occurring in year of origin \( i \).

We assume that the reporting times of contracts of the same year of origin are independent and identically distributed. Let \( T_i \) denote the reporting time of an arbitrary contract with year of origin \( i \). We define the probabilities

\[
p_{ij} = P[T_i = j],
\]
6.4. The hazard rate approach

and the hazard rates

\[ \lambda_{ij} = \frac{P[T_i = j]}{P[T_i \geq j]} \]  \hspace{1cm} (6.4)

So the hazard rate in this context is equal to the conditional probability that an arbitrary claim with year of origin \( i \) is reported at time \( T_i \), given that it had not been reported before. We assume further that the reporting times of contracts with different years of origin are independent.

Next the problem of predicting claim numbers will be put in a hazard functions framework, after which some statistical methods will be considered.

6.4.1 Translating the problem in terms of hazard functions

The random variable representing the number of claims from year of origin \( i \) which are reported in development year \( j \) is denoted by \( N_{ij} \). These claim numbers are observable for \( 0 \leq i + j \leq J + 1 \). The assumptions imply that claim numbers from different years of origin are independent and that the joint distribution of the observable claim numbers from year of origin \( i \) is the multinomial distribution given by

\[
\begin{align*}
\mathbb{P} \left[ \bigcap_{j=1}^{J-i+1} \{ N_{ij} = n_{ij} \} \right] &= \frac{n_i!}{\prod_{j=1}^{J-i+1} n_{ij}!} \cdot \prod_{j=1}^{J-i+1} p_{ij}^{n_{ij}} \cdot \left( 1 - \sum_{k=1}^{J-i+1} p_{ik} \right)^{n_i - \sum_{k=1}^{J-i+1} n_{ik}} \cdot (6.5)
\end{align*}
\]

where

\[
\left( \begin{array}{c}
\frac{n_i!}{\prod_{j=1}^{J-i+1} n_{ij}!} \cdot \frac{n_i}{n_i - \sum_{k=1}^{J-i+1} n_{ik}}
\end{array} \right)
\]

is a multinomial coefficient. This implies that the joint distribution of all observable claim numbers is given by

\[
\begin{align*}
\mathbb{P} \left[ \bigcap_{i=1}^{J-i} \bigcap_{j=1}^{J-i+1} \{ N_{ij} = n_{ij} \} \right] &= \prod_{i=1}^{J-i} \left( \begin{array}{c}
\frac{n_i!}{\prod_{j=1}^{J-i+1} n_{ij}!} \cdot \frac{n_i}{n_i - \sum_{k=1}^{J-i+1} n_{ik}}
\end{array} \right) \cdot \prod_{j=1}^{J-i+1} p_{ij}^{n_{ij}} \cdot \left( 1 - \sum_{k=1}^{J-i+1} p_{ik} \right)^{n_i - \sum_{k=1}^{J-i+1} n_{ik}} \cdot (6.7)
\end{align*}
\]

Hence the observable claim numbers satisfy the (unconditional) multinomial model. We have

\[
p_{ij} = \prod_{k=1}^{j-1} (1 - \lambda_{ik}) \cdot \lambda_{ij} \cdot (6.8)
\]
with, by definition:

$$\prod_{k=1}^{0} (1 - \lambda_{ik}) = 1.$$  \hspace{1cm} (6.9)

Furthermore,

$$1 - \sum_{k=1}^{J-i+1} p_{ik} = \prod_{k=1}^{J-i+1} (1 - \lambda_{ik}).$$ \hspace{1cm} (6.10)

So the joint distribution of all observable claim numbers can also be expressed in terms of the hazard rates of the reporting times and we obtain

$$P \left[ \bigcap_{i=1}^{J} \bigcap_{j=1}^{J-i+1} \{ N_{ij} = n_{ij} \} \right] = \prod_{i=1}^{J} \left\{ \left( \frac{n_{i}}{n_{i,1} \ldots n_{i,J-i+1}} \right) \cdot \prod_{k=1}^{J-i+1} \left( \frac{(1 - \lambda_{ik}) \cdot \lambda_{ij}}{\prod_{k=1}^{J-i+1} (1 - \lambda_{ik})} \right)^{n_{ij}} \right\} \cdot \left( \frac{n_{i}}{n_{i,1} \ldots n_{i,J-i+1}} \right) \prod_{j=1}^{J-i+1} \left( \frac{\lambda_{ij}^{n_{ij}} (1 - \lambda_{ij})^{n_{i} - \sum_{j=0}^{J-i} n_{j,f_j}}}{\prod_{i=1}^{J} \{ N_{ij} = n_{ij} \}} \right).$$ \hspace{1cm} (6.11)

This yields the following log-likelihood function to be maximized:

$$\ln \left[ P \left[ \bigcap_{i=1}^{J} \bigcap_{j=1}^{J-i+1} \{ N_{ij} = n_{ij} \} \right] \right] = \sum_{i=1}^{J} \ln \left[ \left( \frac{n_{i}}{n_{i,1} \ldots n_{i,J-i+1}} \right) \right] + \sum_{i=1}^{J} \sum_{j=1}^{J-i+1} \left( n_{ij} \ln [\lambda_{ij}] + \left( n_{i} - \sum_{j=0}^{J-i} n_{j,f_j} \right) \ln [1 - \lambda_{ij}] \right).$$ \hspace{1cm} (6.12)

The logarithms of the multinomial coefficients do not affect the optimal solutions of the likelihood maximization problem. They will be omitted in the special cases considered below.

**Remark 2 (The censoring concept)** Statistical models based on the hazard rate approach explicitly take into account that the study period, in this case equal to $J$ years, is limited and therefore not all contracts incurring claims the reporting time can be observed. Such contracts are said to be right-censored. The only thing we know about any contract with year of origin $i$, $i \in \{1, \ldots, J\}$, still in force at the end of calendar year $J$, is that it either incurred a claim with reporting time longer than $J - i + 1$, or did not incur a claim at all.
6.4. A non-parametric method considering only development year

One may assume that the relative development pattern of all claims is the same for each year of origin:

$$\lambda_{ij} = p_j \quad \forall i, j \in \{1, \ldots, J\},$$

equivalent to assuming that the variables $T_i$ defined above are identically distributed, $i \in \{1, \ldots, J\}$. By substituting this into (6.12), we get as maximum likelihood estimators of $p_j$, denoted by $\hat{p}_j$, $j \in \{1, \ldots, J\}$:

$$\hat{p}_j = \frac{\sum_{i=1}^{J-j+1} n_{ij}}{\sum_{i=1}^{J-j+1} \left( n_i - \sum_{j_0=1}^{j-1} n_{i,j_0} \right)},$$

with, by definition,

$$\sum_{j_0=1}^{0} n_{i,j_0} = 0.$$  

(6.15)

The right hand side of (6.14) can be seen to be the ratio of the number of "failures" in development year $j$ to the number of contracts "at risk" at the beginning of year $j$. The estimate of the corresponding survivor function, representing the probability that no claims are reported within $j$ years after occurring, is equal to:

$$P[\tilde{T}_i > j] = \prod_{k=1}^{j} (1 - \hat{p}_k) = \prod_{k=1}^{j} \left( \frac{\sum_{i=1}^{J-k+1} \left( n_i - \sum_{j_0=1}^{k} n_{i,j_0} \right)}{\sum_{i=1}^{J-k+1} \left( n_i - \sum_{j_0=1}^{k-1} n_{i,j_0} \right)} \right),$$

also known as the Kaplan-Meier or product limit estimate of the survivor function. The predictions for the future claim numbers, denoted by $\widetilde{n}_{ij}$, $i = 2, \ldots, J$, $j = J-i+2, \ldots, J$, are

$$\widetilde{n}_{ij} = n_i \prod_{k=1}^{j-1} (1 - \hat{p}_k) \cdot \hat{p}_j.$$

(6.17)

**Remark 3 (Comparison with original chain ladder method)** Contrary to the original chain ladder method, which instead has estimates

$$n_{ij} = n_{i,i-1} + \sum_{s=J-i}^{J} \hat{c}_{s,s+1},$$

with

$$\hat{c}_{s,s+1} = \frac{\sum_{i=s}^{J-s} n_{i,s+1}}{\sum_{i=1}^{J-s} n_{i,s}},$$

the above method, like all methods described in the remainder of this chapter, requires the availability of the number of contracts $n_i$, $i \in \{1, \ldots, J\}$, which in most practical cases is not a severe restriction.
Usually, however, one supposes that there are more effects than only the development pattern, so covariates will have to be included. Therefore in the next subsection we introduce a specification satisfying this requirement.

### 6.4.3 The proportional hazards specification

A class of models that has been widely used in many scientific disciplines, including biometrics and (labor) economics, is the proportional hazards model, where, in a continuous time space, the hazard function is factored as

\[ \lambda(t, z) = \phi(z, \beta)p(t), \]  

(6.20)

where \( z \) denotes a vector of covariates, \( \beta \) a vector of unknown coefficients, while \( p(t) \), an unknown function, denotes the baseline hazard function (baseline, since it is equal to the hazard function itself if \( \phi(z, \beta) = 1 \)). For an overview of the characteristics of this specification, see Kiefer (1988). One of the properties mentioned here is that the proportional effect of the hazard function on the covariates is independent of time:

\[ \frac{\partial \ln (\lambda(t, z))}{\partial z} = \frac{\partial \ln (\phi(z, \beta))}{\partial z}. \]  

(6.21)

One can, of course, assume that \( p(t) \) belongs to a parametric family, and then estimate all coefficients by means of maximum likelihood, but this would give no advantage compared to the approaches considered in last section. If however, one assumes \( \phi(z, \beta) \) to be as follows:

\[ \phi(z, \beta) = e^{\gamma z}, \]  

(6.22)

then Cox (1972) showed that specification of \( p(t) \) is not required. In fact, we are then dealing with the semiparametric Cox model, which will be discussed in the next section.

### 6.5 The Cox model applied to prediction of claim numbers

In this section, two specifications of the hazard function are considered which both prove to be special cases of the Cox model. In case time \( t \) is continuous, this model is of the following general shape:

\[ \lambda(t, z) = p(t)\exp(\beta'z), \]  

(6.23)

with \( z \) and \( \beta \) denoting vectors of regressors and regression coefficients, respectively, while \( p(\cdot) \) represents the baseline hazard. This model was introduced in Cox (1972) and discussed extensively in Kalbfleisch & Prentice (1980).

The Cox model has a broad range of applications, in particular in medical statistics. Its popularity is derived from the semiparametric estimation method first exhibited in Cox (1972). The method is semiparametric because
6.5. The Cox model applied to prediction of claim numbers

- it is not parametric: estimation of $\beta$ does not require specification of the baseline hazard function, and
- it is not non-parametric: the hazard function is specified proportionally and log-linearity of the effect of regressors $z$ is assumed (the latter argumentation is indicated by Keiding et al. (1998)).

Because the Cox model is not non-parametric, it is not assumption-free. However, this can hardly be called a drawback, at least when compared to the models of Norberg (1986) and Verbeek (1972), discussed in Section 3, as, regarding the effects, they cannot do without specifications either. The assumptions related to the Cox model are even less restrictive than those related to the parametric method considered in Norberg (1986) and Verbeek (1972), as the latter two require specification of all parameters. Therefore, the probability of a misspecification in the Cox model is smaller than in the above mentioned parametric models.

In general, parametric methods have the advantage that the estimation (usually by means of maximum likelihood) is relatively easy, but the drawback that the model may be misspecified. For non-parametric methods, it is just the opposite: estimation may be cumbersome but there is no chance of a misspecification. We will see in the next subsections that the estimation of the parameters $\beta$ and the function $\phi(t)$ in the semiparametric Cox model - by maximum likelihood - is not very complicated. Therefore, the model considered in this section is not only more accurate than the parametric models discussed in the previous paragraph, but also shares its advantages.

Defining the following survivor functions $F(t; z)$ and $F(t)$, in this context to be interpreted as the probability that no claim is reported before time $t$ by a contract with covariate vector $z$ or by a "baseline" contract characterized by $\exp(\beta'z) = 1$:

\begin{align*}
F(t; z) &= e^{-\int_0^t \lambda(t; z)dt}; \\
F(t) &= e^{-\int_0^t \phi(t)dt},
\end{align*}

we have that the following relationship holds:

\begin{equation}
F(t; z) = (F(t))^{\exp(\beta'z)}.
\end{equation}

By using the effect "development time" as the time variable, the Cox model can be used to predict future claim numbers. However, IBNR-models are discrete time models. By adopting $j \in \{1, \ldots, J\}$ as discrete time variable, like we did before, the two respective relationships between hazard rate and survivor function, displayed in (6.24) and (6.25), respectively, should be altered in the following way:

\begin{align*}
F_j(z) &= \prod_{k < j} (1 - \lambda_k(z)); \\
F_j &= \prod_{k < j} (1 - \mu_k),
\end{align*}
with $F_j(z)$ and $F_j$ denoting the respective probabilities that the development time of a claim is larger than $j$. Replacing $F(t;z)$ and $F(t)$ in (6.26) with $F_j(z)$ and $F_j$, respectively, and then substituting for each $j \in \{1, \ldots, J\}$ the equalities (6.24) and (6.25), results in

$$\lambda_j(z) = 1 - (1 - p_j)^{\exp(\beta'z)},$$  \hspace{1cm} (6.29)

see Kalbfleisch & Prentice (1980, p. 98, formula (4.30)).

**Remark 4 (Difference with the approach in Cox, 1972)** In the original paper of Cox (1972) the discrete time model

$$\frac{\lambda_j(z) d_j}{1 - \lambda_j(z) d_j} = \frac{p_j d_j}{1 - p_j d_j} \exp(\beta'z),$$  \hspace{1cm} (6.30)

is proposed instead of (6.29). It also reduces to the continuous one in case the time intervals become infinitesimally small. Kalbfleisch & Prentice (1980) indicate that the differences between (6.29) and (6.30) tend to be very small if the values of the hazard functions are small as well, but that the above specification is only an approximation in the discrete time framework.

We assume that the values for either the hazard function or the baseline hazard are small. It follows that (6.23) can be used as an approximation:

$$\lambda_j(z) = p_j \exp(\beta'z).$$  \hspace{1cm} (6.31)

In the run-off triangle, each item of data, i.e. a claim number, reveals two pieces of information, one of them being "development year". The other quantity is either "year of origin" or "calendar year of report". This results in the chain ladder method and separation method, respectively, if approximation (6.31) can be used, as both $z$ and $\beta$ are one-dimensional. The two mentioned cases will be considered in the next two subsections.

There will be problems with the numerical computations in this case, because the parameters cannot be defined uniquely. In order to avoid convergence problems, one can select a value of $\beta$ and determine the other values in an analogous manner, as is also done in the classical approach.

Finally, similar to the approach in Kalbfleisch & Prentice (1980, p. 99), the following transformation will be used in order to remove range restrictions on the parameters:

$$\gamma_j = \ln(- \ln(1 - p_j)).$$  \hspace{1cm} (6.32)

### 6.5.1 The chain ladder method

If the chain ladder method is applied, each hazard rate $\lambda_{ij}$ is assumed to depend on two effects representing development year and year of origin, respectively:

$$\lambda_{ij} = p_j \exp(\beta_i).$$  \hspace{1cm} (6.33)
6.5. The Cox model applied to prediction of claim numbers

Using transformation (6.32), the likelihood function, displayed in (6.12), becomes

\[
\ln L = \sum_{i=1}^{J} \left\{ \sum_{j=1}^{J-i+1} n_{ij} \ln \left[ 1 - \exp \left[ -\exp (\gamma_j + \beta_i) \right] \right] \right\} .
\]  

(6.34)

The score statistics (see Kalbfleisch & Prentice (1980, p. 99-100)) are

\[
\frac{\partial \ln L}{\partial \gamma_j} = \sum_{i=1}^{J-j+1} n_{ij} \exp (\gamma_j + \beta_i) e^{-\exp (\gamma_j + \beta_i)} \left[ 1 - e^{-\exp (\gamma_j + \beta_i)} \right]^{-1}
\]

\[ - \sum_{i=1}^{J-j+1} \left( n_i - \sum_{j_0=1}^{j} n_{i,j_0} \right) \exp (\gamma_j + \beta_i) ,
\]  

(6.35)

and

\[
\frac{\partial \ln L}{\partial \beta_i} = \sum_{j=1}^{J-i+1} n_{ij} \exp (\gamma_j + \beta_i) e^{-\exp (\gamma_j + \beta_i)} \left[ 1 - e^{-\exp (\gamma_j + \beta_i)} \right]^{-1}
\]

\[ - \sum_{j=1}^{J-i+1} \left( n_i - \sum_{j_0=1}^{j} n_{i,j_0} \right) \exp (\gamma_j + \beta_i) .
\]  

(6.36)

The Fischer information matrix (see Kalbfleisch & Prentice (1980, p. 100)) is:

\[
H = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix} = \begin{pmatrix} \frac{\partial^2 \ln L}{\partial \gamma \gamma} & \frac{\partial^2 \ln L}{\partial \gamma \beta} \\ \frac{\partial^2 \ln L}{\partial \beta \gamma} & \frac{\partial^2 \ln L}{\partial \beta \beta} \end{pmatrix},
\]  

(6.37)

where both $H_{11}$ and $H_{22}$ are diagonal matrices. The diagonal elements of $H_{11}$ are

\[
\frac{\partial^2 \ln L}{\partial \gamma_j^2} = \sum_{i=1}^{J-j+1} n_{ij} \exp (\gamma_j + \beta_i) e^{-\exp (\gamma_j + \beta_i)} \left[ \exp (\gamma_j + \beta_i) - 1 \right] \left[ 1 - e^{-\exp (\gamma_j + \beta_i)} \right]^{-2}
\]

\[ + \sum_{i=1}^{J-j+1} \left( n_i - \sum_{j_0=1}^{j} n_{i,j_0} \right) \exp (\gamma_j + \beta_i) .
\]  

(6.38)

The elements of $H_{12}$ and $H_{21}$ are

\[
\frac{\partial^2 \ln L}{\partial \gamma_j \partial \beta_i} = n_{ij} \exp (\gamma_j + \beta_i) e^{-\exp (\gamma_j + \beta_i)} \left[ e^{-\exp (\gamma_j + \beta_i)} + \exp (\gamma_j + \beta_i) - 1 \right] \left[ 1 - e^{-\exp (\gamma_j + \beta_i)} \right]^{-2}
\]

\[ + \left( n_i - \sum_{j_0=1}^{j} n_{i,j_0} \right) \exp (\gamma_j + \beta_i) .
\]  

(6.39)
152

Prediction of Claim Numbers Based on Hazard Rates

The diagonal elements of $H_{22}$ are:

\[ -\frac{\partial^2 \ln L}{\partial \beta_i^2} = \sum_{j=1}^{J-i+1} n_{ij} \left\{ \exp (\gamma_j + \beta_i) e^{-\exp(\gamma_j+\beta_i)\ln(1-e^{-\exp(\gamma_j+\beta_i)})} \right\} \cdot \left. \left( e^{-\exp(\gamma_j+\beta_i)} + \exp (\gamma_j + \beta_i) - 1 \right) \right\} \cdot \left( 1 - e^{-\exp(\gamma_j+\beta_i)} \right)^{-2} \]

\[ + \sum_{j=1}^{J-i+1} \left( n_i - \sum_{j_0=1}^{J-i+1} n_{i,j_0} \right). \] (6.40)

We can compute the required estimators by means of the Newton-Raphson approach. Only a matrix of order $J$ has to be inverted numerically. The simple starting value in Kalbfleisch & Prentice (1980, p. 101) is

$\beta_i = 0$ for $i = 1, \ldots, J$, \hspace{1cm} (6.41)

and the maximum likelihood estimator of $\gamma$ at $\beta = 0$:

$\gamma_j = \ln \left[ -\ln \left( \frac{\sum_{i=1}^{J-j+1} n_{ij}}{\sum_{i=1}^{J-j+1} n_{i,j-1}} \right) \right]$, $j = 1, \ldots, J$. \hspace{1cm} (6.42)

**Remark 5** Note that this estimator is actually the same as (6.14).

### 6.5.2 The separation method

If one uses the separation method, one assumes that the hazard rate is the product of two effects, namely the development time and the calendar year:

$\lambda_{ij} = p_j \exp (\beta_{i+j-1})$. \hspace{1cm} (6.43)

In terms of this approach, the random variable $N_{ij}$ can be interpreted as the number of claims with development time $j$ occurring in calendar year $i + j - 1$.

Compared to the previous subsection, formulas (6.34) to (6.36) and (6.38) to (6.40) have to be modified, while (6.41) and (6.42) remain the same. The results are:

\[ \ln L = \sum_{i=1}^{J} \left\{ \sum_{j=1}^{J-i+1} n_{i-j+1,j} \ln \left[ 1 - \exp \left( -\exp (\gamma_j + \beta_i) \right) \right] \right\} - \sum_{i=1}^{J} \left( n_i - \sum_{j_0=1}^{J-i+1} n_{i,j_0} \right) \exp (\gamma_j + \beta_i) \} \] \hspace{1cm} (6.44)

The score statistics are:

\[ \frac{\partial \ln L}{\partial \gamma_j} = \sum_{i=j}^{J} n_{i-j+1,j} \exp (\gamma_j + \beta_i) e^{-\exp(\gamma_j+\beta_i)} \left( 1 - e^{-\exp(\gamma_j+\beta_i)} \right)^{-1} \]

\[ - \sum_{i=j}^{J} \left( n_{i-j+1} - \sum_{j_0=1}^{J-i+1} n_{i-j+1,j_0} \right) \exp (\gamma_j + \beta_i) \] \hspace{1cm} (6.45)
\[ \frac{\partial \ln L}{\partial \beta_i} = \sum_{j=1}^{i} n_{i-j+1,j} \exp \left( \gamma_j + \beta_i \right) e^{-\exp(\gamma_j + \beta_i)} \left( 1 - e^{-\exp(\gamma_j + \beta_i)} \right)^{-1} \]
\[ - \sum_{j=1}^{i} \left( n_{i-j+1} - \sum_{j_0=1}^{j} n_{i-j+1,j_0} \right) \exp \left( \gamma_j + \beta_i \right). \] (6.46)

Again both \( H_{11} \) and \( H_{22} \) are diagonal matrices. The diagonal elements of \( H_{11} \) are:
\[ -\frac{\partial^2 \ln L}{\partial \gamma_j^2} = \sum_{i=j}^{J} n_{i-j+1,j} \left\{ \exp \left( \gamma_j + \beta_i \right) e^{-\exp(\gamma_j + \beta_i)} \left( e^{-\exp(\gamma_j + \beta_i)} + \exp \left( \gamma_j + \beta_i \right) - 1 \right) \right\} \cdot \left( 1 - e^{-\exp(\gamma_j + \beta_i)} \right)^{-2} \]
\[ + \sum_{i=j}^{J} \left( n_{i-j+1} - \sum_{j_0=1}^{j} n_{i-j+1,j_0} \right) \exp \left( \gamma_j + \beta_i \right). \] (6.47)

The elements of \( H_{12} \) and \( H_{21} \) are:
\[ \frac{\partial^2 \ln L}{\partial \gamma_j \partial \beta_i} = n_{i-j+1,j} \exp \left( \gamma_j + \beta_i \right) e^{-\exp(\gamma_j + \beta_i)} \left( e^{-\exp(\gamma_j + \beta_i)} + \exp \left( \gamma_j + \beta_i \right) - 1 \right) \]
\[ \cdot \left( 1 - e^{-\exp(\gamma_j + \beta_i)} \right)^{-2} \]
\[ + \left( n_{i-j+1} - \sum_{j_0=1}^{j} n_{i-j+1,j_0} \right) \exp \left( \gamma_j + \beta_i \right). \] (6.48)

The diagonal elements of \( H_{22} \) are:
\[ -\frac{\partial^2 \ln L}{\partial \beta_i^2} = \delta_{in} \sum_{j=1}^{i} n_{i-j+1,j} \left\{ \exp \left( \gamma_j + \beta_i \right) e^{-\exp(\gamma_j + \beta_i)} \right\} \cdot \left( e^{-\exp(\gamma_j + \beta_i)} + \exp \left( \gamma_j + \beta_i \right) - 1 \right) \cdot \left( 1 - e^{-\exp(\gamma_j + \beta_i)} \right)^{-2} \]
\[ + \delta_{in} \sum_{j=1}^{i} \left( n_{i-j+1} - \sum_{j_0=1}^{j} n_{i-j+1,j_0} \right). \] (6.49)

The starting value is the same as in the chain ladder method.

### 6.6 Numerical example

In this section the methods derived will be applied on a real insurance portfolio. The data are shown below. The number of contracts each year of origin is equal to 70,000 (so \( n_i = 70,000 \) for \( i = 1, \ldots, J \)). The elements of the run-off triangle (see Figure 5.2) have normal font while the actual development of claim numbers is bold faced:
As mentioned in Section 2, one of the parameters of the vector \((\beta_1, \ldots, \beta_{10})\) has to be selected. In this case, we have fixed \(\beta_{10}\) and \(\beta_1\) at 0 for the chain ladder method and the separation method, respectively. After 183 and 134 iterations, respectively, the following results for the estimations of the parameters have been obtained (s.e. denotes standard error):

### Table 5.3a
Results of estimation for chain ladder method; estimates, both absolute and relative with respect to standard error, and corresponding gradient.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(\gamma_1)</td>
<td>-6.8660</td>
<td>-174.901</td>
<td>0.0009</td>
<td>(\beta_1)</td>
<td>0.0258</td>
<td>21.529</td>
<td>0.0009</td>
</tr>
<tr>
<td>(\gamma_2)</td>
<td>-6.8565</td>
<td>-165.256</td>
<td>0.0021</td>
<td>(\beta_2)</td>
<td>0.0827</td>
<td>65.547</td>
<td>-0.0050</td>
</tr>
<tr>
<td>(\gamma_3)</td>
<td>-7.4071</td>
<td>-129.772</td>
<td>-0.0021</td>
<td>(\beta_3)</td>
<td>-0.0191</td>
<td>-14.277</td>
<td>-0.0026</td>
</tr>
<tr>
<td>(\gamma_4)</td>
<td>-7.7782</td>
<td>-106.079</td>
<td>0.0007</td>
<td>(\beta_4)</td>
<td>-0.4567</td>
<td>-319.400</td>
<td>0.0016</td>
</tr>
<tr>
<td>(\gamma_5)</td>
<td>-8.0872</td>
<td>-88.221</td>
<td>0.0004</td>
<td>(\beta_5)</td>
<td>-0.1558</td>
<td>-100.849</td>
<td>0.0008</td>
</tr>
<tr>
<td>(\gamma_6)</td>
<td>-8.5163</td>
<td>-68.129</td>
<td>-0.0001</td>
<td>(\beta_6)</td>
<td>-0.0364</td>
<td>-21.490</td>
<td>0.0007</td>
</tr>
<tr>
<td>(\gamma_7)</td>
<td>-8.5760</td>
<td>-60.042</td>
<td>0.0018</td>
<td>(\beta_7)</td>
<td>-0.2036</td>
<td>-107.633</td>
<td>0.0006</td>
</tr>
<tr>
<td>(\gamma_8)</td>
<td>-9.4489</td>
<td>-38.948</td>
<td>-0.0009</td>
<td>(\beta_8)</td>
<td>-0.1293</td>
<td>-59.197</td>
<td>-0.0002</td>
</tr>
<tr>
<td>(\gamma_9)</td>
<td>-8.8092</td>
<td>-41.310</td>
<td>-0.0009</td>
<td>(\beta_9)</td>
<td>-0.4231</td>
<td>-158.214</td>
<td>0.0010</td>
</tr>
<tr>
<td>(\gamma_{10})</td>
<td>-8.9815</td>
<td>-26.932</td>
<td>-0.0008</td>
<td>(\beta_{10})</td>
<td>-0.0000</td>
<td>-0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>
Table 5.3b

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>γ₁</td>
<td>-7.0268</td>
<td>-179.002</td>
<td>0.0003</td>
<td>β₂</td>
<td>-0.0987</td>
<td>-36.920</td>
<td>0.0005</td>
</tr>
<tr>
<td>γ₂</td>
<td>-7.0359</td>
<td>-169.583</td>
<td>0.0001</td>
<td>β₃</td>
<td>0.3873</td>
<td>177.288</td>
<td>0.0003</td>
</tr>
<tr>
<td>γ₃</td>
<td>-7.5722</td>
<td>-132.668</td>
<td>0.0005</td>
<td>β₄</td>
<td>0.1126</td>
<td>59.492</td>
<td>0.0020</td>
</tr>
<tr>
<td>γ₄</td>
<td>-7.8836</td>
<td>-107.507</td>
<td>0.0000</td>
<td>β₅</td>
<td>0.0754</td>
<td>44.575</td>
<td>0.0032</td>
</tr>
<tr>
<td>γ₅</td>
<td>-8.1574</td>
<td>-88.982</td>
<td>0.0011</td>
<td>β₆</td>
<td>0.0045</td>
<td>-2.904</td>
<td>0.0073</td>
</tr>
<tr>
<td>γ₆</td>
<td>-8.5775</td>
<td>-68.613</td>
<td>0.0001</td>
<td>β₇</td>
<td>0.0682</td>
<td>47.653</td>
<td>0.0011</td>
</tr>
<tr>
<td>γ₇</td>
<td>-8.6158</td>
<td>-60.306</td>
<td>0.0000</td>
<td>β₈</td>
<td>0.0819</td>
<td>61.193</td>
<td>0.0003</td>
</tr>
<tr>
<td>γ₈</td>
<td>-9.3497</td>
<td>-38.566</td>
<td>0.0017</td>
<td>β₉</td>
<td>0.2609</td>
<td>-206.779</td>
<td>0.0017</td>
</tr>
<tr>
<td>γ₉</td>
<td>-8.6037</td>
<td>-40.348</td>
<td>-0.0006</td>
<td>β₁₀</td>
<td>-0.0515</td>
<td>-42.995</td>
<td>-0.0033</td>
</tr>
<tr>
<td>γ₁₀</td>
<td>-8.9031</td>
<td>-26.712</td>
<td>0.0002</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The elements of both the vectors (1, ..., 10) and either (1, ..., 9) or (2, ..., 10), respectively, turned out to be mutually uncorrelated. Therefore only the correlation coefficients between the elements of the two respective vectors will be given:

<table>
<thead>
<tr>
<th>β₁</th>
<th>β₂</th>
<th>β₃</th>
<th>β₄</th>
<th>β₅</th>
<th>β₆</th>
<th>β₇</th>
<th>β₈</th>
<th>β₉</th>
</tr>
</thead>
<tbody>
<tr>
<td>γ₁</td>
<td>-0.004</td>
<td>-0.004</td>
<td>-0.003</td>
<td>-0.003</td>
<td>-0.004</td>
<td>-0.005</td>
<td>-0.004</td>
<td>-0.005</td>
</tr>
<tr>
<td>γ₂</td>
<td>-0.004</td>
<td>-0.004</td>
<td>-0.003</td>
<td>-0.003</td>
<td>-0.004</td>
<td>-0.005</td>
<td>-0.004</td>
<td>-0.005</td>
</tr>
<tr>
<td>γ₃</td>
<td>-0.003</td>
<td>-0.003</td>
<td>-0.003</td>
<td>-0.002</td>
<td>-0.003</td>
<td>-0.004</td>
<td>-0.004</td>
<td>-0.005</td>
</tr>
<tr>
<td>γ₄</td>
<td>-0.003</td>
<td>-0.003</td>
<td>-0.003</td>
<td>-0.002</td>
<td>-0.003</td>
<td>-0.003</td>
<td>-0.003</td>
<td>-0.000</td>
</tr>
<tr>
<td>γ₅</td>
<td>-0.002</td>
<td>-0.003</td>
<td>-0.003</td>
<td>-0.002</td>
<td>-0.003</td>
<td>-0.003</td>
<td>-0.000</td>
<td>-0.000</td>
</tr>
<tr>
<td>γ₆</td>
<td>-0.002</td>
<td>-0.002</td>
<td>-0.002</td>
<td>-0.002</td>
<td>-0.002</td>
<td>-0.000</td>
<td>-0.000</td>
<td>-0.000</td>
</tr>
<tr>
<td>γ₇</td>
<td>-0.002</td>
<td>-0.003</td>
<td>-0.002</td>
<td>-0.002</td>
<td>-0.002</td>
<td>-0.000</td>
<td>-0.000</td>
<td>-0.000</td>
</tr>
<tr>
<td>γ₈</td>
<td>-0.002</td>
<td>-0.002</td>
<td>-0.002</td>
<td>-0.002</td>
<td>-0.002</td>
<td>-0.000</td>
<td>-0.000</td>
<td>-0.000</td>
</tr>
<tr>
<td>γ₉</td>
<td>-0.003</td>
<td>-0.003</td>
<td>-0.000</td>
<td>-0.000</td>
<td>-0.000</td>
<td>-0.000</td>
<td>-0.000</td>
<td>-0.000</td>
</tr>
<tr>
<td>γ₁₀</td>
<td>-0.004</td>
<td>-0.000</td>
<td>-0.000</td>
<td>-0.000</td>
<td>-0.000</td>
<td>-0.000</td>
<td>-0.000</td>
<td>-0.000</td>
</tr>
</tbody>
</table>

Figure 5.4a

Correlation coefficients between estimated factors; chain ladder method.
Prediction of Claim Numbers Based on Hazard Rates

The IBNR-square with estimated claim numbers for the chain ladder method, to be compared with the actual claim numbers in Figure 5.2, is shown in Figure 5.5a. On the other hand, regarding the separation method we have restricted ourselves to the estimation of the claim numbers available. The reason lies in the fact that the pattern of subsequent estimates in the row \( \hat{\beta}_2, \ldots, \hat{\beta}_j \) is irregular, so it is very hard to extrapolate these in order to obtain the estimates \( \beta_{j+1}, \ldots, \beta_{2j} \), if no information about future developments is available.

### Figure 5.5a

IBNR-square with estimated claim numbers for the chain ladder method (to be compared with Figure 5.2).
6.7. Conclusions and recommendations for further research

In this chapter, the IBNR problem of predicting future claim numbers was translated into one concerning the estimation of hazard functions. Such an approach relies on the assumption that a claim occurring can only be reported once. If this holds, the non-life actuary has a variety of statistical methods at his disposal. In this chapter, we discussed a major one, namely the Cox model based on proportional hazard rates. It has, in contrast to the parametric methods which have appeared in the literature up to now, the property that it is semiparametric. In this case, this means that estimation of the coefficients related to the effects (covariates) does not require specification of the baseline hazard function, representing the effect of development year. As a consequence, the chance of a misspecification is not as great for the Cox model as it is for a parametric model.

If one supposes that the hazard functions, or the conditional probabilities that a claim is reported given that it has not been reported before, are small, then, by approximation, each hazard function can be specified as the product of several effects. This reduces to either the chain ladder method or the separation method, depending on whether, besides development year, the single effect of either year of development or the effect of calendar year of report, respectively, is included in the model. For both the special cases, it turned
out that the parameters were relatively easy to estimate by means of maximum likelihood. Both versions were also numerically illustrated.

Recalling formula (6.29)

$$\lambda_j(z) = 1 - (1 - p_j)^{\exp(\beta^T z)}, \quad (6.50)$$

note that the model dealt with in this chapter only allows for $z$ being equal to one of the unit vectors of length $J$. Other exogenous factors, such as region and age, can, however, be included by simply altering the length of both the vectors $z$ and $\beta$, such that to each additional effect, one additional element in both $\beta$ and $z$ corresponds.