Eötvös Loránd University
Corvinus University of Budapest

Zoltán Milotai

Frequency and severity models in reserving

MSc Thesis

Supervisor:

Gabriella Antalffy-Németh

Actuary

Budapest, 2016

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Current reserving methodology and it's shortcomings

The current reserving methodology applied by Hungarian insurance companies are depending on claims triangulational methods, aggregating the observed claim payments as it can be seen in 1.1 below.

Calendar Month	Incurred/Service Month												
	May-01	Jun-01	Jul-01	Aug-01	Sep-01	Oct-01	Nov-01	Dec-01	Jan-02	Feb-02	Mar-02	Apr-02	May-02
0	447,214	452,494	405,550	437,787	417,387	474,083	407,673	442,587	454,181	380,028	423,140	429,631	429,656
1	538,819	681,641	660,011	690,135	617,342	682,192	625,153	675,124	608,263	674,512	709,767	698,883	
2	271,003	204,624	231,426	367,778	256,459	196,974	419,248	231,065	383,807	182,944	191,327		
3	279,777	200,106	107,810	53,168	85,579	102,279	71,569	96,288	154,518	50,315			
4	21,122	17,748	65,640	74,421	66,469	65,293	33,996	10,388	131,772				
5	109,798	17,997	22,049	31,236	5,988	210,728	25,843	7,124					
6	9,762	5,258	17,653	14,744	15,750	15,866	3,088						
7	12,859	3,105	5,212	7,762	4,317	7,862							
8	2,369	3,103	8,725	4,699	5,880								
9	-5,365	1,852	3,390	1,567									
10	87,211	3,969	-618										
11	763	2,635											
12	12,861												

Figure 1.1: An example for a claims development triangle.

This is not a coincidence, current Hungarian legislation regarding insurance technical reserves (43/2015. (III. 12.) Korm. rendelet) states that for IBNR (Incurred but not reported losses) reserve calculation (on lines

of business with at least three years of existence) :73 §(2) b" for claims of insurance contracts the IBNR necessity has to be calculated based on previous years experience with methods using claim triangular data."

Reserving methods based on claims triangulation, are inherently compressing the data resulting in loss of precious information about individual losses, that bars us from deriving an adequate distribution for IBNR and RBNS (Reported but not settled) losses . After triangulation the method applied deriving the reserve amount can be very sophisticated, but since the data loss already present can only be used for point type estimation.

As an example to fathom this let us say, that one has for example 5,000 losses over a period of 12 months, and uses them to build a triangle such as that in figure 1.1, one is left with only

$$12 * \frac{12+1}{2} = 78$$

points to go by to project the losses to ultimate and to estimate the distribution of outstanding claims or reserves. And if we were to look at 10,000 losses over the same period, still the 78 points would be the result of aggregation. Although these estimation methods were understandably very useful in times when calculations were performed by hand, and the triangular approach meant significant ease in calculation, now that we have advanced calculating power due to computers this compression is unnecessary.

There are more problems about the triangle based methods. Current state of the art pricing methodology is already applying stochastic frequency and severity models in calculation whilst triangular methods in reserving. This means that we have two misaligned valuation frameworks for what is ultimately the same risk, but looked from two different points of view: prospectively (pricing) and retrospectively (reserving)! Also for solvency capital to be in accordance to the EU Solvency II standards, the infamous 99.5 percentile needs to be calculated, that can only obtained through distribution estimates not point reserve estimates.

In the following I would like to show based mainly along the lines of the framework elaborated in Pietro Parodi's article [1], that frequency and severity models can improve estimation punctuality in contrast to triangle based methods. I will do so by preparing an IBNR estimation model and aggregated claims distribution for the professional liability portfo-

lio of a Hungarian insurance company, based on the past 16 years claim database(2000-2015).

1.2 Benefits of triangle free model over the chain-ladder

Apart from the question of accuracy and predictive power, the triangle-free approach has several advantages. Some of these are listed below:

- Any other information we have about claims can be easily built into the model e.g. different treatment for claims below or above a given threshold;
- meaningful results may be gained for accident years with scarce or even nil claim counts where chain-ladder would not yield reasonable results;
- the calculation of the tail factor can be done in a more sophisticated fashion rather than in the heuristic expert judgment that is typical of triangle-based approaches;

Chapter 2

Outline

2.1 Base outline of the model

I Estimate the delay distribution, based on the empirical distribution of delays (here the distribution might be biased, as only a limited time window of data is available therefore claims with exceedingly great delay are not represented in the sample data. As a consequence an adjustment might be adequate to counter that e.g. in form of a tail fitting).

II Use the delay distribution in **I** to estimate the number of incurred but not reported (IBNR) claims based on the number of claims reported to date.(This will be done separately for each accident year.) Also determine the most suitable frequency model (e.g. Poisson, Negative Binomial) accordingly.

III Model the severity distribution for the IBNR claims (this may be different for each loss year, or at least depend on claims inflation), also taking IBNER (incurred but not enough reserved/reported) claims into account.

IV Combine the frequency and severity distributions via Monte Carlo simulation or another method (e.g. Fast Fourier Transform, Panjer recursion. . .) to produce an estimate of the aggregate distribution of IBNR losses

As a consequence, after the completion of the model our estimations

are able to provide confidence intervals and percentiles of the future claim amounts, thus granting much more sophisticated results as the triangle based point estimates.

Chapter 3

Model

3.1 Data

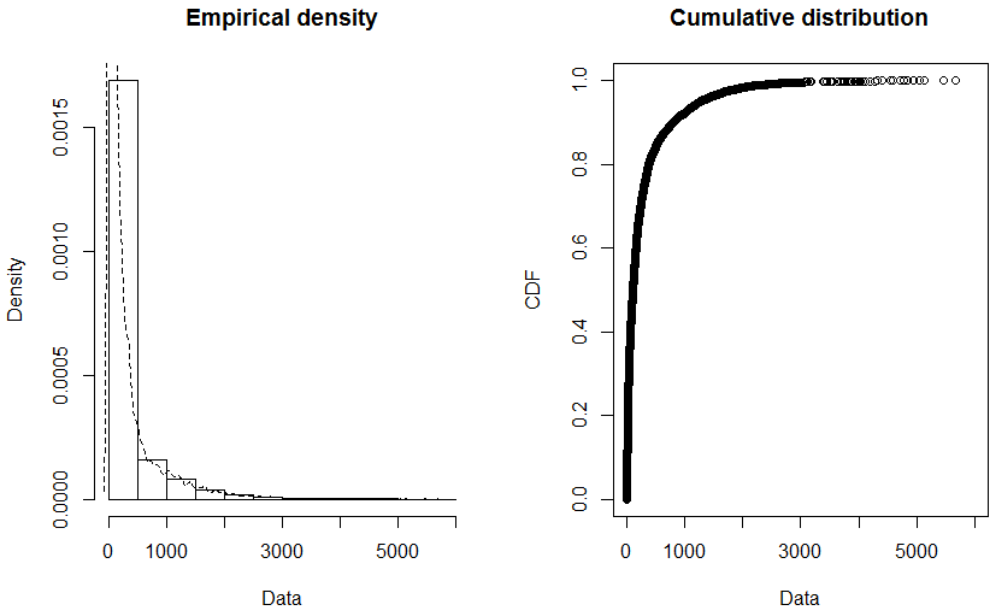
For data we will use a database from a Hungarian insurance company (hereinafter referred to as "the Company"). The database contains 16 years of claim experience (2000-2015) for Professional Liability line of business[1]. The database contained 23 thousand rows of data recording incremental changes (reserve increase/decrease, payment done, recourse) of nearly 13,000 claims over the above mentioned period. This line of business (lob) was chosen due to it's tendency for long run patterns, as it is a good candidate for observing delays. Also according to the Company's Actuary this line of business have seen remarkably little change in its products structure, thus making estimation more reliable. The chosen lob contains motley professional liability insurance, just to name a few type: tax advisory, wind-up companies, accountancy, private investigation, security, financial institution liability etc. (The first candidate for choosing an ideal lob for estimation would have been MTPL [Motor third party liability] as this lob is usually one of the most prominent for non-life insurance companies, and also has long run pattern. However in case of the Company, the products sold in this lob has been greatly altered in recent years, making estimation increasingly cumbersome.)
Luckily the obtained database was very detailed not only containing one payment per case, but recording on a different record each time a payment

[1] Naturally the data have been applied a positive monotone transformation for encryption to protect the privacy of the insurance company

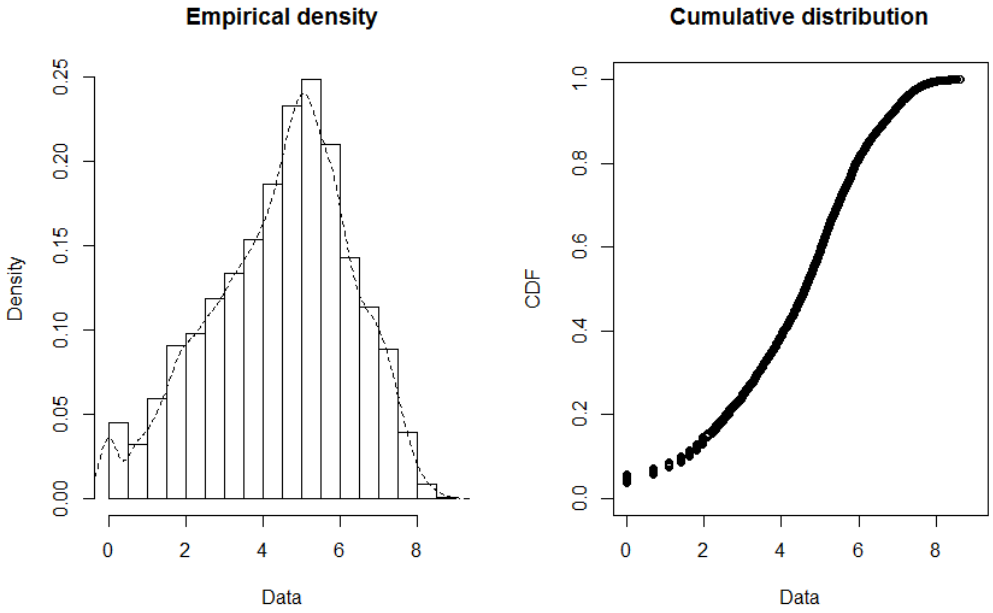
is made or reserve is created/released. Therefore it makes us able to calculate not only IBNR claim frequency and severity of the claim distribution, but the effect of IBNER(Incurred But Not Enough Reserved) as well.

3.2 Delay in reporting

I have used (mostly) R to conduct my analysis. (The R codes used for delivering my results (or most of them) can be seen at appendix section.) The first analysis performed was aimed to assess the delays between the occurrence of insurance events and their reporting date to the insurance company, that in turn will be used to construct frequency distribution. First I have examined empirical density and cumulative distribution of the delays.



Due to the nature of the data(high amount of small observations and a very few large ones), a logarithmic transformation made it much easier to see through. See density and distribution after logarithmic transformation on the below figure.



As delays have non-negative values 0 included, I added one to the delay values as this way the set of feasible distributions for testing fitness will include log-normal distribution(and as all fitting attempts failed with the original delay numbers I decided to use these increased values, and later adjust the result).

The following candidates were considered when looking for distribution best describing claim delays : weibull, pareto, log-normal, gamma, exponential, log-logistic. (As loglogistic and pareto distributions have multiple parameters, they are need to be estimated as well. I have wrote for cycles in r to estimate the best parameters for them.) Based on the resulting AIC and BIC values, the best fit (the lowest value both in AIC and BIC) was produced by weibull curve, second and third being log-normal and loglogistic. I have prepared a table to illustrate (in the above sense) the three best fit distribution together.

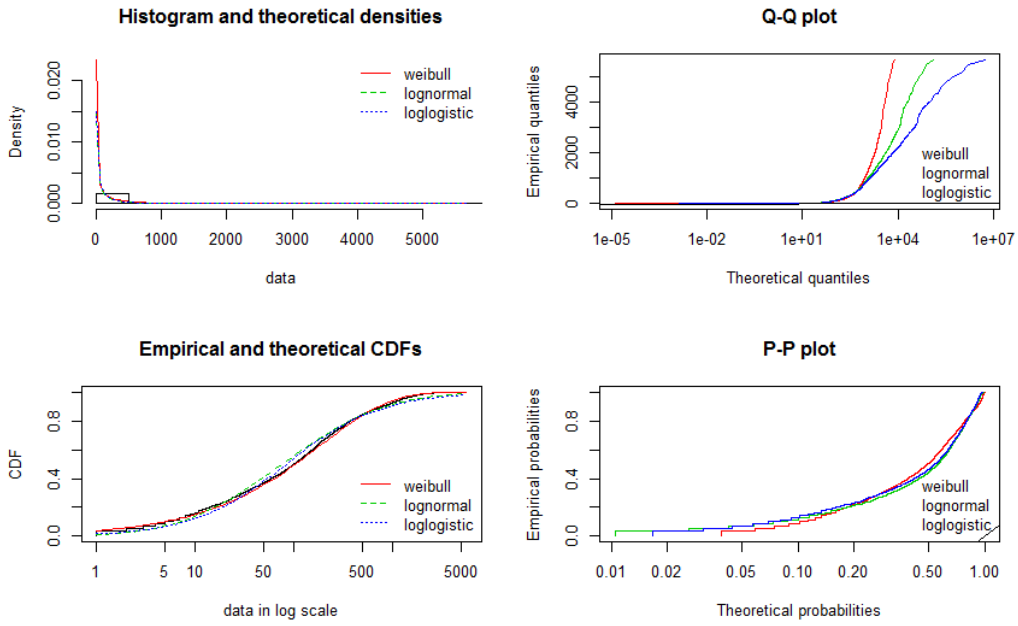


Figure 3.1: The comparison of the three best fitting distributions(by AIC).

The weibull distribution had the following estimated parameters (provided by R fitting):

Table 3.1: Parameters of the selected weibull distribution.

	estimate	error
shape	0.6156942	0.004161964
scale	189.4480040	2.869202594

There is also the question of goodness-of-fit. In that sense all of the above mentioned fit provided by R was poor, resulting in Kolmogorov-Smirnoff values ranging from 3%-20%. Only in case of exponential distribution(which was one of the worse in terms of AIC) have we seen a K-S ratio of greater than 10% namely 20%. Also based on more details elaborated in the following sections(to be able to compute variance and take into account disappearing claims), I decided to predict the claim numbers based on empirical cumulative distribution function with only using the above fittings for tail estimation, and calculating claim numbers for

each accident year separately. Before going into detail on exactly how I estimated claim numbers I summarize the theoretical approach.

3.3 IBNR claim count - point estimate

With the help of the delay cumulative distribution function (hereinafter referred to as $F(t)$), we are able to predict the IBNR claim count for the [0,t] period the following way. Lets denote the (so far unknown) frequency density function of claims with $\nu(t)$, the already reported number of claims with n_t and the total number of claims with N_t. This $\nu(t)$ function varies the same way, as the probability of having a claim, thus allowing us to take into account seasonality. If we know this function, we could easily arrive to the expected total number of occurred claims N_t:

$$\mathbf{E}(N_t) = \int_0^t \nu(\tau)\mathrm{d}\tau \tag{3.1}$$

To the calculation of the already reported part of the former, we can use the delay distribution estimated in the above section:

$$\mathbf{E}(n_t) = \int_0^t \nu(\tau)F_(t-\tau)\mathrm{d}\tau \tag{3.2}$$

In our case where n_t is known, and N_t is searched the following estimation can be used:

$$\hat{N_t} = \frac{\int_0^t \nu(\tau)\mathrm{d}\tau}{\int_0^t \nu(\tau)F_(t-\tau)\mathrm{d}\tau} n_t \tag{3.3}$$

By the assumption of uniform claim frequency density function, this boils down to the following formula:

$$\hat{N_t} = \frac{t}{\int_0^t F_(t-\tau)\mathrm{d}\tau} n_t \tag{3.4}$$

and as a result the number of occurred but not reported claims at t:

$$\hat{N_t} - n_t = \frac{t}{\int_0^t F_(t-\tau)\mathrm{d}\tau} n_t - n_t \tag{3.5}$$

In our case we will need a discrete formula, which based on the above equation takes the form of:

$$\hat{N}_t - n_t = \frac{t}{\sum_{\tau=0}^{t} F_(t-\tau)} n_t - n_t \tag{3.6}$$

(If we were to estimate the outstanding claims as a whole(instead of by accident years), the above formula would be the one used.)

One more thing is necessary for us to consider before calculating the expected number of claims.

3.3.1 Disappearance rate - proportion of reported claims that eventually become zero

The other important factor is the zero claims. In the previous section where the claim count estimation was performed all reported claim were taken into account. However as a natural course of claim reporting, in many cases eventually no payment takes place. This could be due to several reason including fraudulent reporting, court case etc.

In order to us to capture the true nature of claim count distribution, we have to make an estimation on the proportion of these would-be zero claims, and adjust the expected claim numbers.(As the other workaround would be to take this into account at the severity distribution, but finding distribution that is actually zero in many percent of the cases, and has good fit to the positive part of the sample deems highly unlikely.) I have prepared the following table about the numbers and exposures (reserve amount) of disappearing claims subtracted from the available data, summarizing how many years have passed from occurrence before a claim was rejected.

We can observe a high amount of disappearance rates, more than third of all reported claims disappear, both in terms of numbers and exposure.

In liability line of business the claims can be quite high and therefore the Company is more likely to debate claims in court. This high ratio can be interpreted as a "success" ratio for the Company as every rejected claim is money saved. We can also observe on table 3.2 that while in claim numbers 63% remains, in exposure only 62%. We can state based on the data, that bigger claims are more likely to disappear, not just because of the final remainder values. In numbers we can see a higher amounts for the

first two years, but in case of exposure a heavier tail is observable. This means that while smaller claims tend to nullify in their first few years, more substantial claims entailing court case persist for longer periods.

Table 3.2: The ratio of claims becoming zero.

years till declared zero	number ratio	exposure ratio
0	19.86%	7.88%
1	7.46%	6.80%
2	3.16%	4.02%
3	2.12%	4.09%
4	1.15%	2.67%
5	1.18%	3.30%
6	0.79%	3.50%
7	0.51%	2.25%
8	0.30%	1.27%
9	0.13%	0.51%
10	0.12%	0.44%
11	0.09%	0.39%
12	0.10%	0.32%
13	0.08%	0.45%
14	0.03%	0.10%
15	0.03%	0.15%
non-zero claims	62.87%	61.87%

However in our case we need to calculate the ratios from a different point of view, as we need to apply it to only late claims. (The above tables contain disappearance rates by time passed from occurrence, without taking into account the delay) Therefore I also calculated ratios of how big part of claims (in exposure) ultimately disappears based on delay in reporting.(As the last delay years(12-15) have had very small sample size, they were not used in the estimation)

Table 3.3: Disappearance rates by delay of reporting in years.

Delay years	Ultimately disappearing ratio
0	33.7
1	32.31
2	39.12
3	29.19
4	49.33
5	25.62
6	34.75
7	20.05
8	35.03
9	52.62
10	36.54
11	48.63
12	82.41
13	19.12
14	0
15	0

3.3.2 Claim count estimation by vintages

In order to use the above information, we not only need an aggregate number of expected claims, we need them by vintages. And based on the section 3.2 unsuccessful distribution estimation I decided to use the empirical cumulative distribution function of the sample with some modifications.

As empirical CDF-s do not inherently contain tails, and in our case liability is a line of business prone to long tails, we have to make a tail estimation to our vintages, otherwise we would certainly underestimate the number of claims. I decided to use the tail from the best fitting exponential distribution of the whole sample(in all of my endeavors to produce an agreeable goodness-of-fit for the claim frequency this was the only distribution with Kolmogorov-Smirnoff results greater than 10%, namely 20%.).

For estimation by vintages we have to slightly alter the 3.6 formula to estimate the claim numbers. (Because e.g. for the 2000 vintage we now don't estimate the late claims reported after 2000, but the late claims re-

ported after 2015). Our estimated ECDF function (let's call it E) uses 16 years of data, or in days 5844 and after that it only takes the value of 1. And also denote probability of delay being greater than 5844 days by $D(5844)$.(Which is calculated from the fitted exponential distribution with parameter $\lambda = 0.003554269$.)

The amended formula for outstanding number of claims (in case of claims occurred in year 2000) is the following:

$$\hat{N}_t - n_t = \frac{365}{\sum_{i=0}^{365}(E_(5844 - i)(1 - D(5844)) + D(5844))} n_t - n_t \tag{3.7}$$

where

$$D(5844) = 1 - (1 - e^{-\lambda * 5844}) = e^{-0.003554269 * 5844} \approx 9.53246... \times 10^{-}10$$

I also prepared the modified formulas for the another vintages and the following table contains the results before and after the application of disappearance rates.

Table 3.4: Point estimate claim count by years with and without disappearance rate.

Accident year	Before disappearance rates	After disappearance rates
2015	422.28	285.83
2014	160.14	97.49
2013	70.49	49.91
2012	32.24	16.33
2011	18.4	13.69
2010	15.77	10.29
2009	7.52	6.01
2008	3.72	2.42
2007	3.49	1.65
2006	2.26	1.43
2005	1.25	0.64
2004	1.08	0.19
2003	0.52	0.42
2002	0.22	0.22
2001	0.15	0.15
2000	0.03	0.03

3.4 IBNR claim count - distribution

As we use frequency-severity model for claim forecast, we cannot use a single point estimate for the claim numbers, for greater accuracy we need a distribution. The two most commonly used distribution for claim frequency are the Poisson and the negative binomial. The ubiquitous Poisson's great advantage is that only requires one parameter, the mean. Wright [2] argues in his paper that if at the estimation of parameters any of the following four parameter uncertainty is present in our model, then they account for increase in variance.

- Estimation uncertainty:
- Heterogeneity
- Contagion
- exposure uncertainty

As multiple of the above applies in our examined case, the variance have to be greater than the mean, therefore making Poisson distribution inappropriate. Therefore we will use negative binomial in our calculation. For this all we need to have is the variance, as the mean was already estimated in section 3.3.

More precisely we will estimate mean-to-variance ratio. After calculating the projected claims numbers for each year separately, calculating the variance from these ultimate claim numbers then dividing with the average of the claim numbers. (This is only adequate with taking uniform exposure over the years granted. As we had no unbiased exposure to use we had to accept this supposition.)

I have computed the yearly ultimate claim numbers and aggregated into the following table.

Table 3.5: Ultimate claim numbers by accident years.

Year	Ultimate claim
2000	743.03
2001	917.15
2002	606.22
2003	603.52
2004	599.08
2005	567.25
2006	653.26
2007	772.49
2008	994.72
2009	799.52
2010	760.77
2011	770.4
2012	936.24
2013	1298.49
2014	1325.14
2015	1235.28

Based on the above ultimate claim numbers I have calculated (with R) the variance-to-mean ratio to be 73.73125. This concludes our search for frequency distribution with negative binomial of the following parameters:

$$\text{Mean} = \frac{rp}{1-p} = 486.7 \quad \text{Variance} = \frac{rp}{(1-p)^2} = 35,885 \tag{3.8}$$

3.5 IBNR claim severity - preparation

Before setting on to estimating claim severity there are several issues that need to be addressed. The claim amount depends on many factors, here we mention the three most important are:

- claim inflation;
- IBNER;
- business mix;

The first and second item will be elaborated below.

Regarding the last one, meaning the change of the composition of business mix written year-to-year, we unfortunately have no detailed historical data. According the Company's actuary this line has been mainly unchanged throughout the years and as no other information available we did not examine this effect in our analysis.

3.5.1 Claim inflation

The first is the application of proper claim inflation. As our data takes up an extensive period in time - 16 years. Since Hungary experienced sometimes as high as 10% inflation during that period an adjustment in claim size is necessary to make the 2015 year claims comparable to claims taken place in 2000. I have used the most widely available inflation measure Customer Price Index (CPI) retrieved from the Hungarian national bank. It was available in monthly granularity, and it was applied to the data also on a monthly basis. One may find the yearly (accumulated) values in the following table.

Table 3.6: Yearly CPI from Hungarian National Bank.

year	inflation
2000	10.08%
2001	6.82%
2002	4.99%
2003	5.84%
2004	5.62%
2005	3.55%
2006	6.56%
2007	7.65%
2008	3.75%
2009	5.83%
2010	4.58%
2011	3.96%
2012	4.88%
2013	0.39%
2014	-1.00%
2015	0.89%

3.5.2 IBNER

A portion of the claims in the obtained data set are not fully developed yet, and the IBNER ratios are necessary to calculate to arrive at the ultimate value of claims. We will examine the year to year change in already reported claim amounts to determine year-to-year change in claim amounts. Some differentiation however, needs to be made.

The claims in the database in terms of reporting delay are highly vary, there are even claims with 15 years of delay. Assuming that the future IBNER ratios for a claim reported few months after occurrence, and a claim reported 10 years after occurrence are the same, is not a reasonable assumption in my opinion. Therefore we will differentiate IBNER ratios based on reporting tardiness. So claims that were reported within 1 year of occurrence will be group delay 0, claims that were reported between 1 and 2 years after occurrence will be delay group 1 etc. Of course this entails that for higher reporting delay categories we will have fewer data and thus less robust result, however the inherent difference between the categories makes this differentiation pivotal.

Table 3.7: IBNER - The observed exposure amounts per delay group.

Delay	Initial Claim Amount
0	1,846
1	440
2	306
3	231
4	103
5	93
6	34
7	107
8	21
9	4
10	13
11	7
12	1
13	2
14	1
15	2

First I checked whether enough data will be available for all years. You can see the claim amounts on table 3.7 in millions. As expected for claims with highly delayed reporting period there are very little exposure. Therefore some of the higher categories will need to be unified.

As you can see in the table above, all delay groups after year seven have limited amount of exposure. Keeping that in mind, let us have a look at the result at table 3.8

The analysis have been performed for observing the changes throughout all the 15 years period, however no observable change has taken place in the amounts after ten years of development after the reporting of the claim, so the following table is cropped to ten years.

Table 3.8: IBNER - The observed changes in claim amounts per delay group.

Delay	Initial	year 1	year 2	year 3	year 4	year 5	year 6	year 7	year 8	year 9	year 10
0	100	90.73	90.97	88.97	88.65	82.94	79.59	78.07	72.21	72.55	72.32
1	100	115.25	114.97	111.37	111.87	103.35	97.57	96.97	96.89	96.89	96.89
2	100	77.01	79.44	84.31	84.27	66.64	64.8	64.59	64.08	64.06	64.06
3	100	113.22	110.29	107.78	89.96	89.96	66.43	49.84	49.01	49.45	49.45
4	100	85.01	83.81	70.7	73.77	73.77	74.12	74.12	65.09	65.09	65.09
5	100	154.22	143.99	143.99	143.99	138.89	138.81	138.81	138.81	138.81	138.81
6	100	80.99	87.31	87.03	90.26	88.8	87.47	87.47	82.77	82.77	82.77
7	100	95.47	93.8	93.8	94.93	94.93	94.93	94.08	94.05	94.08	94.08
8	100	83.84	75.12	83.84	90.94	90.94	90.94	90.94	90.94	90.94	90.94
9	100	82.93	82.93	82.93	105.53	105.53	105.53	105.53	105.53	105.53	105.53
10	100	100	101.67	101.67	101.67	76.65	103.41	103.41	103.41	103.41	103.41
11	100	100	55.12	48.63	48.63	48.63	48.63	48.63	48.63	48.63	48.63
12	100	100	100	100	100	100	100	100	100	100	100
13	100	108.48	108.48	108.48	108.48	108.48	108.48	108.48	108.48	108.48	108.48
14	100	9.34	9.34	9.34	9.34	9.34	9.34	9.34	9.34	9.34	9.34
15	100	100	100	100	100	100	100	100	100	100	100

Due to the above mentioned, I decided to group the observed IBNER factors into the following six categories: 0, 1, 2, 3, 4-9, 9+. After regrouping the IBNER factors can be seen at table 3.9

Table 3.9: IBNER - The observed changes in claim amounts per condensed delay group.

Group	Initial	year 1	year 2	year 3	year 4	year 5	year 6	year 7	year 8	year 9	year 10
1	100	90.73	90.97	88.97	88.65	82.94	79.59	78.07	72.21	72.55	72.32
2	100	115.25	114.97	111.37	111.87	103.35	97.57	96.97	96.89	96.89	96.89
3	100	77.01	79.44	84.31	84.27	66.64	64.8	64.59	64.08	64.06	64.06
4	100	113.22	110.29	107.78	89.96	89.96	66.43	49.84	49.01	49.45	49.45
5	100	105.37	102	98.75	100.95	99.51	99.46	99.21	96.2	96.2	96.2
6	100	97.75	86.94	85.24	85.24	72	86.16	86.16	86.16	86.16	86.16

The ratios gained from the database shows some level of prudence in estimation of the claims by the Company. In some cases, for group 2,4 and 5 we can see, some initial increase in claims amount, but eventually after 6 years all group's ratios slump below the initial value, meaning that in all groups potential claim payments set at the reporting of the claim is invariably higher than the ultimate payment. This in turn will mean, that overall at the severity computation we will have to scale downward the claims.

After consulting with the actuary it turned out this is not a coincidence. The guidelines for setting the RBNS reserves for claims is artificially made high, so that run-off results almost always end up positive.

3.6 IBNR claim severity - computation

Based on the above sections, in order to calculate severity we will use the ultimate value of claims reached by adjusting the claim amounts with the above calculated IBNER ratios, and naturally only taking into account eventually non-zero claims. Regarding the claim inflation I have first examined the average claim amounts per accident year to get a basic idea how the amount of average claim payment varied over the 16 year observing period.

The average claim per year can be seen in the table below without inflation adjustment.(I have adjusted the claims for IBNER before creating the comparison, otherwise it would have been misleading to compare old fully developed claims with more recent just registered ones.)

Based on table 3.10 we can not see a definite trend in claims. Due to the inflation observed in the Hungarian economy, we should observe an increase in claim amounts, which is not present. After consulting with the

Table 3.10: Average claim per accident year.

Accident year	Average claim
2000	153478
2001	176672
2002	77814
2003	150785
2004	105229
2005	86328
2006	88290
2007	150465
2008	194323
2009	245123
2010	213879
2011	162428
2012	122706
2013	149961
2014	211339
2015	202402

actuary, She assured me that this is not a coincidence, and they have performed a similar average claim test not just for the professional liability portfolio, but the whole liability line and observed that the average claim size remains considerably stable over the years. Therefore I concluded that application of inflation rates(that even reaches as high compounded value as 2.2) might not be suitable for my data. Therefore at the calculation of severity distribution I did not apply the calculated inflation rates.

Also I filtered out zero claims as they have been accounted for at the application of disappearances rates in case of claim numbers, see 3.3.1. Therefore my analysis was bent on computing the severity of the actual non-zero claim amounts. The method to work out the claim distribution was very similar to the case of the delay distribution, first I examined the empirical density and cumulative distribution. I again applied logarithmic scaling to get meaningful figures. The result can be seen on figure 3.2 below.

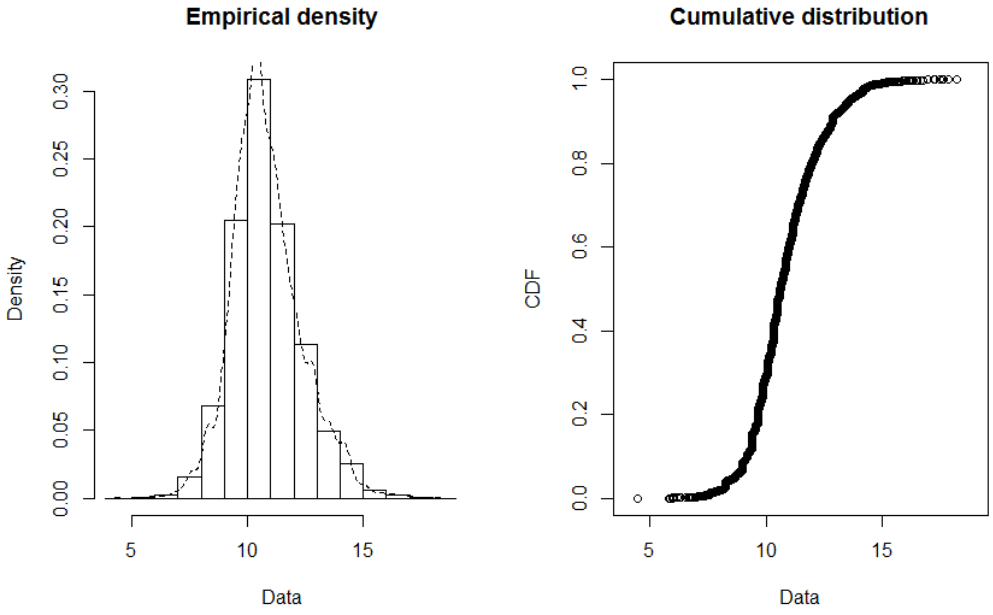


Figure 3.2: Observed empirical density and CDF.

Afterwards I have performed distribution fitting. The possible candidates were: log-normal, exponential, gamma, weibull, log-logistic and Pareto. The result was once again disappointing only gamma, log-normal and weibull did produce a fit, with goodness-of-fit results (Kolmogorov-Smirnoff) respectively 21.4%, 6.8% and 12.7%. As a conclusion I have decided to use the best-fitting gamma distribution.

So the result of severity model was a gamma distribution with the following parameters:

$$\begin{aligned} \text{Shape: } \alpha &= 0.403411 \\ \text{Rate(=1/scale): } \beta &= 0.000001556572 \end{aligned} \tag{3.9}$$

The mean and variance of the chosen distribution can be reached via the following formulas:

$$\begin{aligned} \text{Mean: } \frac{\alpha}{\beta} &= \frac{0.403411}{0.000001556572} \approx 259,166 \\ \text{Variance: } \frac{\alpha}{\beta^2} &= \frac{0.403411}{0.000001556572^2} \approx 166,498,110,074 \end{aligned} \tag{3.10}$$

3.7 Monte Carlo model

As we estimated both the frequency and severity distribution for our IBNR claims, we can now prepare Monte-Carlo simulation to model the distribution of outstanding claims. As (very conveniently) R have built-in random number generator for all the known distributions (including gamma and negative binomial) this made our model construction much easier. For simulation size I have decided to use a distribution based on 10,000,000 runs. This size would still run under 20 minutes, and yield robust results.(By robust I mean that when testing the received cdf the resulting probabilities for the same total claim amounts have differed less than 0.0001). A summary of the most important percentiles can be seen below.

Table 3.11: Percentiles of final IBNR distribution.

value	percentiles
270 937 958	0.995
251 634 528	0.99
179 948 905	0.9
145 052 811	0.75
111 794 450	0.5
83 933 802	0.25

The problem with these results, that we cannot compare it to anything. As the most important part of an estimation is to see whether it is a good estimate of real life values, I have decided to make adjusted results as well, where we have a real life data for comparison.

3.8 Comparison of results

The used database comprised of 16 years data. Therefore my idea was to use only the first 15 years of the data to prepare a model, then estimate the next one year IBNR, and this way I can compare the result with last years data. As the method I used to derive the "comparable" estimation is nearly the same as the one that was applied to derive the full database estimation, I will not go into details here just summarize the main points.(Luckily the overwhelming majority of my work was done

with R codes therefore the new results were gained with a few alterations from the original code) One important difference was that while in case of the full database we had to estimate all future IBNR claims from a point on, in case of the "comparable" estimation we need to estimate the IBNR claims surfacing in the next one year. So to derive the number of IBNR claims expected in the next one year first I computed all the expected IBNR claims from the 2014 year end, then I computed all the expected IBNR claims with the same model after 2015 year end then I subtracted the two value from each other. Apart from this the method was the same, only the calculated distributions parameters have differed. The newly estimated parameters were as follows:

$$\begin{aligned} &\text{Frequency distribution: Negative binomial} \\ &\text{Mean: } 424.1212 \\ &\text{Variance-to-mean ratio: } 64.2446 \\ &\text{Severity distribution: Gamma} \\ &\text{Shape: } \alpha = 0.4075858 \\ &\text{Rate(=1/scale): } \beta = 0.000001564 \end{aligned} \tag{3.11}$$

The mean of the distribution was 110,533,504 and the percentiles where the following:

Table 3.12: Percentiles of "comparable" IBNR distribution.

value	percentiles
255 052 102	0.995
236 759 145	0.99
169 251 500	0.9
136 283 863	0.75
104 925 159	0.5
78 647 158	0.25

The result gained from the actual last year database was (after adjusting to disappearances and IBNER) 179,580,212. Based on the above percentiles, a result at least as high as this will only occur in less than 10% of the cases. This is an indication that my model might underestimate the actual data.
I have analyzed the data, to see if I can find the reason.

I have found that two very high amount claim occurred in the actual data. These two claims out of the 433 account for the 1/6th of the whole actual claim amount(one of them was nearly the biggest claim in the whole 16 year database). By taking these outlying value out, the result shrink to less than 150,000,000 in value that is below the 80th percentile. Another reason could be the way how I derived claim severity. My frequency model predicted not just an aggregate number, but yearly outstanding claim numbers, so I can see that in both the total, and the "comparable" estimation about 94-95% of future claims are coming from the last three accident years. But when I calculated severity I did not weighted claim amounts according to these proportions. Below I calculated an amended result when I only take into account the last 3 years claims when calculating severity.

Table 3.13: Percentiles of last 3 years claim severity based IBNR.

value	percentiles
228 787 326	0.995
212 473 982	0.99
151 810 171	0.9
122 246 203	0.75
94 079 771	0.5
70 542 765	0.25

As we can see this does not help in our case, as the result are even lower, than in the all-years severity case.
One more idea would be the severity distribution. Even though only two distribution was able to produce a fit on our sample without error, and of these two gamma was the one with better K-S ratios, it does not have a heavy enough tail. I have calculated, that what are the chances that a 433 sample of gamma variables with our estimated parameters have the maximum gamma variable equal or greater to the observed (outlying) maximum, and the observed number of successful cases was zero(calculated on a 10,000,000 sample). Therefore my opinion is that a distribution with a heavier tail would have been better to represent the actual claim distribution.

Chapter 4

Acknowledgments

I would like to thank my supervisor, Gabriella Antalffy-Németh for providing me the very detailed database and much insight on it's structure. And also for the many advice and consultation which helped me to construct my model.

Bibliography

[1] Pietro Parodi: *Triangle-free reserving : a non-traditional framework for estimating reserves and reserve uncertainty*. London, 2013.

[2] Wright, Thomas: *A general framework for forecasting number of claims*. Actuarial studies in non-life insurance, Astin, 2007.

[3] Philip E. Heckman, Glenn G. Meyers: *The calculation of aggregate loss distributions from claims severity and claim count distributions*. Proceedings of the Casualty Actuarial Society, 1983

[4] John P. Robertson: *The computation of aggregate loss distribution*. Proceedings of the Casualty Actuarial Society, 1992

Appendix A

R codes

A.1 Delay computation

```
#sources, and output location
library("fitdistrplus")
library("actuar")
data_path<-"D:/other/MSc szakdoga/data/IBNR_szakdoga_adatok_sent_tempered.csv"
output_path<-"D:/other/MSc szakdoga/results/one_year_less/"

#reading, and ordering the database

database<-read.csv(data_path, sep =  ';', dec ='.')

#removing unnecessary coloumns and duplicates
colnames(database)
nrow(database)
database$delay<-database$eltérés.bejelentés.és.bekövetkezés.között.napokban
database <- subset(database, select = c(Kár.ID,Kárdátum.period,delay,Bejelentés.dátum.period))
database<-unique(database)
nrow(database)
plotdist(log(database$delay), histo = TRUE, demp = TRUE)

#first estimating an aggregate distribution for the database

minta<-database$delay

fln<-fitdist(minta,"lnorm", method = "mle")
summary(fln)

fe<-fitdist(minta,"exp")
summary(fe)

fg<-fitdist(minta,"gamma", method = "mle", lower = c(0, 0))
summary(fg)

fw<-fitdist(minta,"weibull")
summary(fw)
```

```
fll<-fitdist(minta, "llogis", start = list(shape = 1, scale = 50))
summary(fll)

#goodness-of-fit test
gofstat(fw)
gofstat(fe)
gofstat(fln)
gofstat(fg)
gofstat(fll)

#creation of vintages
database$vintage<-ceiling(database$Kárdátum.period/12)
#unique(database$vintage)
vintage_2000<-database[database$vintage==1,]
vintage_2001<-database[database$vintage==2,]
vintage_2002<-database[database$vintage==3,]
vintage_2003<-database[database$vintage==4,]
vintage_2004<-database[database$vintage==5,]
vintage_2005<-database[database$vintage==6,]
vintage_2006<-database[database$vintage==7,]
vintage_2007<-database[database$vintage==8,]
vintage_2008<-database[database$vintage==9,]
vintage_2009<-database[database$vintage==10,]
vintage_2010<-database[database$vintage==11,]
vintage_2011<-database[database$vintage==12,]
vintage_2012<-database[database$vintage==13,]
vintage_2013<-database[database$vintage==14,]
vintage_2014<-database[database$vintage==15,]
vintage_2015<-database[database$vintage==16,]

#constructing the empirical cumulative distribution function based on the database
#and adding an exponential tail

CDF<-ecdf(database$delay)
CDF_exp_tail <- function(x){
  z<-CDF(x)*(1-0.0000000009532461)+0.0000000009532461
  return(z)
}
outstanding<-c(1:16)
#CDF_exp_tail(100)

#vintage_2000

# using the ecdf with the exponential tail to estimate outstanding no of claims
sum<-0

for (i in 1:365) {
  sum<-sum+CDF_exp_tail(16*365.25-i)
}
outstanding[1]<-365/sum*nrow(vintage_2000)-nrow(vintage_2000)


#vintage_2001

# using the ecdf with the exponential tail to estimate outstanding no of claims
sum<-0
```

```
for (i in 1:365) {
  sum<-sum+CDF_exp_tail(15*365.25-i)
}
outstanding[2]<-365/sum*nrow(vintage_2001)-nrow(vintage_2001)


#vintage_2002

# using the ecdf with the exponential tail to estimate outstanding no of claims
sum<-0

for (i in 1:365) {
  sum<-sum+CDF_exp_tail(14*365.25-i)
}
outstanding[3]<-365/sum*nrow(vintage_2002)-nrow(vintage_2002)


#vintage_2003

# using the ecdf with the exponential tail to estimate outstanding no of claims
sum<-0

for (i in 1:365) {
  sum<-sum+CDF_exp_tail(13*365.25-i)
}
outstanding[4]<-365/sum*nrow(vintage_2003)-nrow(vintage_2003)


#vintage_2004

# using the ecdf with the exponential tail to estimate outstanding no of claims
sum<-0

for (i in 1:365) {
  sum<-sum+CDF_exp_tail(12*365.25-i)
}
outstanding[5]<-365/sum*nrow(vintage_2004)-nrow(vintage_2004)



#vintage_2005

# using the ecdf with the exponential tail to estimate outstanding no of claims
sum<-0

for (i in 1:365) {
  sum<-sum+CDF_exp_tail(11*365.25-i)
}
outstanding[6]<-365/sum*nrow(vintage_2005)-nrow(vintage_2005)

#vintage_2006

# using the ecdf with the exponential tail to estimate outstanding no of claims
sum<-0

for (i in 1:365) {
```

```
  sum<-sum+CDF_exp_tail(10*365.25-i)
}
outstanding[7]<-365/sum*nrow(vintage_2006)-nrow(vintage_2006)


#vintage_2007

# using the ecdf with the exponential tail to estimate outstanding no of claims
sum<-0

for (i in 1:365) {
  sum<-sum+CDF_exp_tail(9*365.25-i)
}
outstanding[8]<-365/sum*nrow(vintage_2007)-nrow(vintage_2007)


#vintage_2008

# using the ecdf with the exponential tail to estimate outstanding no of claims
sum<-0

for (i in 1:365) {
  sum<-sum+CDF_exp_tail(8*365.25-i)
}
outstanding[9]<-365/sum*nrow(vintage_2008)-nrow(vintage_2008)


#vintage_2009

# using the ecdf with the exponential tail to estimate outstanding no of claims
sum<-0

for (i in 1:365) {
  sum<-sum+CDF_exp_tail(7*365.25-i)
}
outstanding[10]<-365/sum*nrow(vintage_2009)-nrow(vintage_2009)


#vintage_2010

# using the ecdf with the exponential tail to estimate outstanding no of claims
sum<-0

for (i in 1:365) {
  sum<-sum+CDF_exp_tail(6*365.25-i)
}
outstanding[11]<-365/sum*nrow(vintage_2010)-nrow(vintage_2010)



#vintage_2011

# using the ecdf with the exponential tail to estimate outstanding no of claims
sum<-0

for (i in 1:365) {
  sum<-sum+CDF_exp_tail(5*365.25-i)
```

```
}
outstanding[12]<-365/sum*nrow(vintage_2011)-nrow(vintage_2011)

#vintage_2012

# using the ecdf with the exponential tail to estimate outstanding no of claims
sum<-0

for (i in 1:365) {
  sum<-sum+CDF_exp_tail(4*365.25-i)
}
outstanding[13]<-365/sum*nrow(vintage_2012)-nrow(vintage_2012)


#vintage_2013

# using the ecdf with the exponential tail to estimate outstanding no of claims
sum<-0

for (i in 1:365) {
  sum<-sum+CDF_exp_tail(3*365.25-i)
}
outstanding[14]<-365/sum*nrow(vintage_2013)-nrow(vintage_2013)


#vintage_2014

# using the ecdf with the exponential tail to estimate outstanding no of claims
sum<-0

for (i in 1:365) {
  sum<-sum+CDF_exp_tail(2*365.25-i)
}
outstanding[15]<-365/sum*nrow(vintage_2014)-nrow(vintage_2014)



#vintage_2015

# using the ecdf with the exponential tail to estimate outstanding no of claims
sum<-0

for (i in 1:365) {
  sum<-sum+CDF_exp_tail(1*365.25-i)
}
outstanding[16]<-365/sum*nrow(vintage_2015)-nrow(vintage_2015)
outstanding

write.csv(outstanding, file = paste0(output_path,"Yearly_outstanding_claims_before_disappearance.csv"))

#aggregating ultimate claim number per year
yearly_claim<-c(rep(0,16))
yearly_claim[1]=nrow(vintage_2000)
yearly_claim[2]=nrow(vintage_2001)
yearly_claim[3]=nrow(vintage_2002)
yearly_claim[4]=nrow(vintage_2003)
yearly_claim[5]=nrow(vintage_2004)
```

```
yearly_claim[6]=nrow(vintage_2005)
yearly_claim[7]=nrow(vintage_2006)
yearly_claim[8]=nrow(vintage_2007)
yearly_claim[9]=nrow(vintage_2008)
yearly_claim[10]=nrow(vintage_2009)
yearly_claim[11]=nrow(vintage_2010)
yearly_claim[12]=nrow(vintage_2011)
yearly_claim[13]=nrow(vintage_2012)
yearly_claim[14]=nrow(vintage_2013)
yearly_claim[15]=nrow(vintage_2014)
yearly_claim[16]=nrow(vintage_2015)

ultimate_claim<-c(rep(0,16))
for (i in 1:16) {
  ultimate_claim[i]=outstanding[i]+yearly_claim[i]
}

exposure_path<-"D:/other/MSc szakdoga/data/GWP_per_year.csv"

exposure<-read.csv(exposure_path, sep =  ';', dec ='.')


exposure_adjusted_ultimate<-c(rep(0,16))
for (i in 1:16) {
  exposure_adjusted_ultimate[i]=ultimate_claim[i]/exposure[i,2]*1000000
}

var(ultimate_claim)
ave(ultimate_claim)
var(exposure_adjusted_ultimate)

ultimate_and_exp=matrix(c(rep(0,64)),nrow = 16,ncol = 4)

for (i in 1:16) {
  ultimate_and_exp[i,1]=exposure[i,1]
  ultimate_and_exp[i,2]=exposure[i,2]
  ultimate_and_exp[i,3]=ultimate_claim[i]
  ultimate_and_exp[i,4]=exposure_adjusted_ultimate[i]

  }
write.csv(ultimate_claim, file = paste0(output_path,"ultimate_claim_numbers.csv"))
write.csv(ultimate_and_exp, file = paste0(output_path,"ultimate_claim_numbers_and_exp.csv"))
```

A.2 Disappearance rate derivation

```
#sources, and output location
library("fitdistrplus")
library("actuar")
claim_path<-"D:/other/MSc szakdoga/results/Yearly_outstanding_claims_before_disappearance.csv"
data_path<-"D:/other/MSc szakdoga/data/IBNR_szakdoga_adatok_sent_tempered_disappear.csv"
output_path<-"D:/other/MSc szakdoga/results/"

#reading, and ordering the database

outstanding<-read.csv(claim_path, sep =  ',', dec ='.')
database<-read.csv(data_path, sep =  ';', dec ='.')
```

```
#gsub("-", "0", database$Kifizetés, fixed = TRUE)
#database$Kifizetés<-as.numeric(database$Kifizetés)

#removing unnecessary coloumns and duplicates
colnames(database)
nrow(database)
#database$delay<-database$eltérés.bejelentés.és.bekövetkezés.között.napokban
#database <- subset(database, select = c(Kár.ID,Kárdátum.period,delay))
#database<-unique(database)
#nrow(database)
#plotdist(log(database$delay), histo = TRUE, demp = TRUE)

#creation of vintages by lateness in years
database$vintage<-floor(-(database$Kárdátumperiod-database$Bejelentésdátumperiod)/12 )
unique(database$vintage)

#finding the claims that became zero, sorting them into delay groups, and calculating total reserve
became_zero<-integer(12843) #12843 darab kárid van x=0 ha nem vált 0vá, x=egy ha igen
delay_group<-integer(12843) #késettség a 12843 elemre
reserve<-integer(12843) # az összege a pozitív tartalékoknak
for (i in 1:12843) {
  current_data<-database[database$KárID==i,]
  if ( sum(current_data$Kifizetés)==0 & sum(current_data$Tartalék)==0) {
    became_zero[i]=1
  }
  delay_group[i]=min(current_data$vintage)
  current_data_positive<-current_data[current_data$Tartalék>0,]
  reserve[i]<-sum(current_data_positive$Tartalék)
}
unique(delay_group)
#summarizing the above to the 16 delay group
full_reserve<-c(rep(0,16))
disappearing_reserve<-integer(16)

for (i in 1:12843) {
full_reserve[delay_group[i]+1]=full_reserve[delay_group[i]+1]+reserve[i]
  if  (became_zero[i]==1) {
disappearing_reserve[delay_group[i]+1]=disappearing_reserve[delay_group[i]+1]+reserve[i]
  }
}

#calculating disappearance rates to the 16 delay group
disappearance_rates<-c(rep(0,16))
for (i in 1:16) {
  disappearance_rates[i]=disappearing_reserve[i]/full_reserve[i]
}

#applying disappearance rates to the ultimate IBNR claim numbers computed in previous r code
result=c(rep(0,16))
for (i in 0:14) {
  result[i+1]=outstanding[16-i,2]*(1-disappearance_rates[i+2] )
}
result

write.csv(disappearance_rates, file = paste0(output_path,"disappearance_from_r.csv"))
write.csv(result, file = paste0(output_path,"Yearly_outstanding_claims_after_disappearance.csv"))
```

A.3 Claim severity and Monte Carlo

```
#sources, and output location
library("fitdistrplus")
library("actuar")
library("mcsm")
data_path<-"D:/other/MSc szakdoga/data/IBNR_szakdoga_adatok_sent_tempered_severity.csv"
output_path<-"D:/other/MSc szakdoga/results/"
IBNER_path<-"D:/other/MSc szakdoga/IBNER_percentage_v2.csv"
#reading the database that was manually ordered in excel
version
database<-read.csv(data_path, sep =  ';', dec ='.')
IBNER<-read.csv(IBNER_path, sep =  ';', dec ='.')

#removing measuring average claim per year
colnames(database)
nrow(database)
ncol(database)
database <- subset(database, select =
c(KárID,Kárdátumperiod,Bejelentésdátumperiod,Könyvelésihóperiod,Tartalék,Kifizetés))
#data_array<-matrix(c(rep(0,142860)), nrow = 23810, ncol = 6)
#data_array=database
#data_array[1,1:6]
database[3,"KárID"]
#creation of vintages
database$vintage<-ceiling(database$Kárdátumperiod/12)
database$delay_group=floor((database$Bejelentésdátumperiod-database$Kárdátumperiod)/12)

#unique(database$vintage) 1-16ig terjednek a vintage számok
#tail(database, n=1)
#computation of avereage claims per accident year groups
average_claim<-c(1:16)

#vintage_2000 average claim


tail(database$delay_group, n=1)
IBNER[tail(database$delay_group, n=1),5]
nrow(IBNER)
ncol(IBNER)
in_year<-c(rep(0,12843))

sum_per_claim<-c(rep(0,12843))
sum_per_vintage<-c(rep(0,16))
no_per_vintage<-c(rep(0,16))
vintage_group=integer(12843)
typeof(sum_per_claim)
sum_per_claim_base<-c(rep(0,12843))

for (i in 1:12843) {
  current_data<-database[database$KárID==i,]
  vintage_group[i]=min(current_data$vintage)
  in_year[i]=min(12,18-ceiling(tail(current_data$Bejelentésdátumperiod,n=1)/12))
  kifiz=as.double(sum(current_data$Kifizetés))
  tart=as.double(sum(current_data$Tartalék))
  sum_per_claim[i]<-kifiz+tart*IBNER[1+tail(current_data$delay_group, n=1),12]/
  IBNER[1+tail(current_data$delay_group, n=1),in_year[i]]
```

```
  #sum_per_claim_base[i]=kifiz+tart
}
#counting the number of elements in a vintage group

IBNER[1+tail(current_data$delay_group, n=1),in_year[6818]]/100

sum_per_claim[6818]
sum_per_claim_base[6818]
max(database$Tartalék)
max(database$Kifizetés)
IBNER
unique(in_year)
unique(1+database$delay_group)
typeof(IBNER[1+tail(current_data$delay_group, n=1),in_year])

for (i in 1:16) {
  no_per_vintage[i]=sum(vintage_group == i)
}

#summarizing claim amounts per vintage
for (i in 1:16) {
  for (j in 1:12843) {
    if (vintage_group[j]==i){ sum_per_vintage[i]=sum_per_vintage[i]+sum_per_claim[j]}
     }
  }

#and the averige claim per year
for (i in 1:16) {
 average_claim[i]=sum_per_vintage[i]/no_per_vintage[i]
}

average_claim

#the actual severity analysis and curve fitting starts here

write.csv(average_claim, file = paste0(output_path,"average_claim_per_acc_year.csv"))
minta=sum_per_claim[sum_per_claim>10]
length(minta)
plotdist(log(minta), histo = TRUE, demp = TRUE)
#plotdist(minta, histo = TRUE, demp = TRUE) not meaningful

fln<-fitdist(minta,"lnorm", method = "mle")
summary(fln)

fe<-fitdist(minta,"exp")
summary(fe)

fg<-fitdist(minta,"gamma", method = "mle", lower = c(0, 0))
summary(fg)

fw<-fitdist(minta,"weibull")
summary(fw)

fll<-fitdist(minta, "llogis", start = list(shape = 1, scale = 50))
summary(fll)
```

```
fp <- fitdist(minta, "pareto", start = list(shape = 2, scale = 500))
summary(fp)

#fb<- fitdist(minta, "burr", start = list(shape1 = 0.6, shape2 = 2))
#summary(fb)

#goodness-of-fit test
gofstat(fw)
gofstat(fe)
gofstat(fln)
gofstat(fg)
gofstat(fll)
gofstat(fp)

#best fitting is gamma with shape 0.4157577515 and rate(=1/scale)  0.0000017187 or I could use ecdf

#monte carlo

#nb parameters mean: 486.7 var-to-mean: 73.73125
#qnbinom(p=0.5,size = 6.691758,mu=486.7) quantiles for negative binomial
total_claim=c(rep(0,10000000))

# to check performance time system.time()
for (i in 1:10000000) {
  number=rnbinom(n=1,size = 6.691758,mu=486.7) #size is the same as r in the r,p parametrization
  total_claim[i]=sum(rgamma(n=number, shape=0.4157577515 , rate=0.0000017187))
}

#és a végs? teljes kárnagyságunkból csinálunk eloszlást
aggregate_loss_distribution=ecdf(total_claim)

#computation of percentiles

percentiles <- function(x){
  i=1
  while (aggregate_loss_distribution(i)<x) {
    if(aggregate_loss_distribution(2*i)<x){ i=2*i}
    if(aggregate_loss_distribution(1.1*i)<x){ i=i*1.1}
    if(aggregate_loss_distribution(1.01*i)<x){ i=i*1.01}
    if(aggregate_loss_distribution(1.001*i)<x){ i=i*1.001}
    if(aggregate_loss_distribution(1.0001*i)<x){ i=i*1.0001}
    i=i+1
  }
                return(i)
}


important_percentiles=matrix(data =
 c(percentiles(0.995),percentiles(0.99),percentiles(0.9),percentiles(0.75),
 percentiles(0.5),percentiles(0.25),0.995,0.99,0.9,0.75,0.5,0.25),
 ncol = 2,dimnames = list(NULL,c("value","percentile")) )

system.time(aggregate_loss_distribution(120010000))

min(rgamma(n=1000, shape=0.403411, rate=0.000001556572))

write.csv(important_percentiles, file = paste0(output_path,"final_result_percentiles.csv"))
```