"Science Stays True Here"
Journal of Mathematics and Statistical Science (ISSN 2411-2518), Vol.6, 123-139 | Science Signpost Publishing

A Chain-Ladder Analysis of P&I Claims

Nicholas G. Berketis*

Keywords: claim, CL method, run-off triangle, claim reserving, P. & I.

According to the latest Annual Review for 2018/19 prepared by FP Marine:

″As we write our report this year, the International Group (I.G.) of P&I Club's are starting to announce their intentions for the 20th February 2019 renewal, with the majority again not applying a general increase.

Following a period of zero general increases, coupled with the impact of the churn effect on rates, the I.G. Clubs have, during the course of 2018, been making noises that rates have now fallen as far as they are able to fall.″.

The insurance industry, unlike other industries, does not sell products as such but promises. An insurance policy is a promise by the insurer to the policyholder to pay for future claims for an upfront received premium.

As a result, Insurers don't know the upfront cost for their service, but rely on historical data analysis and judgement to predict a sustainable price for their offering. In General Insurance (or Non-Life Insurance, e.g. motor, property and casualty insurance) most Policies run for a period of 12 months. However, the claims payment process can take years or even decades. Therefore, often not even the delivery date of their product is known to Insurers.

In particular, losses arising from casualty insurance can take a long time to settle and even when the claims are acknowledged it may take time to establish the extent of the claims' settlement cost. Claims can take years to materialize. A complex and costly example are the claims from asbestos liabilities, particularly those in connection with mesothelioma and lung damage arising from prolonged exposure to asbestos. A research report by a working party of the Institute and Faculty of Actuaries estimated that the un-discounted cost of UK mesothelioma-related claims to the UK Insurance Market for the period 2009 to 2050 could be around £10bn. The cost for asbestos related claims in the US for the worldwide insurance industry was estimated to be around $120bn in 2002.

Thus, it should come as no surprise that the biggest item on the liabilities side of an Insurer's balance sheet is often the provision or reserves for future claims payments. Those reserves can be broken down in case reserves (or outstanding claims), which are losses already reported to the insurance company and losses that are incurred but not reported (IBNR) yet.

The analysis is based on *R* (Version 3.6.2 – 12th December, 2019), an integrated language and environment for statistical computing and graphics. *R* provides a wide variety of statistical and graphical techniques.

---

* Manager of J.Kouroutis & Co. Ltd. Insurance and Reinsurance Brokers, Piraeus, Greece. Visiting Lecturer of Marine Insurance for undergraduates at Frederick University, Cyprus and for postgraduates at the MSc in International Shipping, Finance and Management of the Athens University of Economics and Business, Athens, Greece.

The estimated cost of notified pool claims (in USD 000,000) is as follows:

YEAR	No.OF CLAIMS	12M	24M	36M	48M	60M	72M	84M	96M	108M	120M
2008/09	22	876	1162	1063	1220	1200	1195	1229	1249	1245	1245
2009/10	22	2263	2218	2235	2195	2469	2667	2639	2607	2604	NA
2010/11	14	1791	2411	2669	2525	2506	2590	2599	2540	NA	NA
2011/12	22	2310	2779	2808	2896	2893	2887	2844	NA	NA	NA
2012/13	14	3686	4539	4670	4651	4463	4186	NA	NA	NA	NA
2013/14	17	2798	3270	3640	3649	4116	NA	NA	NA	NA	NA
2014/15	14	1796	1936	2045	2158	NA	NA	NA	NA	NA	NA
2015/16	10	1984	2766	2840	NA	NA	NA	NA	NA	NA	NA
2016/17	6	840	1259	NA	NA	NA	NA	NA	NA	NA	NA
2017/18	7	2272	NA	NA	NA	NA	NA	NA	NA	NA	NA

This triangle shows the known values of loss from each origin year and of annual evaluations thereafter. For example, the known values of loss originating from the 2013/14 exposure period are 2798, 3270, and 3640 as of year ends 2013, 2012, and 2011, respectively. The latest diagonal – i.e., the vector 1245, 2604, . . . 2272 from the upper right to the lower left – shows the most recent evaluation available.

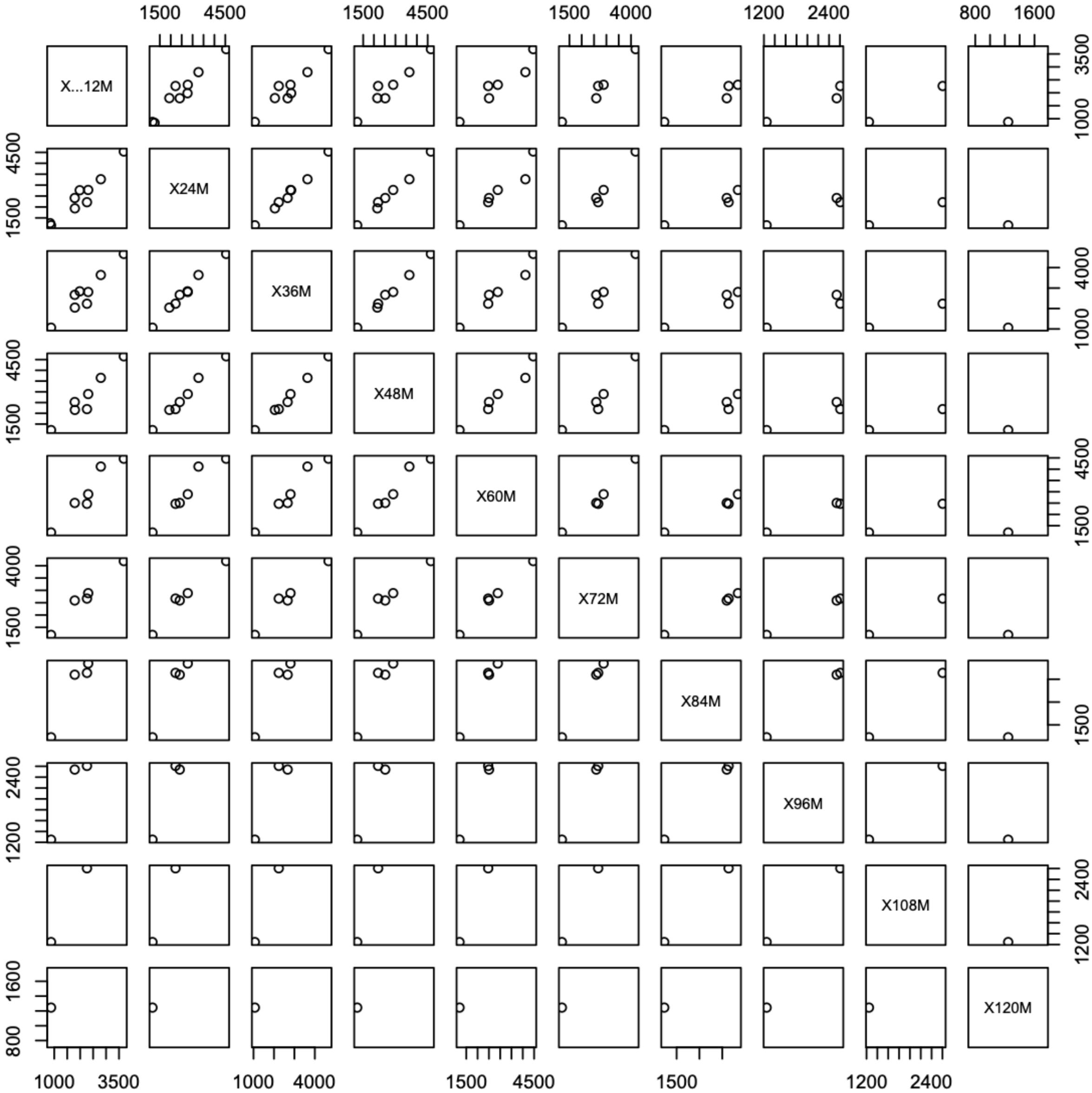
The column headings – 1, 2, . . . , 10 – hold the ages (in years) of the observations in the column relative to the beginning of the exposure period. For example, for the 2014/15 origin year, the age of the 2045 value, evaluated as of 20/02/2018, is three years.

The objective of a reserving exercise is to forecast the future claims development in the bottom right corner of the triangle and potential further developments beyond development age 10. Eventually all claims for a given origin period will be settled, but it is not always obvious to judge how many years or even decades it will take.

We speak of long and short tail business depending on the time it takes to pay all claims.

In order proceed with our analysis, we first plotted the data to get an overview. Figure 1 that follows shows the claims development chart for the past 10 years.

Figure 1


X...12M
X24M
X36M
X48M
X60M
X72M
X84M
X96M
X108M
X120M
1500
4500
4000
1200
2400
800
1600
3500
1000

Chain-ladder methods

The classical chain-ladder is a deterministic algorithm to forecast claims based on historical data. It assumes that the proportional developments of claims from one development period to the next are the same for all origin years.

Basic idea

Most commonly as a first step, the age-to-age link ratios are calculated as the volume weighted average development ratios of a cumulative loss development triangle from one development period to the next C_{ik}, i, k = 1, . . . , n.

$$f_k = \frac{\sum_{i=1}^{n-k} C_{i,k+1}}{\sum_{i=1}^{n-k} C_{i,k}}$$

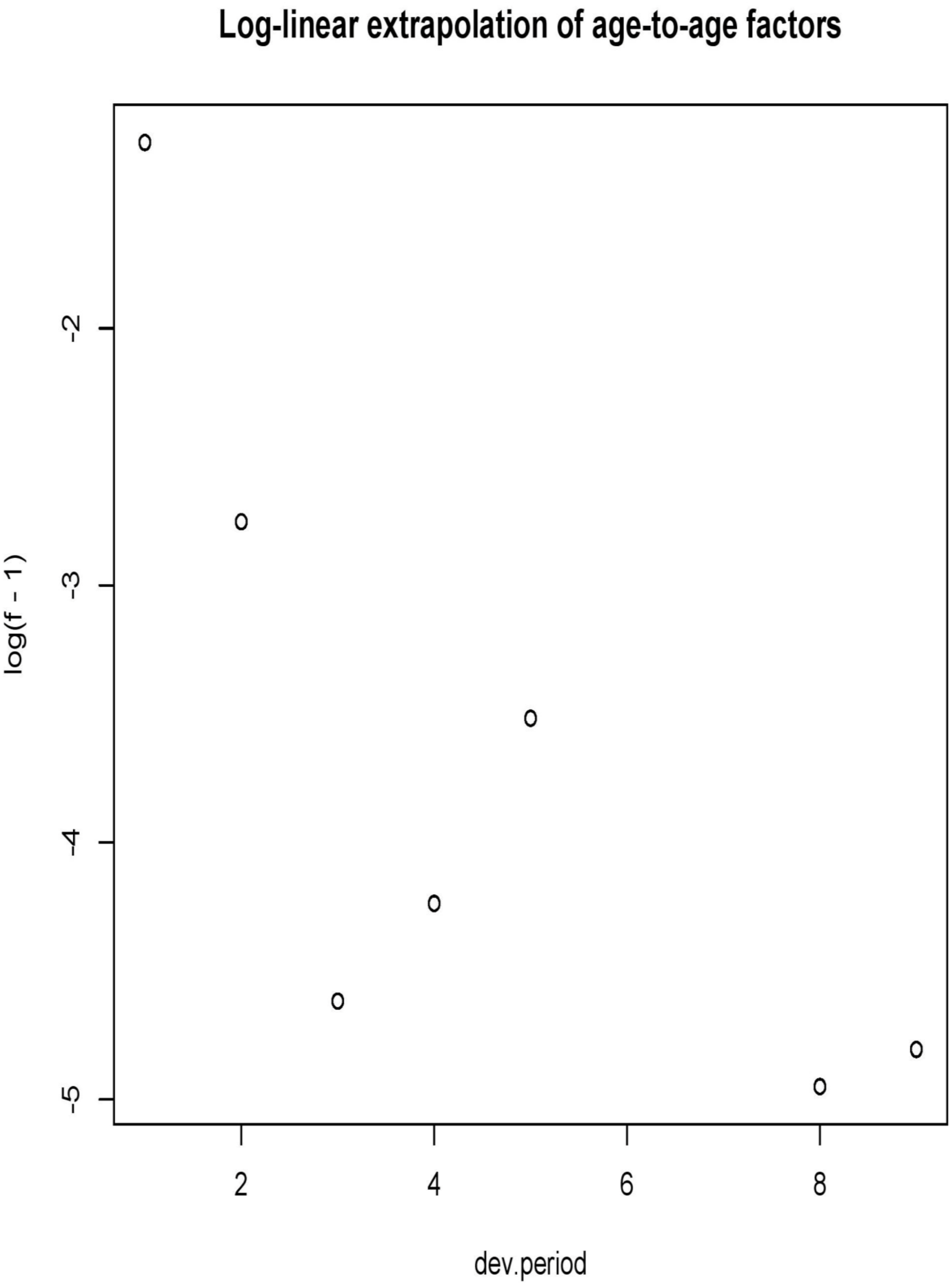
[1] 1.2178369 1.0421707 1.0085729 1.0298203 0.9995566 0.9970018 0.9890212 0.9981846 1.0000000

Often it is not suitable to assume that the oldest origin year is fully developed. A typical approach is to extrapolate the development ratios, e.g. assuming a log-linear model.

[1] 1.0127889

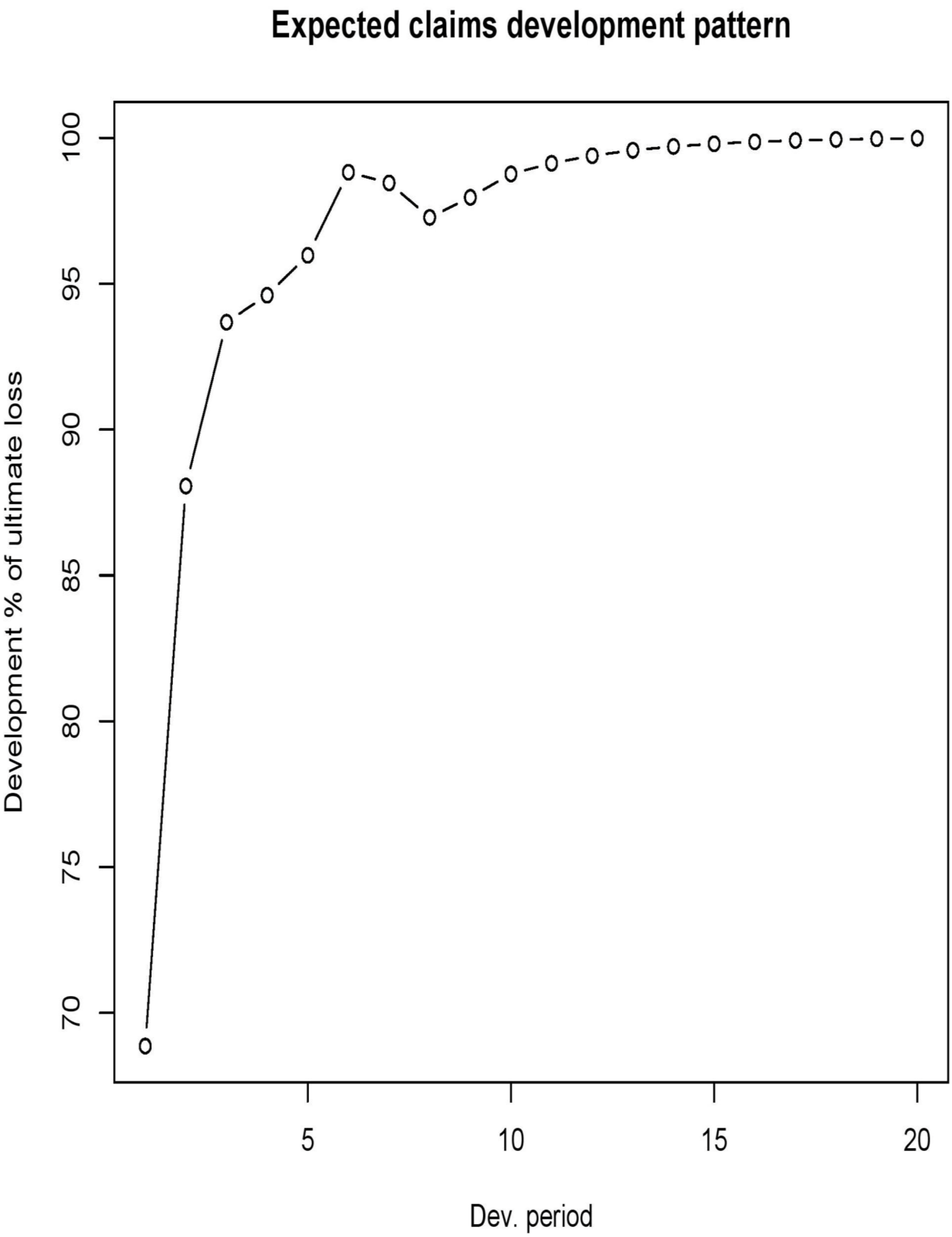
Figure 2 below shows the Log-linear extrapolation of age-to-age factors.

Figure 2



The age-to-age factors allow us to plot the expected claims development patterns. This is shown on Figure 3

Figure 3


Expected claims development pattern
Development % of ultimate loss
70
75
80
85
90
95
100
5
10
15
20
Dev. period

The link ratios are then applied to the latest known cumulative claims amount to forecast the next development period. The squaring of the triangle is calculated below, where an ultimate column is appended to the right to accommodate the expected development beyond the oldest age (10) of the triangle due to the tail factor (1.0127889) being greater than unity.

	X12M	X24M	X36M	X48M	X60M	X72M	X84M	X96M	X108M	X120M	Ult
1	876	1162	1063	1220	1200	1195	1229	1249	1245	1245	1261
2	2263	2218	2235	2195	2469	2667	2639	2607	2604	2604	2637
3	1791	2411	2669	2525	2506	2590	2599	2540	2535	2535	2568
4	2310	2779	2808	2896	2893	2887	2844	2813	2808	2808	2844
5	3686	4539	4670	4651	4463	4186	4173	4128	4120	4120	4173
6	2798	3270	3640	3649	4116	4114	4102	4057	4049	4049	4101
7	1796	1936	2045	2158	2222	2221	2215	2190	2186	2186	2214
8	1984	2766	2840	2864	2950	2948	2940	2907	2902	2902	2939
9	840	1259	1312	1323	1363	1362	1358	1343	1341	1341	1358
10	2272	2767	2884	2908	2995	2994	2985	2952	2947	2947	2984

The total estimated outstanding loss under this method is about 27,000. In particular, it was calculated as 27,079.

This approach is also called Loss Development Factor (LDF) method. More generally, the factors used to square the triangle need not always be drawn from the dollar weighted averages of the triangle. Other sources of factors from which the actuary may select link ratios include simple averages from the triangle, averages weighted toward more recent observations or adjusted for outliers, and benchmark patterns based on related, more credible loss experience. Also, since the ultimate value of claims is simply the product of the most current diagonal and the cumulative product of the link ratios, the completion of interior of the triangle is usually not displayed in favor of that multiplicative calculation.

<u>Mack chain-ladder</u>

Thomas Mack published in 1993[1] a method which estimates the standard errors of the chain-ladder forecast without assuming a distribution under three conditions.

Following the notation of Mack[2] let C_{ik} denote the cumulative loss amounts of origin period (e.g. accident year) $i = 1, \ldots, m$, with losses known for development period (e.g. development year) $k \leq n + 1 - i$.

In order to forecast the amounts C_{ik} for $k > n+1-i$ the Mack chain-ladder-model assumes:

CL1: $E[F_{ik} I C_{i1}, C_{i2}, \ldots, C_{ik}] = f_k$ with $F_{ik} = \frac{C_{i,k+1}}{C_{ik}}$ (2)

[1] Mack, Thomas, (1993), "Distribution-free Calculation of the Standard Error of Chain Ladder Reserve Estimates", ASTIN Bulletin, Vol. 23(2): 213– 225.

[2] Mack, Thomas, (1999), "The Standard Error of Chain Ladder Reserve Estimates: Recursive Calculation and Inclusion of a Tail Factor". ASTIN Bulletin, Vol. 29(2): 361-366.

CL2: $Var(\frac{C_{i,k+1}}{C_{ik}} \mathrm{I} C_{i1}, C_{i2},..., C_{ik}) = \frac{\sigma_k^2}{w_{ik} C_{ik}^{\alpha}}$ (3)

CL3: $\{C_{i1},..., C_{in}\}$, $\{C_{j1},..., C_{jn}\}$, are independent for origin period $i \neq j$ (4)

with $w_{ik} \in [0; 1]$, $\alpha \in \{0, 1, 2\}$. If these assumptions hold, the Mack chain-ladder model gives an unbiased estimator for IBNR (Incurred But Not Reported) claims.

The Mack chain-ladder model can be regarded as a weighted linear regression through the origin for each development period: lm(y ~ x + 0, weights=w/x^(2- alpha)), where y is the vector of claims at development period k + 1 and x is the vector of claims at development period k.

The Mack method is implemented in the ChainLadder package via the function MackChainLadder.

We therefore applied the MackChainLadder function to our triangle:

	Latest	**Dev. To. Date**	**Ultimate**	**IBNR**	**Mack. S.E.**	**CV (IBNR)**
1	1,245	1.000	1,245	0.00	0.00	NaN
2	2,604	1.000	2,604	0.00	0.40	Inf
3	2,540	1.002	2,535	-4.61	3.89	-0.844
4	2,844	1.013	2,808	-36.33	51.14	-1.408
5	4,186	1.016	4,120	-65.86	89.60	-1.360
6	4,116	1.016	4,049	-66.56	233.54	-3.509
7	2,158	0.987	2,186	28.42	258.63	9.102
8	2,840	0.979	2,902	62.06	338.32	5.451
9	1,259	0.939	1,341	81.77	242.76	2.969
10	2,272	0.771	2,947	674.62	513.58	0.761

	Totals
Latest:	26,064.00
Dev:	0.97
Ultimate:	26,737.51
IBNR:	673.51
Mack. S.E.:	888.59
CV(IBNR):	1.32

Executing Mack Chain Ladder will print the following columns of information per accident year (origin period):

1. **Latest**: the claim amount for the last development period
2. **Dev.To.Date**: the development to date or the ratio of the latest over the predicted ultimate
3. **Ultimate**: predicted ultimate claim
4. **IBNR**: the predicted IBNR reserve
5. **Mack.S.E.**: the standard error, or the standard deviation of the bounds for the predicted ultimate and IBNR since the estimate is unbiased (shown in Mack's 1999 paper). In other words, since

the S.E given is equal to one standard deviation, a confidence interval for the true ultimate value can be found using the standard error and the predicted ultimate.

6. **CV(IBNR)**: coefficient of variation, or the ratio of the standard error over the predicted IBNR

The bottom output gives a total or sum of the latest, ultimates, IBNRs. It also gives the standard error of the total ultimate (this is not the total of the standard errors). The development to date factor is the ratio of the total latest against the total ultimate, and the CV(IBNR) is the percentage of the total standard error in the total IBNR.

If the CV (absolute value) is greater than 25%, then another model or a log linear regression should be used.

We can access the loss development factors and the full triangle via:

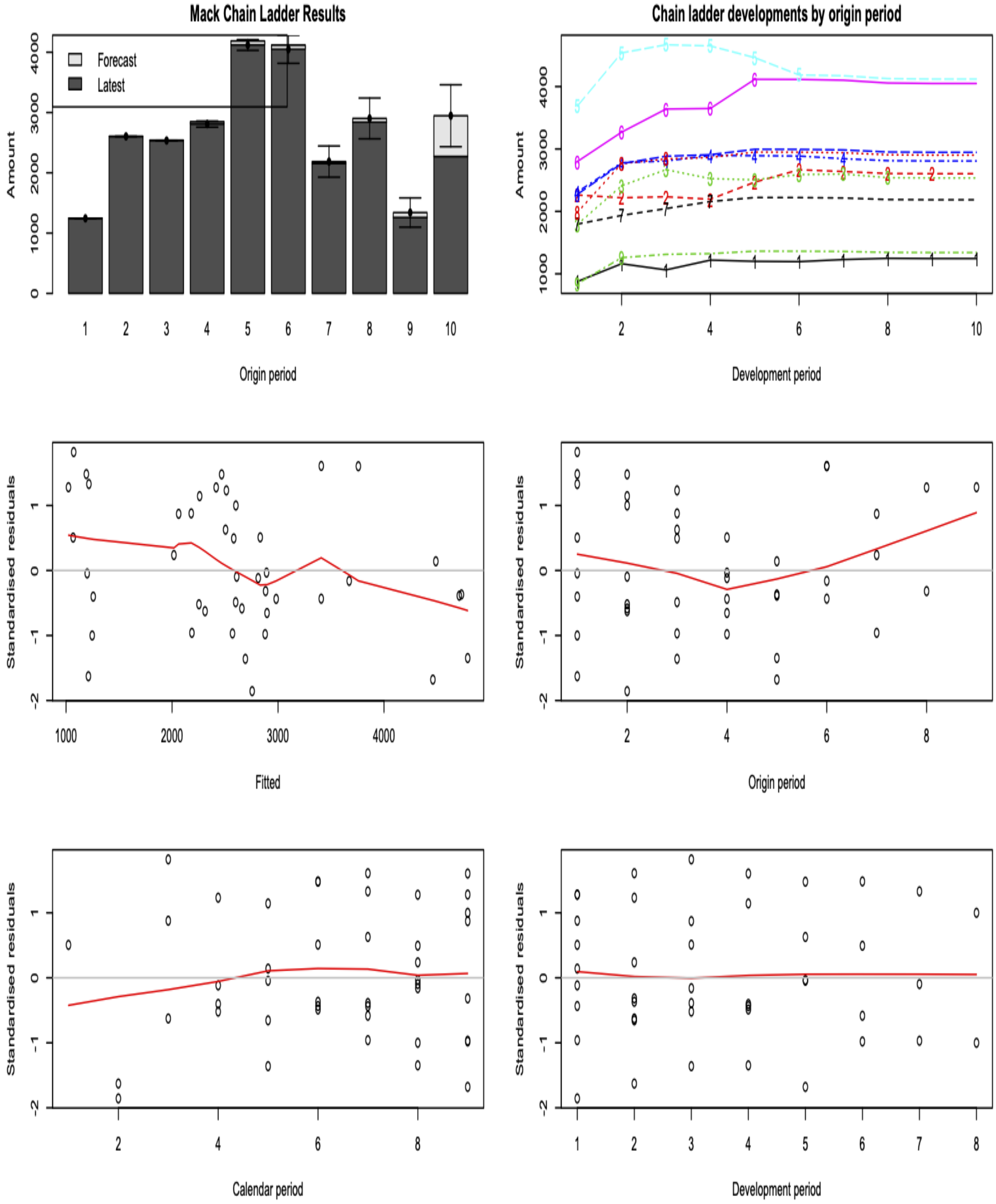
[1] 1.2178369 1.0421707 1.0085729 1.0298203 0.9995566 0.9970018 0.9890212 0.9981846 1.0000000

[10] 1.0000000

origin	X12M	X24M	X36M	X48M	X60M	X72M	X84M	X96M	X108M	X120M
1	876	1162.000	1063.000	1220.000	1200.000	1195.000	1229.000	1249.000	1245.000	1245.000
2	2263	2218.000	2235.000	2195.000	2469.000	2667.000	2639.000	2607.000	2604.000	2604.000
3	1791	2411.000	2669.000	2525.000	2506.000	2590.000	2599.000	2540.000	2535.389	2535.389
4	2310	2779.000	2808.000	2896.000	2893.000	2887.000	2844.000	2812.776	2807.670	2807.670
5	3686	4539.000	4670.000	4651.000	4463.000	4186.000	4173.450	4127.630	4120.137	4120.137
6	2798	3270.000	3640.000	3649.000	4116.000	4114.175	4101.840	4056.806	4049.442	4049.442
7	1796	1936.000	2045.000	2158.000	2222.352	2221.367	2214.707	2190.392	2186.415	2186.415
8	1984	2766.000	2840.000	2864.347	2949.763	2948.455	2939.615	2907.341	2902.063	2902.063
9	840	1259.000	1312.093	1323.341	1362.804	1362.199	1358.115	1343.205	1340.766	1340.766
10	2272	2766.925	2883.609	2908.329	2995.057	2993.729	2984.753	2946.625	2946.625	2946.625

To plot that Mack's assumption are valid review the residual plots, we see no trends in either of them. Please refer to the Figure 4 that follows:

Figure 4


Mack Chain Ladder Results
Forecast
Latest
Amount
Origin period
Chain ladder developments by origin period
Amount
Development period
Standardised residuals
Fitted
Standardised residuals
Origin period
Standardised residuals
Calendar period
Standardised residuals
Development period

Bootstrap chain-ladder

The BootChainLadder function uses a two-stage bootstrapping/simulation approach following the paper by England and Verrall[3]. In the first stage an ordinary chain-ladder method is applied to the cumulative claims' triangle. From this we calculate the scaled Pearson residuals which we bootstrap R times to forecast future incremental claims payments via the standard chain-ladder method. In the second stage we simulate the process error with the bootstrap value as the mean and using the process distribution assumed. The set of reserves obtained in this way forms the predictive distribution, from which summary statistics such as mean, prediction error or quantiles can be derived.

BootChainLadder (Triangle = GGG21, R = 999, process.distr = "gamma")

	Latest Mean	**Ultimate Mean**	**IBNR**	**IBNR S.E.**	**IBNR 75%**	**IBNR 95%**
1	1,245	1,245	0.0	0	0.00e+00	0.0
2	2,604	2,604	0.0	0	0.00e+00	0.0
3	2,540	2,553	12.9	485	4.17e-34	12.9
4	2,844	2,858	14.0	1,157	6.92e-05	706.1
5	4,186	4,164	-21.5	2,025	2.96e-01	1,726.7
6	4,116	4,057	-58.6	1,613	8.59e-01	1,404.4
7	2,158	2,270	112.4	3,334	1.03e+01	1,791.5
8	2,840	3,324	483.7	6,905	1.45e+02	2,727.0
9	1,259	1,597	338.3	2,856	5.65e+01	2,417.0
10	2,272	16,437	14,165.1	418,400	1.37e+03	6,117.8

	Totals
Latest:	26,064
Mean Ultimate:	41,110
Mean IBNR:	15,046
IBNR S.E.	423,875
Total IBNR 75%:	3,071
Total IBNR 95%:	11,028

The BootChainLadder is a model that provides a predicted distribution for the IBNR values for a claims' triangle. However, this model predicts IBNR values by a different method than the previous model. First, the development factors are calculated and then they are used in a backwards recursion to predict values for the past loss triangle. Then the predicted values and the actual values are used to calculate Pearson residuals.

Using the adjusted residuals and the predicted losses from before, the model solves for the actual losses in the Pearson formula and forms a new loss triangle. The steps for predicting past losses and residuals are then repeated for this new triangle. After that, the model uses chain ladder ratios to predict the future losses then calculates the ultimate and IBNR values like in the previous Mack model. This

[3] England, P. D., & Verrall, Richard J., (2002), "Stochastic Claims Reserving in General Insurance", Presented to the Institute of Actuaries, 28 January.

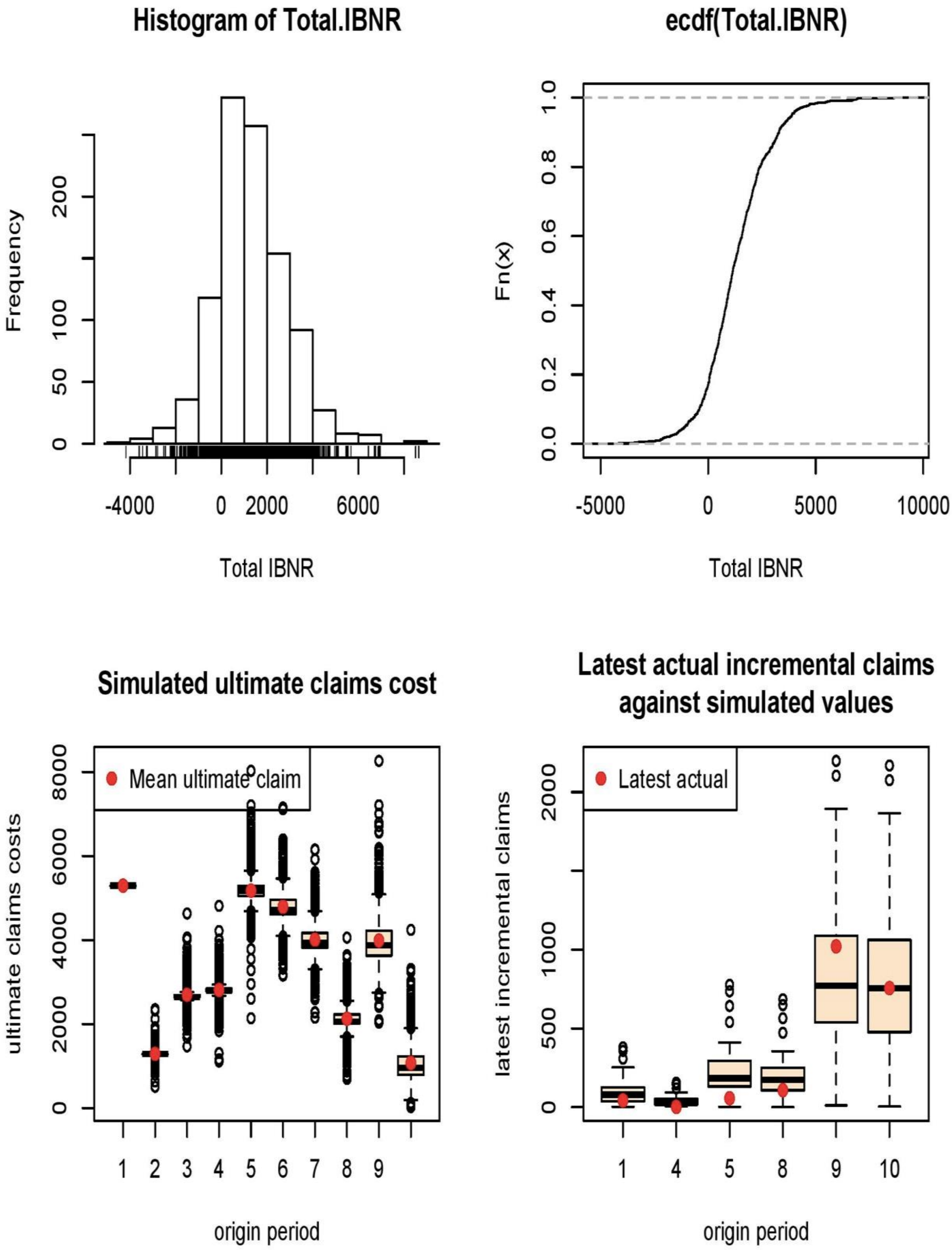
cycle is performed R times, depending on the argument values in the model (default is 999 times). The IBNR for each origin period is calculated from each triangle (the default 999) and used to form a predictive distribution, from which summary statistics are obtained such as mean, prediction error, and quantiles.

The output has some of the same values as the Munich and Mack models did. The Mean and SD IBNR is the average and the standard deviation of the predictive distribution of the IBNRs for each origin year.

The output also gives the 75% and 95% quantiles of the predictive distribution of IBNRs, in other words 95% or 75% of the predicted IBNRs lie at or below the given values.

The above also appear on following Figure 5:

Figure 5


Histogram of Total.IBNR
Frequency
0
50
100
200
-4000
0
2000
6000
Total IBNR
ecdf(Total.IBNR)
Fn(x)
0.0
0.2
0.4
0.6
0.8
1.0
-5000
0
5000
10000
Total IBNR
Simulated ultimate claims cost
Mean ultimate claim
ultimate claims costs
0
2000
4000
6000
8000
1
2
3
4
5
6
7
8
9
origin period
Latest actual incremental claims
against simulated values
Latest actual
latest incremental claims
0
500
1000
2000
1
4
5
8
9
10
origin period

The above Figure 5 shows four graphs, starting with a histogram of the total simulated IBNRs over all origin periods, including a rug plot; a plot of the empirical cumulative distribution of the total IBNRs over all origin periods; a box-whisker plot of simulated ultimate claims costs against origin periods; and a box-whisker plot of simulated incremental claims cost for the latest available calendar period against actual incremental claims of the same period. In the last plot the simulated data should follow the same trend as the actual data, otherwise the original data might have some intrinsic trends which are not reflected in the model.

Quantiles of the bootstrap IBNR can be calculated via the quantile function:

$ByOrigin

	IBNR 75%	**IBNR 95%**	**IBNR 99%**	**IBNR 99.5%**
1	0.000000e+00	0.0000	0.000	0.000
2	0.000000e+00	0.0000	0.000	0.000
3	4.166187e-34	12.8683	1,369.092	2,415.410
4	6.924761e-05	706.1427	3,323.095	4,571.740
5	2.961730e-01	1,726.6523	7,029.690	9,961.165
6	8.591069e-01	1,404.4233	6,222.872	7,847.880
7	1.028578e+01	1,791.4871	5,264.044	6,996.378
8	1.452764e+02	2,726.9522	8,220.398	10,135.458
9	5.645704e+01	2,417.0369	7,422.080	10,228.090
10	1.366388e+03	6,117.8134	12,867.554	18,276.164

	Totals
IBNR 75%:	3,070.501
IBNR 95%:	11,028.228
IBNR 99%:	28,929.597
IBNR 99.5%:	49,596.226

The distribution of the IBNR appears to follow a log-normal distribution, so let’s fit it:

meanlog	sdlog
7.32375643	1.74804729
(0.06835406)	(0.04833362)

Figure 6

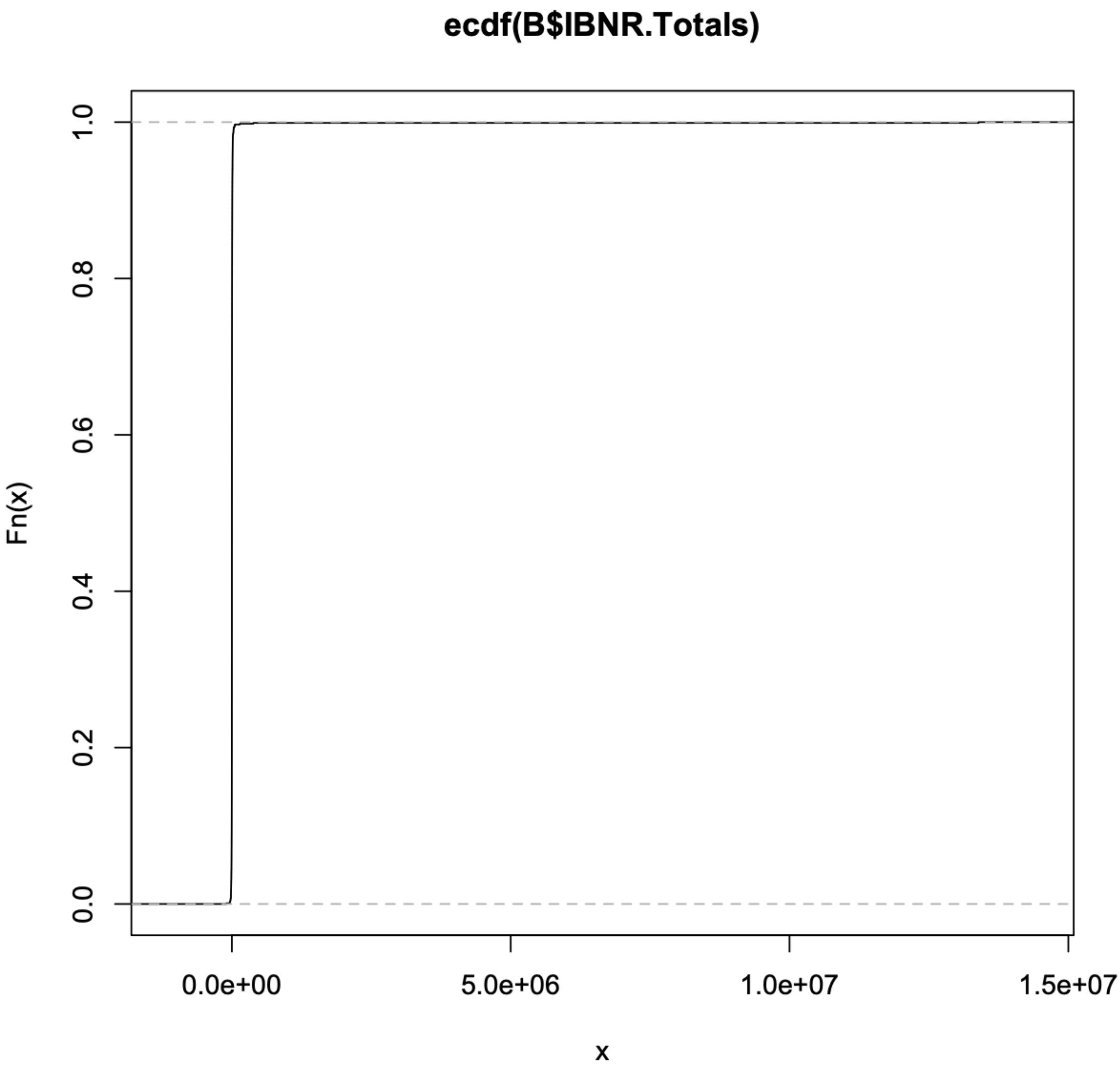
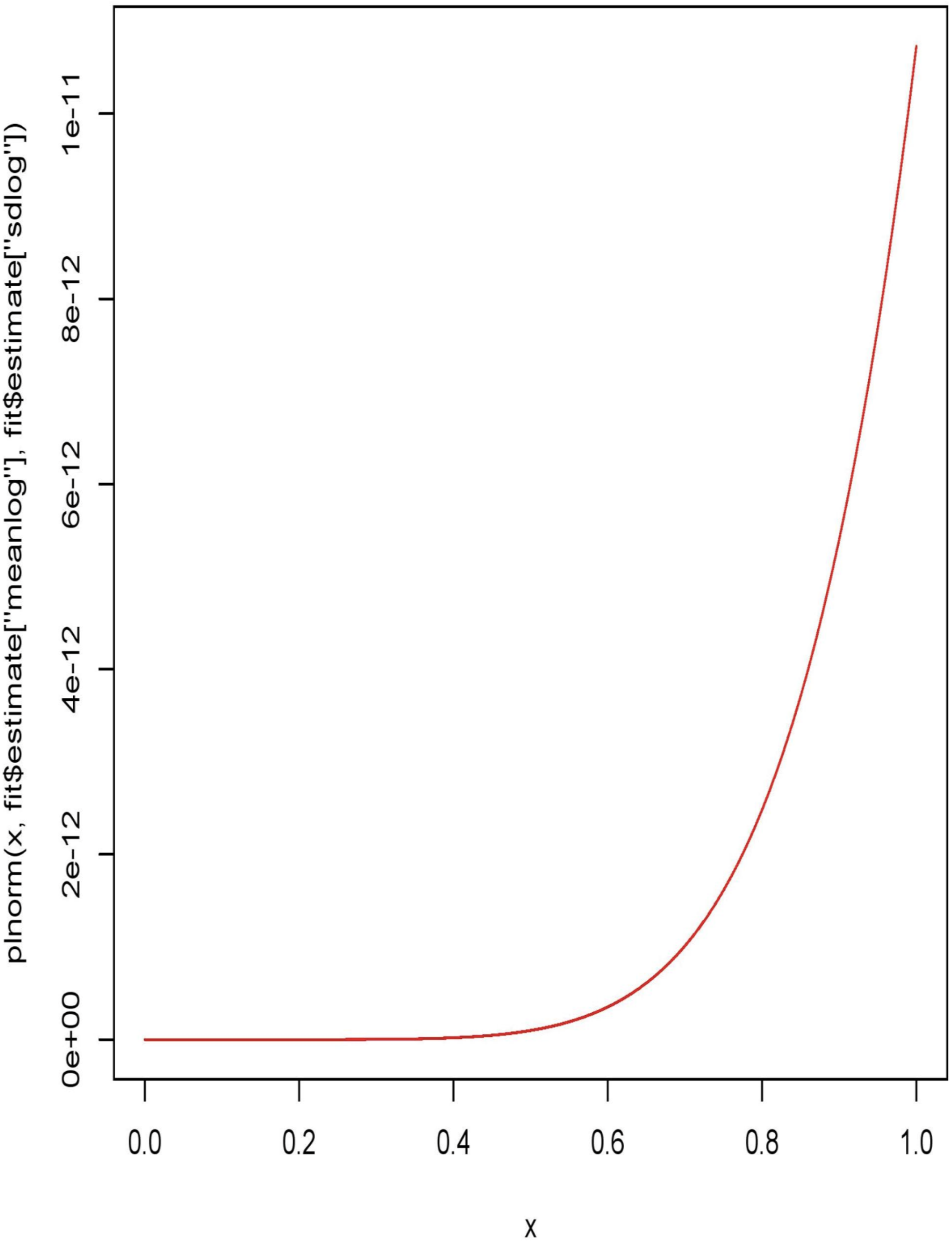

ecdf(B$IBNR.Totals)
Fn(x)
1.0
0.8
0.6
0.4
0.2
0.0
0.0e+00
5.0e+06
1.0e+07
1.5e+07
x

Figure 7


plnorm(x, fit$estimate["meanlog"], fit$estimate["sdlog"])
0e+00
2e-12
4e-12
6e-12
8e-12
1e-11
0.0
0.2
0.4
0.6
0.8
1.0
x

Conclusions

1. The Loss Development Factor (LDF) is above unity, i.e. 1.0127889, which shows an increasingly positive trend for I.B.N.R.'s;
2. The claim amount for the last development period is estimated by both Mack and Bootstrap chain ladder methods at 26,064;
3. The predicted ultimate claim is estimated 27,079 under chain ladder method, Mack chain ladder estimated it at 26,737.51, while Bootstrap chain ladder method showed 41,110;
4. The predicted I.B.N.R. reserve was estimated at 673.51 under the Mack chain ladder method and 15,046 under Bootstrap chain ladder method;
5. Since the coefficient of variation of I.B.N.R.'s was estimated in absolute value well above 25%, i.e. 132%, we followed the Bootstrap chain ladder method, which also justified the increasingly positive trend of I.B.N.R.'s.
6. Hence, the results do not follow the trend for no general increase in the 2019 renewal.

Sources:

1. Carrato, Alessandro, Concina, Fabio, Gesmann, Markus, Murphy, Dan, Wuthrich, Mario & Zhang, Wayne, (2018), "Claims Reserving with R: ChainLadder-0.2.9 Package Vignette", 6, December.
2. England, P. D., & Verrall, Richard J., (2002), "Stochastic Claims Reserving in General Insurance", Presented to the Institute of Actuaries, 28 January.
3. FP Marine Risks, (2019), "P&I Report 2018-2019".
4. International Group of P&I Clubs, (2019), "Annual Review 2018/19", October.
5. Mack, Thomas, (1993), "Distribution-free Calculation of the Standard Error of Chain Ladder Reserve Estimates", ASTIN Bulletin, Vol. 23(2):213– 225.
6. Mack, Thomas, (1999), "The Standard Error of Chain Ladder Reserve Estimates: Recursive Calculation and Inclusion of a Tail Factor". ASTIN Bulletin, Vol. 29(2): 361-366.
7. Merz, Michael and Mario V. W¨uthrich, (2014), "Claims Run-off Uncertainty: The Full Picture", SSRN Manuscript, 2524352.
8. http://opensourcesoftware.casact.org/chain-ladder
9. Weindorfer, Bjorn, (2012), "A Practical Guide to the Use of the Chain-Ladder Method for Determining Technical Provisions for Outstanding Reporting Claims in Non-Life Insurance", Working papers series by University of Applied science BFI Vienna, October, No. 77/2012.