ACSC/STAT 4703, Actuarial Models II

FALL 2023
Toby Kenney

Homework Sheet 7

Model Solutions

Basic Questions

1. *The following table shows the paid losses (in thousands) on claims from one line of business of an insurance company over the past 5 years.*

Accident	*Earned*	*Development year*				
year	*premiums*	*0*	*1*	*2*	*3*	*4*
2018	*6914*	*1771*	*1730*	*1244*	*493*	*592*
2019	*10659*	*2658*	*2651*	*1818*	*861*	
2020	*10698*	*2502*	*2897*	*1917*		
2021	*16863*	*3818*	*4073*			
2022	*18971*	*4488*				

Assume that all payments on claims arising from accidents in 2018 have now been settled. Estimate the future payments arising each year from open claims arising from accidents in each calendar year using

(a) The chain-ladder method

We first compute the cumulative run-off triangle:

Accident	Development year				
year	0	1	2	3	4
2018	1771	3501	4745	5238	5830
2019	2658	5309	7127	7988	
2020	2502	5399	7316		
2021	3818	7891			
2022	4488				

Then we compute the loss development factors:

0/1	$\frac{22100}{10749} = 2.05600520979$
1/2	$\frac{19188}{14209} = 1.35041171089$
2/3	$\frac{13226}{11872} = 1.11404986523$
3/4	$\frac{5830}{5238} = 1.11302023673$

Using these values to complete the table gives the following cumulative losses:

Accident	Development year				
year	0	1	2	3	4
LDF		2.05600520979	1.35041171089	1.11404986523	1.11302023673
2019				8891	
2020			8150	9072	
2021		10656	11871	13213	
2022	9227	12461	13882	15451	

The future payments are the differences between consecutive years:

Accident	Development year					
year	0	1	2	3	4	
2019						903
2020					834	921
2021				2765	1215	1342
2022			4739	3233	1421	1569

(b) The Bornhuetter-Ferguson method with expected loss ratio 0.78.

Using the LDFs from part (a), we get the following:

Development Year	Cumulative proportion of losses paid	Proportion of losses paid
0	0.2904704	0.29047041
1	0.5972087	0.30673827
2	0.8064776	0.20926892
3	0.8984563	0.09197866
4	1.0000000	0.10154374

This gives the following reserves:

Accident	Development year				
year	0	1	2	3	4
2019					844
2020				768	847
2021			2753	1210	1336
2022		4539	3097	1361	1503

(c) The Bühlmann-Straub estimate.

We first recalculate the per-premium loss development factors and $\hat{\gamma}_j$, using the same method as parts (a) and (b):

```
HW7Q1pp<-HW7Q1/HW7Q1ep
HW7Q1cumpp<-t(apply(HW7Q1pp,1,cumsum)) #cumulative run-off
HW7Q1cumcum<-apply(HW7Q1cumpp,2,cumsum)
HW7Q1LDFpp<-HW7Q1cumcum[5+4*seq_len(4)]/HW7Q1cumcum[4*seq_len(4)]
beta<-1/c(rev(cumprod(rev(HW7Q1LDFpp))),1)
gamma<-beta-c(0,beta[-5])
HW7Q1est_ult_pp<-c(1,cumprod(rev(HW7Q1LDFpp)))*rev(HW7Q1cumpp[1+4*seq_len(5)])
```

This gives the following values for γ_j:

j	γ_j
0	0.29208836
1	0.30583625
2	0.20986286
3	0.09066879
4	0.10154374

and for $\widehat{C_{i,J}}$:

i	$\widehat{C_{ij}}$
2018	0.8432167
2019	0.8341125
2020	0.8465917
2021	0.7826197
2022	0.8099316

We use the formulae:

$$\hat{v} = \frac{1}{I}\sum_{i=0}^{I-1}\frac{1}{I-i}\sum_{j=0}^{I-i}\hat{\gamma}_j\left(\frac{X_{ij}}{\hat{\gamma}_j} - \hat{C_{i,J}}\right)^2 \qquad \hat{a} = \frac{\sum_{i=0}^{I}\hat{\beta}_{I-i}\left(\hat{C}_{i,J} - \overline{C}\right)^2 - I\hat{v}}{\sum_{i=0}^{I}\hat{\beta}_{I-i} - \frac{1}{\sum_{i=0}^{I}\hat{\beta}_{I-i}}\sum_{i=0}^{I}\hat{\beta}^2_{I-i}}$$

```
gamma.matrix<-(rep(1,5)%*%t(gamma))

hatv<-mean(rowSums(gamma.matrix*(HW7Q1pp/gamma.matrix-HW7Q1est_ult_pp)^2,na.rm=TRUE)/(4:0

Cbar<-sum(HW7Q1cumpp[1+4*seq_len(5)])/sum(beta)

hata<-(sum(rev(beta)*(HW7Q1est_ult_pp-Cbar)^2)-5*hatv)/(sum(beta)-sum(beta^2)/sum(beta))

Z<-rev(beta)/(rev(beta)+hatv/hata)

muhat<-sum(Z*HW7Q1est_ult_pp)/sum(Z)

Chat_BS<-Z*HW7Q1est_ult_pp+(1-Z)*muhat

Chat_BS%*%t(gamma)

Est_Future<-HW7Q1ep*Chat_BS%*%t(gamma)

Est_Future[rep(seq_len(5),5)+rep(seq_len(5),each=5)<7]<-NA

sum(Est_Future,na.rm=TRUE)
```

This gives $\overline{C} = 0.8289218$, $\hat{v} = 0.0002504783$, $\hat{a} = 0.0002207077$ and the following per-premium payments:

Accident year	Development year 0	1	2	3	4
2019					0.08429815
2020				0.07572391	0.08480635
2021			0.1703485	0.07359705	0.08242440
2022		0.2518681	0.1728303	0.07466930	0.08362525

which correspond to the following reserves:

Accident year	Development year 0	1	2	3	4
2019					898.5339
2020				810.0944	907.2583
2021			2872.586	1241.0671	1389.9226
2022		4778.190	3278.763	1416.5513	1586.4546

This gives the total outstanding claims reserves as $19,179.42.

2. *The file* `HW7_data.txt` *contains a run-off triangle. An actuary is planning to use the chain-ladder method to estimate future reserves. Test whether losses in different years are correlated, and whether there are any calendar year effects.*

   We first calculate correlation between vectors of development factors in each year.

```
run.off<-read.table("../HW7_data.txt")
cum.run.off<-t(apply(run.off,1,cumsum))
dev.fact<-cum.run.off[,-1]/cum.run.off[,-10]

### pairwise correlation matrix
cor(dev.fact,use="pairwise.complete")

### extract correlations between adjacent development years
r<-cor(dev.fact,use="pairwise.complete")[2+10*(seq_len(10)-1)]

### last three years don't have enough data.
### calculate t statistic

tstat<-r*sqrt((6:(-3))/(1-r^2))

### calculate p-values
pt(-abs(tstat),df=6:(-3))*2 # multiply by 2 for 2-sided test.
```

This suggests that none of the development factors are correlated.

To check for Calendar Year effects, we rank the development factor estimates.

```
rank.df<-apply(dev.fact,2,rank)/rep(9:1,each=10)
rank.df[rank.df>1]<-NA

### take values in calendar year.
### matrices in R are listed by columns, so we can achieve this by
### changing the dimensions
cy.dy<-matrix(rank.df,9,10)

###Count number of times above median, and number of times not equal to median
over.median<-rowSums(cy.dy>0.5,na.rm=TRUE)
not.median<-rowSums(cy.dy!=0.5,na.rm=TRUE)
cbind(over.median,not.median)

###Calculate p-values
pmin(pbinom(over.median-0.5,not.median,0.5,lower.tail=FALSE)*2,
     pbinom(over.median,not.median,0.5)*2)
###
rowMeans(matrix(rank.df,9,10),na.rm=TRUE)

### standardise columns of rank
rank.sd<-sqrt(colMeans((matrix(rank.df,9,10)-0.5)^2,na.rm=TRUE))
((matrix(rank.df,9,10)-0.5)/(rep(1,9)%*%t(rank.sd)))

rowMeans(((matrix(rank.df,9,10)-0.5)/(rep(1,9)%*%t(rank.sd))),na.rm=TRUE)

### Standardise row means
rowMeans((matrix(rank.df,9,10)-0.5)/(rep(1,9)%*%t(rank.sd)),na.rm=TRUE)*sqrt(seq_len(9))
```

Mack's test, gives the following numbers of development factors above the median.

Calendar Year	Number of observations above median	Number of observations not equal to median	p-value
2	1	1	1.000000
3	0	1	1.000000
4	2	3	1.000000
5	2	4	1.000000
6	3	5	1.000000
7	3	6	1.000000
8	5	7	0.453125
9	4	7	1.000000
10	5	7	0.453125

As we can see, none of the calendar years are significant.

The standardised ranks are also not significant.

3. *For the run-off table in Question 1, use Mack's model to estimate the MSE of the estimated outstanding losses.*

We first compute the estimates $\hat{\sigma}_j^2$.

```
beta<-1/c(rev(cumprod(rev(HW7Q1.LDF))),1)
gamma<-beta-c(0,beta[-5])

sigmahat<-rowMeans(HW7Q1.cum[,-5]*(fij-rep(1,5)%*%t(HW7Q1.LDF))^2,na.rm=TRUE)*
    (5-seq_len(5))/(4-seq_len(5))

### Use Mack's suggestion
sigmahat[4]<-min(sigmahat[2],sigmahat[3],sigmahat[3]^2/sigmahat[2])
```

j	$\hat{\sigma}_j^2$
0	3.890008
1	4.901336
2	26.080671
3	4.901336

Next we estimate process variance:

```
cumsum(rev(sigmahat[-5]/beta[-5]/HW7Q1.LDF^2))*HW7Q1.Est.Ult[-1]
```

This gives the following values.

j	$\widehat{\mathrm{Var}}(C_{i,J}\|C_{i,J-i})$
1	39151.87
2	276321.18
3	461939.90
4	589119.86

We then compute the variance of our estimates:

```
S<-colSums(HW7Q1.cum,na.rm=TRUE)-HW7Q1.cum[1+4*seq_len(5)]

HW7Q1.Est.Ult[-1]^2*cumsum(rev(sigmahat[-5]/(HW7Q1.LDF^2*S[-5])))
```

This gives the following values.

j	$\mathbb{E}\left(\widehat{C}_{i,J}-\mathbb{E}\left(\widehat{C}_{i,J}\middle\|D\right)\right)$
1	59706.98
2	207821.95
3	473924.46
4	668473.91

Finally, we compute the covariance

```
CovarianceMSE<-(HW7Q1.Est.Ult[-1]%*%t(HW7Q1.Est.Ult[-1]))*
     pmax(cumsum(rev(sigmahat[-5]/(HW7Q1.LDF^2*S[-5])))%*%t(rep(1,4)),
           rep(1,4)%*%t(cumsum(rev(sigmahat[-5]/(HW7Q1.LDF^2*S[-5])))))

### Total estimation error.
sum(CovarianceMSE)+sum(HW7Q1.Est.Ult[-1]^2*cumsum(rev(sigmahat[-5]/(HW7Q1.LDF^2*S[-5]))))
```

This gives the total

7213348

4. *The files* `HW7Q4_reported.txt`, `HW7Q4_settled.txt` *and* `HW7Q4_aggregate.txt` *give numbers of claims reported and settled, and aggregate claim amounts for each accident year and development year. By using the chain-ladder method to project number of settled claims, propotions of settled claims and average aggregate losses per claim, estimate the reserves needed.*

We first calculate the LDFs and project the number of reported claims:

```
Run.Off.Reported<-read.table("HW7Q4_reported.txt")

Cum.Reported<-t(apply(Run.Off.Reported,1,cumsum))
Cum.Cum.Reported<-apply(Cum.Reported,2,cumsum)
Reported.DF<-(Cum.Cum.Reported[,-1]/Cum.Cum.Reported[,-10])[9*seq_len(9)]
Latest.Cum.Reported<-Cum.Reported[rev(1+9*seq_len(10))]

Est.Ultimate.Reported<-c(1,cumprod(rev(Reported.DF)))*Latest.Cum.Reported
Est.Cum.Reported<-Est.Ultimate.Reported%*%t(rev(1/c(1,cumprod(rev(Reported.DF)))))
```

We then calculate LDFs for settled claims, and use them to project the number of settled claims.

```
Run.Off.Settled<-read.table("HW7Q4_settled.txt")
Cum.Settled<-t(apply(Run.Off.Settled,1,cumsum))
Cum.Cum.Settled<-apply(Cum.Settled,2,cumsum)
Settled.DF<-(Cum.Cum.Settled[,-1]/Cum.Cum.Settled[,-10])[9*seq_len(9)]

beta<-c(rev(cumprod(rev(1/Settled.DF))),1)
gamma<-beta-c(0,beta[-10])

Projected.Settled<-Est.Ultimate.Reported%*%t(gamma)

### Cover past values
Projected.Settled[seq_len(10)%*%t(rep(1,10))+rep(1,10)%*%t(seq_len(10))<12]<-NA

### Merge Observed and Projected
Observed.Projected.Settled<-
    matrix(rowMeans(cbind(as.vector(as.matrix(Run.Off.Settled)),
                          as.vector(as.matrix(Projected.Settled))),
                    na.rm=TRUE),10,10)
```

Finally, we calculate the average claim severity for each development year.

```
Aggregate.Payments<-read.table("HW7Q4_aggregate.txt")


Ave.Settlement.Ammount<-colMeans(Aggregate.Payments/Run.Off.Settled,na.rm=TRUE)
Exp.Agg.Payments<-Projected.Settled*(rep(1,10)%*%t(Ave.Settlement.Ammount))

### Estimate total reserves
sum(Exp.Agg.Payments,na.rm=TRUE)
```

This gives the total outstanding claims reserve as 174927.9.

Standard Questions

5. *An actuary is using a Poisson model to analyse the run-off triangle from Question 1.*

   (a) *Show that the following aggregate losses are within a 0.05 likelihood interval under the Poisson model. (That is, show that the likelihood for these parameter values is at least 0.05 times the maximum likelihood for the data.)*

i	μ_i
2018	*5830*
2019	*9026*
2020	*9119*
2021	*13540*
2022	*15885*

The log-likelihood is

$$l(\mu, \gamma) = \sum_{i+j \leqslant I} X_{ij} \log(\mu_i \gamma_j) - \mu_i \gamma_j$$

Setting the derivative with respect to γ_j to a constant for all γ_j, we get

$$\frac{\sum_{i \leqslant I-j} X_{ij}}{\gamma_j} = \sum_{i \leqslant I-j} \mu_i + C$$

for all j. This gives us

$$\gamma_j = \frac{\sum_{i \leqslant I-j} X_{ij}}{\sum_{i \leqslant I-j} \mu_i + C}$$

Substituting the values gives

$$\gamma_0 = \frac{15237}{53400 + C}$$
$$\gamma_1 = \frac{11351}{37515 + C}$$
$$\gamma_2 = \frac{4979}{23975 + C}$$
$$\gamma_3 = \frac{1354}{14856 + C}$$
$$\gamma_4 = \frac{592}{5830 + C}$$

We see that numerically, to achieve $\gamma_0 + \gamma_1 + \gamma_2 + \gamma_3 + \gamma_4 = 1$, we must have $C = -251.272$, which gives the following values of γ_i:

i	γ_i
0	0.2866861
1	0.3046126
2	0.2098743
3	0.0927097
4	0.1061174

This gives the following Poisson means for the observed data:

Accident	Earned	Development year				
year	premiums	0	1	2	3	4
2018	1671.380	1775.891	1223.567	540.4976	618.6643	
2019	2587.628	2749.433	1894.325	836.7978		
2020	2614.290	2777.762	1913.843			
2021	3881.729	4124.454				
2022	4554.008					

This gives the log likelihood:

$$l(\mu, \gamma) = 229880.0$$

For the MLE estimates obtained using the chain ladder method, the log-likelihood is

$$l(\mu, \gamma) = 229882.9$$

Thus, the log-likelihood ratio between the given values of μ_i and the MLEs is $229882.9 - 229880.0 = 2.9$, so the relative likelihood is $e^{-2.9} = 0.0550232200564$ so the MLEs given are in a 5% relative maximum likelihood interval.

(b) *For the values in the above table, what is the probability that outstanding claims exceed $20,000,000?*

Under the Poisson model, the outstanding claims follow a Poisson distribution, with mean

$$\sum_{i=1}^{I} \mu_i - \sum_{i+j \leqslant I} X_{ij} = 53400 - 33513 = 19887$$

We can approximate this by a normal distribution with mean 19887 and variance 19887. The probability of this exceeding 20000 is

$$1 - \Phi\left(\frac{20000 - 19887}{\sqrt{19887}}\right) = 1 - \Phi\left(0.801297534767\right) = 0.2114797$$