Chapter 8
Credibility theory

Credibility, as developed by American actuaries, has provided us with a very powerful and intuitive formula. The European Continental School has contributed to its interpretation. The concepts are fundamental to insurance and will continue to be most important in the future. I find it deplorable that the world of finance has not yet realized the importance of collateral knowledge far beyond insurance and the power of credibility-type formulae — Hans Bühlmann, 1999

8.1 Introduction

In insurance practice it often occurs that one has to set a premium for a group of insurance contracts for which there is some claim experience regarding the group itself, but a lot more on a larger group of contracts that are more or less related. The problem is then to set up an experience rating system to determine next year's premium, taking into account not only the individual experience with the group, but also the collective experience. Two extreme positions can be taken. One is to charge the same premium to everyone, estimated by the overall mean $\overline{X}$ of the data. This makes sense if the portfolio is homogeneous, which means that all risk cells have identical mean claims. But if this is not the case, the 'good' risks will take their business elsewhere, leaving the insurer with only 'bad' risks. The other extreme is to charge to group j its own average claims $\overline{X}_j$ as a premium. Such premiums are justified if the portfolio is heterogeneous, but they can only be applied if the claims experience with each group is large enough. As a compromise, one may ask a premium that is a weighted average of these two extremes:

$$z_j\overline{X}_j + (1 - z_j)\overline{X}. \tag{8.1}$$

The factor z_j that expresses how 'credible' the individual experience of cell j is, is called the *credibility factor*; a premium such as (8.1) is called a *credibility premium*. Charging a premium based on collective as well as individual experience is justified because the portfolio is in general neither completely homogeneous, nor completely heterogeneous. The risks in group j have characteristics in common with the risks in other groups, but they also possess unique group properties.

One would choose z_j close to one under the following circumstances: the risk experience with cell j is vast, it exhibits only little variation, or the variation between groups is substantial. There are two methods to find a meaningful value for z_j. In *limited fluctuation credibility theory*, a cell is given full credibility $z_j = 1$ if the experience with it is large enough. This means that the probability of having at least a certain relative error in the individual mean does not exceed a given threshold. If the experience falls short of full credibility, the credibility factor is taken as the ratio of the experience actually present and the experience needed for full credibility. More interesting is the *greatest accuracy credibility theory*, where the credibility factors are derived as optimal coefficients in a Bayesian model with variance components. This model was developed in the 1960's by Bühlmann.

Note that apart from claim amounts, the data can also concern *loss ratios*, that is claims divided by premiums, or claims as a percentage of the sum insured, and so on. Quite often, the claims experience in a cell relates to just one contract, observed in a number of periods, but it is also possible that a cell contains various 'identical' contracts.

In practice, one should use credibility premiums only if one only has very few data. If one has additional information in the form of collateral variables, for example, probably using a generalized linear model (GLM) such as described in the following chapter is indicated, or a mixed model. The main problem is to determine how much virtual experience, see Remark 8.2.7 and Exercise 8.4.7, one should incorporate.

In Section 8.2 we present a basic model to illustrate the ideas behind credibility theory. In this model the claims total X_{jt} for contract j in period t is decomposed into three separate components. The first component is the overall mean m. The second a deviation from this mean that is specific for this contract. The third is a deviation for the specific time period. By taking these deviations to be independent random variables, we see that there is a covariance structure between the claim amounts, and under this structure we can derive estimators of the components that minimize a certain sum of squares. In Section 8.3 we show that exactly these covariance structures, and hence the same optimal estimators, also arise in more general models. Furthermore, we give a short review of possible generalizations of the basic model. In Section 8.4, we investigate the Bühlmann-Straub model, in which the observations are measured in different precision. In Section 8.5 we give an application from motor insurance, where the numbers of claims are Poisson random variables with as a parameter the outcome of a structure parameter that is assumed to follow a gamma distribution.

8.2 The balanced Bühlmann model

To clarify the ideas behind credibility theory, we study in this section a stylized credibility model. Consider the random variable X_{jt}, representing the claim figure of cell j, $j = 1,2,\dots,J$, in year t. For simplicity, we assume that the cell contains a single contract only, and that every cell has been observed during T observation periods. So for each j, the index t has the values $t = 1,2,\dots,T$. Assume that this claim statistic is the sum of a cell mean m_j plus 'white noise', that is, that all X_{jt} are independent and $\mathrm{N}(m_j, s^2)$ distributed, with possibly unequal mean m_j for each cell, but with the same variance $s^2 > 0$. We can test for equality of all group means using the familiar statistical technique of *analysis of variance* (ANOVA). If the null-hypothesis that all m_j are equal fails to hold, this means that there will be more variation between the cell averages $\overline{X}_j$ around the overall average $\overline{X}$ than can be expected in view of the observed variation within the cells. For this reason we look at the following random variable, called the *sum-of-squares-between*:

$$SSB = \sum_{j=1}^{J} T(\overline{X}_j - \overline{X})^2. \tag{8.2}$$

One may show that, under the null-hypothesis that all group means m_j are equal, the random variable *SSB* has mean $(J-1)s^2$. Since s^2 is unknown, we must estimate this parameter separately. This estimate is derived from the *sum-of-squares-within*, defined as

$$SSW = \sum_{j=1}^{J} \sum_{t=1}^{T} (X_{jt} - \overline{X}_j)^2. \tag{8.3}$$

It is easy to show that the random variable *SSW* has mean $J(T-1)s^2$. Dividing *SSB* by $J-1$ and *SSW* by $J(T-1)$ we get two random variables, each with mean s^2, called the *mean-square-between* (*MSB*) and the *mean-square-within* (*MSW*) respectively. We can perform an F-test now, where large values of the *MSB* compared to the *MSW* indicate that the null-hypothesis that all group means are equal should be rejected. The test statistic to be used is the so-called *variance ratio* or F-ratio:

$$F = \frac{MSB}{MSW} = \frac{\frac{1}{J-1}\sum_j T(\overline{X}_j - \overline{X})^2}{\frac{1}{J(T-1)}\sum_j \sum_t (X_{jt} - \overline{X}_j)^2}. \tag{8.4}$$

Under the null-hypothesis, *SSB* divided by s^2 has a $\chi^2(J-1)$ distribution, while *SSW* divided by s^2 has a $\chi^2(J(T-1))$ distribution. Furthermore, it is possible to show that these random variables are independent. Therefore, the ratio F has an $F(J-1, J(T-1))$ distribution. Proofs of these statements can be found in many texts on mathematical statistics, under the heading 'one-way analysis of variance'. The critical values of F can be found in an F-table (Fisher distribution).

Example 8.2.1 (A heterogeneous portfolio)
Suppose that we have the following observations for 3 groups and 5 years:

	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$\overline{X}_j$
$j=1$	99.3	93.7	103.9	92.5	110.6	100.0
$j=2$	112.5	108.3	118.0	99.4	111.8	110.0
$j=3$	129.2	140.9	108.3	105.0	116.6	120.0

As the reader may verify, the *MSB* equals 500 with 2 degrees of freedom, while the *MSW* is 109 with 12 degrees of freedom. This gives a value $F = 4.588$, which is significant at the 95% level, the critical value being 3.885. The conclusion is that the data show that the mean claims per group are not all equal.

To get R to do the necessary calculations, do the following:

```
J <- 3; K <- 5; X <- scan(n=J*K)
 99.3  93.7 103.9  92.5 110.6
112.5 108.3 118.0  99.4 111.8
129.2 140.9 108.3 105.0 116.6
j <- rep(1:J, each=K); j <- as.factor(j)
X.bar <- mean(X); Xj.bar <- tapply(X, j, mean)
MSB <- sum((Xj.bar-X.bar)^2) * K / (J-1)
MSW <- sum((X-rep(Xj.bar,each=K))^2)/J/(K-1)
MSB/MSW; qf(0.95, J-1, J*(K-1))  ## 4.588 and 3.885
```

The use of `K` instead of `T` to denote time avoids problems with the special identifiers `T` and `t` in R. The vector `Xj.bar` is constructed by applying the `mean` function to all groups of elements of `X` with the same value of `j`.

It is also possible to let R do the analysis of variance. Use a linear model, explaining the responses `X` from the group number `j` (as a factor). This results in:

```
> anova(lm(X~j))
Analysis of Variance Table

Response: X
          Df  Sum Sq Mean Sq F value  Pr(>F)
j          2 1000.00  500.00  4.5884 0.03311 *
Residuals 12 1307.64  108.97
```

The probability of obtaining a larger F-value than the one we observed here is 0.03311, so the null-hypothesis that the group means are all equal is rejected at the 5% level. ∇

If the null-hypothesis fails to be rejected, there is apparently no convincing statistical evidence that the portfolio is heterogeneous. So there is no reason not to ask the same premium for each contract. In case of rejection, apparently there is variation between the cell means m_j. In this case one may treat these numbers as fixed unknown numbers, and try to find a system behind these numbers, for example by doing a regression on collateral data. Another approach is to assume that the numbers m_j have been produced by a chance mechanism, hence by 'white noise' similar to the one responsible for the deviations from the mean within each cell. This means that we can decompose the claim statistics as follows:

$$X_{jt} = m + \Xi_j + \Xi_{jt}, \quad j = 1,\dots,J,\ t = 1,\dots,T, \tag{8.5}$$

with Ξ_j and Ξ_{jt} independent random variables for which

$$\mathrm{E}[\Xi_j] = \mathrm{E}[\Xi_{jt}] = 0, \quad \mathrm{Var}[\Xi_j] = a, \quad \mathrm{Var}[\Xi_{jt}] = s^2. \tag{8.6}$$

Because the variance of X_{jt} in (8.5) equals the sum of the variances of its components, models such as (8.5) are called *variance components models*. Model (8.5) is a simplified form of the so-called classical Bühlmann model, because we assumed independence of the components where Bühlmann only assumes the correlation to be zero. We call our model that has equal variance for all observations, as well as equal numbers of policies in all cells, the *balanced Bühlmann model*.

The interpretation of the separate components in (8.5) is the following.

1. m is the overall mean; it is the expected value of the claim amount for an arbitrary policyholder in the portfolio.
2. Ξ_j denotes a random deviation from this mean, specific for contract j. The conditional mean, given $\Xi_j = \xi$, of the random variables X_{jt} equals $m + \xi$. It represents the long-term average of the claims each year if the length of the observation period T goes to infinity. The component Ξ_j describes the risk quality of this particular contract; the mean $\mathrm{E}[\Xi_j]$ equals zero, its variation describes differences between contracts. The distribution of Ξ_j depicts the risk structure of the portfolio, hence it is known as the *structure distribution*. The parameters m, a and s^2 characterizing the risk structure are called the *structural parameters*.
3. The components Ξ_{jt} denote the deviation for year t from the long-term average. They describe the within-variation of a contract. It is the variation of the claim experience in time through good and bad luck of the policyholder.

Note that in the model described above, the random variables X_{jt} are *dependent* for fixed j, since they share a common risk quality component Ξ_j. One might say that stochastically independent random variables with the same probability distribution involving unknown parameters in a sense are dependent anyway, since their values all depend on these same unknown parameters.

In the next theorem, we are looking for a predictor of the as yet unobserved random variable $X_{j,T+1}$. We require this predictor to be a linear combination of the observable data $X_{11},\dots,X_{JT}$ with the same mean as $X_{j,T+1}$. Furthermore, its mean squared error must be minimal. We prove that under model (8.5), this predictor has the credibility form (8.1), so it is a weighted average of the individual claims experience and the overall mean claim. The theorem also provides us with the optimal value of the credibility factor z_j. We want to know the optimal predictor of the amount to be paid out in the next period $T+1$, since that is the premium we should ask for this contract. The distributional assumptions are assumed to hold for all periods $t = 1,\dots,T+1$. Note that in the theorem below, normality is not required.

Theorem 8.2.2 (Balanced Bühlmann model; homogeneous estimator)
Assume that the claim figures X_{jt} for contract j in period t can be written as the sum of stochastically independent components, as follows:

$$X_{jt} = m + \Xi_j + \Xi_{jt}, \quad j = 1,\dots,J,\ t = 1,\dots,T+1, \tag{8.7}$$

where the random variables Ξ_j are iid with mean $\mathrm{E}[\Xi_j] = 0$ and $\mathrm{Var}[\Xi_j] = a$, and also the random variables Ξ_{jt} are iid with mean $\mathrm{E}[\Xi_{jt}] = 0$ and $\mathrm{Var}[\Xi_{jt}] = s^2$ for all j and t. Furthermore, assume the random variables Ξ_j to be independent of the Ξ_{jt}. Under these conditions, the homogeneous linear combination $g_{11}X_{11} + \cdots + g_{JT}X_{JT}$ that is the best unbiased predictor of $X_{j,T+1}$ in the sense of minimal mean squared error (MSE)

$$\mathrm{E}[\{X_{j,T+1} - g_{11}X_{11} - \cdots - g_{JT}X_{JT}\}^2] \tag{8.8}$$

equals the credibility premium

$$z\overline{X}_j + (1-z)\overline{X}, \tag{8.9}$$

where

$$z = \frac{aT}{aT+s^2} \tag{8.10}$$

is the resulting best credibility factor (which in this case is equal for all j),

$$\overline{X} = \frac{1}{JT}\sum_{j=1}^{J}\sum_{t=1}^{T} X_{jt} \tag{8.11}$$

is the collective estimator of m, and

$$\overline{X}_j = \frac{1}{T}\sum_{t=1}^{T} X_{jt} \tag{8.12}$$

is the individual estimator of m.

Proof. Because of the independence assumptions and the equal distributions, the random variables X_{it} with $i \neq j$ are interchangeable. By convexity, (8.8) has a unique minimum. In the optimum, all values of $g_{it}, i \neq j$ must be identical, for reasons of symmetry. If not, by interchanging coefficients we can show that more than one extremum exists. The same goes for all values $g_{jt}, t = 1, \ldots, T$. Combining this with the unbiasedness restriction, we see that the homogeneous linear estimator with minimal MSE must be of the form (8.9) for some z. We only have to find its optimal value.

Since X_{jt}, $\overline{X}_j$ and $\overline{X}$ all have mean m, we can rewrite the MSE (8.8) as:

$$\begin{aligned}
&\mathrm{E}[\{X_{j,T+1} - (1-z)\overline{X} - z\overline{X}_j\}^2] = \mathrm{E}[\{X_{j,T+1} - \overline{X} - z(\overline{X}_j - \overline{X})\}^2] \\
&= \mathrm{E}[\{X_{j,T+1} - \overline{X}\}^2] - 2z\,\mathrm{E}[\{X_{j,T+1} - \overline{X}\}\{\overline{X}_j - \overline{X}\}] + z^2\,\mathrm{E}[\{\overline{X}_j - \overline{X}\}^2] \\
&= \mathrm{Var}[X_{j,T+1} - \overline{X}] - 2z\,\mathrm{Cov}[X_{j,T+1} - \overline{X}, \overline{X}_j - \overline{X}] + z^2\,\mathrm{Var}[\overline{X}_j - \overline{X}].
\end{aligned} \tag{8.13}$$

This quadratic form in z is minimal for the following choice of z:

$$z = \frac{\mathrm{Cov}[X_{j,T+1} - \overline{X}, \overline{X}_j - \overline{X}]}{\mathrm{Var}[\overline{X}_j - \overline{X}]} = \frac{aT}{aT+s^2}, \tag{8.14}$$

where it is left to the reader (Exercise 8.2.1) to verify the final equality in (8.13) by proving and filling in the necessary covariances:

$$\begin{aligned}
&\mathrm{Cov}[X_{jt}, X_{ju}] = a \text{ for } t \neq u; \\
&\mathrm{Var}[X_{jt}] = a + s^2; \\
&\mathrm{Cov}[X_{jt}, \overline{X}_j] = \mathrm{Var}[\overline{X}_j] = a + \frac{s^2}{T}; \\
&\mathrm{Cov}[\overline{X}_j, \overline{X}] = \mathrm{Var}[\overline{X}] = \frac{1}{J}(a + \frac{s^2}{T}).
\end{aligned} \tag{8.15}$$

So indeed predictor (8.9) leads to the minimal MSE (8.8) for the value of z given in (8.10). ∇

Remark 8.2.3 (Asymptotic properties of the optimal credibility factor)
The credibility factor z in (8.10) has plausible asymptotic properties:

1. If $T \to \infty$, then $z \to 1$. The more claims experience there is, the more faith we can have in the individual risk premium. This asymptotic case is not very relevant in practice, as it assumes that the risk does not change over time.
2. If $a \downarrow 0$, then $z \downarrow 0$. If the expected individual claim amounts are identically distributed, there is no heterogeneity in the portfolio. But then the collective mean m, when known, or its best homogeneous estimator $\overline{X}$ are optimal linear estimators of the risk premium. See (8.16) and (8.9).
3. If $a \to \infty$, then $z \to 1$. This is also intuitively clear. In this case, the result on the other contracts does not provide information about risk j.
4. If $s^2 \to \infty$, then $z \to 0$. If for a fixed risk parameter, the claims experience is extremely variable, the individual experience is not especially useful for estimating the real risk premium. ∇

Note that (8.9) is only a *statistic* if the ratio s^2/a is known; otherwise its distribution will contain unknown parameters. In Example 8.2.5 below we show how this ratio can be estimated as a by-product of the ANOVA. The fact that the credibility factor (8.14) does not depend on j is due to the simplifying assumption we have made that the number of observation periods is the same for each j, as well as that all observations have the same variance.

If we allow our linear estimator to contain a constant term, looking in fact at the best *inhomogeneous* linear predictor $g_0 + g_{11}X_{11} + \cdots + g_{JT}X_{JT}$, we get the next theorem. Two things should be noted. One is that it will prove that the unbiasedness restriction is now superfluous. The other is that (8.16) below looks just like (8.9), except that the quantity $\overline{X}$ is replaced by m. But this means that the inhomogeneous credibility premium for group j does not depend on the data from other groups $i \neq j$. The homogeneous credibility premium assumes the ratio s^2/a to be known; the inhomogeneous credibility premium additionally assumes that m is known.

Theorem 8.2.4 (Balanced Bühlmann model; inhomogeneous estimator)
Under the same distributional assumptions about X_{jt} as in the previous theorem, the inhomogeneous linear combination $g_0 + g_{11}X_{11} + \cdots + g_{JT}X_{JT}$ to predict next year's claim total $X_{j,T+1}$ that is optimal in the sense of mean squared error is the credibility premium

$$z\overline{X}_j + (1-z)m, \tag{8.16}$$

where z and $\overline{X}_j$ are as in (8.10) and (8.12).

Proof. The same symmetry considerations as in the previous proof tell us that the values of $g_{it}, i \neq j$ are identical in the optimal solution, just as those of $g_{jt}, t = 1,\ldots,T$. So for certain g_0, g_1 and g_2, the inhomogeneous linear predictor of $X_{j,T+1}$ with minimal MSE is of the following form:

$$g_0 + g_1\overline{X} + g_2\overline{X}_j. \tag{8.17}$$

The MSE can be written as variance plus squared bias, as follows:

$$\begin{aligned} &\mathrm{E}[\{X_{j,T+1} - g_0 - g_1\overline{X} - g_2\overline{X}_j\}^2] \\ &= \mathrm{Var}[X_{j,T+1} - g_1\overline{X} - g_2\overline{X}_j] + \{\mathrm{E}[X_{j,T+1} - g_0 - g_1\overline{X} - g_2\overline{X}_j]\}^2. \end{aligned} \tag{8.18}$$

The second term on the right hand side is zero, and hence minimal, if we choose $g_0 = m(1 - g_1 - g_2)$. This entails that the estimator we are looking for is necessarily unbiased. The first term on the right hand side of (8.18) can be rewritten as

$$\begin{aligned} &\mathrm{Var}[X_{j,T+1} - (g_2 + g_1/J)\overline{X}_j - g_1(\overline{X} - \overline{X}_j/J)] \\ &= \mathrm{Var}[X_{j,T+1} - (g_2 + g_1/J)\overline{X}_j] + \mathrm{Var}[g_1(\overline{X} - \overline{X}_j/J)] + 0, \end{aligned} \tag{8.19}$$

because the covariance term vanishes since $g_1(\overline{X} - \overline{X}_j/J)$ depends only of X_{it} with $i \neq j$. Hence any solution (g_1, g_2) with $g_1 \neq 0$ can be improved, since a lower value of (8.19) is obtained by taking $(0, g_2 + g_1/J)$. Therefore choosing $g_1 = 0$ is optimal. So all that remains to be done is to minimize the following expression for g_2:

$$\mathrm{Var}[X_{j,T+1} - g_2\overline{X}_j] = \mathrm{Var}[X_{j,T+1}] - 2g_2\mathrm{Cov}[X_{j,T+1}, \overline{X}_j] + g_2^2\mathrm{Var}[\overline{X}_j], \tag{8.20}$$

which has as an optimum

$$g_2 = \frac{\mathrm{Cov}[X_{j,T+1}, \overline{X}_j]}{\mathrm{Var}[\overline{X}_j]} = \frac{aT}{aT + s^2}, \tag{8.21}$$

so the optimal g_2 is just z as in (8.10). The final equality can be verified by filling in the relevant covariances (8.15). This means that the predictor (8.16) for $X_{j,T+1}$ has minimal MSE. ∇

Example 8.2.5 (Credibility estimation in Example 8.2.1)
Consider again the portfolio of Example 8.2.1. It can be shown (see Exercise 8.2.8), that in model (8.5) the numerator of F in (8.4) (the MSB) has mean $aT + s^2$, while the denominator MSW has mean s^2. Hence $1/F$ will be close to $s^2/\{aT + s^2\}$, which means that we can use $1 - 1/F$ to estimate z. Note that this is not an unbiased estimator, since $\mathrm{E}[1/MSB] \neq 1/\mathrm{E}[MSB]$. The resulting credibility factor is $z = 0.782$ for each group. So the optimal forecasts for the claims next year in the three groups are $0.782\overline{X}_j + (1 - 0.782)\overline{X}$, $j = 1, 2, 3$, resulting in 102.18, 110 and 118.82. Notice the 'shrinkage effect': the credibility estimated premiums are closer together than the original group means 100, 110 and 120. ∇

Remark 8.2.6 (Estimating the risk premium)
One may argue that instead of aiming to predict next year's claim figure $X_{j,T+1}$, including the fluctuation $\Xi_{j,T+1}$, we actually should estimate the *risk premium* $m + \Xi_j$ of group j. But whether we allow a constant term in our estimator or not, in each case we get the same optimum. Indeed for every random variable Y:

$$\begin{aligned}&\mathrm{E}[\{m+\Xi_j+\Xi_{j,T+1}-Y\}^2]\\&=\mathrm{E}[\{m+\Xi_j-Y\}^2]+\mathrm{Var}[\Xi_{j,T+1}]+2\mathrm{Cov}[m+\Xi_j-Y,\Xi_{j,T+1}].\end{aligned}\tag{8.22}$$

If Y depends only on the X_{jt} that are already observed, hence with $t \le T$, the covariance term must be equal to zero. Since it follows from (8.22) that the MSEs for Y as an estimator of $m+\Xi_j$ and of $X_{j,T+1} = m+\Xi_j+\Xi_{j,T+1}$ differ only by a constant $\mathrm{Var}[\Xi_{j,T+1}] = s^2$, we conclude that both MSEs are minimized by the same estimator Y. ∇

The credibility premium (8.16) is a weighted average of the estimated individual mean claim, with as a weight the credibility factor z, and the estimated mean claim for the whole portfolio. Because we assumed that the number of observation years T for each contract is the same, by asking premium (8.16) on the lowest level we receive the same premium income as when we would ask $\overline{X}$ as a premium from everyone. For $z = 0$ the individual premium equals the collective premium. This is acceptable in a homogeneous portfolio, but in general not in a heterogeneous one. For $z = 1$, a premium is charged that is fully based on individual experience. In general, this individual information is scarce, making this estimator unusable in practice. Sometimes it even fails completely, like when a prediction is needed for a contract that up to now has not produced any claim.

The quantity $a > 0$ represents the heterogeneity of the portfolio as depicted in the risk quality component Ξ_j, and s^2 is a global measure for the variability within the homogeneous groups.

Remark 8.2.7 (Virtual experience)
Write $X_{j\Sigma} = X_{j1} + \cdots + X_{jT}$, then an equivalent expression for the credibility premium (8.16) is the following:

$$\frac{s^2 m + aT\overline{X}_j}{s^2+aT} = \frac{ms^2/a + X_{j\Sigma}}{s^2/a+T}.\tag{8.23}$$

So if we extend the number of observation periods T by an extra s^2/a periods and also add ms^2/a as virtual claims to the actually observed claims $X_{j\Sigma}$, the credibility premium is nothing but the average claims, adjusted for virtual experience. ∇

8.3 More general credibility models

In model (8.5) of the previous section, we assumed the components Ξ_j and Ξ_{jt} to be independent random variables. But from (8.14) and (8.15) one sees that actually only the covariances of the random variables X_{jt} are essential. We get the same results if we impose a model with weaker requirements, as long as the covariance structure remains the same. An example is to only require independence and identical distributions of the Ξ_{jt}, conditionally given Ξ_j, with $\mathrm{E}[\Xi_{jt}|\Xi_j = \xi] = 0$ for all ξ. If the joint distribution of Ξ_j and Ξ_{jt} is like that, the Ξ_{jt} are not necessarily independent, but they are uncorrelated, as can be seen from the following lemma:

Lemma 8.3.1 (Conditionally iid random variables are uncorrelated)
Suppose that given Ξ_j, the random variables $\Xi_{j1}, \Xi_{j2}, \ldots$ are iid with mean zero. Then we have

$$\mathrm{Cov}[\Xi_{jt}, \Xi_{ju}] = 0,\ t \neq u; \quad \mathrm{Cov}[\Xi_j, \Xi_{jt}] = 0. \tag{8.24}$$

Proof. Because of the decomposition rule for conditional covariances, see Exercise 8.3.1, we can write for $t \neq u$:

$$\mathrm{Cov}[\Xi_{ju}, \Xi_{jt}] = \mathrm{E}\big[\mathrm{Cov}[\Xi_{ju}, \Xi_{jt} | \Xi_j]\big] + \mathrm{Cov}\big[\mathrm{E}[\Xi_{ju}|\Xi_j], \mathrm{E}[\Xi_{jt}|\Xi_j]\big]. \tag{8.25}$$

This equals zero since, by our assumptions, $\mathrm{Cov}[\Xi_{ju}, \Xi_{jt}|\Xi_j] \equiv 0$ and $\mathrm{E}[\Xi_{ju}|\Xi_j] \equiv 0$. Clearly, $\mathrm{Cov}[\Xi_j, \Xi_{jt}|\Xi_j] \equiv 0$ as well. Because

$$\mathrm{Cov}[\Xi_j, \Xi_{jt}] = \mathrm{E}\big[\mathrm{Cov}[\Xi_j, \Xi_{jt}|\Xi_j]\big] + \mathrm{Cov}\big[\mathrm{E}[\Xi_j|\Xi_j], \mathrm{E}[\Xi_{jt}|\Xi_j]\big], \tag{8.26}$$

the random variables Ξ_j and Ξ_{jt} are uncorrelated as well. ∇

Note that in the model of this lemma, the random variables X_{jt} are not marginally uncorrelated, let alone independent.

Example 8.3.2 (Mixed Poisson distribution)
Assume that the X_{jt} random variables represent the numbers of claims in a year on a particular motor insurance policy. The driver in question has a number of claims in that year that has a Poisson(Λ_j) distribution, where the parameter Λ_j is a drawing from a certain non-degenerate structure distribution. Then the first component of (8.5) represents the expected number of claims $m = \mathrm{E}[X_{jt}] = \mathrm{E}[\Lambda_j]$ of an arbitrary driver. The second is $\Xi_j = \Lambda_j - m$; it represents the difference in average numbers of claims between this particular driver and an arbitrary driver. The third term $\Xi_{jt} = X_{jt} - \Lambda_j$ equals the annual fluctuation around the mean number of claims of this particular driver. In this case, the second and third component, though uncorrelated in view of Lemma 8.3.1, are not independent, for example because $\mathrm{Var}[X_{jt} - \Lambda_j | \Lambda_j - m] \equiv \mathrm{Var}[X_{jt}|\Lambda_j] \equiv \Lambda_j$. See also Section 8.5. ∇

Remark 8.3.3 (Parameterization through risk parameters)
The variance components model (8.5), even with relaxed independence assumptions, sometimes is too restricted for practical applications. Suppose that X_{jt} as in (8.5) now represents the annual claims total of the driver from Example 8.3.2, and also suppose that this has a compound Poisson distribution. Then apart from the Poisson parameter, there are also the parameters of the claim size distribution. The conditional variance of the noise term, given the second term (mean annual total claim costs), is now no longer a function of the second term. To remedy this, Bühlmann studied slightly more general models, having a latent random variable Θ_j, that might be vector-valued, as a structure parameter. The risk premium is the conditional mean $\mu(\Theta_j) := \mathrm{E}[X_{jt}|\Theta_j]$ instead of simply $m + \Xi_j$. If $\mathrm{E}[X_{jt}|\Theta_j]$ is not a one-to-one function of Θ_j, it might occur that contracts having the same Ξ_j in the

basic model above, have a different pattern of variation $\mathrm{Var}[\Xi_{jt}|\Theta_j]$ in Bühlmann's model. Therefore the basic model is insufficient here. But it can be shown that in this case the same covariances, and hence the same optimal estimators, are found.

An advantage of using a variance components model over Bühlmann's Bayesian way of describing the risk structure is that the resulting models are technically as well as conceptually easier, at only a slight cost of generality and flexibility. ∇

It is possible to extend credibility theory to models that are more complicated than (8.5). Results resembling the ones from Theorems 8.2.2 and 8.2.4 can be derived for such models. In essence, to find an optimal predictor in the sense of least squares one minimizes the quadratic MSE over its coefficients, if needed with an additional unbiasedness restriction. Because of the symmetry assumptions in the balanced Bühlmann model, only a one-dimensional optimization was needed there. But in general we must solve a system of linear equations that arises by differentiating either the MSE or a Lagrange function. The latter situation occurs when there is an unbiasedness restriction. One should not expect to obtain analytical solutions such as above.

Some possible generalizations of the basic model are the following.

Example 8.3.4 (Bühlmann-Straub model; varying precision)
Credibility models such as (8.5) can be generalized by looking at X_{jt} that are averages over a number of policies. It is also conceivable that there are other reasons to assume that not all X_{jt} have been measured equally precisely, therefore have the same variance. For this reason, it may be expedient to introduce weights in the model. By doing this, we get the Bühlmann-Straub model. In principle, these weights should represent the total number of observation periods of which the figure X_{jt} is the mean *(natural weights)*. Sometimes this number is unknown. In that case, one has to make do with approximate relative weights, like for example the total premium paid. If the actuary deems it appropriate, he can adjust these numbers to express the degree of confidence he has in the individual claims experience of particular contracts. In Section 8.4 we prove a result, analogous to Theorem 8.2.2, for the homogeneous premium in the Bühlmann-Straub model. ∇

Example 8.3.5 (Jewell's hierarchical model)
A further generalization is to subdivide the portfolio into sectors, and to assume that each sector p has its own deviation from the overall mean. The claims experience for contract j in sector p in year t can then be decomposed as follows:

$$X_{pjt} = m + \Xi_p + \Xi_{pj} + \Xi_{pjt}. \tag{8.27}$$

This model is called Jewell's hierarchical model. Splitting up each sector p into subsectors q, each with its own deviation $\Xi_p + \Xi_{pq}$, and so on, leads to a hierarchical chain of models with a tree structure. ∇

Example 8.3.6 (Cross classification models)
It is conceivable that X_{pjt} is the risk in sector p, and that index j corresponds to some other general factor to split up the policies, for example if p is the region and j the gender of the driver. For such two-way cross classifications it does not make

sense to use a hierarchical structure for the risk determinants. Instead, one could add to (8.27) a term Ξ'_j, to describe the risk characteristics of group j. In this way, one gets

$$X_{pjt} = m + \Xi_p + \Xi'_j + \Xi_{pj} + \Xi_{pjt}. \tag{8.28}$$

This is a cross-classification model. In Chapter 9, we study similar models, where the row and column effects are fixed but unknown, instead of being modeled as random variables such as here. ∇

Example 8.3.7 (De Vijlder's credibility model for IBNR)
Credibility models are also useful to tackle the problem of estimating IBNR reserves to be held, see also Chapter 10. These are provisions for claims that are not, or not fully, known to the insurer. In a certain calendar year T, realizations are known for random variables X_{jt} representing the claim figure for policies written in year j, in their tth year of development, $t = 0, 1, \ldots, T-j$. A credibility model for this situation is

$$X_{jt} = (m + \Xi_j)d_t + \Xi_{jt}, \tag{8.29}$$

where the numbers d_t are development factors, for example with a sum equal to 1, that represent the fraction of the claims paid on average in the tth development period, and where $m + \Xi_j$ represents the claims, aggregated over all development periods, on policies written in year j. ∇

Example 8.3.8 (Regression models; Hachemeister)
We can also generalize (8.5) by introducing collateral data. If for example y_{jt} represents a certain risk characteristic of contract j, like for example the age of the policy holder in year t, Ξ_j might be written as a linear, stochastic, function of y_{jt}. Then the claims in year t are equal to

$$\{m^{(1)} + \Xi_j^{(1)}\} + \{m^{(2)} + \Xi_j^{(2)}\}y_{jt} + \Xi_{jt}, \tag{8.30}$$

which is a credibility-regression model. Classical one-dimensional regression arises when $\Xi_j^{(k)} \equiv 0, k = 1, 2$. This means that there are no latent risk characteristics. Credibility models such as (8.30) were first studied by Hachemeister. ∇

8.4 The Bühlmann-Straub model

Just as in (8.7), in the Bühlmann-Straub model the observations can be decomposed as follows:

$$X_{jt} = m + \Xi_j + \Xi_{jt}, \quad j = 1, \ldots, J,\ t = 1, \ldots, T+1, \tag{8.31}$$

where the unobservable risk components $\Xi_j, j = 1, 2, \ldots, J$ are iid with mean zero and variance a; the Ξ_{jt} are also independent with mean zero. The components Ξ_j and Ξ_{jt} are assumed to be independent, too. The difference between the Bühlmann and the Bühlmann-Straub models is that in the latter the variance of the Ξ_{jt} components is s^2/w_{jt}, where w_{jt} is the *weight* attached to observation X_{jt}. This weight represents the relative precision of the various observations. Observations with variances like this arise when X_{jt} is an average of w_{jt} replications, hence $X_{jt} = \sum_k X_{jtk}/w_{jt}$ where $X_{jtk} = m + \Xi_j + \Xi_{jtk}$ with Ξ_{jtk} iid with zero mean and variance s^2. The random variables Ξ_{jtk} then denote deviations from the risk premium $m + \Xi_j$ for the kth individual contract in time period t and group j. In this case, the weights are called *natural weights*. Sometimes these weights are unknown, or there is another mechanism that leads to differing variances. In that case we can, for example, approximate the volume by the total premium for a cell.

To find the best homogeneous unbiased linear predictor $\sum h_{it}X_{it}$ of the *risk premium* $m + \Xi_j$ (see Remark 8.2.6), we minimize its MSE. In Theorem 8.4.1 below, we derive the optimal values in (8.33) for the coefficients h_{it}, under the unbiasedness restriction. The following notation will be used, see (8.10)–(8.12):

$$
\begin{aligned}
&w_{j\Sigma} = \sum_{t=1}^{T} w_{jt}; && w_{\Sigma\Sigma} = \sum_{j=1}^{J} w_{j\Sigma}; \\
&z_j = \frac{a w_{j\Sigma}}{s^2 + a w_{j\Sigma}}; && z_{\Sigma} = \sum_{j=1}^{J} z_j; \\
&X_{jw} = \sum_{t=1}^{T} \frac{w_{jt}}{w_{j\Sigma}} X_{jt}; && X_{ww} = \sum_{j=1}^{J} \frac{w_{j\Sigma}}{w_{\Sigma\Sigma}} X_{jw}; \qquad X_{zw} = \sum_{j=1}^{J} \frac{z_j}{z_{\Sigma}} X_{jw}.
\end{aligned}
\tag{8.32}
$$

Notice the difference between for example X_{jw} and X_{ju}. If a w appears as an index, this indicates that there has been a weighted summation over this index, using the (natural or other) weights of the observations. An index z denotes a weighted summation with credibility weights, while a Σ is used for an unweighted summation. The simplest way to allow for different numbers of observation periods T_j is to include some observations with weight zero when necessary.

Theorem 8.4.1 (Bühlmann-Straub model; homogeneous estimator)
The MSE-best homogeneous unbiased predictor $\sum_{i,t} h_{it}X_{it}$ of the risk premium $m + \Xi_j$ in model (8.31), that is, the solution to the following restricted minimization problem

$$
\begin{aligned}
&\min_{h_{it}} \mathrm{E}[\{m + \Xi_j - \sum_{i,t} h_{it}X_{it}\}^2] \\
&\text{subject to } \mathrm{E}[m + \Xi_j] = \sum_{i,t} h_{it}\mathrm{E}[X_{it}],
\end{aligned}
\tag{8.33}
$$

is the following credibility estimator (see also (8.9)):

$$
z_j X_{jw} + (1 - z_j) X_{zw}. \tag{8.34}
$$

Here X_{jw} as in (8.32) is the individual estimator of the risk premium, X_{zw} is the credibility weighted collective estimator and z_j the credibility factor for contract j.

Proof. To prove that of all the linear combinations of the observations to estimate $m+\Xi_j$ with the same mean, (8.34) has the smallest MSE, we could do a Lagrange optimization, solving the first order conditions to find an extremum. But it is simpler to prove the result by making use of the result that linear combinations of uncorrelated random variables with a given mean have minimal variance if the coefficients are inversely proportional to the variances; see Exercise 8.4.1. First we derive the optimal 'mix' $h_{it}/h_{i\Sigma}$ of the contracts in group i. The best choice proves to be $h_{it}/h_{i\Sigma} = w_{it}/w_{i\Sigma}$; from this we see that the observations X_{it} have to appear in (8.33) in the form X_{iw}. Then we derive that the totals $h_{i\Sigma}$ of the coefficients with group $i \neq j$ are best taken proportional to z_j. Finally, the optimal value of $h_{j\Sigma}$ is derived.

From (8.33) we see that the following problem must be solved to find the best predictor of $m+\Xi_j$:

$$\min_{h_{it}:h_{\Sigma\Sigma}=1} \mathrm{E}\Big[\{m+\Xi_j - \sum_{i,t} h_{it}X_{it}\}^2\Big]. \tag{8.35}$$

The restriction $h_{\Sigma\Sigma}=1$ is the unbiasedness constraint in (8.33). By this constraint, the expectation in (8.35) is also the variance. Substituting decomposition (8.31) for X_{it}, we get from (8.35):

$$\min_{h_{it}:h_{\Sigma\Sigma}=1} \mathrm{Var}\Big[(1-h_{j\Sigma})\Xi_j - \sum_{i\neq j} h_{i\Sigma}\Xi_i - \sum_{i,t} h_{it}\Xi_{it}\Big], \tag{8.36}$$

or, what is the same because of the variances of the components Ξ_j and Ξ_{jt} and the independence of these components:

$$\min_{h_{it}:h_{\Sigma\Sigma}=1} (1-h_{j\Sigma})^2 a + \sum_{i\neq j} h_{i\Sigma}^2 a + \sum_i h_{i\Sigma}^2 \sum_t \frac{h_{it}^2}{h_{i\Sigma}^2}\frac{s^2}{w_{it}}. \tag{8.37}$$

First we optimize the inner sum, extending over t. Because of Exercise 8.4.1 the optimal values of $h_{it}/h_{i\Sigma}$ are $w_{it}/w_{i\Sigma}$. So we can replace the observations $X_{it}, t=1,2,\ldots,T$ by their weighted averages X_{iw}. We see that the credibility estimator has the form $\sum_i h_{i\Sigma}X_{iw}$, where the values of $h_{i\Sigma}$ are still to be found.

The minimal value for the inner sum equals $s^2/w_{i\Sigma}$. From (8.32) we see that $a + s^2/w_{i\Sigma} = a/z_i$. So we can rewrite (8.37) in the form

$$\min_{h_{i\Sigma}:h_{\Sigma\Sigma}=1} (1-h_{j\Sigma})^2 a + h_{j\Sigma}^2 \frac{s^2}{w_{j\Sigma}} + (1-h_{j\Sigma})^2 \sum_{i\neq j} \frac{h_{i\Sigma}^2}{(1-h_{j\Sigma})^2}\frac{a}{z_i}. \tag{8.38}$$

As $h_{\Sigma\Sigma}=1$, we have $\sum_{i\neq j} h_{i\Sigma}/(1-h_{j\Sigma}) = 1$. So again because of Exercise 8.4.1, the optimal choice in (8.38) for the factors $h_{i\Sigma}$, $i\neq j$ is

$$\frac{h_{i\Sigma}}{1-h_{j\Sigma}} = \frac{z_i}{z_\Sigma - z_j}. \tag{8.39}$$

The minimal value for the sum in (8.38) is $a/(z_\Sigma - z_j)$, so (8.38) leads to

$$\min_{h_{j\Sigma}} (1 - h_{j\Sigma})^2 (a + \frac{a}{z_\Sigma - z_j}) + h_{j\Sigma}^2 \frac{s^2}{w_{j\Sigma}}. \tag{8.40}$$

The optimal value for $h_{j\Sigma}$, finally, can be found by once again applying Exercise 8.4.1. This optimal value is, as the reader may verify,

$$\begin{aligned} h_{j\Sigma} &= \frac{w_{j\Sigma}}{\frac{s^2}{a+a/(z_\Sigma - z_j)} + w_{j\Sigma}} = \frac{1}{\frac{1/z_j - 1}{1+1/(z_\Sigma - z_j)} + 1} \\ &= \frac{z_j(z_\Sigma - z_j + 1)}{(1-z_j)(z_\Sigma - z_j) + z_j(z_\Sigma - z_j + 1)} = z_j + (1-z_j)\frac{z_j}{z_\Sigma}. \end{aligned} \tag{8.41}$$

Because of (8.39) we see that $h_{i\Sigma} = (1-z_j)z_i/z_\Sigma$, which implies that (8.34) is indeed the MSE-optimal homogeneous unbiased linear predictor of the risk premium $m + \Xi_j$. ∇

Notice that if we replace Ξ_j in (8.31) by the constant ξ_j, that is, we take $a = 0$, we get the classical weighted mean X_{ww}. This is because in that case the relative weight $w_{j\Sigma}$ for X_{jw} is equal to the credibility weight z_j.

The *inhomogeneous* estimator of $m + \Xi_j$ contains a constant h, next to the homogeneous linear combination of the X_{jt} in (8.33). One may show, just as in Theorem 8.2.4, that the unbiasedness restriction is superfluous in this situation. The inhomogeneous estimator is equal to the homogeneous one, except that X_{zw} in (8.34) is replaced by m. The observations outside group j do not occur in the estimator. For the inhomogeneous estimator, both the ratio s^2/a and the value of m must be known. By replacing m by its best estimator X_{zw} under model (8.31), we get the homogeneous estimator again. Just as in Remark 8.2.6, the optimal predictor of $m + \Xi_j$ is also the optimal predictor of $X_{j,T+1}$. The asymptotic properties of (8.34) are analogous to those given in Remark 8.2.3. Also, the credibility premium can be found by combining the actual experience with virtual experience, just as in Remark 8.2.7. See the exercises.

8.4.1 Parameter estimation in the Bühlmann-Straub model

The credibility estimators of this chapter depend on the generally unknown structure parameters m, a and s^2. To be able to apply them in practice, one has to estimate these portfolio characteristics. Some unbiased estimators (not depending on the structure parameters that are generally unknown) are derived in the theorem below. We can replace the unknown structure parameters in the credibility estimators by these estimates, hoping that the quality of the resulting estimates is still good. The estimators of s^2 and a are based on the weighted sum-of-squares-within:

$$SSW = \sum_{j,t} w_{jt}(X_{jt} - X_{jw})^2, \tag{8.42}$$

and the weighted sum-of-squares-between

$$SSB = \sum_j w_{j\Sigma}(X_{jw} - X_{ww})^2. \tag{8.43}$$

Note that if all weights w_{jt} are taken equal to one, these expressions reduce to (8.2) and (8.3), defined in the balanced Bühlmann model.

Theorem 8.4.2 (Unbiased parameter estimates)
In the Bühlmann-Straub model, the following statistics are unbiased estimators of the corresponding structure parameters:

$$\begin{aligned}
\widetilde{m} &= X_{ww}, \\
\widetilde{s^2} &= \frac{1}{J(T-1)} \sum_{j,t} w_{jt}(X_{jt} - X_{jw})^2, \\
\widetilde{a} &= \frac{\sum_j w_{j\Sigma}(X_{jw} - X_{ww})^2 - (J-1)\widetilde{s^2}}{w_{\Sigma\Sigma} - \sum_j w_{j\Sigma}^2 / w_{\Sigma\Sigma}}.
\end{aligned} \tag{8.44}$$

Proof. The proof of $\mathrm{E}[X_{ww}] = m$ is easy. Using the covariance relations (8.15), we get for $\widetilde{s^2}$:

$$\begin{aligned}
J(T-1)\mathrm{E}[\widetilde{s^2}] &= \sum_{j,t} w_{jt} \left\{ \mathrm{Var}[X_{jt}] + \mathrm{Var}[X_{jw}] - 2\mathrm{Cov}[X_{jt}, X_{jw}] \right\} \\
&= \sum_{j,t} w_{jt} \left\{ a + \frac{s^2}{w_{jt}} + a + \frac{s^2}{w_{j\Sigma}} - 2(a + \frac{s^2}{w_{j\Sigma}}) \right\} \\
&= J(T-1)s^2.
\end{aligned} \tag{8.45}$$

For $\widetilde{a}$ we have

$$\begin{aligned}
&\mathrm{E}\Big[\sum_j w_{j\Sigma}(X_{jw} - X_{ww})^2 \Big] \\
&= \sum_j w_{j\Sigma} \left\{ \mathrm{Var}[X_{jw}] + \mathrm{Var}[X_{ww}] - 2\mathrm{Cov}[X_{jw}, X_{ww}] \right\} \\
&= \sum_j w_{j\Sigma} \Big\{ a + \frac{s^2}{w_{j\Sigma}} + a \sum_k \frac{w_{k\Sigma}^2}{w_{\Sigma\Sigma}^2} + \frac{s^2}{w_{\Sigma\Sigma}} - 2\Big(\frac{s^2}{w_{\Sigma\Sigma}} + \frac{a w_{j\Sigma}}{w_{\Sigma\Sigma}} \Big) \Big\} \\
&= a \sum_j w_{j\Sigma} \Big(1 + \sum_k \frac{w_{k\Sigma}^2}{w_{\Sigma\Sigma}^2} - 2\frac{w_{j\Sigma}}{w_{\Sigma\Sigma}} \Big) + s^2 \sum_j w_{j\Sigma} \Big(\frac{1}{w_{j\Sigma}} - \frac{1}{w_{\Sigma\Sigma}} \Big) \\
&= a \Big(w_{\Sigma\Sigma} - \sum_j \frac{w_{j\Sigma}^2}{w_{\Sigma\Sigma}} \Big) + (J-1)s^2.
\end{aligned} \tag{8.46}$$

Taking $\mathrm{E}[\widetilde{a}]$ in (8.44), using (8.45) and (8.46) we see that $\widetilde{a}$ is unbiased as well. ∇

Remark 8.4.3 (Negativity of estimators)
The estimator $\widetilde{s^2}$ is non-negative, but $\widetilde{a}$ might well be negative. Although this may be an indication that $a = 0$ holds, it can also happen if $a > 0$. Let us elaborate on Example 8.2.1, returning to the balanced Bühlmann model where all weights w_{jt} are equal to one. In that case, defining MSW and MSB as in (8.4), the estimators of s^2 and a in Theorem 8.4.2 reduce to

$$\widetilde{s^2} = MSW; \qquad \widetilde{a} = \frac{MSB - MSW}{T}. \tag{8.47}$$

To estimate z, we substitute these estimators into $z = \frac{aT}{aT+s^2}$, and we get the following statistic:

$$\widetilde{z} = 1 - \frac{MSW}{MSB}. \tag{8.48}$$

Using $X_{jt} = m + \Xi_j + \Xi_{jt}$ and defining $\overline{\Xi}_j = \frac{1}{T}\sum_t \Xi_{jt}$, we see that the SSW can be written as

$$SSW = \sum_{j=1}^{J}\sum_{t=1}^{T}(X_{jt} - \overline{X}_j)^2 = \sum_{j=1}^{J}\sum_{t=1}^{T}(\Xi_{jt} - \overline{\Xi}_j)^2. \tag{8.49}$$

Under the assumption that the Ξ_{jt} are iid $\mathrm{N}(0, s^2)$, the right hand side divided by s^2 has a $\chi^2(J(T-1))$ distribution. It is independent of the averages $\overline{\Xi}_j$, and hence also of the averages $\overline{X}_j = m + \Xi_j + \overline{\Xi}_j$. So MSW is independent of the $\overline{X}_j$, hence also of MSB.

Assuming that the components Ξ_j are iid $\mathrm{N}(0, a)$, we find in similar fashion that

$$\frac{SSB}{a + s^2/T} = \frac{J-1}{aT + s^2} MSB \tag{8.50}$$

is $\chi^2(J-1)$ distributed. So under the normality assumptions made, if it is multiplied by the constant $s^2/(aT + s^2) = 1 - z$, the variance ratio MSB/MSW of Section 8.2 is still $F(J-1, J(T-1))$ distributed. Thus,

$$(1-z)\frac{MSB}{MSW} = \frac{1-z}{1-\widetilde{z}} \sim F(J-1, J(T-1)). \tag{8.51}$$

In this way, $\Pr[\widetilde{a} < 0]$ can be computed for different values of J, T and s^2/a, see for example Exercise 8.4.9.

Note that by (8.47), the event $\widetilde{a} < 0$ is the same as $MSB/MSW < 1$. In Section 8.2 we established that the data indicate rejection of equal means, which boils down to $a = 0$ here, only if MSB/MSW exceeds the right-hand $F(J-1, J(T-1))$ critical value, which is larger than one for all J, T. Thus we conclude that, although $\Pr[\widetilde{a} < 0] > 0$ for every $a > 0$, obtaining such a value means that a Fisher test for $a = 0$ based on these data would not have led to rejection. This in turn means that there is in fact no statistical reason not to charge every contract the same premium.

In order to estimate $a = \mathrm{Var}[\Xi_j]$, one would use $\max\{0, \widetilde{a}\}$ in practice, but, though still consistent, this is not an unbiased estimator. ∇

Remark 8.4.4 (Credibility weighted mean and ordinary weighted mean)
The best unbiased estimator of m in model (8.31) is not X_{ww}, but X_{zw}. This does not contradict Exercise 8.4.1, since both X_{ww} and X_{zw} are linear combinations of the random variables X_{jw}, the variances of which are not proportional to the original weights $w_{j\Sigma}$, but rather to the credibility adjusted weights z_j. So a lower variance is obtained if we estimate m by the credibility weighted mean X_{zw} instead of by the ordinary weighted mean X_{ww}. A problem is that we do not know the credibility factors z_j to be used, as they depend on the unknown parameters that we are actually estimating. One way to achieve better estimators is to use iterative *pseudo-estimators*, which find estimates of the structure parameters by determining a fixed point of certain equations. See Example 8.4.6, as well as the more advanced literature on credibility theory. ∇

Example 8.4.5 (Computing the estimates in the Bühlmann-Straub model)
First, we generate a dataset consisting of `J=10` contracts, with `K=5` years of exposure each, satisfying the distributional assumptions (8.31) of the Bühlmann-Straub model. For that, we execute the following `R`-statements.

```
J <- 10; K <- 5; j <- rep(1:J, each=K); j <- as.factor(j)
m <- 100; a <- 100; s2 <- 64;
set.seed(6345789)
w <- 0.50 + runif(J*K)
X <- m + rep(rnorm(J, 0, sqrt(a)), each=K) +
         rnorm(J*K, 0, sqrt(s2/w))
```

Note that we attach a random weight in the interval $(0.5, 1.5)$ to each observation. In the last line, the second term is a vector of `J` independent $\mathrm{N}(0,a)$ random drawings Ξ_j, replicated `K` times each, the last a vector of independent $\mathrm{N}(0,s^2/w_{jk})$ random drawings Ξ_{jk}, $j=1,\dots,J, k=1,\dots,K$.

Just as in Example 8.2.1, we apply ANOVA to determine if there is any significant variation in the group means; if not, there is no heterogeneity between contracts in the portfolio, therefore no reason to apply credibility theory.

```
> anova(lm(X~j,weight=w))
Analysis of Variance Table

Response: X
          Df Sum Sq Mean Sq F value    Pr(>F)
j          9 5935.0   659.4  14.836 3.360e-10 ***
Residuals 40 1778.0    44.5
```

In this example the data clearly exhibit an effect of the factor group number. The `Sum Sq` values 5935 and 1778 are the SSB and the SSW, respectively; see (8.42) and (8.43), and see also below. The MSB and MSW, see (8.4), arise by dividing by `Df`.

In these laboratory conditions, the parameter values a, m and s^2 are known. Therefore we can directly compute the credibility premiums (8.34) for this case. First we compute the quantities in (8.32).

```
w.js <- tapply(w, j, sum); w.ss <- sum(w.js)
z.j <- 1 / (1 + s2/(a*w.js)); z.s <- sum(z.j)
X.jw <- tapply(X*w, j, sum)/w.js
X.ww <- sum(X.jw * w.js) / w.ss
X.zw <- sum(X.jw * z.j) / z.s
pr.j <- z.j * X.jw + (1-z.j) * X.zw #(8.34)
```

In the real world, these parameters m, s^2 and a are of course unknown and have to be estimated from the data. In (8.42)–(8.44), we find formulas for unbiased estimators $\widetilde{m}$, $\widetilde{s^2}$ and $\widetilde{a}$. Using R, they can be found as follows:

```
m.tilde <- X.ww
SSW <- sum(w*(X-X.jw[j])^2)
s2.tilde <- SSW/J/(K-1)
SSB <- sum(w.js*(X.jw-X.ww)^2)
a.tilde <- (SSB - (J-1)*s2.tilde) / (w.ss - sum(w.js^2)/w.ss)
```

Using the statements:

```
z.j.tilde <- 1 / (1 + s2.tilde / (a.tilde * w.js))
z.s.tilde <- sum(z.j.tilde)
X.zw.tilde <- sum(X.jw * z.j.tilde)/ z.s.tilde
pr.j.tilde <- z.j.tilde * X.jw + (1-z.j.tilde) * X.zw.tilde
```

we can recompute the credibility premiums (8.34) using the unbiased parameter estimates and (8.32). ∇

Example 8.4.6 (A pseudo-estimator for the heterogeneity parameter)
The estimator $\widetilde{a}$ of the heterogeneity parameter a given in (8.44) is unbiased, but it is also awkward looking and unintuitive. Consider the unbiased estimate of s^2, the heterogeneity in time of the results of the contract, in (8.44):

$$\widetilde{s^2} = \frac{1}{J(T-1)} \sum_{j,t} w_{jt}(X_{jt} - X_{jw})^2. \tag{8.52}$$

It adds up the squared differences of the observations with the contract mean, weighted by the natural weight w_{jt} that was used to construct X_{jw}. To get an unbiased estimate of s^2, we divide by the total experience JT, corrected for the fact that J means have been estimated so only $J(T-1)$ independent terms remain.

To get an analogous estimate of the between-groups heterogeneity a, consider

$$A = \frac{1}{J-1} \sum_{j} z_j (X_{jw} - X_{zw})^2. \tag{8.53}$$

In this case, there are J groups, and one mean is estimated. The squared differences between group mean and the best estimate X_{zw} for the overall mean m have weight proportional to the credibility weight z_j, the same set of weights that produces the minimal variance estimator X_{zw}. The reader is invited to show that $\mathrm{E}[A] = a$, see the exercises.

There is, however, a problem with the random variable A. If we fill in X_{zw} in the previous equation we get:

$$A = \frac{1}{J-1}\sum_j z_j\Big(X_{jw} - \sum_i \frac{z_i}{z_\Sigma}X_{iw}\Big)^2 \quad \text{with} \quad z_j = \Big(1 + \frac{s^2}{a w_{j\Sigma}}\Big)^{-1}. \tag{8.54}$$

So the right hand side depends on the unknown structure parameters a and s^2 (actually only on the ratio s^2/a), let us say as $A = f(a, s^2)$ with f the appropriate function. As a result, the random variable A is not a statistic, hence not an estimator. In such cases, we speak of *pseudo-estimators*. So we look at the following estimate of A:

$$A_1 = f(\widetilde{a}, \widetilde{s^2}). \tag{8.55}$$

But A_1 does not use the 'best' credibility weights available, therefore we look at $A_2 = f(A_1, \widetilde{s^2})$. Optimizing, we then look iteratively at $A_{n+1} = f(A_n, \widetilde{s^2})$, $n = 2, 3, \ldots$. Taking the limit for $n \to \infty$, and calling the limiting random variable $\lim_{n\to\infty} A_n =: \widehat{a}$, we see that the random variable $\widehat{a}$ is the solution to the following implicit equation:

$$\widehat{a} = \frac{1}{J-1}\sum_j \widehat{z}_j\Big(X_{jw} - \sum_i \frac{\widehat{z}_i}{\widehat{z}_\Sigma}X_{iw}\Big)^2 \quad \text{with} \quad \widehat{z}_j = \Big(1 + \frac{\widetilde{s^2}}{\widehat{a} w_{j\Sigma}}\Big)^{-1}. \tag{8.56}$$

The exact statistical properties of this estimator $\widehat{a}$ are hard to determine; even proving existence and uniqueness of the solution is a problem. But it is very easy to solve this equation by successive substitution, using the following `R` statements:

```
a.hat <- a.tilde
repeat {
  a.hat.old <- a.hat
  z.j.hat <- 1/(1+s2.tilde/(w.js*a.hat))
  X.zw.hat <- sum(z.j.hat * X.jw) / sum(z.j.hat)
  a.hat <- sum(z.j.hat*(X.jw-X.zw.hat)^2)/(J-1)
  if (abs((a.hat-a.hat.old)/a.hat.old) < 1e-6) break}
```

Here `a.tilde`, assumed positive, and `s2.tilde` are the unbiased estimates, `a..` is the current best guess A_n of $\widehat{a}$, and `a.` is the next one A_{n+1}. ∇

8.5 Negative binomial model for the number of car insurance claims

In this section we expand on Example 8.3.2 by considering a driver with a random accident proneness drawn from a non-degenerate distribution, and, given that his accident proneness equals λ, a Poisson(λ) distributed number of claims in a year. Charging a credibility premium in this situation leads to an experience rating system that resembles the bonus-malus systems we described in Chapter 6.

If for a motor insurance policy, all relevant variables for the claim behavior of the policyholder can be observed as well as used, the number of claims still is generated by a stochastic process. Assuming that this process is a Poisson process, the rating factors cannot do more than provide us with the exact Poisson intensity, that is, the Poisson parameter of the number of claims each year. Of the claim size, we know the probability distribution. The cell with policies sharing common values for all the risk factors would be homogeneous, in the sense that all policy holders have the same Poisson parameter and the same claims distribution. In reality, however, some uncertainty about the parameters remains, because it is impossible to obtain all relevant information on these parameters. So the cells are heterogeneous. This heterogeneity is the actual justification of using a bonus-malus system. In case of homogeneity, each policy represents the same risk, and there is no ground for asking different premiums within a cell.

The heterogeneity of the claim frequency can be modeled by assuming that the Poisson parameter λ has arisen from a structure variable Λ, with distribution $U(\lambda) = \Pr[\Lambda \leq \lambda]$. Just as in (8.5), we decompose the number of claims X_{jt} for driver $j = 1, \ldots, J$ in time period $t = 1, \ldots, T_j$ as follows:

$$X_{jt} = \mathrm{E}[\Lambda] + \{\Lambda_j - \mathrm{E}[\Lambda]\} + \{X_{jt} - \Lambda_j\}. \qquad (8.57)$$

Here $\Lambda_j \sim \Lambda$ iid. The last two components are uncorrelated, but not independent; see Exercise 8.5.6. Component $\Lambda_j - \mathrm{E}[\Lambda]$ has variance $a = \mathrm{Var}[\Lambda]$; for $X_{jt} - \Lambda_j$, just as in Example 3.3.1, $\mathrm{Var}[X_{jt}] - \mathrm{Var}[\Lambda_j] = \mathrm{E}[\Lambda]$ remains. The structural parameters m and s^2 coincide because of the Poisson distributions involved.

Up to now, except for its first few moments, we basically ignored the structure distribution. Several models for it are possible. Because of its mathematical properties and good fit (see later on for a convincing example), we will prefer the gamma distribution. Another possibility is the structure distribution that produces a 'good' driver, having claim frequency λ_1, with probability p, or a 'bad' driver with claim frequency $\lambda_2 > \lambda_1$. The number of claims of an arbitrary driver then has a mixed Poisson distribution with a two-point mixing distribution. Though one would expect more than two types of drivers to be present, this 'good driver/bad driver' model quite often fits rather closely to data found in practice.

For convenience, we drop the index j, except when we refer back to earlier sections. It is known, see again Example 3.3.1, that if the structure distribution of the Poisson parameter is gamma(α, τ), the marginal distribution of the number of claims X_t of driver j in time period t is negative binomial$(\alpha, p = \tau/(\tau+1))$. In Lemaire (1985), we find data from a Belgian portfolio with $J = 106\,974$ policies, see Table 8.1. The numbers n_k, $k = 0, 1, \ldots$, denote the number of policies with k accidents. If $X_t \sim$ Poisson(λ) for all j, the maximum likelihood estimate $\widehat{\lambda}$ for λ equals the average number of claims over all j. In Section 3.9 we showed how to find the negative binomial parameter estimates $\widehat{\alpha}$ and $\widehat{p}$ by maximum likelihood, solving (3.76) and (3.77). Then $\widehat{\tau} = \widehat{p}/(1-\widehat{p})$ in the Poisson-gamma mixture model follows from the invariance property of ML-estimators in case of reparameterization. Equation (3.76) ensures that the first moment of the estimated structure distribution, hence also of the marginal distribution of the number of claims, coincides with the first sample moment. The parameters p, λ_1 and λ_2 of the good driver/bad driver model have been estimated by the method of moments. Note that this method might fail to

Table 8.1 Observed numbers of accidents in some portfolio, and fitted values for a pure Poisson model and a negative binomial model fitted with ML, and a mixed Poisson model fitted by the method of moments.

k	n_k	$\widehat{n}_k$ (Poisson)	$\widehat{n}_k$ (Neg.Bin.)	$\widehat{n}_k$ (good/bad)
0	96 978	96 689.5	96 980.8	96 975.1
1	9 240	9 773.4	9 230.9	9 252.0
2	704	494.0	708.6	685.0
3	43	16.6	50.0	56.9
4	9	0.4	3.4	4.6
5+	0	0.0	0.2	0.3
χ^2		191.	0.1	2.1

produce admissible estimates $\widehat{\lambda}_i \geq 0$ and $0 \leq \widehat{p} \leq 1$. The resulting estimates for the three models considered were

$$\begin{aligned} &\widehat{\lambda} = 0.1010806; \\ &\widehat{\alpha} = 1.631292, \quad \widehat{\tau} = 16.13852; \\ &\widehat{\lambda}_1 = 0.07616114, \quad \widehat{\lambda}_2 = 0.3565502, \quad \widehat{p} = 0.8887472. \end{aligned} \tag{8.58}$$

Observed and estimated frequencies are in Table 8.1. The bottom row contains $\chi^2 = \sum_k (n_k - \widehat{n}_k)^2 / \widehat{n}_k$. When computing such χ^2-statistics, one usually combines cells with estimated numbers less than 5 with neighboring cells. So the last three rows are joined together into one row representing 3 or more claims. The two mixed models provide an excellent fit; in fact, the fit of the negative binomial model is almost too good to be true. Note that we fit 4 numbers using 2 or 3 parameters. But homogeneity for this portfolio is rejected without any doubt whatsoever.

Though the null-hypothesis that the numbers of claims for each policy holder are independent Poisson random variables with the same parameter is rejected, while the mixed Poisson models are not, we cannot just infer that policy holders have a fixed unobservable risk parameter, drawn from a structure distribution. It might well be that the numbers of claims are just independent negative binomial random variables, for example because the number of claims follows a Poisson process in which each year a new intensity parameter is drawn independently from a gamma structure distribution.

With the model of this section, we want to predict as accurately as possible the number of claims that a policy holder produces in the next time period $T+1$. This number is a Poisson(λ) random variable, with λ an observation of Λ, of which the prior distribution is known to be, say, gamma(α, τ). Furthermore, observations $X_1, \ldots, X_T$ from the past are known. The posterior distribution of Λ, given $X_1 = x_1, \ldots, X_T = x_T$, is also a gamma distribution, with adjusted parameters $\tau' = \tau + T$ and $\alpha' = \alpha + x_\Sigma$ with $x_\Sigma = x_1 + \cdots + x_T$; see Exercise 8.5.2. Assuming a quadratic loss function, in view of Exercise 8.2.9 the best predictor of the number of claims next year is the posterior expectation of Λ:

$$\lambda_{T+1}(x_1,\ldots,x_T) = \frac{\alpha + x_\Sigma}{\tau + T}. \tag{8.59}$$

This is just the observed average number of claims per time unit, provided we include a virtual prior experience of α claims in a time period of length τ. See also Remark 8.2.7. The forecasted premium (8.59) is also a credibility forecast, being a linear combination of a priori premium and policy average, because, see (8.10):

$$\frac{\alpha + x_\Sigma}{\tau + T} = z\frac{x_\Sigma}{T} + (1-z)\frac{\alpha}{\tau} \quad \text{for} \quad z = \frac{T}{\tau + T}. \tag{8.60}$$

Remark 8.5.1 (Non-linear estimators; exact credibility)
In Theorems 8.2.2 and 8.2.4 it was required that the predictors of $X_{j,T+1}$ were linear in the observations. Though such linear observations are in general the easiest to deal with, one may also look at more general functions of the data. Without linearity restriction, the best predictor in the sense of MSE for $X_{j,T+1}$ is the so-called *posterior Bayes estimator*, which is just the conditional mean $\mathrm{E}[X_{j,T+1}|X_{11},\ldots,X_{JT}]$. See also (8.59). If the Ξ_j and the Ξ_{jt} are independent *normal* random variables, the optimal linear estimator coincides with the Bayes estimator. In the literature, this is expressed as 'the credible mean is exact Bayesian'. Also combining a gamma prior and a Poisson posterior distribution gives such 'exact credibility', because the posterior Bayes estimator happens to be linear in the observations. See Exercise 8.5.2. The posterior mean of the claim figure is equal to the credibility premium (8.60). ∇

If we split up the premium necessary for the whole portfolio according to the mean value principle, we get an experience rating system based on credibility, which is a solid system for the following reasons:

1. The system is fair. Upon renewal of the policy, every insured pays a premium that is proportional to his estimated claim frequency (8.59), taking into account all information from the past.
2. The system is balanced financially. Write $X_\Sigma = X_1 + \cdots + X_T$ for the total number of claims generated, then $\mathrm{E}[X_\Sigma] = \mathrm{E}\big[\mathrm{E}[X_\Sigma|\Lambda]\big] = T\mathrm{E}[\Lambda]$, so

$$\mathrm{E}\left[\frac{\alpha + X_\Sigma}{\tau + T}\right] = \frac{\alpha + T\frac{\alpha}{\tau}}{\tau + T} = \frac{\alpha}{\tau}. \tag{8.61}$$

   This means that for every policy, the mean of the proportionality factor (8.59) is equal to its overall mean α/τ. So the expected value of the premium to be paid by an arbitrary driver remains constant over the years.
3. The premium only depends on the number of claims filed in the previous T years, and not on how these are distributed over this period. So for the premium next year, it makes no difference if the claim in the last five years was in the first or in the last year of this period. The bonus-malus system in Section 6.2 does not have this property. But it is questionable if this property is even desirable. If one assumes, like here, the intensity parameter λ to remain constant, K is a sufficient statistic. In practice, however, the value of λ is not constant. People get past their

Table 8.2 Optimal estimates (8.62) of the claim frequency next year compared with a new driver

	Number of years T										
Nr. of claims x_Σ	0	1	2	3	4	5	6	7	8	9	10
0	100	94	89	84	80	76	73	70	67	64	62
1		153	144	137	130	124	118	113	108	104	100
2		212	200	189	180	171	164	157	150	144	138
3		271	256	242	230	219	209	200	192	184	177
4		329	311	295	280	267	255	243	233	224	215

youth or past their prime, or the offspring gets old enough to drive the family car. Following this reasoning, later observations should count more heavily than old ones.

4. Initially, at time $t = 0$, everyone pays the same premium, proportional to α/τ. If T tends to ∞, the difference between the premium $(\alpha + x_\Sigma)/(\tau + T)$ asked and the actual average payments on the policy x_Σ/T vanishes. The variance $(\alpha + x_\Sigma)/(\tau + T)^2$ of the posterior distribution converges to zero. So in the long run, everyone pays the premium corresponding to his own risk; the influence of the virtual experience vanishes.

Using the values $\alpha = 1.6$ and $\tau = 16$, see (8.58), we have constructed Table 8.2 giving the optimal estimates of the claim frequencies in case of various lengths T of the observation period and numbers $k = x_\Sigma$ of claims observed. The initial premium is set to 100%, the a posteriori premiums are computed as:

$$100\frac{\lambda_{T+1}(x_1,\dots,x_T)}{\lambda_1} = \frac{100\frac{\alpha+x_\Sigma}{\tau+T}}{\alpha/\tau} = 100\frac{\tau(\alpha+x_\Sigma)}{\alpha(\tau+T)} \tag{8.62}$$

One sees that in Table 8.2, a driver who caused exactly one claim in the past ten years represents the same risk as a new driver, who is assumed to carry with him a virtual experience of 1.6 claims in 16 years. A person who drives claim-free for ten years gets a discount of $1 - \tau/(\tau + 10) = 38\%$. After a claims experience of 16 years, actual and virtual experience count just as heavily in the premium.

Example 8.5.2 (Comparison with the bonus-malus system of Chapter 6)
As an example, look at the premiums to be paid in the 6th year of insurance by a driver who has had one claim in the first year of observation. In Table 8.2, his premium next year equals 124%. In the system of Table 6.1, his path on the ladder was $2 \to 1 \to 2 \to 3 \to 4$, so now he pays the premium of step 5, that is, 70%. The total of the premiums paid (see Table 8.2) is $100 + 153 + 144 + 137 + 130 + 124 = 788\%$ of the premium for a new entrant. In the system of Table 6.1, he has paid only $100 + 120 + 100 + 90 + 80 + 70 = 560\%$. Note that for the premium next year in Table 8.2, it makes no difference if the claim occurred in the first or the last year of observation, though this affects the total claims paid. ∇

Remark 8.5.3 (Overlapping claim frequencies)
Consider a policyholder with T years of claims experience. The posterior distribution of the expected number of claims Λ is gamma$(\alpha+x_\Sigma, \tau+T)$ if x_Σ claims were filed. If $T=3$, in case $x_\Sigma=0$ and $x_\Sigma=2$, the premium to be paid next year differs by a factor $189/84=2.25$. But the posterior distributions of both claim frequencies overlap to a large extent. Indeed, in the first case, the probability is 60.5% to have a claim frequency lower than the average $\alpha/(\tau+T)=0.0842$ for drivers with a similar claims experience, since $G(0.0842;\alpha,\tau+T)=0.605$. But in the second case, there also is a substantial probability to have a better Poisson parameter than the average of drivers as above, since $G(0.0842;\alpha+x_\Sigma,\tau+T)=0.121$ for $x_\Sigma=2$ and $T=3$. Experience rating by any bonus-malus system may be quite unfair for 'good' drivers who are unlucky enough to produce claims. ∇

8.6 Exercises

Section 8.2

1. Finish the proofs of Theorems 8.2.2 and 8.2.4 by filling in and deriving the relevant covariance relations (8.15). Use and verify the linearity properties of covariances: for all random variables X,Y and Z, we have $\mathrm{Cov}[X,Y+Z]=\mathrm{Cov}[X,Y]+\mathrm{Cov}[X,Z]$, while for all real α, $\mathrm{Cov}[X,\alpha Y]=\alpha\mathrm{Cov}[X,Y]$.
2. Let $X_1,\dots,X_T$ be uncorrelated random variables with mean m and variance s^2. Consider the weighted average $X_w=\sum_t w_t X_t$, where the weights $w_t\geq 0, t=1,\dots,T$ satisfy $\sum_t w_t=1$. Show that $\mathrm{E}[X_w]=m$, $\mathrm{Cov}[X_t,X_w]=w_t s^2$ and $\mathrm{Var}[X_w]=\sum_t w_t^2 s^2$.
   [If especially $w_t\equiv\frac{1}{T}$, we get $X_w=\overline{X}$ and $\mathrm{E}[\overline{X}]=m; \mathrm{Cov}[X_t,\overline{X}]=\mathrm{Var}[\overline{X}]=\frac{s^2}{T}$.]
3. Show that the sample variance $S^2=\frac{1}{T-1}\sum_1^T\{X_t-\overline{X}\}^2$ is an unbiased estimator of s^2.
4. Show that the best predictor of $X_{j,T+1}$ is also the best estimator of the risk premium $m+\Xi_j$ in the situation of Theorem 8.2.2. What is the best linear unbiased estimator (BLUE) of Ξ_j?
5. Determine the variance of the credibility premium (8.9). What is the MSE? Also determine the MSE of (8.9) as an estimator of $m+\Xi_j$.
6. Determine the credibility estimator if the unbiasedness restriction is not imposed in Theorem 8.2.2. Also investigate the resulting bias.
7. Show that if each contract pays the homogeneous premium, the sum of the credibility premiums equals the average annual outgo in the observation period.
8. Show that in model (8.5), the MSB has mean $aT+s^2$, while the MSW has mean s^2.
9. Prove that for each random variable Y, the real number p that is the best predictor of it in the sense of MSE is $p=\mathrm{E}[Y]$.
10. Let $\vec{X}=(X_{11},\dots,X_{1T},X_{21},\dots,X_{2T},\dots,X_{J1},\dots,X_{JT})^T$ be the vector containing the observable random variables in (8.7). Describe the covariance matrix $\mathrm{Cov}[\vec{X},\vec{X}]$.

Section 8.3

1. Derive the formula $\mathrm{Cov}[X,Y] = \mathrm{E}[\mathrm{Cov}[X,Y|Z]] + \mathrm{Cov}[\mathrm{E}[X|Z],\mathrm{E}[Y|Z]]$ for the decomposition of covariances into conditional covariances.

Section 8.4

1. Let $X_1,\dots,X_T$ be independent random variables with variances $\mathrm{Var}[X_t] = s^2/w_t$ for certain positive numbers (weights) w_t, $t = 1,\dots,T$. Show that the variance $\sum_t \alpha_t^2 s^2/w_t$ of the linear combination $\sum_t \alpha_t X_t$ with $\alpha_\Sigma = 1$ is minimal when we take $\alpha_t \propto w_t$, where the symbol $\propto$ means 'proportional to'. Hence the optimal solution has $\alpha_t = w_t/w_\Sigma$. Prove also that the minimal value of the variance in this case is s^2/w_Σ.
2. Prove that in model (8.31), we have $\mathrm{Var}[X_{zw}] \leq \mathrm{Var}[X_{ww}]$. See Remark 8.4.4.
3. Determine the best homogeneous linear estimator of m.
4. Show that in determining the best inhomogeneous linear estimator of $m+\Xi_j$, the unbiasedness restriction is superfluous.
5. Show that, just as in Remark 8.2.6, the optimal predictors of $X_{j,T+1}$ and $m+\Xi_j$ coincide in the Bühlmann-Straub model.
6. Describe the asymptotic properties of z_j in (8.32); see Remark 8.2.3.
7. In the same way as in Remark 8.2.7, describe the credibility premium (8.34) as a mix of actual and virtual experience.
8. Show that (8.9) follows from (8.34) in the special case (8.5)–(8.6) of the Bühlmann-Straub model given in (8.31).
9. In the situation of Remark 8.4.3, for $s^2/a = 0.823$, $J = 5$ and $T = 4$, show that the probability of the event $\widetilde{a} < 0$ equals 0.05.
10. Estimate the credibility premiums in the Bühlmann-Straub setting when the claims experience for three years is given for three contracts, each with weight $w_{jt} \equiv 1$. Find the estimates both by hand and by using R, if the claims on the contracts are as follows:

	$t=1$	$t=2$	$t=3$
$j=1$	10	12	14
$j=2$	13	17	15
$j=3$	14	10	6

11. Show that the pseudo-estimator A in (8.53) has indeed mean a.
12. Compare the quality of the iterative estimator $\widehat{a}$ in (8.56) and the unbiased one $\widetilde{a}$, by generating a large sample (say, 100 or 1000 replications of the laboratory portfolio as above). Look at sample means and variances, and plot a histogram. Count how often the iterative estimate is closer to the real value `a`.

Section 8.5

1. Verify that the parameters estimates given in (8.58) are as they should be.

2. [♠] Suppose that Λ has a gamma(α,τ) prior distribution, and that given $\Lambda = \lambda$, the annual numbers of claims $X_1,\ldots,X_T$ are independent Poisson(λ) random variables. Prove that the posterior distribution of Λ, given $X_1 = x_1,\ldots,X_T = x_T$, is gamma$(\alpha + x_\Sigma, \tau + T)$, where $x_\Sigma = x_1 + \cdots + x_T$.
3. By comparing $\Pr[X_2 = 0]$ with $\Pr[X_2 = 0|X_1 = 0]$ in the previous exercise, show that the numbers of claims X_t are not marginally independent. Also show that they are not uncorrelated.
4. Show that the mode of a gamma(α,τ) distribution, that is, the argument where the density is maximal, is $(\alpha-1)_+/\tau$.
5. [♠] Determine the estimated values for n_k and the χ^2-test statistic if α and τ are estimated by the method of moments.
6. Show that in the model (8.57) of this section, Λ_j and $X_{jt} - \Lambda_j$ are uncorrelated. Taking $\alpha = 1.6$ and $\tau = 16$, determine the ratio $\mathrm{Var}[\Lambda_j]/\mathrm{Var}[X_{jt}]$. [Since no model for X_{jt} can do more than determine the value of Λ_j as precisely as possible, this ratio provides an upper bound for the attainable 'percentage of explained variation' on an individual level.]
7. [♠] What is the Loimaranta efficiency of the system in Table 8.2? What is the steady state distribution?
8. Verify the estimated values for n_k and the χ^2-test statistic if the estimates $\widehat{\lambda}_1$, $\widehat{\lambda}_2$, $\widehat{p}$ in (8.58) are determined by maximum likelihood.