Bühlmann-Straub Model

The Bühlmann’s model cannot be applied to group insurances because that model does not allow for changes in the number of insured members of the group. Therefore we appeal to the Bühlmann-Straub model for such cases. In the Bühlmann-Straus Model we assume that there are \( n \) policy years and for each year \( t \) there are \( m_t \) exposures, and that \( X_t \) is the claim size, number of claims, or...

per unit of exposure during period \( t \). Note that the (loss, claim size, or number of claims) “per unit of exposure” is used because the exposure can vary through time and from risk to risk.

So, if the aggregate claim size in year \( t \) is \( Y_t \), then we actually have \( X_t = \frac{Y_t}{m_t} \). Note that since \( X_t \) measures a quantity per unit of exposure, the \( X_t \)'s are no longer assumed to have the same distribution.

<table>
<thead>
<tr>
<th>Risk</th>
<th>Periods of Experience</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( X_{11} ) ( X_{12} ) ( \cdots ) ( X_{1N_1} )</td>
</tr>
<tr>
<td></td>
<td>( m_{11} ) ( m_{12} ) ( \cdots ) ( m_{1N_1} )</td>
</tr>
<tr>
<td>2</td>
<td>( X_{21} ) ( X_{22} ) ( \cdots ) ( \cdots ) ( X_{2N_2} )</td>
</tr>
<tr>
<td></td>
<td>( m_{21} ) ( m_{22} ) ( \cdots ) ( \cdots ) ( m_{2N_2} )</td>
</tr>
<tr>
<td>( \vdots )</td>
<td>( \vdots ) ( \vdots ) ( \cdots ) ( \cdots ) ( \vdots )</td>
</tr>
<tr>
<td>R</td>
<td>( X_{R1} ) ( X_{R2} ) ( \cdots ) ( \cdots ) ( X_{RN_R} )</td>
</tr>
<tr>
<td></td>
<td>( m_{R1} ) ( m_{R2} ) ( \cdots ) ( \cdots ) ( m_{RN_R} )</td>
</tr>
</tbody>
</table>

The number of periods of experience can vary by risk, and that the experience periods do not have to start at the same time either.

**Example (from the Dean’s lectures).** ABC Insurance, Inc. sells dental insurance plans to companies with fewer than one hundred employees. An actuary is analyzing the number of claims per employee. Looking at the first company in her file, she sees that the company has three full years of plan coverage. In the first year there were 40 employee-years with 84 claims, in the second year there were 44 employee-years with 88 claims, and in the third year there were 42 employee-years with 105 claims. Designating this selected company as Risk 1, then:

\[ X_{11} = \frac{84 \text{ claims}}{40 \text{ employee-years}} = 2.1 \text{ claims/employee-year} \]
\[ X_{12} = 88 \text{ claims} / 44 \text{ employee-years} = 2.0 \text{ claims/employee-year} \]

\[ X_{13} = 105 \text{ claims} / 42 \text{ employee-years} = 2.5 \text{ claims/employee-year} \]

The exposures are \( m_{11} = 40 \text{ employee-years} \), \( m_{12} = 44 \text{ employee-years} \), and \( m_{13} = 42 \text{ employee-years} \).
Bühlmann-Straub Model for one policyholder when underlying probabilities are known

As in the Bühlmann’s model we assume that the conditional random variables \( X_1|\theta, X_2|\theta, \ldots \) are independent. Further assumption is that the process variances, \( \text{Var}(X_i|\theta) \), are inversely proportional to the size (i.e., exposure) of the risk during each observation period, in other words, the product

\[
\sigma^2(\theta) := m_i \text{Var}(X_i|\theta)
\]

is constant (for all \( t \)).

Now we define the following quantities:

1. Hypothetical Mean for risk \( \theta \) per unit of exposure:
   \[
   \mu(\theta) = E(X_1|\theta) = E(X_2|\theta) = \cdots
   \]

2. Process Variance for risk \( \theta \):
   \[
   \text{Var}(X_1|\theta) = \frac{\sigma^2(\theta)}{m_1} \quad \cdots \quad \text{Var}(X_i|\theta) = \frac{\sigma^2(\theta)}{m_i} \quad \cdots
   \]

3. Population mean: \( \mu = E[\mu(\theta)] = E[E[X_i|\theta]] \)

4. Expected Value of Process Variance: \( \text{EPV} = E[\sigma^2(\theta)] \)

5. Variance of Hypothetical Means: \( \text{VHM} = \text{Var}[\mu(\theta)] \)

Example (Dean’s notes page 12). The annual numbers of claims for truck drivers in a homogeneous population are independently and identically distributed. [The population might represent the work force of a large trucking company with strict hiring standards and good safety training for each driver.] For each driver the number of claims per year has a mean of \( \mu(\theta) \) and a variance of \( \sigma^2(\theta) \). (The \( \theta \) parameter applies to every driver in the group.)
A group of 10 drivers is selected from the larger population.

1. What is the expected annual claims frequency for the group of 10 drivers?
2. What is the variance of the annual claims frequency for the group?

**Solution (from the Dean’s notes).** Let \( X_{1t}, X_{2t}, \ldots, X_{10t} \) be random variables representing the number of claims in year \( t \) for each of the ten selected drivers. Then \( X_t = \frac{1}{10} \sum_{i=1}^{10} X_{it} \) is the annual claims frequency for the group; that is, it is the annual number of claims per driver. The exposure is \( m_t = 10 \) and the unit of exposure is one driver. The expected value and variance for the annual claims frequency for the group are

\[
E[X_t] = E\left[ \frac{1}{10} \sum_{i=1}^{10} X_{it} \right] = \frac{1}{10} \sum_{i=1}^{10} E[X_{it}] = \frac{1}{10} \sum_{i=1}^{10} \mu(\theta) = \mu(\theta)
\]

\[
\text{Var}[X_t] = \text{Var}\left[ \frac{1}{10} \sum_{i=1}^{10} X_{it} | \theta \right] = \frac{1}{(10)^2} \sum_{i=1}^{10} \text{Var}[X_{it} | \theta] = \frac{1}{100} \sum_{i=1}^{10} \sigma^2(\theta) = \frac{\sigma^2(\theta)}{10}
\]

In this example, the exposure is the number of drivers in the group, which is 10. The expected claims frequency is the same whether there is one driver, 10 drivers, or 100 drivers in the group; however, the variance in the groups claims frequency is inversely proportional to the number of drivers in the group.

\[\blacksquare\]

In the Bühlmann-Straub model one seeks a point estimation for \( E[\mu(\theta) | X_1 = x_1, \ldots, X_n = x_n] \). But as we have argued before, this conditional expectation is the same as the conditional expectation

\[
E[X_{n+1} | X_1 = x_1, \ldots, X_n = x_n] = E[X_{n+1} | X_1 = x_1, \ldots, X_n = x_n]
\]

We set:

\[
\hat{X} = \sum_{i=1}^{N} \left( \frac{m_i}{m} \right) X_i \quad \text{where} \quad m = \sum_{i=1}^{N} m_i
\]

4
Then
\[
E(\bar{X}|\theta) = E \left( \sum_{t=1}^{N} \left( \frac{m_t}{m} \right) X_t | \theta \right) = \sum_{t=1}^{N} \left( \frac{m_t}{m} \right) E(X_t | \theta) = \sum_{t=1}^{N} \left( \frac{m_t}{m} \right) \mu(\theta) = \mu(\theta)
\]
\[
\text{Var}(\bar{X}|\theta) = \text{Var} \left( \sum_{t=1}^{N} \left( \frac{m_t}{m} \right) X_t | \theta \right) = \sum_{t=1}^{N} \left( \frac{m_t}{m} \right)^2 \text{Var}(X_t | \theta)
\]
\[
= \sum_{t=1}^{N} \left( \frac{m_t}{m} \right)^2 \frac{\sigma^2(\theta)}{m_t} = \frac{\sigma^2(\theta)}{m}
\]

The unconditional mean and variance of \( \bar{X} \) are:
\[
E[\bar{X}] = E[E[\bar{X}|\theta]] = E[\mu(\theta)] = \mu
\]
\[
\text{Var}[\bar{X}] = \text{Var}[E[\bar{X}|\theta]] + E[\text{Var}[\bar{X}|\theta]] = \text{Var}[\mu(\theta)] + \frac{E[\sigma^2(\theta)]}{m} = VHM + \frac{EPV}{m}
\]

In Bühlmann-Straub model, the credibility assigned to \( \bar{X} \) (to estimate \( \mu(\theta) \)) is
\[
Z = \frac{\text{Variance of the Hypothetical Means}}{\text{Total Variance of the Estimator}}
\]

Upon simplifying, we get:
\[
Z = \frac{m}{m + K}
\]
where the value \( K \) is defined by:
\[
K = \frac{EPV}{VHM}
\]

The credibility estimate is
\[
\hat{\mu}(\theta) = Z \cdot \bar{X} + (1 - Z) \cdot \mu
\]

**Note.** The Bühlmann’s Model is a special case of the Bühlmann-Straub Model with \( m_t = 1 \) for all time \( t \).
Bühlmann-Straub Model for more-than-one policyholder (nonparametric estimation)

Here we have $r$ group policyholders and for each group $i$ we have $n_i$ policy years; the start of the years for different groups may differ. We adopt the following notations:

$X_{it} =$ the average loss/claim for policyholder $i$ in year $t$:

$X_i = (X_{i1}, \ldots, X_{in_i})$

$m_i$ denote the number of exposure units for policyholder $i$ in year $t$:

The total number of exposure units over all years for each group $i$ is

$$m_i = \sum_{t=1}^{T} m_{it}$$

The total exposure units for all policyholders over all years is

$$m = \sum_{i=1}^{r} m_i$$

The average loss experience of policyholder $i$ over all the years

$$\bar{X}_i = \frac{1}{m_i} \sum_{t=1}^{n_i} m_{it} X_{it}$$

The overall average losses is

$$\bar{X} = \frac{1}{m} \sum_{i=1}^{r} m_i \bar{X}_i$$

Assumptions:

1. The random vectors $\{X_1, \ldots, X_r\}$ are assumed to be mutually statistically independent.

2. The distribution of each vector $X_i$ depends on a risk parameter $\theta_i$, and we assume that the random variables $\{\theta_1, \ldots, \theta_r\}$ form an i.i.d.
3. within any group \( i \), the variables
\[
X_{i1|\theta_i}, \ldots, X_{in_i|\theta_i}
\]
are independent.

Set
\[
\mu(\theta_i) = E[X_{it|\theta_i}]
\]
so, for each group, the hypothetical means are constant over time. Here we have:

\[
\sigma^2(\theta_i) = m_{it} \text{Var}(X_{it|\theta_i})
\]

\[
\begin{align*}
\mu &= E[\mu(\theta_i)] \\
\text{EPV} &= E[\sigma^2(\theta_i)] \\
\text{VHM} &= \text{Var}[\mu(\theta_i)]
\end{align*}
\]

We are going to estimate these parameters, which are called **structural parameters**.

**Unbiased Estimation for \( \mu \):**

\[
\hat{\mu} = \bar{X}
\]

**Unbiased Estimation for \( \sigma^2(\theta_i) \):**

\[
\hat{\sigma_i}^2 = \frac{1}{n_i-1} \sum_{t=1}^{n_i} m_{it}(X_{it} - \bar{X}_i)^2
\]

**Unbiased Estimation for \( \text{EPV} \):**

\[
\hat{\text{EPV}} = \sum_{i=1}^{r} w_i \hat{\sigma_i}^2 = \frac{1}{\sum_{i=1}^{r}(n_i-1)} \sum_{i=1}^{r} \sum_{t=1}^{n_i} m_{it}(X_{it} - \bar{X}_i)^2 \\
w_i = \frac{n_i-1}{\sum_{i=1}^{r}(n_i-1)}
\]
Unbiased Estimation for VHM:

\[ \bar{VHM} = \frac{m}{m^2 - \sum m_i^2} \left\{ \sum_{i=1}^{r} m_i(\bar{X}_i - \bar{X})^2 - (r - 1)\bar{EPV} \right\} \]

If we set

\[ \hat{k} = \frac{\bar{EPV}}{\bar{VHM}} \quad \hat{z}_i = \frac{m_i}{m_i + \hat{k}} \]

then the credibility estimate for the credibility premium

\[ E[X_{i,n+1}|X_{i,1} = x_{i,1}, \ldots, X_{n,i} = x_{n,i}] \]

and for \( \mu (\theta_i) \) is

\[ \bar{Z}_i \bar{X}_i + (1 - \bar{Z}_i)\bar{X} \]

Estimate of the premium for policyholder \( i \) is:

\[ m_{i,n+1} \left( \hat{Z}_i \bar{X}_i + (1 - \hat{Z}_i)\bar{X} \right) \]

(i) Determine the number of periods \( n_i \) for each of the policyholders.

(ii) Determine the exposure measure \( m_{it} \) for each policyholder \( i \) during each period \( t \).

(iii) Calculate the claim amounts \( x_{it} \).

(iv) Calculate the average claim amounts \( \bar{x}_i \) for each policyholder over all periods.

(v) Calculate the estimated \( \bar{\mu} = \bar{x} \).

(vi) Calculate the estimated \( \bar{EPV} = \sum_{i=1}^{r} w_i \hat{\sigma}_i^2 \)

\[ w_i = \frac{n_i - 1}{\sum_{i=1}^{n} (n_i - 1)} \]

(vii) Calculate the estimated \( \bar{VHM} = \frac{m}{m^2 - \sum m_i^2} \left\{ \sum_{i=1}^{r} m_i(\bar{X}_i - \bar{X})^2 - (r - 1)\bar{EPV} \right\} \)

(viii) Calculate \( \hat{k} = \frac{\bar{EPV}}{\bar{VHM}} \)
(ix) Calculate the credibility factors: $\hat{z}_i = \frac{m_i}{m_i + k}$

(x) Calculate the average claim amount per exposure unit for policyholder $i$:

$$\hat{z}_i \bar{X}_i + (1 - \hat{z}_i) \bar{X}$$

(xi) Calculate the aggregate claim amount for (policyholder) group $i$:

$$m_{i,n_{i+1}} \left( \hat{z}_i \bar{X}_i + (1 - \hat{z}_i) \bar{X} \right)$$

**Example.** The aggregate claim amount for two groups over three years are given in the following table:

<table>
<thead>
<tr>
<th>Group</th>
<th>Policy Year</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Aggregate Claim</td>
<td>8,000</td>
<td>11,000</td>
<td>15,000</td>
<td>?</td>
</tr>
<tr>
<td></td>
<td>size of group</td>
<td>40</td>
<td>50</td>
<td>70</td>
<td>75</td>
</tr>
<tr>
<td>1</td>
<td>Aggregate Claim</td>
<td>20,000</td>
<td>24,000</td>
<td>19,000</td>
<td>?</td>
</tr>
<tr>
<td></td>
<td>size of group</td>
<td>100</td>
<td>120</td>
<td>115</td>
<td>95</td>
</tr>
</tbody>
</table>

Estimate the aggregate claim amount to be observed during the fourth year for each group.

**Solution.**

**Group 1.** Exposure measures

$m_{11} = 40$ , $m_{12} = 50$ , $m_{13} = 70$.

$m_1 = 40 + 50 + 70 = 160$

Average claim amounts:

$$x_{11} = \frac{8,000}{40} = 200 \quad x_{12} = \frac{11,000}{50} = 220 \quad x_{13} = \frac{15,000}{70} = 214.29$$

$$\bar{x}_1 = \frac{8,000 + 11,000 + 15,000}{160} = 212.50$$
Group 2. Exposure measures

\( m_{21} = 100 \), \( m_{22} = 120 \), \( m_{23} = 115 \).

\[ m_2 = 100 + 120 + 115 = 335 \]

Average claim amounts:

\[ x_{21} = \frac{20,000}{100} = 200 \quad x_{22} = \frac{24,000}{120} = 200 \quad x_{23} = \frac{19,000}{115} = 165.22 \]

\[ \bar{x}_2 = \frac{20,000 + 24,000 + 19,000}{335} = 188.06 \]

Overall exposure units for the first three years:

\[ m = m_1 + m_2 = 160 + 335 = 495 \]

Estimate for overall mean:

\[ \hat{\mu} = \bar{x} = \frac{m_1 \bar{x}_1 + m_2 \bar{x}_2}{m} = \frac{(160)(212.50) + (335)(188.06)}{495} = 195.96 \]

Estimate of the EPV:

\[ \hat{\text{EPV}} = \frac{\sum_{i=1}^{2} \sum_{j=1}^{3} m_{ij} (x_{ij} - \bar{x}_i)^2}{\sum_{i=1}^{2} (3 - 1)} \]

\[ = \frac{40(200 - 212.5)^2 + 50(212 - 212.5)^2 + 70(214.29 - 212.5)^2 + 100(200 - 188.06)^2 + \cdots}{2 + 2} \]

\[ = 25160.58 \]

\[ \sum_{i=1}^{r} m_i (\bar{X}_i - \bar{X})^2 = (160)(212.5 - 195.96)^2 + 335(188.06 - 195.96)^2 = 64678.806 \]
\[
\frac{m}{m^2 - \sum m_i^2} = \frac{1}{m - \frac{1}{m} \sum m_i^2} = \frac{1}{495 - \frac{1}{495} \left\{ (160)^2 + (335)^2 \right\}} = 0.0046
\]

\[
\hat{VHM} = \frac{m}{m^2 - \sum m_i^2} \left\{ \sum_{i=1}^{\ell} m_i (\bar{X}_i - \bar{X})^2 - (r - 1) \hat{EPV} \right\} = 0.0046 \left\{ 64678.806 - (1)(25160.58) \right\}
\]

\[= 182.48 \]

\[
\hat{k} = \frac{\hat{EPV}}{\hat{VHM}} = \frac{25160.58}{182.48} = 137.88
\]

the credibility factors for the two policyholders:

\[
\hat{z}_1 = \frac{m_1}{m_1 + \hat{k}} = \frac{160}{160 + 137.88} = 0.537
\]

\[
\hat{z}_2 = \frac{m_2}{m_2 + \hat{k}} = \frac{335}{335 + 137.88} = 0.708
\]

Bühlmann-Straub estimates of the average claim amounts per exposure unit:

\[
\hat{Z}_1 \bar{X}_1 + (1 - \hat{Z}_1) \bar{X} = (0.537)(212.50) + (0.463)(195.96) = 204.84
\]

\[
\hat{Z}_2 \bar{X}_2 + (1 - \hat{Z}_2) \bar{X} = (0.708)(188.06) + (0.292)(195.96) = 190.37
\]

The aggregate claim amount for each of the two groups:

\[
m_{1,4} \left( \hat{Z}_1 \bar{X}_1 + (1 - \hat{Z}_1) \bar{X} \right) = (75)(204.84) = 15363.00
\]

\[
m_{2,4} \left( \hat{Z}_2 \bar{X}_2 + (1 - \hat{Z}_2) \bar{X} \right) = (95)(190.37) = 18085.15
\]

**Example.** The aggregate claim amount for two groups over three years are given in the following table:
<table>
<thead>
<tr>
<th>Group \downarrow</th>
<th>Policy Year</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Aggregate Claim</td>
<td>——</td>
<td>11,000</td>
<td>15,000</td>
<td>?</td>
</tr>
<tr>
<td>size of group</td>
<td>——</td>
<td>50</td>
<td>70</td>
<td>75</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Aggregate Claim</td>
<td>20,000</td>
<td>24,000</td>
<td>19,000</td>
<td>?</td>
</tr>
<tr>
<td>size of group</td>
<td>100</td>
<td>120</td>
<td>115</td>
<td>95</td>
<td></td>
</tr>
</tbody>
</table>

Estimate the aggregate claim amount to be observed during the fourth year for each group.

**Solution.**

Note that there is no data available for policyholder 1 for the first year, so the calculations would start like this:

\[ m_{11} = 50, \; m_{12} = 70. \]

\[ m_1 = 50 + 70 = 120 \]

Average claim amounts:

\[ x_{11} = \frac{11,000}{50} = 220 \]

\[ x_{12} = \frac{15,000}{70} = 214.29 \]

\[ \bar{x}_1 = \frac{11,000 + 15,000}{120} = 216.67 \]

students will do the rest.

**Note.** In some situations we might have \( \sqrt{VHM} \leq 0 \). In this case we set \( \sqrt{VHM} \leq 0 \) which then results in \( k = +\infty \) and then \( \hat{Z} = 0 \).

**Example (from the Dean’s notes - page 20).** Two risks were selected at random from a population.

Risk 1 had 0 claims in year one, 3 claims in year two, and 0 claims in year three: \((0, 3, 0)\). The claims
by year for Risk 2 were (2, 1, 2). In this case, R = 2 and N = 3. Use the Buhlmann’s model to estimate the expected number of claims per year for each risk for the fourth year.

Solution.

\[
\begin{align*}
\bar{x}_1 &= \frac{0 + 3 + 0}{3} = 1 \\
\bar{x}_2 &= \frac{2 + 1 + 2}{3} = \frac{5}{3} \\
\bar{x} &= \frac{1 + \left(\frac{5}{3}\right)}{2} = \frac{4}{3}
\end{align*}
\]

\[
\begin{align*}
\hat{\sigma}_1^2 &= \frac{(0-1)^2 + (3-1)^2 + (0-1)^2}{3-1} = 3 \\
\hat{\sigma}_2^2 &= \frac{(2-\frac{5}{3})^2 + (1-\frac{5}{3})^2 + (2-\frac{5}{3})^2}{3-1} = \frac{1}{3}
\end{align*}
\]

\[
\hat{EPV} = \frac{\hat{\sigma}_1^2 + \hat{\sigma}_2^2}{2} = \frac{3 + \frac{1}{3}}{2} = \frac{5}{3}
\]

\[
\hat{VHM} = \frac{1}{2-1} \left\{ (1 - \frac{4}{3})^2 + \left(\frac{5}{3} - \frac{4}{3}\right)^2 \right\} - \frac{5}{3} = \frac{-1}{3}
\]

this happened to be negative, so we make it zero

Then

\[\hat{Z} = 0\]