Individual Stochastic Loss Reserving

Mathieu Pigeon

UCL
Property Casualty Insurance

- Personal lines
  - Homeowners – theft, burglary, fire, wind, water
  - Auto – first-party physical damage, third-party liability

- Commercial lines
  - Commercial property – fire, wind, business interruption
  - General liability – premises and operations, products
  - Workers compensation – injury to employees
  - Professional liability – medical malpractice

- Reinsurance
The jobs of P&C actuaries

- Pricing and ratemaking
  - Determine what rates (or premiums) to charge for insurance.
  - Underwriting models are used for risk selection.
  - Fundamental insurance equation:
    \[
    \text{Premium} = \text{Losses} + \text{Expenses} + \text{Profit}.
    \]

- Actuarial loss reserving
  - Determine the outstanding liability for policies issued in the past.
  - An estimate of the outstanding liability needs to be recorded in the annual statement and it represents the largest liability amount on the balance sheet.
Actuarial Science and Statistical Models

Source: adapted from www.actuaryzhang.com
### Time Line

**Occurrence**
- Notification

**Loss payments**
- $t_1$
- $t_2$
- $t_3$
- $t_4$
- $t_5$

**Closure**
- $t_6$

- IBNR
- RBNP
- RBNS
Claim Evolution

![Diagram showing claim evolution over years with claim numbers on the y-axis and years on the x-axis.]
Claim Evolution

Evaluation date: January 1st 2003
Notation

- Let $i$ be the occurrence period ($i = 1, \ldots, I$).
- Let $j$ be the development period ($j = 1, \ldots, J$).
- Let $X_{ij}$ be the incremental claim amount at the end of development period $j$ of the occurrence period $i$.
- Let $C_{ij}$ be the cumulative claim amount at the end of development period $j$ of the occurrence period $i$.
- Let $C_{ij} = \sum_{k=1}^{j} X_{ik}$. 
### A Triangular Point of View

<table>
<thead>
<tr>
<th>Occurrence period</th>
<th>Development period</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (1997)</td>
<td>$C_{11}$ $C_{12}$ $C_{13}$ $C_{14}$ $C_{15}$ $C_{16}$</td>
</tr>
<tr>
<td>2 (1998)</td>
<td>$C_{21}$ $C_{22}$ $C_{23}$ $C_{24}$ $C_{25}$</td>
</tr>
<tr>
<td>3 (1999)</td>
<td>$C_{31}$ $C_{32}$ $C_{33}$ $C_{34}$</td>
</tr>
<tr>
<td>4 (2000)</td>
<td>$C_{41}$ $C_{42}$ $C_{43}$ IBNR +</td>
</tr>
<tr>
<td>5 (2001)</td>
<td>$C_{51}$ $C_{52}$ RBNP +</td>
</tr>
<tr>
<td>6 (2002)</td>
<td>$C_{61}$ RBNS</td>
</tr>
</tbody>
</table>

**Table:** Cumulative claims amounts
Main hypotheses:

- $E\left[ \frac{C_{i(j+1)}}{C_{ij}} \mid C_{i1}, \ldots, C_{1j} \right] = \lambda_j$;

- $\text{Var}\left[ \frac{C_{i(j+1)}}{C_{ij}} \mid C_{i1}, \ldots, C_{1j} \right] = \frac{\sigma_j^2}{w_{ij}C_{ij}^\alpha}$, with $w_{ij} \in [0, 1]$ and $\alpha \in \{0, 1, 2\}$;

and

- $\{C_{i1}, \ldots, C_{in}\}$ and $\{C_{i'1}, \ldots, C_{i'n}\}$ are independent if $i \neq i'$.
Collective Stochastic Models: Issues

- Lots of information about claims (status, policy, policy holder, past development process) remains unused in collective model.
- The heterogeneity of the claim process is misrepresented (different treatment of small and large claims).
- Instability in predictions for recent occurrence periods.
General Individual Model

- Claims arrive at random moments $0 < M_k$, $k = 1, 2, \ldots$
- The $k$th claim causes a cluster of payments from the insurer to the insured.
- The $j$th payment for the $k$th claim is a positive r.v. $X_{kj}$ executed at time

$$M_{kj} = M_k + \sum_{t=1}^{j} M_{kt}, \quad 1 \leq j \leq U_k,$$

where $(M_{kt})_{t \geq 1}$ is a sequence of positive r.v. and $U_k$ is a positive integer-valued r.v.
Each arrival time is marked by the random element

\[ ((M_{kt})_{t \geq 1}, (X_{kt})_{t \geq 1}, U_k), \quad k = 1, 2, \ldots \]

taking values in

\[ (0, \infty)^\infty \times (0, \infty)^\infty \times \mathbb{N}. \]
The individual link ratios are given by

$$\lambda_j^{(k)} = \frac{\sum_{t=1}^{j+1} X_{kt}}{\sum_{t=1}^{j} X_{kt}},$$

for $j = 1, \ldots, U_k - 1$.

The individual final amount is given by

$$X_k^{TOT} = X_{k1}\lambda_1^{(k)}\lambda_2^{(k)} \cdots \lambda_{U_k-1}^{(k)}.$$
Model I

- The r.v. $M_{kj}$ is decomposed as follow

$$M_{kj} = M_k + T_k + Q_k + \sum_{t=1}^{j-1} N_{kt}, \quad 1 \leq j \leq U_k,$$

with distributions $f_1(t; \nu)$, $f_2(q; \psi)$, $f_3(u; \beta)$ and $f_4(n; \phi)$.

- The number of claims that occurred in occurrence period $i$ is given by

$$K_i = \sum_k I_{\{M_k=i\}}$$

and is supposed to follow a Poisson process with occurrence measure $\theta w(i)$ where $w(i)$ denote the exposure measure for the occurrence period $i$. 

Model I

- For each claim $k$, the logarithm of the development vector

$$\Lambda_{u_k+1}^{(k)} = \left[ \ln(X_{k1}) \quad \ln(\lambda_{1}^{(k)}) \quad \ldots \quad \ln(\lambda_{u_k}^{(k)}) \right]$$

is supposed to have a Multivariate Skew Symmetric Distribution.
Multivariate Skew Symmetric Distributions

A random vector \( \mathbf{x} \) \((k \times 1)\) has Multivariate Skew Symmetric distribution if it has pdf

\[
\text{MSS} \left( \mathbf{x}; \mu, \Sigma^{1/2}, \Delta \right) = \frac{2^k}{\det(\Sigma)^{1/2}} g^{*} \left( \Sigma^{-1/2} (\mathbf{x} - \mu) \right) \\
\times \prod_{j=1}^{k} H \left( \delta_j \mathbf{e}_j' \Sigma^{-1/2} (\mathbf{x} - \mu) \right),
\]

where \( \mu \in \mathbb{R}^k \), \( \Sigma \) be a \((k \times k)\) positive definite symmetric matrix and \( \Delta = [\delta_1 \ldots \delta_k]' \). \( g^{*}(\mathbf{x}) = \prod_{j=1}^{k} g(x_j) \) with \( g(\cdot) \) a pdf symmetric around 0 and \( H(\cdot) \) a cdf function with \( H'(\cdot) \) symmetric around 0.
Likelihood Function

- For completed claims \((C)\) likelihood function is given by

\[
L^C \propto \prod_k \text{MSS} \left( \Lambda_{u_k+1}; \mu_{u_k+1}, \Sigma^{1/2}_{u_k+1}, \Delta_{u_k+1} \mid u_{ik} \right)
\times f_1(t_k; \nu \mid T_k \leq t^*_k - 1) f_2(q_k; \psi \mid Q_k \leq t^*_k - t_k - 1)
\times f_3(u_k; \delta \mid U_k \leq t^*_k - q_k - t_k - 1).
\]

- For Reported But Not Settled claims \((RBNS)\), the likelihood expression is

\[
L^{RBNS} \propto \prod_k \text{MSS} \left( \Lambda_{u^*_k+1}; \mu_{u^*_k+1}, \Sigma^{1/2}_{u^*_k+1}, \Delta_{u^*_k+1} \mid u^*_k \right)
\times f_1(t_k; \nu \mid T_k \leq t^*_k - 1) f_2(q_k; \psi \mid Q_k \leq t^*_k - t_k - 1)
\times (1 - F_3(u^*_k - 1; \beta)).
\]
Finally, for Reported But Not Paid claims ($RBNP$), the likelihood function is

$$L^{RBNP} \propto \prod_k f_1(t_k; \nu | T_k \leq t^*_k - 1)(1 - F_2(t^*_k - t_k - 1; \psi)).$$
Model I - Multivariate Skew Normal Case

Under MSN assumption, the $IBNR + RBNP$ reserve is given by

\[
E\left[ TOT^{IBNR+RBNP} \right] = (E[K] + OC) \\
\times E \left[ 2^{U+1} \exp \left( t_1' \mu_{U+1} + 0.5 t_1' \Sigma_{U+1}^{1/2} \left( \Sigma_{U+1}^{1/2} \right)' t_1 \right) \right] \\
\times \prod_{j=1}^{U+1} \Phi \left( \frac{\delta_j \left( \left( \Sigma_{U+1}^{1/2} \right)' t_1 \right)}{\sqrt{1 + \delta_j^2}} \right) \right]_U
\]

with the $((U + 1) \times 1)$ vector $t_n = [n \ n \ \ldots \ n]'$, $K = \sum_{i=1}^{l} K_i$ and $OC$ is the number of claims reported but not paid in the database.
Model I - Multivariate Skew Normal Case

- Under MSN assumption, the RBNS reserve is given by (without subscript/superscript for simplicity)

\[
E \left[ \text{TOT}^{RBNS} \bigg| \textbf{l} \right] = \sum x_1 l_1 \ldots l_{u_1-1} \\
\times E \left[ 2^U \exp \left( h_1^* \mu_{U+1}^* + 0.5h_1^* \left( \Sigma_{U+1}^* \right)^{1/2} \left( \left( \Sigma_{U+1}^* \right)^{1/2} \right)' h_1 \right) \right] \\
\times \prod_{j=1}^{U} \Phi \left( \frac{\delta_j^* \left( \left( \Sigma_{U+1}^* \right)^{1/2} \right)' h_1}{\sqrt{1 + \left( \delta_j^* \right)^2}} \right)_{U \geq u_1-1}
\]

where \( \textbf{l} \) denote the observed information for all claim, \( \textbf{h}_n = [n \ n \ \ldots \ n]' \), \( \mu_{U+1}^*, \Sigma_{U+1}^* \) and \( \Delta_{U+1}^* \) are parameters of a MSN conditional distribution and \( u_1 \) is the length of observed part.
Data set

- The data set used in this example is a random sample from a large database which comes from an European insurance company.
- It contains 5,000 claims related to bodily injury (BI) from 1997 to 2009 with annual information.
- The effect of inflation has been removed.
## A Triangular Point of View

<table>
<thead>
<tr>
<th>Arrival year</th>
<th>Development year 1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1997</td>
<td>261</td>
<td>614</td>
<td>359</td>
<td>526</td>
<td>546</td>
<td>137</td>
<td>130</td>
<td>339</td>
</tr>
<tr>
<td>1998</td>
<td>202</td>
<td>473</td>
<td>307</td>
<td>336</td>
<td>269</td>
<td>56</td>
<td>179</td>
<td>78</td>
</tr>
<tr>
<td>1999</td>
<td>238</td>
<td>569</td>
<td>393</td>
<td>270</td>
<td>249</td>
<td>286</td>
<td>132</td>
<td>97</td>
</tr>
<tr>
<td>2000</td>
<td>237</td>
<td>557</td>
<td>429</td>
<td>496</td>
<td>406</td>
<td>365</td>
<td>247</td>
<td>275</td>
</tr>
<tr>
<td>2001</td>
<td>389</td>
<td>628</td>
<td>529</td>
<td>559</td>
<td>446</td>
<td>375</td>
<td>147</td>
<td>239</td>
</tr>
<tr>
<td>2002</td>
<td>260</td>
<td>570</td>
<td>533</td>
<td>444</td>
<td>132</td>
<td>122</td>
<td>332</td>
<td>1,082</td>
</tr>
<tr>
<td>2003</td>
<td>236</td>
<td>743</td>
<td>558</td>
<td>237</td>
<td>217</td>
<td>205</td>
<td>171</td>
<td></td>
</tr>
<tr>
<td>2004</td>
<td>248</td>
<td>794</td>
<td>401</td>
<td>236</td>
<td>254</td>
<td>98</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table:** Cumulative claims amounts (in thousands)
Estimation Results

The parameter of the distribution of the random variable $K$ (occurrence of claim) is $\hat{\theta}_{BI} = 0.7445$ (s.e. 0.02).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>N</th>
<th>s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>5.9226</td>
<td>(0.03)</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>1.3968</td>
<td>(0.02)</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table: Estimation results for the logarithm of the severity of the first payment in case there is only one payment.
## Estimation Results

\[
f_1(t; \nu) = \sum_{i=0}^{p} \nu_i l_i(t) + \left( 1 - \sum_{i=0}^{p} \nu_i \right) f_{T|T>p}(t),
\]

where \( l_i = 1 \) for the \( i^{th} \) period after occurrence time and \( l_i = 0 \) otherwise.

<table>
<thead>
<tr>
<th>( i )</th>
<th>Report delay</th>
<th>First pmt delay</th>
<th>Number partial pmt</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.8953 (&lt; 0.001)</td>
<td>0.7127 (&lt; 0.001)</td>
<td>0.6177 (&lt; 0.001)</td>
</tr>
<tr>
<td>1</td>
<td>0.0819 (0.003)</td>
<td>0.2522 (0.003)</td>
<td>0.2611 (0.003)</td>
</tr>
<tr>
<td>2</td>
<td>0.5144 (0.064)</td>
<td>0.6431 (0.052)</td>
<td>0.5461 (0.046)</td>
</tr>
<tr>
<td>( p = 1 )</td>
<td>( p = 1 )</td>
<td>( p = 1 )</td>
<td></td>
</tr>
</tbody>
</table>

**Table:** Estimation results for a Geometric distribution, combined with degenerate components.
## Estimation Results

<table>
<thead>
<tr>
<th>Loc. par. ($\mu$)</th>
<th>Sc. par. ($\Sigma_c^{1/2}$)</th>
<th>Sh. par. ($\Delta$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_1 = 5.77$</td>
<td>$\sigma_1 = 1.34$</td>
<td>$\delta_1 = 0.30$</td>
</tr>
<tr>
<td>(0.05)</td>
<td>$\sigma_2 = 1.28$</td>
<td>$\delta_2 = 2.64$</td>
</tr>
<tr>
<td>$\mu_2 = 0.50$</td>
<td>$\sigma_3 = 1.11$</td>
<td>$\delta_3 = 2.29$</td>
</tr>
<tr>
<td>(0.03)</td>
<td>$\sigma_4 = 1.09$</td>
<td>$\delta_4 = 7.37$</td>
</tr>
<tr>
<td>$\mu_3 = 0.56$</td>
<td>$\sigma_5 = 0.73$</td>
<td>$\delta_5 = -0.01$</td>
</tr>
<tr>
<td>(0.05)</td>
<td>$\rho_1 = -0.32$</td>
<td></td>
</tr>
<tr>
<td>$\mu_4 = 0.19$</td>
<td>$\rho_2 = -0.20$</td>
<td></td>
</tr>
<tr>
<td>(0.06)</td>
<td>$\rho_3 = -0.01$</td>
<td></td>
</tr>
<tr>
<td>$\mu_5 = 1.17$</td>
<td>$\rho_4 = -0.30$</td>
<td></td>
</tr>
<tr>
<td>(0.13)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table:** Estimation results for logarithms of development factors for claims with more than one period with payment.
## Prediction Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Class</th>
<th>Expected value</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>CL model</td>
<td>TOTAL</td>
<td>9,082,114</td>
<td>1,184,546</td>
</tr>
<tr>
<td>Ind. MSN (theo.)</td>
<td>IBNR + RBNP</td>
<td>3,187,306</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RBNS</td>
<td>4,978,684</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TOTAL</td>
<td>8,165,991</td>
<td></td>
</tr>
<tr>
<td>Ind. MSN (sim.)</td>
<td>IBNR + RBNP</td>
<td>3,205,949</td>
<td>1,180,077</td>
</tr>
<tr>
<td></td>
<td>RBNS</td>
<td>5,009,486</td>
<td>1,673,145</td>
</tr>
<tr>
<td></td>
<td>TOTAL</td>
<td>8,215,435</td>
<td>2,034,198</td>
</tr>
<tr>
<td>Ind. MSN (sim. + unc.)</td>
<td>TOTAL</td>
<td>8,199,629</td>
<td>2,139,896</td>
</tr>
</tbody>
</table>

**Table:** Prediction results. Observed value is 7,684,000.
Prediction Results

BI: Total claims reserve

[Graph showing distribution of total claims reserve]
Time Line: Claim and Reserve Processes

- Occurrence
  - Notification
  - $t_1$, $t_2$

- Loss payments
  - $t_3$, $t_4$, $t_5$
  - Adjustments

- Closure
  - $t_{5.5}$, $t_6$
  - Closure

Initial Reserve
Claim and Reserve Processes

Ex. 1

Ex. 2
Matrix Variate Skew Symmetric Distributions

A random matrix $\mathbf{x}$ ($k \times n$) has matrix variate skew symmetric distribution if it has pdf

$$MVSS(\mathbf{x}; \mu, A, B, \Delta) = \frac{2^{nk}}{\det(A)^n \det(B)^k} g^* \left( (\mathbf{A}^{-1}(\mathbf{x} - \mu)\mathbf{B}^{-1}) \right)$$

$$\times \prod_{i=1}^{n} \prod_{j=1}^{k} H \left( \delta_{ji} \mathbf{e}'_j (\mathbf{A}^{-1}(\mathbf{x} - \mu)\mathbf{B}^{-1}) \mathbf{c}_i \right),$$

where $\mu \in \mathbb{R}^{k \times n}$, $A$ and $B$ are nonsingular definite symmetric matrix (order $k$ and $n$) and $\delta_{ji}$ are real scalars. $g^*(\mathbf{x}) = \prod_{i=1}^{n} \prod_{j=1}^{k} g(x_{ij})$ with $g(\cdot)$ a pdf symmetric around 0 and $H(\cdot)$ a cdf function with $H'(\cdot)$ symmetric around 0.
Discussion

- Refine the individual model by including dependence between some elements ($T$, $Q$, $X$, etc.).
- Inclusion of inflation in the model.
- Joint modelisation of the claim process and reserve process.


Thank you

Models are to be used, but never believed.
(Henri Thiel)