Risk model with dependent frequency and severity for Liability and Housing Insurance

Renata Gomes Alcoforado\textsuperscript{1,2} & Alfredo D. Egídio dos Reis\textsuperscript{2}

\textsuperscript{1}UNIVERSIDADE FEDERAL DE PERNAMBUCO
\textsuperscript{2}CEMAPRE & ISEG - UNIVERSIDADE DE LISBOA
Summary

1. Introduction
   - Motivation
   - Objectives

2. Dependent Risks in Real Data
   - Real Data
   - Distribution Adequacy
   - Dependence
   - GAMLSS Modelling
     - Median Severity of Claims per Day
     - Mean Severity of Claims per Day

3. Final Remarks
Introduction - Common Assumption

The random variables $N(t)$ and $X_i$, are independent, in which

- $N(t)$ is the number of claims up to time $t$
- $X_i$ is the severity of claim $i$
Motivation

Examples

- Car insurance claims with a €200 franchise;
- Drivers who file several claims per year are typically involved in minor accidents;
- Home insurance claims due to sewer backup of flooding tend to be both large and frequent in problematic neighborhoods (Garrido et al., 2016).

Need to adapt the aggregate claims model to account for potential association between claim frequency and severity.
Objectives

The main goals of this paper are:

- Spot dependence between claims frequency and severity in the same group in the insurer’s portfolio and also across groups;
- Analyse this dependence and observe how the frequency affects the severity of claims.
The data consists in two different branches that we work as three groups:

- The total portfolio - with all data of the company
- Group 1 - liability insurance
- Group 2 - housing insurance
Real Data

- Period: Every claim that happened in the period of 01/01/2015 until 31/12/2019;
- Frequency of claims: for each group represents the number of claims per day during this five year period (1826 days);
- Deflation: the severity of claims were deflated using inflation data from the World Bank and were all deflated for the first day of the period in study (01/01/2015)
Real Data

<table>
<thead>
<tr>
<th></th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Severity</strong></td>
<td>0.0194</td>
<td>5820</td>
<td>14910</td>
<td>28820</td>
<td>33270</td>
<td>569918</td>
</tr>
<tr>
<td><strong>Frequency</strong></td>
<td>0</td>
<td>5</td>
<td>7</td>
<td>8.579</td>
<td>10</td>
<td>103</td>
</tr>
<tr>
<td><strong>Severity 1</strong></td>
<td>1.35</td>
<td>4347.08</td>
<td>7897.25</td>
<td>11972.21</td>
<td>12711.83</td>
<td>315022.59</td>
</tr>
<tr>
<td><strong>Frequency 1</strong></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.529</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td><strong>Severity 2</strong></td>
<td>0.0194</td>
<td>4381</td>
<td>12125</td>
<td>22884</td>
<td>26582</td>
<td>462884</td>
</tr>
<tr>
<td><strong>Frequency 2</strong></td>
<td>0</td>
<td>5</td>
<td>7</td>
<td>8.05</td>
<td>9</td>
<td>101</td>
</tr>
</tbody>
</table>

Table 1: Summary of the variables claim Severity and Frequency
Boxplot for the Frequency and Severity of Claims

- **Introduction**
- **Motivation**
- **Objectives**
- **Dependent Risks in Real Data**
- **Real Data Distribution Adequacy Dependence**
- **GAMLSS Modelling**
- **Median Severity of Claims per Day**
- **Mean Severity of Claims per Day**
- **Final Remarks**
- **References**
Poisson Inverse Gaussian was used in Tremblay (1992) to fit the number of car accidents in order to build a bonus-malus system that minimizes the insurer’s risk. In addition, Denuit et al. (2007) said that this distribution is an ideal candidate for modelling positive, right-skewed data.
Distribution Adequacy: Total Portfolio - Severity

Weibull distribution with parameters, shape = 0.790821 and scale = $1.125513 \times 10^4$ (p-value = 0.1443)
Distribution Adequacy: Total Portfolio - Frequency

Poisson Inverse Gaussian with parameters mean = 8.578861 and shape = 2.117495 (p-value = 0.2283)
Distribution Adequacy: Group 1 - Severity

LogNormal \((\text{meanlog} = 8.76, \text{sdlog} = 1.119992)\) and \(p\)-value 0.3678
Distribution Adequacy: Group 1 - Frequency

Negative Binomial \((\mu = 0.5290252, \theta = 1.716889)\) with \(p\)-value equal to 0.9096
Weibull with parameters, shape $= 7.92 \times 10^{-1}$ and scale $= 1.98 \times 10^4$, and $p$-value $= 0.1904$
Distribution Adequacy: Group 2 - Frequency

Poisson Inverse Gaussian with parameters (mean = 8.049836, shape = 1.922438) with $p$-value equal to 0.1368
Dependence

In this case we used the variables

- Frequency
- Severity per day
- Mean of the Severity per day
- Median of the Severity per day

Since none of the data samples were drawn from a Normal universe, we use Kendall and Spearman coefficients.
Spearman’s correlation coefficient ($\rho$) with $p$-value under 0.05
Dependence - Some additional results

- Statistically Significant correlation between Frequency and aggregate severity of claims per day in both groups;
- Correlation between Severity Day and both Median and Mean shows us that when the aggregate amount of severity per day is not only due to the increase in the Frequency but also due to the severity per claim on that day. This is valid for both groups.
GAMLSS Modelling

We model the **median** and the **mean** of severity of claims per day. We use the median in addition to the mean because the mean is a measure that is very sensitive to extreme values. So we aim to show the results from these two measurements.
GAMLSS Modelling - But first, why GAMLSS?

- We use Generalized Additive Models for Location, Scale and Shape (briefly known as GAMLSS) because it is a general framework for univariate regression type statistical problems (Stasinopoulos and Rigby, 2007);
We use Generalized Additive Models for Location, Scale and Shape (briefly known as GAMLSS) because it is a general framework for univariate regression type statistical problems (Stasinopoulos and Rigby, 2007);

It allows flexibility in specifying the distribution of the response variable and it also allows others explanatory variables to be easily included;
We use Generalized Additive Models for Location, Scale and Shape (briefly known as GAMLSS) because it is a general framework for univariate regression type statistical problems (Stasinopoulos and Rigby, 2007);

- It allows flexibility in specifying the distribution of the response variable and it also allows others explanatory variables to be easily included;

- As Denuit et al. (2019) states, in the GAMLSS approach, the exponential dispersion distribution assumption for the response is relaxed, resulting in removing the restriction that the actuarial analysis had to the distributions used in the classical GLM/GAM setting;
GAMLSS Modelling - But first, why GAMLSS?

- We use Generalized Additive Models for Location, Scale and Shape (briefly known as GAMLSS) because it is a general framework for univariate regression type statistical problems (Stasinopoulos and Rigby, 2007);
- It allows flexibility in specifying the distribution of the response variable and it also allows others explanatory variables to be easily included;
- As Denuit et al. (2019) states, in the GAMLSS approach, the exponential dispersion distribution assumption for the response is relaxed, resulting in removing the restriction that the actuarial analysis had to the distributions used in the classical GLM/GAM setting;
- In our case we have one main explanatory variable and then we add some dummies representing years of data.
GAMLSS Modelling - fitting Distribution Families

We fitted the variables in order to know which GAMLSS family is best fit for each sample.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Family</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>Delaporte</td>
<td>10275.7</td>
</tr>
<tr>
<td>Severity</td>
<td>Box-Cox Power Exponential</td>
<td>342791</td>
</tr>
<tr>
<td>Median_Severity_Day 1</td>
<td>Exponential Gaussian</td>
<td>38092.5</td>
</tr>
<tr>
<td>Mean_Severity_Day 1</td>
<td>Box-Cox-(t)</td>
<td>39281.1</td>
</tr>
<tr>
<td>Frequency 1</td>
<td>Double Poisson</td>
<td>3592.03</td>
</tr>
<tr>
<td>Severity 1</td>
<td>Box-Cox Power Exponential</td>
<td>19506.7</td>
</tr>
<tr>
<td>Median_Severity_Day 2</td>
<td>Pareto Type 2</td>
<td>-58173.4</td>
</tr>
<tr>
<td>Mean_Severity_Day 2</td>
<td>Box-Cox Power Exponential</td>
<td>13430.9</td>
</tr>
<tr>
<td>Frequency 2</td>
<td>Delaporte</td>
<td>10079.5</td>
</tr>
<tr>
<td>Severity 2</td>
<td>Generalized beta 2 (i.e. of the second kind)</td>
<td>322628</td>
</tr>
<tr>
<td>Median_Severity_Day 3</td>
<td>Exponential Gaussian</td>
<td>38397.7</td>
</tr>
<tr>
<td>Mean_Severity_Day 3</td>
<td>Box-Cox-(t)</td>
<td>39429.9</td>
</tr>
</tbody>
</table>
Using Exponential Gaussian family of distributions, we have three GAMLSS models.

<table>
<thead>
<tr>
<th>Link Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>For location ($\mu$) is the identity;</td>
</tr>
<tr>
<td>For Scale ($\sigma$) is the “log function”;</td>
</tr>
<tr>
<td>And for Shape ($\nu$) is the “log function”.</td>
</tr>
</tbody>
</table>
GAMLSS Modelling Median Severity - Total Portfolio

<table>
<thead>
<tr>
<th>Link function: identity</th>
<th>( \mu ) coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>5758.22</td>
</tr>
<tr>
<td>freq</td>
<td>-79.91</td>
</tr>
<tr>
<td>d.2015</td>
<td>-924.26</td>
</tr>
<tr>
<td>d.2016</td>
<td>-863.89</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>t value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>20.598</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td>-3.124</td>
<td>0.00181</td>
</tr>
<tr>
<td>-2.395</td>
<td>0.01674</td>
</tr>
<tr>
<td>-2.143</td>
<td>0.03221</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Link function: log</th>
<th>( \sigma ) coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>7.80849</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Std Error</th>
<th>t value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.06537</td>
<td>119.4</td>
<td>&lt;2e-16</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Link function: log</th>
<th>( \nu ) coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>9.20654</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Std Error</th>
<th>t value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.02985</td>
<td>308.4</td>
<td>&lt;2e-16</td>
</tr>
</tbody>
</table>
GAMLSS Modelling Median Severity - Group 1

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std Error</th>
<th>t value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>µ coefficients</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>940.5</td>
<td>395.9</td>
<td>2.376</td>
<td>0.017799</td>
</tr>
<tr>
<td>freq</td>
<td>985.2</td>
<td>207.3</td>
<td>4.753</td>
<td>2.46E-06</td>
</tr>
<tr>
<td>d.2015</td>
<td>1873.7</td>
<td>482.1</td>
<td>3.887</td>
<td>0.000112</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>σ coefficients</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>7.54157</td>
<td>0.07961</td>
<td>&lt;2e-16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9.0607</td>
<td>0.04486</td>
<td>202</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td><strong>ν coefficients</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
GAMLSS Modelling Median Severity - Group 2

<table>
<thead>
<tr>
<th>µ coefficients</th>
<th>Estimate</th>
<th>Std Error</th>
<th>t value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>5931.11</td>
<td>299.07</td>
<td>19.832</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td>freq</td>
<td>-102.39</td>
<td>28.25</td>
<td>-3.625</td>
<td>0.000297</td>
</tr>
<tr>
<td>d.2015</td>
<td>-1236.88</td>
<td>378.18</td>
<td>-3.271</td>
<td>0.001093</td>
</tr>
<tr>
<td>d.2016</td>
<td>-1100.27</td>
<td>373.64</td>
<td>-2.945</td>
<td>0.003273</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>σ link function: log</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ν link function: log</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
</tr>
</tbody>
</table>
Mean Severity of Claims per Day

Using Box-Cox-$t$ (BCTo)$(\mu, \sigma, \nu, \tau)$ Family - see Rigby and Stasinopoulos (2006).

- For location $(\mu)$ is the “log function”;
- For Scale $(\sigma)$ is the “log function”;
- For Skewness $(\nu)$ is the identity;
- And for Kurtosis $(\tau)$ is the “log function”.

Link Function
GAMLSS Modelling Mean Severity - Total Portfolio

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std Error</th>
<th>t value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>µ</strong> coefficients (Intercept)</td>
<td>9.967018</td>
<td>0.023144</td>
<td>430.644</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td>freq</td>
<td>-0.005709</td>
<td>0.002009</td>
<td>-2.841</td>
<td>0.00455</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>σ</strong> link function: log</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>σ</strong> coefficients (Intercept)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>ν</strong> link function: identity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ν</strong> coefficients (Intercept)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>τ</strong> link function: log</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>τ</strong> coefficients (Intercept)</td>
</tr>
</tbody>
</table>
GAMLSS Modelling Mean Severity - Group 1

\begin{table}[h]
\begin{tabular}{lcccc}
\hline
& \multicolumn{2}{c}{\mu, link function: log} & & \\
& \mu coefficients & Estimate & Std Error & t value & p-value \\
(Intercept) & 8.54302 & 0.05857 & 145.851 & <2e-16 \\
freq & 0.12785 & 0.02676 & 4.777 & 2.20e-06 \\
d1.2015 & 0.43243 & 0.05987 & 7.223 & 1.41e-12 \\
d1.2016 & 0.38798 & 0.06 & 6.466 & 1.96e-10 \\
d1.2017 & 0.20763 & 0.05794 & 3.583 & 0.000365 \\
\hline
\end{tabular}
\end{table}

\begin{table}[h]
\begin{tabular}{lcccc}
\hline
& \multicolumn{2}{c}{\sigma, link function: log} & & \\
& \sigma coefficients & Estimate & Std Error & t value & p-value \\
(Intercept) & -0.88981 & 0.06765 & -13.15 & <2e-16 \\
\hline
\end{tabular}
\end{table}

\begin{table}[h]
\begin{tabular}{lcccc}
\hline
& \multicolumn{2}{c}{\nu, link function: identity} & & \\
& \nu coefficients & Estimate & Std Error & t value & p-value \\
(Intercept) & 0.1809 & 0.06812 & 2.656 & 0.00811 \\
\hline
\end{tabular}
\end{table}

\begin{table}[h]
\begin{tabular}{lcccc}
\hline
& \multicolumn{2}{c}{\tau, link function: log} & & \\
& \tau coefficients & Estimate & Std Error & t value & p-value \\
(Intercept) & 0.3421 & 0.1142 & 2.996 & 0.00284 \\
\hline
\end{tabular}
\end{table}
GAMLSS Modelling Mean Severity - Group 2

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std Error</th>
<th>t value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu ) coefficients</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>10.00813</td>
<td>0.023016</td>
<td>434.836</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td>freq</td>
<td>-0.00713</td>
<td>0.002088</td>
<td>-3.414</td>
<td>0.00455</td>
</tr>
<tr>
<td>( \sigma ) coefficients</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>-0.60743</td>
<td>0.02542</td>
<td>-23.9</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td>( \nu ) coefficients</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>0.17252</td>
<td>0.03669</td>
<td>4.703</td>
<td>2.76e-06</td>
</tr>
<tr>
<td>( \tau ) coefficients</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>2.485</td>
<td>0.231</td>
<td>10.76</td>
<td>&lt;2e-16</td>
</tr>
</tbody>
</table>
Final Remarks

Regarding the distribution adequacy

- The discrete variables were fitted to Poisson-Inverse Gaussian and Negative Binomial Distribution;

Regarding the dependence calculation

- We used Kendall's and Spearman's correlation coefficients;
- We detected dependence between variables in the same groups and variables across groups.
Final Remarks

Regarding the distribution adequacy

- The discrete variables were fitted to Poisson-Inverse Gaussian and Negative Binomial Distribution;
- The continuous to Weibull and LogNormal Distribution;

Regarding the dependence calculation

- We used Kendall's and Spearman's correlation coefficients;
- We detected dependence between variables in the same groups and variables across groups.
Final Remarks

**Regarding the distribution adequacy**

- The discrete variables were fitted to Poisson-Inverse Gaussian and Negative Binomial Distribution;
- The continuous to Weibull and LogNormal Distribution;
- The Total Portfolio and the Housing Group were fitted to the same distribution - out of the 15,665 claims in our data, 14,699 are from Group 2, representing 93.83%.

**Regarding the dependence calculation**
Final Remarks

Regarding the distribution adequacy

- The discrete variables were fitted to Poisson-Inverse Gaussian and Negative Binomial Distribution;
- The continuous to Weibull and LogNormal Distribution;
- The Total Portfolio and the Housing Group were fitted to the same distribution - out of the 15,665 claims in our data, 14,699 are from Group 2, representing 93.83%.

Regarding the dependence calculation

- We used Kendall’s and Spearman’s correlation coefficients;
Final Remarks

Regarding the distribution adequacy

- The discrete variables were fitted to Poisson-Inverse Gaussian and Negative Binomial Distribution;
- The continuous to Weibull and LogNormal Distribution;
- The Total Portfolio and the Housing Group were fitted to the same distribution - out of the 15,665 claims in our data, 14,699 are from Group 2, representing 93.83%.

Regarding the dependence calculation

- We used Kendall’s and Spearman’s correlation coefficients;
- We detected dependence between variables in the same groups and variables across groups.
Final Remarks

Regarding the GAMLSS Modelling

- For the median of severity - Exponential Gaussian Family;

  For the mean - Box-Cox-t Family in all three groups;

  For the Total Portfolio and for the Housing group, the frequency of claims impacts negatively in the median/mean of severity;

  Different result from Garrido et al. (2016);

Conjectures why this might happen.
Final Remarks

Regarding the GAMLSS Modelling

- For the median of severity - Exponential Gaussian Family;
- For the mean - Box-Cox-t Family in all three groups;
Final Remarks

Regarding the GAMLSS Modelling

- For the median of severity - Exponential Gaussian Family;
- For the mean - Box-Cox-\(t\) Family in all three groups;
- For the Total Portfolio and for the Housing group, the frequency of claims impacts **negatively** in the median/mean of severity;
Final Remarks

Regarding the GAMLSS Modelling

- For the median of severity - Exponential Gaussian Family;
- For the mean - Box-Cox-\(t\) Family in all three groups;
- For the Total Portfolio and for the Housing group, the frequency of claims impacts negatively in the median/mean of severity;
- Different result from Garrido et al. (2016);
Final Remarks

Regarding the GAMLSS Modelling

- For the median of severity - Exponential Gaussian Family;
- For the mean - Box-Cox- $t$ Family in all three groups;
- For the Total Portfolio and for the Housing group, the frequency of claims impacts **negatively** in the median/mean of severity;
- Different result from Garrido et al. (2016);
- For the Liability Group, we detected an opposite effect;
Final Remarks

Regarding the GAMLSS Modelling

- For the median of severity - Exponential Gaussian Family;
- For the mean - Box-Cox-t Family in all three groups;
- For the Total Portfolio and for the Housing group, the frequency of claims impacts **negatively** in the median/mean of severity;
- Different result from Garrido et al. (2016);
- For the Liability Group, we detected an opposite effect;
- Conjectures why this might happen.
Final Remarks

Regarding Novelty
- Type of Insurance;

Regarding Future Works
Final Remarks

Regarding Novelty
- Type of Insurance;
- Results.

Regarding Future Works
Final Remarks

Regarding Novelty
- Type of Insurance;
- Results.

Regarding Future Works
- Premiums and Ruin Probability Calculations;
Final Remarks

Regarding Novelty
- Type of Insurance;
- Results.

Regarding Future Works
- Premiums and Ruin Probability Calculations;
- Delay of data.


The end... Thanks (:)

Special thanks to the anonymous insurer that provided the data.

Project CEMAPRE - UID/MULTI/00491/2019 financed by FCT/MCTES through national funds.