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

Renata Gomes Alcoforado [1,2] & Alfredo D. Egídio dos Reis [2]

[1]UNIVERSIDADE FEDERAL DE PERNAMBUCO
[2] CEMAPRE & ISEG - UNIVERSIDADE DE LISBOA

Actuarial Colloquia
22 June 2022

Summary

R. Alcoforado
A. D. Egídio dos Reis

Introduction - Common Assumption

R. Alcoforado
A. D. Egídio
dos Reis

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

R. Alcoforado
A. D. Egídio
dos Reis

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

R. Alcoforado
A. D. Egídio dos Reis

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.

Real Data

R. Alcoforado
A. D. Egídio dos Reis

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

R. Alcoforado
A. D. Egídio
dos Reis

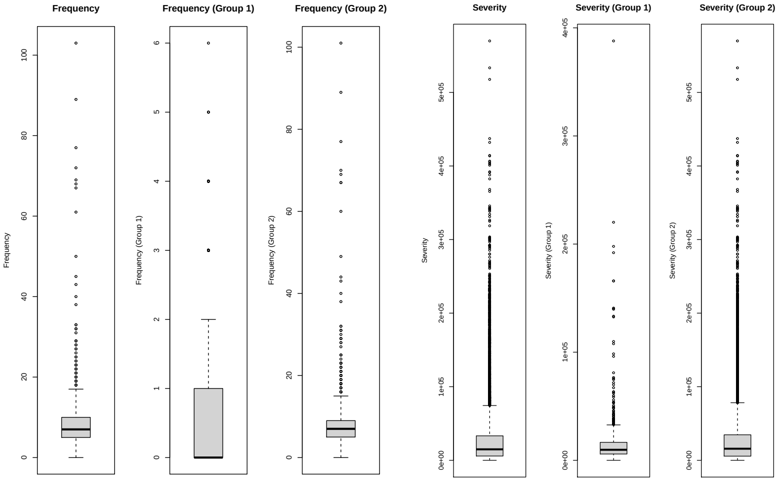
- Period: Every claim that happened in the period of 01/01/2015 until 31/12/2019;
- Frequency of claims: for each group represents the number of claims per day during this five year period (1826 days);
- Deflation: the severity of claims were deflated using inflation data from the World Bank and were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

Real Data

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

Table 1: Summary of the variables claim Severity and Frequency

Boxplot for the Frequency and Severity of Claims



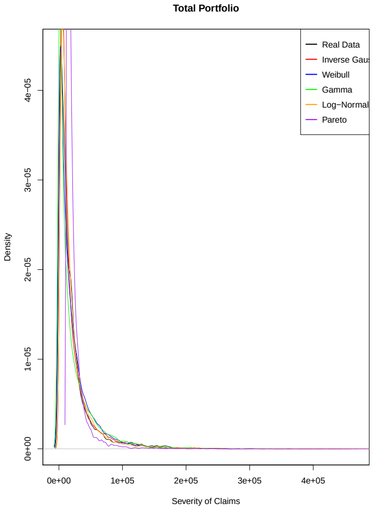
Distribution Adequacy

Tested Distributions

- For the continuous variables (severity of claims) - Log Normal, Inverse Gaussian, Gamma, Pareto and Weibull
- For the discrete variables - Poisson, Negative Binomial and a Poisson Inverse Gaussian

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.

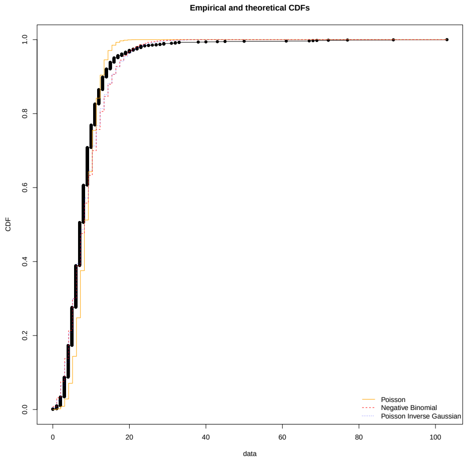
Distribution Adequacy: Total Portfolio - Severity



Weibull distribution with parameters, shape $= 0.790821$ and scale $= 1.125513 \times 10^4$ (p-value $= 0.1443$)

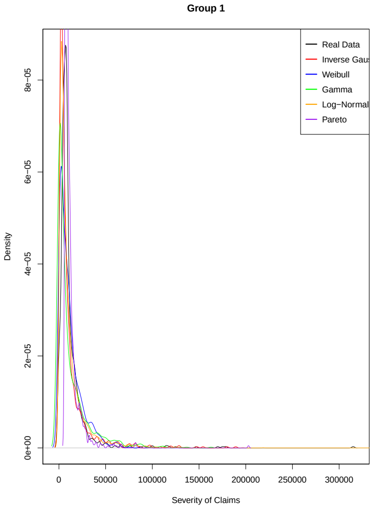
R. Alcoforado
A. D. Egídio dos Reis

Distribution Adequacy: Total Portfolio - Frequency



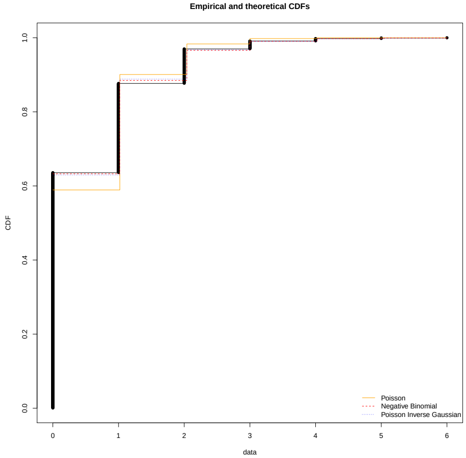
Poisson Inverse Gaussian with parameters mean $= 8.578861$ and shape $= 2.117495$ (p-value $= 0.2283$)

Distribution Adequacy: Group 1 - Severity



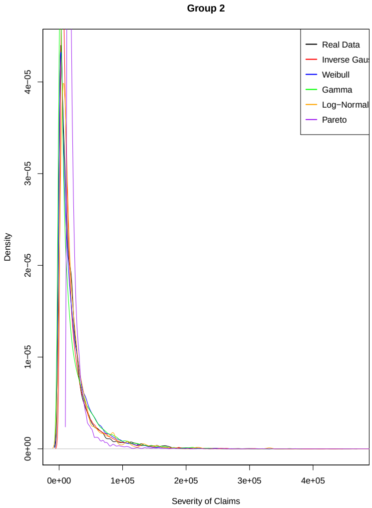
LogNormal $(meanlog = 8.76, \; sdlog = 1.119992)$ and p-value 0.3678

Distribution Adequacy: Group 1 - Frequency



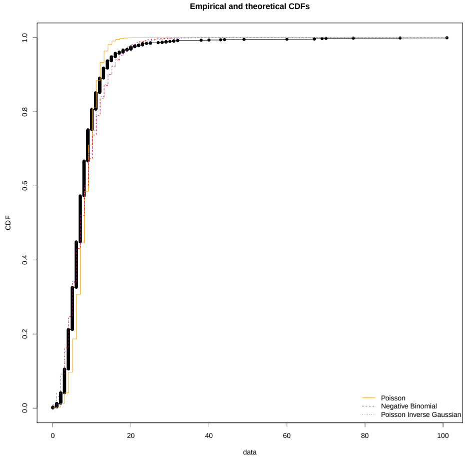
Negative Binomial ($\mu = 0.5290252, \theta = 1.716889$) with p-value equal to 0.9096

Distribution Adequacy: Group 2 - Severity



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 sample were drawn form a Normal universe we use Kendall and Spearman coefficients.

Spearman's correlation coefficient (ρ) with p-value under 0.05

R. Alcoforado
A. D. Egídio dos Reis

	Frequency 1	Severity_Day 1	Mean 1	Median 1	Frequency 2	Severity_Day 2	Mean 2	Median 2
Frequency 1	-	0.5865267	0.1574933	0.08879794	0.11289	0.06923181	-	-
Severity_Day 1	0.5865267	-	0.8737314	0.7254146	0.1242605	0.07649307	-	-
Mean_Severity_Day 1	0.1574933	0.8737314	-	0.9576098	0.125299	0.07262132	-	-
Median_Severity_Day 1	0.08879794	0.7254146	0.9576098	-	0.09635208	0.05295117	-	-
Frequency 2	0.11289	0.1242605	0.125299	0.09635208	-	0.6551026	-	-0.08434438
Severity_Day 2	0.06923181	0.07649307	0.07262132	0.05295117	0.6551026	-	0.7167361	0.2734767
Mean_Severity_Day 2	-	-	-	-	-	0.7167361	-	0.5077223
Median_Severity_Day 2	-	-	-	-	-0.08434438	0.2734767	0.5077223	-

Dependence - Some aditional results

R. Alcoforado
A. D. Egídio
dos Reis

- 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);

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;

GAMLSS Modelling - But first, why GAMLSS?

R. Alcoforado
A. D. Egídio dos Reis

- 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

R. Alcoforado
A. D. Egídio
dos Reis

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

Variable	Family	AIC
Frequency	Delaporte	10275.7
Severity	Box-Cox Power Exponential	342791
Median_Severity_Day	Exponential Gaussian	38092.5
Mean_Severity_Day	Box-Cox-t	39281.1
Frequency 1	Double Poisson	3592.03
Severity 1	Box-Cox Power Exponential	19506.7
Median_Severity_Day 1	Pareto Type 2	-58173.4
Mean_Severity_Day 1	Box-Cox Power Exponential	13430.9
Frequency 2	Delaporte	10079.5
Severity 2	Generalized beta 2 (i.e. of the second kind)	322628
Median_Severity_Day 2	Exponential Gaussian	38397.7
Mean_Severity_Day 2	Box-Cox-t	39429.9

Median Severity of Claims per Day

R. Alcoforado
A. D. Egídio
dos Reis

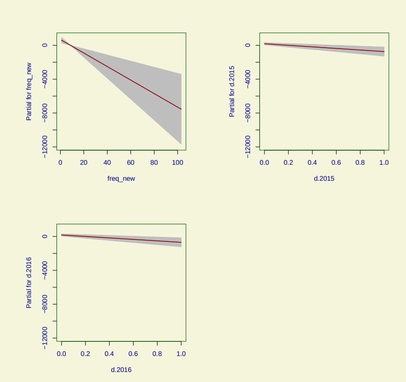
Using Exponential Gaussian family of distributions, we have three GAMLSS models.

Link Function

- For location (μ) is the identity;
- For Scale (σ) is the "log function";
- And for Shape (ν) is the "log function".

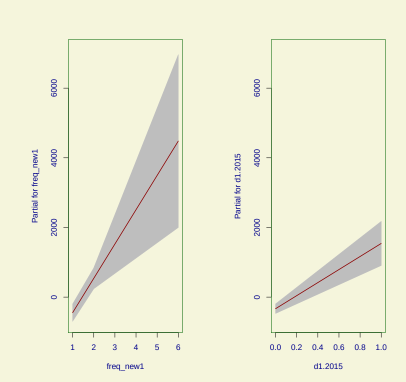
R. Alcoforado
A. D. Egídio
dos Reis

GAMLSS Modelling Median Severity - Total Portfolio



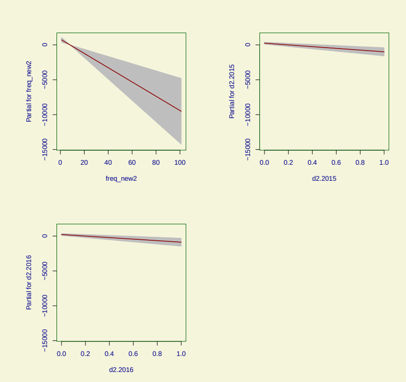
		μ link function: identity		
μ coefficients	**Estimate**	**Std Error**	**t value**	**p-value**
(Intercept)	5758.22	279.55	20.598	<2e-16
freq	-79.91	25.58	-3.124	0.00181
d.2015	-924.26	385.96	-2.395	0.01674
d.2016	-863.89	403.05	-2.143	0.03221
		σ link function:log		
σ coefficients	**Estimate**	**Std Error**	**t value**	**p-value**
(Intercept)	7.80849	0.06537	119.4	<2e-16
		ν link function:log		
ν coefficients	**Estimate**	**Std Error**	**t value**	**p-value**
(Intercept)	9.20654	0.02985	308.4	<2e-16

GAMLSS Modelling Median Severity - Group 1



μ link function: identity				
μ coefficients	**Estimate**	**Std Error**	**t value**	**p-value**
(Intercept)	940.5	395.9	2.376	0.017799
freq	985.2	207.3	4.753	2.46E-06
d.2015	1873.7	482.1	3.887	0.000112
σ link function:log				
σ coefficients	**Estimate**	**Std Error**	**t value**	**p-value**
(Intercept)	7.54157	0.07961	94.73	$<$2e-16
ν link function:log				
ν coefficients	**Estimate**	**Std Error**	**t value**	**p-value**
(Intercept)	9.0607	0.04486	202	$<$2e-16

GAMLSS Modelling Median Severity - Group 2



μ link function: identity				
μ coefficients	**Estimate**	**Std Error**	**t value**	**p-value**
(Intercept)	5931.11	299.07	19.832	$<$2e-16
freq	-102.39	28.25	-3.625	0.000297
d.2015	-1236.88	378.18	-3.271	0.001093
d.2016	-1100.27	373.64	-2.945	0.003273
σ link function:log				
σ coefficients	**Estimate**	**Std Error**	**t value**	**p-value**
(Intercept)	7.86251	0.07243	108.6	$<$2e-16
ν link function:log				
ν coefficients	**Estimate**	**Std Error**	**t value**	**p-value**
(Intercept)	9.2921	0.03035	306.1	$<$2e-16

Mean Severity of Claims per Day

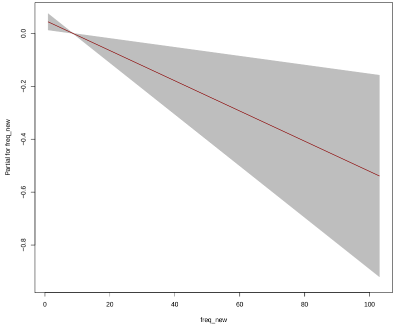
Using Box-Cox-t (BCTo)(μ, σ, ν, τ) Family - see Rigby and Stasinopoulos (2006).

Link Function

- For location (μ) is the "log function";
- For Scale (σ) is the "log function";
- For Skewness (ν) is the identity;
- And for Kurtosis (τ) is the "log function".

GAMLSS Modelling Mean Severity - Total Portfolio

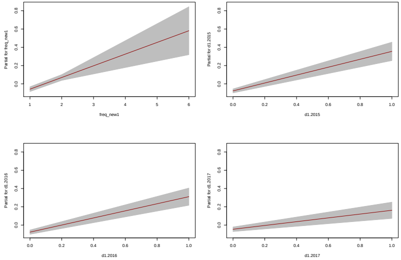
R. Alcoforado
A. D. Egídio dos Reis



		μ **link function: log**		
μ **coefficients**	**Estimate**	**Std Error**	**t value**	**p-value**
(Intercept)	9.967018	0.023144	430.644	<2e-16
freq	-0.005709	0.002009	-2.841	0.00455
		σ **link function:log**		
σ **coefficients**	**Estimate**	**Std Error**	**t value**	**p-value**
(Intercept)	-0.64357	0.02557	-25.16	<2e-16
		ν **link function: identity**		
ν **coefficients**	**Estimate**	**Std Error**	**t value**	**p-value**
(Intercept)	0.14887	0.03842	3.875	0.00011
		τ **link function:log**		
τ **coefficients**	**Estimate**	**Std Error**	**t value**	**p-value**
(Intercept)	2.3454	0.2036	11.52	<2e-16

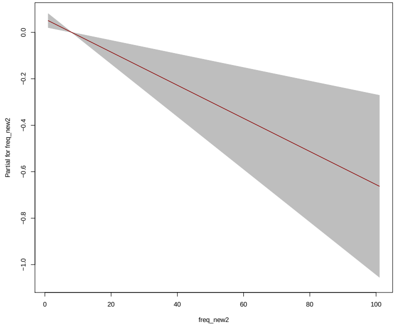
GAMLSS Modelling Mean Severity - Group 1

R. Alcoforado
A. D. Egídio dos Reis



		μ **link function: log**		
μ **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
		σ **link function:log**		
σ **coefficients**	**Estimate**	**Std Error**	**t value**	**p-value**
(Intercept)	-0.88981	0.06765	-13.15	<2e-16
		ν **link function: identity**		
ν **coefficients**	**Estimate**	**Std Error**	**t value**	**p-value**
(Intercept)	0.1809	0.06812	2.656	0.00811
		τ **link function:log**		
τ **coefficients**	**Estimate**	**Std Error**	**t value**	**p-value**
(Intercept)	0.3421	0.1142	2.996	0.00284

GAMLSS Modelling Mean Severity - Group 2



		μ **link function: log**		
μ **coefficients**	**Estimate**	**Std Error**	**t value**	**p-value**
(Intercept)	10.00813	0.023016	434.836	<2e-16
freq	-0.00713	0.002088	-3.414	0.00455
		σ **link function:log**		
σ **coefficients**	**Estimate**	**Std Error**	**t value**	**p-value**
(Intercept)	-0.60743	0.02542	-23.9	<2e-16
		ν **link function: identity**		
ν **coefficients**	**Estimate**	**Std Error**	**t value**	**p-value**
(Intercept)	0.17252	0.03669	4.703	2.76e-06
		τ **link function:log**		
τ **coefficients**	**Estimate**	**Std Error**	**t value**	**p-value**
(Intercept)	2.485	0.231	10.76	<2e-16

Final Remarks

Regarding the distribution adequacy

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

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;

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

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

R. Alcoforado
A. D. Egídio
dos Reis

Regarding the GAMLSS Modelling

- For the median of severity - Exponential Gaussian Family;

Final Remarks

R. Alcoforado
A. D. Egídio dos Reis

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

R. Alcoforado
A. D. Egídio dos Reis

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

R. Alcoforado
A. D. Egídio dos Reis

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.

Denuit, M., Hainaut, D., and Trufin, J. (2019). *Effective Statistical Learning Methods for Actuaries I: Tree-based Methods and extensions*.

Denuit, M., Marchal, X., Pitrebois, S., and Walhin, J.-F. (2007). *Actuarial Modelling of Claim Counts*.

Garrido, J., Genest, C., and Schulz, J. (2016). Generalized linear models for dependent frequency and severity of insurance claims. *Insurance: Mathematics and Economics*, 70:205–215.

Rigby, R. A. and Stasinopoulos, D. M. (2006). Using the Box-Cox t distribution in GAMLSS to model skewness and kurtosis. *Statistical Modelling*, 6(3):209–229.

Stasinopoulos, D. M. and Rigby, R. a. (2007). Generalised additive models for Location Scale and Shape (GAMLSS) in R. *Journal Of Statistical Software*, 23(7):1–46.

Tremblay, L. (1992). Using the Poisson inverse Gaussian in bonus-malus systems. *ASTIN Bulletin: The Journal of the IAA*, 22(1):97–106.

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.