Master in ctuarial and Financial Sciences

MASTER THESIS

Title: Modelling SME Commercial Lines

Author: Javier Torres Sendra

Promoter: Catalina Bolancé Losilla

Course: Actuarial and Finance Sciences

Universitat de Barcelona
Facultat d' Economia i Empresa

Faculty of Economics and Business

Universitat de Barcelona

Master thesis

Master in Actuarial and Financial Sciences

Modelling SME Commercial Lines

Author: Javier Torres Sendra

Promoter: Catalina Bolancé Losilla

The contents of this document are the sole responsibility of the author, who declares that he has not committed plagiarism and that all references to other authors have been expressed in the text.

ABSTRACT

This study examines an insurer database coming from Small and Medium Enterprises, SME, and suggests different types for modelling their claim cost. Due to data has a small number of policies compared to Motor line of business, a transformation is done and number of claims and claim costs are grouped depending on which part of SME is affected by the claim: Building or Content. Moreover, it is created an Aggregate claim data as the sum of the previous two. After doing this transformation to data; Building, Content and Aggregate claim cost are analysed using General Linear Models, GLM. For this study, the possibility to have a claim, the claim number and the claim cost are taking as a dependent random variables in the proposed models.

KEYWORDS: GLM, SME, Claim cost, Occupancy.

INDEX

FIGURE INDEX

TABLE INDEX

1. INTRODUCTION.

Small and Medium Enterprises, SME, represents the second group by premiums in the non-life multi-risk insurance sector in Spain, based on ICEA quarterly book review (2015). Their premiums represent 18,3% of the total premiums. A large literature has been published for Automobile, Household or Healthcare but there is not much literature for SME modelling.

The objective of this piece of work is to analyse the claim cost behaviour for SME. Claim cost data is grouped in claims that affects Building or Content, adding an Aggregate claim cost based on the sum of previous ones.

The model analysis is based on the study of three random variables: if a claim occurs or not, claim number and claim cost. Depending on the selection of the variables, the investigation is classified into two:

- Binomial and cost models.
- Frequency-severity models.

Previous approaches are called two-part model. Nevertheless this analysis is vast used, an Aggregate model, taking claim number and cost at same time is analysed and compared with two-part model. The models used in this study are Generalized Linear Models, GLM, introduced by Nelder and Weddeburn (1972).

Doing this analysis, this work wants to answer how SME claim cost could be modeled and which covariates are statistically significant under different approaches.

So Methodology chapter will explain the different GLM that are used for the study of SME claim cost. Data chapter explains data that has been used for modelling and also some statistical description of covariates. Chapters 4, 5 and 6 explains the analysis done for Content, Building and Aggregate claim cost. Chapter 7 explains the conclusions obtained. Appendix A contains results for models that has been studied but are not the main used in this work. Appendix B is the R code used for obtaining analysis and results done in this work.

2. METHODOLOGY.

The aim of this work is analyse data coming from SME insurance company and analyse several models for claim cost using covariates that are beyond typically factors of household or automobile covariates, and are specifically for SME policies like SME Occupancy or Employees number.

To study it, data is grouped depending on claims affection. Two groups are defined, Building and Content, and a new group is created based on the sum of previous ones. So it is studied models for Building, Content and the Aggregated one.

To analyze SME risks for Building, Content and Aggregated cost we will use General Linear Models, GLM, introduced by Nelder and Weddeburn (1972) and obtain a model for each type of risks and the aggregated one.

GLM as they were explained by McCullagh and Nelder (1989) took a response dependent variable Y and a matrix of covariates X related as follows:

- Y belongs to an exponential family distribution with formula:

$$f_Y(y,\theta,\phi) = exp\left\{\frac{y\theta - b(\theta)}{(\phi/\omega)} + c(y,\phi)\right\}$$

  Where ϕ and ω are previously known, b(), c() are determined functions and E(Y)=μ.

- A predictor η where there exist coefficients β that η=X*β
- A link function g where $\mu = g^{-1}(\eta)$.

I will use this type of models in order to replicate the total cost for the different risk that I want to study. For each group of data: Building, Content and Aggregated; it is analysed using three different kind of random variables: the possibility if a claim occurs or not, claims number and claim amount..

Depending on the approach used for studying this claim cost, we could differentiate in two main techniques:

The first approach, could be understood as it follows:

(a) A random discrete variable to study if a claim happens or not, following a binomial model.

(b) A random continuous variable to study the claims cost.

If a claim happens or not, it is used two binomial exponential distribution models: Probit and Logit.

Probit model was introduced by Bliss (1935) as an interpretation of the dosage- mortality curve. McCullagh and Nelder (1989) englobed this Probit model inside binomial models on GLM defining link function as:

- $\eta = \Phi^{-1}(\mu)$, where Φ^{-1} is the cumulative distribution function of the canonical Normal distribution.

To obtain a model for studying if a claim happens or not, also it is analysed the logit model that was mainly introduced by Berkson (1944) in another biological assay as Probit model did under Bliss (1935). This model is also examined by McCullagh and Nelder (1989) under binomial models and link function:

- $\eta = \ln\{\frac{\mu}{1-\mu}\}$.

The claim cost model is obtained analysing continuous exponential family distributions as it has been assumed before. It is used Gamma, Lognormal and Inverse Gaussian as it is explained in McCullagh and Nelder (1989).

- When Y is a Gamma, i.e. $Y \sim \Gamma(\mu, \nu)$, the exponential distribution model is characterized by:
  - $\phi = \nu^{-1}$,
  - $b(\theta) = -\log(-\theta)$,
  - $c(y, \nu) = \nu \log(\nu y) - \log(y) - \log(\Gamma(y))$,
  - $\mu = \eta^{-1}$.
- When Y is a lognormal it is considered the Normal model and then rescaled by log function. If $\log Y \sim N(\mu, \sigma^2)$ then logY is characterized as:
  - $\phi = \sigma^2$,
  - $b(\theta) = \frac{\theta^2}{2}$,
  - $c(y, \phi) = -\frac{1}{2}\left(\frac{y^2}{\phi} + \log(2\pi\phi)\right)$,
  - $\mu = \eta$.
- Finally, if Y is an Inverse Gaussian, $Y \sim IG(\mu, \sigma^2)$, then it is characterized as:
  - $\phi = \sigma^2$,
  - $b(\theta) = -(-2\theta)^{\frac{1}{2}}$,
  - $c(y, \phi) = -\frac{1}{2}\left\{\log(2\pi\phi y^3) + \frac{1}{\phi y}\right\}$,

- $\mu = (-2\eta)^{-\frac{1}{2}}$.

Nevertheless, I split the model in two compounds, I will also use the model introduced by Tobin in 1958 for studying the aggregated claim cost. This model, called Tobit model, was introduced by Tobin (1958) to take the information that a Probit model is evaluating plus the information of claim cost value. So, this model considers the aggregated cost for each risk instead of splitting it in two parts.

Tobit model is a model to stablish a lineal relationship between an independent set of covariates called X and a dependent variable S*, taking into account that the observed variable S, is the censored result of the dependent variable S^* as follows:

$$S = \begin{cases} S^* \; if \; y > c \\ 0 \; if \; y \leq c \end{cases}, \qquad c = 0 \;\; normally$$

where c is the hedge value and S^* has a linear relation with X covariates using β parameters, defined by the equation:

$$S^* = X\beta + \varepsilon$$

where ε is a normal random variable, $\varepsilon \sim N(0, \sigma^2)$ considered as the random differences between S^* and the linear combination of the X covariates.

As it could be understood, the structure taken to describe the error is considered following a normal distribution as mentioned in Frees (2010). However these restrictions could be considered a limitation, it has been widely applied to studies for Health as Jun Mo Lee et alters (2014) or Kuan-Chia Lin and Su-Fen Cheng (2011), Motor Insurance as Panagiotis et alters (2011) or Household Goods as Tobin himself did in its article published in 1958.

So on the first approach it is examined a two-part model composed by a binomial model and a continuous exponential distribution, and it is added a Tobit model study as a way to examine the aggregated cost.

The second approach is called a frequency-severity model, and it could be understood as it follows:

(a) A random discrete variable to study the claim frequency, understanding frequency by the following equation: $frequency = \frac{number\ of\ claims}{policy\ exposure}$.

(b) A random continuous variable to study the claims average cost as it is described here: $claim\ average\ cost = \frac{claim\ cost}{number\ of\ claims}$.

For studying the claim frequency, it is basically used the Poisson model that is a discrete exponential distribution as it is pointed out in McCullagh and Nelder (1989). In this case the characteristics for N a random variable that is a number of claims, it follows a Poisson distribution, i.e. N~$Pois(\lambda)$, are:

- $\phi = 1$,
- $b(\theta) = exp(\theta)$,
- $c(y, \phi) = -\log(y!)$,
- $\mu = exp(\eta)$.

On the frequency study, it is measured the over dispersion analysis too, understood as E(N) ≠ VAR(N) like McCullagh and Nelder (1989) understood. In this case, it is considered a Quasi Poisson model in order to determine if ϕ parameter, called the dispersion parameter, is not equal to 1 to verify if there is some kind of dispersion in the data set.

For considering over-dispersion, two models are used in order to know which one fix best the frequency model. The two models are mixtures of Poisson model. In first instance, it is used a Negative Binomial that is described as a Poiss(λ) where lambda is a Gamma random variable that $\lambda \sim \Gamma(\mu, \nu)$. Also, it is analysed a Poisson Inverse Gaussian where lambda parameter of Poisson variable follows an Inverse Gaussian $\lambda \sim IG(\mu, \sigma^2)$.

Moreover, as the model for frequency has a large number of zeros it is analysed the Hurdle model and the Zero Inflated Negative Binomial to take the excess number of 0's into account on modelling. The Hurdle model as Frees (2010) says, it is defined as a decision chain process for reporting or not a claim by different policyholder groups. In this study, the group is identified by sum insured. The Zero Inflated Negative Binomial is also explained in Frees(2010) as a model to take into account the excess of zeros as a combined model of point-mass at zero and a negative binomial distribution.

For the study of the claim average cost it is used the same models as in the binomial and cost models for claim cost analysis. So a Gamma, Lognormal and Inverse Gaussian are chosen in order to model this figure.

In the same way as in the binomial and cost models, it could be analysed the whole data instead of dividing it in two part model. In this case a Tweedie model is studied. Tweedie model was introduced by Tweedie (1984) and developed by Jorgensen et alt. (1994) for Automobile insurance claims. Considering N a random variable for claims number and and X_i is the cost of claim at i realization. It could be defined a random variable S as follows:

$$S = \begin{cases} X_1 + X_2 + \cdots + X_N \; if \; N > 0 \\ 0 \; if \; N = 0 \end{cases},$$

Under Tweedie model N follows a Poisson distribution and X_i a Gamma distribution. On this case S follows a *Tweedie*(μ,ϕ,p) where Var[S] = $\phi \, \mu^{p}$. The parameter p is defined p∈(1,2) to ensure that it is a mixture of discrete and continuous variable.

When all the models are done, AIC criteria is used to determine which model is the best one for database claim cost.

This study is done using R as software. It could be found the code used for this study in APPENDIX B.

3. DATA.

To obtain an answer for questions introduced before, I will use a database from an Insurance company from 2012 – 2014 related to SME insurance. This data base contains number of claims, exposure and claims cost due to different perils from years 2012 to 2014. Perils are defined as Sun (2011) did in his thesis, as the root cause of a loss, adding to this definition if it affects at Building or Content. This implies that this database has more granulated information about number of claims and costs than peril level but I will use the same peril word. It also contains variables related to policyholder, business or policy as it is described in Table 1.

Table 1: Database fields description.

Data from policy	
PAYFREQ	Payment frequency. Annual, biannual, quarterly, in one payment.
BUILDINGSI	Building Sum Insured
INSBUTYPE	Type of building insured: Whole building or renovation
BUFORM	Type of building sum insured option: Total or first loss.
DEEXGA	Extension coverage deductible.
THEFTFORM	Type of theft insured option: Total or first loss.
THEFTBUILDSI	Theft damages to building first loss sum insured.
GLASSSI	Glasses damage first loss sum insured.
ELECTSI	Electrical damage first loss sum insured.
DEELECT	Electrical deductible.
DEEOW	Water damage deductible.
CONTENTSI	Content Sum Insured.
YEAR	Policy underwriting year.
Data from Object	
PROVINCE	Province where the risk is.
BUTYPE	Type of building.
RISKLOC	Risk location.
INHABITANTS	Inhabitants number
BUILDYEAR	Building year
PHYSPROTECT	Physical protections against theft.
EMPLOYEES	Number of employees
HYDRANT	Fire hydrant
FIREXT	Fire extinguishers
DETECTOR	Smoke detector
CONALARM	Alarm connected indicator
VIGILANCE	24-hour security
GLASSTYPE	Type of glasses: one layer, two layers, three layers.
SECBOX	Safe-deposit box

THEFTNORM	Meets theft regulations
OCCUPANCY	Group occupancy of the business.
Data from owner	
OWNER	Indicates if it is owner of the building or only rents it.
Data related with claims and policy exposition	
BUILDINGEXP	Building exposure
CONTENTEXP	Content exposure
EXPOSURE	Policy exposure
BUILDINGNUM	Number of claims that affects building
CONTENTNUM	Number of claims that affects content
CLAIMNUM	Number of claims
BUILDINGCOST	Cost of claims that affect building
CONTENTCOST	Cost of claims that affect content
CLAIMCOST	Cost of claim
INDBUILDING	Building affection indicator
INDCONTENT	Content affection indicator
INDCLAIM	Claim indicator
BUILDINGCOSTEXP	Building cost divided by exposure
BUILDINGFREQ	Building frequency
BUILDEXPLN	Building exposure logarithm
CONTENTCOSTEXP	Content cost divided by exposure
CONTENTFREQ	Content frequency
CONTENTEXPLN	Content exposure logarithm
CLAIMCOSTEXP	Claim cost divided by exposure
CLAIMFREQ	Claim frequency
CLAIMEXPLN	Claim exposure logarithm

Database contain claims coming from the following perils:

- CONTENT
- BUILDING
- ELECTRICAL BUILDING
- ELECTRICAL CONTENT
- WATER BUILDING
- WATER CONTENT
- THEFT CONTENT
- THEFT BUILDING
- EXTENSION CONTENT
- EXTENSION BUILDING
- GLASS BUILDING
- GLASS CONTENT

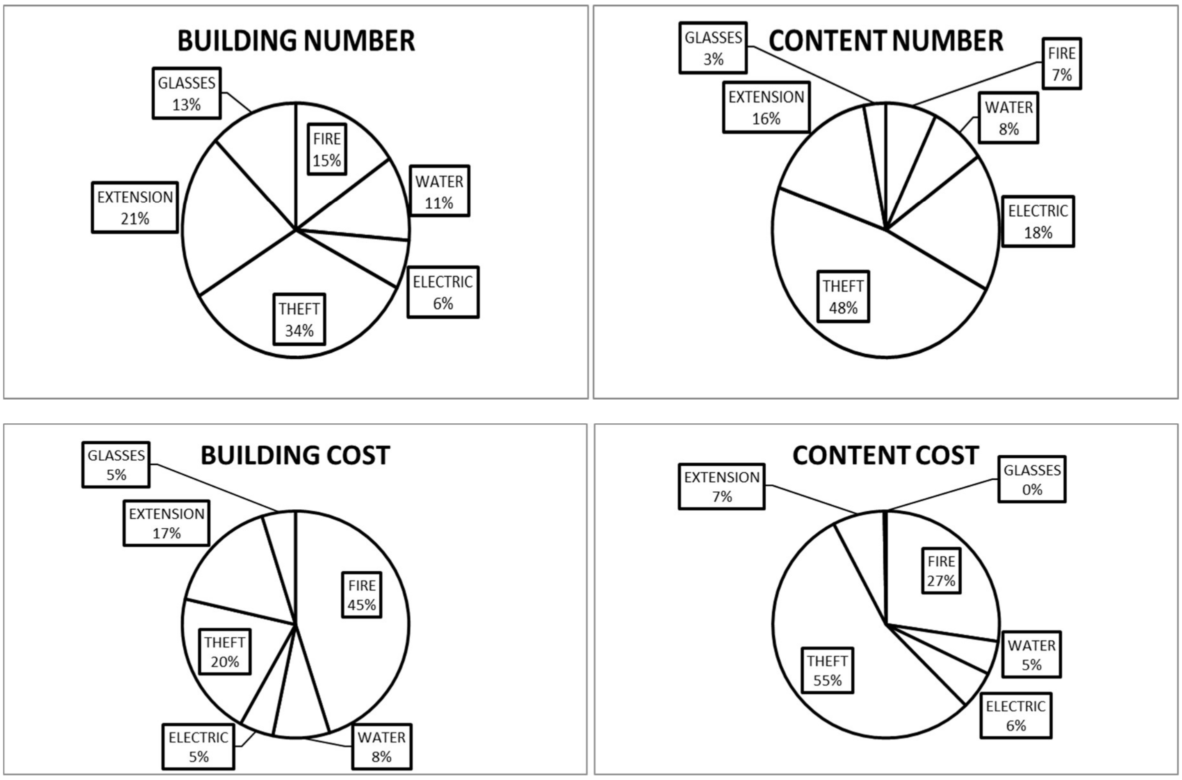
These perils are grouped in two main categories Content and Building, so it is taken into consideration not what causes the claim, I will focus on what is affected. This approach is similar at work done by Frees, Meyers and Cummings (2010) on Household Insurance, so it will not be followed in this case the approach from Veilleux (2007) because database is not so big as in Household or Automotive. As we are interested in Content and Building relationship I will only choose policies that have already Content and Building on their policy. This reduces data base from 171894 policies to 118898 policies.

Claims number and cost are grouped as follows:

- CONTENT: Fire, Electrical content, water content, theft content, extension content.
- BUILDING: Fire, Electrical building, water building, theft building, extension building.

The Content and Building composition by its peril and number of claims or claim cost could be observed on Figure 1. It shows that Building is composed basically by Theft, Fire and Extension and Content by half Theft:

Figure 1: Building and Content Composition.



Own made figure

The characteristics of policy contracts on the given data base are reflected in tables 2, 3 and 4. Mainly the policies are annual, where it is insured the whole building for the total building amount. Table 4 shows what percentage of the Content Sum Insured is covered on Theft.

Table 2: Payment Frequency.

PAYMENT FREQUENCY			
ANNUAL	BIANNUAL	QUATERLY	UNIQUE
64%	23%	13%	0%

Table 3: Building Insurance form.

BUILDING INSURANCE FORM			
INSURANCE BUILDING TYPE		BUILDING FORM	
WHOLE BUILDING	84%	TOTAL SUM INSURED	93%
WORK REFORM	16%	FIRST LOSS	7%

Table 4: Theft Insurance form.

THEFT INSURED FORM	
THEFT INSURED FORM	FREQUENCY
TOTAL VALUE	86%
25% OF TOTAL VALUE	7%
20% OF TOTAL VALUE	3%
10% OF TOTAL VALUE	2%
FIRST LOSS	1%
5% OF TOTAL VALUE	1%

The characteristics of object insured are reflected in tables 5, 6, 7. Insured objects are mainly located in Barcelona, Valencia and Madrid. Moreover, they are usually located in the urban core and inside an industrial park.

Table 5: Province.

PROVINCE	
PROVINCE	FREQUENCY
BARCELONA	14.6%
VALENCIA	10.3%
MADRID	8.6%
ALICANTE	4.9%
LLEIDA	4.1%
ZARAGOZA	3.8%
BALEARES	3.6%
MÁLAGA	3.4%
SEVILLA	2.8%
CÓRDOBA	2.8%
GIRONA	2.8%
GUIPUZCOA	2.2%

CASTELLÓN	2.0%
NAVARRA	2.0%
BADAJOZ	2.0%
TENERIFE	2.0%
TARRAGONA	1.8%
GRAN CANARIA	1.7%
MÚRCIA	1.6%
VIZCAYA	1.4%
CÁDIZ	1.3%
ALBACETE	1.3%
ALMERIA	1.2%
JAEN	1.2%
ASTURIAS	1.2%
GRANADA	1.2%
TOLEDO	1.2%
PONTEVEDRA	1.1%
A CORUÑA	1.0%
CIUDAD REAL	0.9%
LEON	0.9%
HUESCA	0.8%
TERUEL	0.8%
CUENCA	0.8%
VALLADOLID	0.8%
LA RIOJA	0.7%
BURGOS	0.6%
ALAVA	0.5%
LUGO	0.5%
CANTABRIA	0.5%
CÁCERES	0.5%
HUELVA	0.4%
SALAMANCA	0.4%
PALENCIA	0.3%
SEGOVIA	0.3%
GUADALAJARA	0.3%
ZAMORA	0.2%
OURENSE	0.2%
AVILA	0.2%
ANDORRA	0.1%
SORIA	0.1%
MELILLA	0.1%
CEUTA	0.0%

Table 6: Risk location.

RISK LOCATION		
URBAN CORE	TOWN	OUTBACK
78%	17%	4%

Table 7: Building type.

BUILDING TYPE	
BUILDING TYPE	FREQUENCY
INDUSTRIAL PARK	91%
DWELLING	5%
COMMERCIAL SHOP	2%
OFFICE	1%
COMMERCIAL CENTRE	0%
PUBLIC MARKET	0%

Database Policyholders are mainly owners and just a quarter of whole policiholders are tenants as shown in Table 8.

Table 8: Ownership.

OWNERSHIP	
OWNER	TENANT
79%	21%

Table 9 shows the fire measures that are present in the policies of the database. It seems that the vast majority of insured have some fire measures. This could indicate that only policyholders that have some fire preventions are insured.

Table 9: Fire measures.

FIRE MEASURES			
FIRE EXTINGISHERS	HYDRANTS	SMOKE DETECTOR	VIGILANCE
99%	83%	92%	97%

Table 10 and 11 describe the theft protections measures, shown percentages are done under the total policies that have theft insured. There is around 8% of data that has not theft peril insured. On Table 10 it could be seen that glasses have mainly one layer and there is no an extensive utilization of security box. Table 11 shows that alarm connection does not seem a compulsory factor for insure theft coverage. It seems more important to have physical protections like security doors or gratings.

Table 10: Theft protections measures (I).

THEFT PROTECTIONS (I)			
GLASS TYPE		SECURITY BOX	
1	90%	WITHOUT SB	65%

2	9%	SB UNDER 100Kg	31%
3	1%	SB ABOVE 100Kg	4%

Table 11: Theft protection measures (II).

THEFT PROTECTIONS (II)		
PHYSICAL PROTECTIONS	ALARM CONNECTION	THEFT INTERNAL REGULATION
89%	62%	94%

The most extended deductible used on the database is a 10% of the amount of claim with a minimum of 200€. Also it is used another nearly figures like 300€ or 150€.

Table 12: Deductibles.

DEDUCTIBLES			
DEDUCTIBLE VALUE	ELECTRICAL	EXTENSION	WATER
10% min 200€	79%	72%	68%
150 €	7%	13%	13%
300 €	4%	14%	14%
Other	9%	1%	5%

On Table 13 it could be observed the distribution of occupancies in database. Mainly the SME insured belong to warehouses, automotive and metallurgy followed by farm industry.

Table 13: Occupancy.

OCCUPANCY	
OCCUPPANCY	FREQUENCY
WAREHOUSES WITHOUT FOOD	21%
AUTOMOTIVE WITH WORKSHOP	18%
METALLURGY	15%
AUTOMOTIVE WITHOUT WORKSHOP	8%
FARMS	6%
FOOD WAREHOUSES	5%
LOCAL WITOUT ACTIVITY	5%
FOOD HANDLING	5%
LEISURE OR RECREATIONAL	3%
PAPER/LEATHER	3%
WOOD MANUFACTURING	3%
CLOTHES MANUFACTURING	2%
CHEMICAL MANUFACTURING	1%
ENERGY PRODUCTION	1%
LABORATORIES	1%
STONE MANUFACTURING	1%

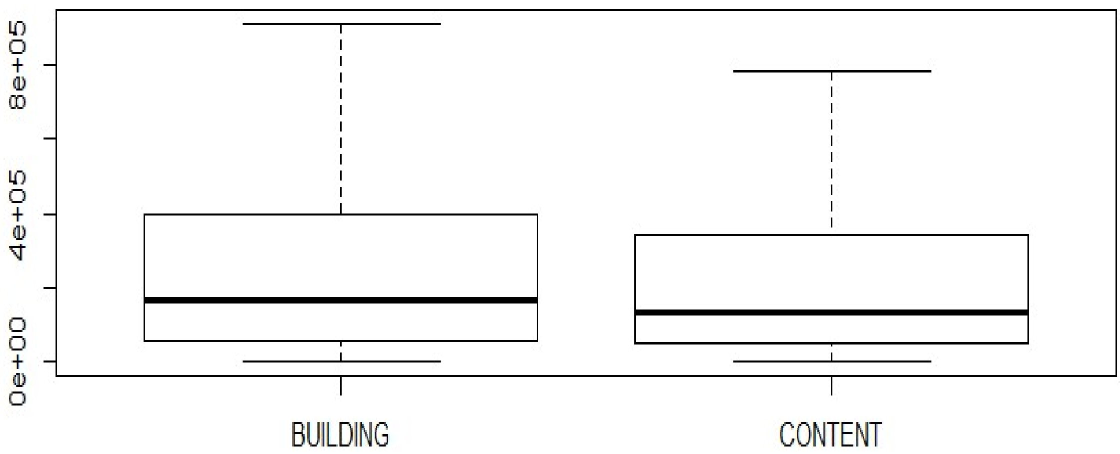
Table 14 shows a statistical description of continuous variables representing sum insured for some perils.

Table 14: Sum insured.

SUM INSURED						
COVER	MIN	1^{ST} Q.	MEDIAN	MEAN	3^{RD} Q.	MAX
BUILDING	1	56370	164100	487500	400000	12850000
CONTENT	0	50000	134000	428700	345000	40270000
THEFT BUILDING	0	1789	3091	4534	6000	318300
GLASS	0	617.9	1200	2142	2578	420000
ELECTRIC	0	2321	5000	8031	6901	1140000

Figure 2 uses a box plot diagram to visualize the difference between Content Sum Insured and Building Sum Insured without taking into account some outliers. It could be perceived that Building Sum Insured has more variability than Content Sum Insured as it was expected due to many policyholders where Owners and not just Tenants.

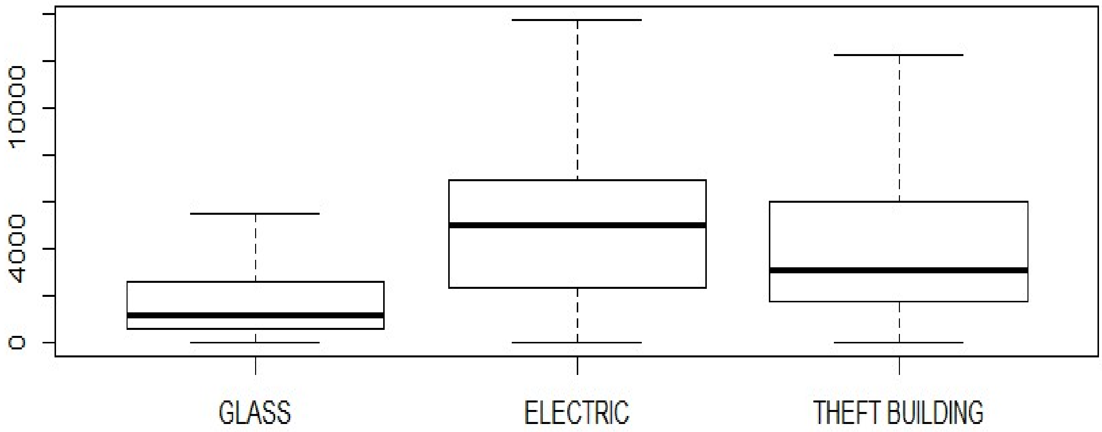
Figure 2: Building and Content Sum Insured.



Own made figure.

Figure 3 shows dispersion of Glasses, Electrical and Theft Building Sum Insured. This three insured sums are lesser than Content and Building and must be taken apart from analysis of Content and Building. As Graphic shows Glasses Sum Insured is the lowest sum insured. Electrical and Theft Building seems to have a similar dispersion values but it must to be highlighted that Theft Building peril is less insured than Electrical and has lesser values than Electrical ones.

Figure 3: Glass, Electrical, Theft Building sum insured.



Own made figure.

4. MODELING CLAIM COST FOR AGGREGATED PERILS.

On this section I compare different types of modelling aggregated claims cost based on GLM models.

4.1. BINOMIAL AND COST MODELS.

Table 15 shows the results of statistical significant covariates under a TOBIT model applied to total claim cost. The results describe that Occupancy are one of the key covariates for TOBIT model. The riskiest activities are LEISURE AND RECREATIONAL and COMMERCIAL CENTRE followed by AUTOMOTIVE WITHOUT WORKSHOPS and FARMS. Related with the SME occupancy is the number of employees. If the SME has more than 3 employees it is riskier than other SME. This result is expected because this covariate is used as a control covariate. Usually, the more number of employees, bigger is the SME and more claims it has. Other covariates that are significant are the Continent and Building Sum Insured (SI), these two are also control covariates as explained before. There is a linear relation between the sum insured and the cost of claim. The other amounts from other embedded perils follow the same rule. For example, Glass Sum Insured has a discount if it is not covered and has a recharge if the amount is above 3000€. Further characteristics related with amount figure are if Building is insured whole building or just a reform work has their difference on model. Insuring the whole building is riskier in this case. If the policyholder is a tenant is riskier than if it is an owner of the building. Also a geographical component has been found, model points out that islands and some north communities are less risky than the other ones. If policyholder pays biannually, model shows that is riskier than other ways of payment and that annual payment is less risky than other.

Table 15: TOBIT fit summary for total claim cost.

PARAMETERS ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT 1	-102105.65	1300.00
INTERCEPT 2	10.96	0.01
BIANNUAL FREQUENCY	3665.93	720.00
GUIPUZCOA PROVINCE	-6789.16	2110.00
NAVARRA PROVINCE	-10668.04	2410.00
BALEARS PROVINCE	-6530.90	1730.00
GRAN CANARIA PROVINCE	-15128.85	2600.00
TENERIFE PROVINCE	-15220.57	2450.00
WORK REFORM ONLY	-6523.24	1310.00
TENNANT	3154.48	1140.00
THEFT NOT INSURED	-7226.40	1400.00
BETWEEN 250000€-350000€ SI	3098.91	1120.00
BETWEEN 350000€-535000€ SI	8048.75	1100.00
BETWEEN 535000€-1050000€	11653.19	1120.00
OVER 1050000€ SI	17130.77	1220.00
OVER 3 EMPLOYEES	5610.44	712.00
BETWEEN 1500€ - 3000€ GLASS SI	3016.68	817.00
OVER 3000€ GLASS SI	4745.78	831.00
GLASS NOT COVERED	-7932.75	1180.00
OVER >9000€ ELECTRICAL SI	4955.15	783.00
ELECTRICAL NOT INSURED	-9083.95	1530.00
BETWEEN 130000€-190000€ CONTENT SI	9565.29	1200.00
BETWEEN 190000€-275000€ CONTENT SI	13361.76	1200.00
BETWEEN 275000€-450000€ CONTENT SI	14120.16	1210.00
BETWEEN 450000€-850000€ CONTENT SI	15838.20	1290.00
BETWEEN 60000€-90000€ CONTENT SI	7560.88	1240.00
BETWEEN 90000€-130000€ CONTENT SI	7414.92	1220.00

OVER 850000€ CONTENT SI	18064.39	1370.00
AUTOMOTIVE OCCUPANCY	14017.54	1090.00
FARM OCCUPANCY	11204.48	1540.00
FOOD HANDLING OCCUPANCY	6328.16	1280.00
FOOD WAREHOUSE OCCUPANCY	4689.60	1320.00
LEISURE AND RECREATIONAL OCCUPANCY	19673.22	1500.00
COMMERCIAL CENTRE OCCUPANCY	14504.47	1370.00

PERSON RESIDUALS				
MIN	1^{ST} Q	MEDIAN	3^{RD} Q	MAX
-6826.79	-0.18	-0.12	-0.08	9.49
-0.80	-0.26	-0.22	-0.18	3527.93

AIC
342338.1

On Table 16 it could be seen the AIC for two methods of binomial variable. As LOGIT model have a less AIC, it is the one chosen.

Table 16: AIC figure for Binomial.

AIC	
LOGIT	PROBIT
73914	74073

On Table 17 are shown the results obtained using a LOGIT model for the Binomial claim model. Significant covariates are similar at TOBIT model and many of them have the same behaviour of those model. It could be seen that Occupancy has more granularity on this model than in TOBIT model. The three riskiest occupancies on this model are the same as in TOBIT model. Employee is a more granular variable on this case, having more significant intervals and being riskier as more employees are. Other perils sum insured are also present on the model. Particularly Theft, Glass and Electrical. It has to be remembered that Theft is an important peril on the weight of Content and Building but there is no Extension covariate presence. Also it has to be pointed out the presence of security box covariate, with a discount if there is no security box. It could be that SME with security box have more money to be stolen and more probability of having a theft claim. Other Geographical covariates are present on this model as it was in TOBIT model but also it is present Inhabitants covariate. As more inhabitants more risky except in medium cities (above 130000 inhabitants). As in TOBIT model the Content and Building Sum Insured have a direct relation with the claim probability.

Table 17: LOGIT parameters estimation.

PARAMETERS ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT	-3.154	0.060
BIANNUAL FREQUENCY	0.112	0.023
NAVARRA PROVINCE	-0.302	0.080
GUIPUZCOA PROVINCE	-0.232	0.056
GRAN CANARIA PROVINCE	-0.554	0.085
TENERIFE PROVINCE	-0.588	0.082
WORK REFORM ONLY	-0.096	0.035
WITHOUT SECURITY BOX	-0.139	0.022
THEFT NOT COVERED	-0.373	0.050
AUTOMOTIVE OCCUPANCY	0.642	0.035
AUTOMOTIVE WITH WORKSHOP OCCUPANCY	0.113	0.029
ENERGY OCCUPANCY	0.407	0.122
FARM OCCUPANCY	0.546	0.052
FOOD HANDLING OCCUPANCY	0.303	0.041
FOOD WAREHOUSE OCCUPANCY	0.240	0.043
LEISURE AND RECREATIONAL OCCUPANCY	0.824	0.046
COMMERCIAL CENTRE OCCUPANCY	0.695	0.043
OTHER MACHINERY OCCUPANCY	0.533	0.205
PAPER AND LEATHER OCCUPANCY	0.211	0.055
STONE MANUFACTURING OCCUPANCY	0.401	0.096
BETWEEN 15000€-130000€ BUILDING SI	0.139	0.042
BETWEEN 130000€-185000€ BUILDING SI	0.174	0.053
BETWEEN 185000€-250000€ BUILDING SI	0.229	0.052
BETWEEN 250000€-350000€ BUILDING SI	0.300	0.052
BETWEEN 350000€-535000€ BUILDING SI	0.464	0.052
BETWEEN 535000€-1050000€ BUILDING SI	0.567	0.052
ABOVE 1050000€ BUILDING SI	0.787	0.054
1 EMPLOYEE	-0.126	0.034
BETWEEN 4 AND 9 EMPLOYEES	0.158	0.028
MORE THAN 9 EMPLOYEES	0.138	0.030
GLASS NOT COVERED	-0.308	0.041
BETWEEN 1500€ - 3000€ GLASS SI	0.107	0.026
ABOVE 3000€ GLASS SI	0.193	0.026
ELECTRIC NOT INSURED	-0.337	0.054
ABOVE 9000€ ELECTRIC SI	0.160	0.024
BETWEEN 7500 <= 30000 INHABITANTS	0.110	0.027
BETWEEN 30000 <=130000 INHABITANTS	0.127	0.029
ABOVE 130000 INHABITANTS	0.088	0.028
BETWEEN 35000€-60000€ CONTENT SI	0.128	0.048
BETWEEN 60000€-90000€ CONTENT SI	0.313	0.046
BETWEEN 90000€-130000€ CONTENT SI	0.297	0.046
BETWEEN 130000€-190000€ CONTENT SI	0.347	0.045

BETWEEN 190000€-275000€ CONTENT SI	0.461	0.045
BETWEEN 275000€-450000€ CONTENT SI	0.493	0.045
BETWEEN 450000€-850000€ CONTENT SI	0.501	0.047
ABOVE 850000€ CONTENT SI	0.564	0.049

RESIDUAL DEVIANCE				
MIN	1ST Q	MEDIAN	3RD Q	MAX
-1.059	-0.498	-0.393	-0.309	2.890

Results for the PROBIT model could be found on APENDIX A Table 56.

Table 18 shows the AIC value for claim cost modelling using several continuous exponential model. As it could be verified for the AIC the better model in this case is LOGNORMAL one. It has to be highlighted that there are two few covariates on this model compared with binomial or TOBIT one.

Table 18: AIC for claim cost.

AIC		
GAMMA	LOGNORMAL	IG
229029	221509	224831

LOGNORMAL model gives the lowest AIC but it is R-squared figure is very low around 0.0395, meaning that it is not well adjusted. It has to be pointed out that Building Sum Insured is not significant on this model and only Occupancy, Inhabitant and Content Sum Insured are present as it could be looked at Table 19. Moreover it appears the variable risk location that points out that outback locations must be recharged.

Table 19: LOGNORMAL parameter estimation claim cost.

PARAMETERS ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT	6.892	0.028
OUTBACK RISK LOCATION	0.371	0.085
AUTOMOTIVE WITH WORKSHOP OCCUPANCY	-0.149	0.042
ENERGY OCCUPANCY	0.774	0.196
FARM OCCUPANCY	0.344	0.087
LEISURE AND RECREATIONAL OCCUPANCY	0.282	0.064
OTHER MACHINERY OCCUPANCY	0.880	0.334
PAPER AND LEATHER OCCUPANCY	-0.245	0.086
STONE MANUFACTURING OCCUPANCY	0.685	0.147
ABOVE 130000 INHABITANT	-0.156	0.036
ABOVE 190000€-275000€ CONTENT SI	0.231	0.052
BETWEEN 275000€-450000€ CONTENT SI	0.282	0.048
BETWEEN 450000€-850000€ CONTENT SI	0.414	0.049
ABOVE 850000€ CONTENT SI	0.774	0.043

R-SQUARE	AIC
0.0395	221509

RESIDUAL DEVIANCE				
MIN	1ST Q	MEDIAN	3RD Q	MAX
-11.643	-1.224	-0.002	1.100	7.698

Results for GAMMA and IG models could be found on APENDIX A Tables 57 and 58.

4.2. FREQUENCY-SEVERITY MODELS.

To use TWEEDIE model, first it has been estimated different p as dispersion parameter. It has been chosen the one with less deviance as a criterion. Table 21 shows different p parameter and their deviance.

Table 20: Dispersion parameter estimation for TWEEDIE model for Aggregated claim cost.

P PARAMETER	NULL DEVIANCE
1.4	40249369
1.5	33001295
1.6	11899209

TWEEDIE model as could be seen on Table 22 is mainly based on sum insured of Content and Building. Payment frequency, insure only a part of the building, have some fire measures or be a tenant has similar behaviour than in TOBIT model. On the other hand, it seems that variables that indicates if a certain peril is covered or not, and SME own variables like Employees number or Occupancy are not significant for this model.

Table 21: TWEEDIE parameters estimation for agreggated claim cost.

PARAMETERS ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT	5.142	0.095
QUARTERLY PAYMENT FREQUENCY	0.311	0.115
RENOVATION WORK ONLY	-0.545	0.171
TENNANT	0.492	0.149
FIRE DETECTOR PRESENT	-0.406	0.153
BETWEEN 535000€ AND 1050000€ BUILDING SI	0.599	0.140
ABOVE 1050000€ BUILDING SI	0.808	0.150
BETWEEN 60000€ AND 90000€ CONTENT SI	0.677	0.168

BETWEEN 90000€ AND 130000€ CONTENT SI	0.755	0.163
BETWEEN 130000€ AND 190000€ CONTENT SI	1.050	0.157
BETWEEN 190000€ AND 275000€ CONTENT SI	1.449	0.154
BETWEEN 275000€ AND 450000€ CONTENT SI	1.476	0.152
BETWEEN 450000€ AND 850000€ CONTENT SI	1.808	0.157
ABOVE 850000€ CONTENT SI	2.046	0.165

RESIDUAL DEVIANCE				
MIN	1^{ST} Q	MEDIAN	3^{RD} Q	MAX
-9.677	-6.968	-6.254	-5.584	177.179

AIC
354931

Continuing on the two part model analysis. It is modelled frequency using discrete models based mainly on POISSON. It has been studied POISSON as the first model but also it has been analysed the QUASI POISSON model for looking at over dispersion. As when we freed the parameter of dispersion in QUASI POISSON model it did not stands to 1, another models that take dispersion into account where analysed. The two models taken for dispersion where Negative Binomial and Poisson Inverse Gaussian. The results at Table 23 shows that a PIG is a better model under Akaike criterion. Finally, it has been studied if a Zero Inflated model is better than the previous models or there is a combination of dispersion model and Zero Inflated that suits frequency modelling. As it could be checked on Table 22 , the model using Zero Inflated Negative Binomial is the one that suits better frequency model. For doing that the model has been split in a Zero model using Building and Content Sum Insured and the rest of variable for doing the main model.

Table 22: AIC figure for frequency.

AIC				
POISSON	NB	PIG	HURDLE	ZINB
243831	94225	94177	102110	94006

As it could be observed in Table 24, there are a major presence of covariates not related with sum insured and they are more granular than in other models. It could be checked that LEISURE AND RECREATIONAL and COMMERCIAL CENTRE occupancies are riskier than others as it has been reflected in previous models. As it has been pointed out covariates are more granular than in previous models like it could be seen in payment frequency or in geographical covariate. Also it appear new covariates that are statistically significant like risk location or glasses layers. Nevertheless more granularity the behaviour of estimate parameters and covariates related with other perils that are not Building and Content remain the same as in previous models.

Table 23: ZINB parameter estimation for frequency.

PARAMETER ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT	-0.706	0.071
BIANNUAL FREQUENCY	0.105	0.025
QUARTERLY FREQUENCY	0.120	0.028
UNIQUE FREQUENCY	0.957	0.187
CUENCA PROVINCE	0.379	0.119
ALBACETE PROVINCE	-0.270	0.100
BALEARES PROVINCE	-0.267	0.058
GRAN CANARIA PROVINCE	-0.443	0.081
TENERIFE PROVINCE	-0.462	0.081
RENOVATION WORK ONLY	-0.163	0.033
URBAN CORE RISK LOCATION	-0.197	0.057
TOWN RISK LOCATION	-0.206	0.061
PARTIAL VALUE 10% THEFT	-0.185	0.059
THEFT NOT COVERED	-0.338	0.053
2 LAYERS GLASS TYPE	0.116	0.033
3 LAYERS GLASS TYPE	0.239	0.072
AUTOMOTIVE OCCUPANCY	0.485	0.036
FARM OCCUPANCY	0.404	0.060
FOOD HANDLING OCCUPANCY	0.198	0.042
FOOD WAREHOUSE OCCUPANCY	0.190	0.044
LEISURE AND RECREATIONAL OCCUPANCY	0.769	0.046
COMMERCIAL CENTRE OCCUPANCY	0.669	0.042
METALLURGY OCCUPANCY	-0.127	0.031
OTHER MACHINERY OCCUPANCY	0.674	0.204
STONE MACHINERY OCCUPANCY	0.274	0.096
1 EMPLOYEE	-0.118	0.036
BETWEEN 4 AND 9 EMPLOYEES	0.151	0.029
ABOVE 9 EMPLOYEES	0.225	0.030
THEFT NOT COVERED	-0.134	0.050
GLASS NOT COVERED	-0.278	0.042
BETWEEN 1500€ - 3000€ GLASS SI	0.111	0.027
ABOVE 3000€ GLASS SI	0.250	0.027
ELECTRIC NOT COVERED	-0.278	0.056
AVOBE 9000€ ELECTRICAL	0.275	0.024
BETWEEN 7500 AND 30000 INHABITANTS	0.138	0.028
BETWEEN 30000 <=130000 INHABITANTS	0.166	0.029
ABOVE 130000 INHABITANTS	0.137	0.029
Log(theta)	-0.569	0.050
MODEL FOR ZEROES		
(Intercept)	0.688	0.059
BETWEEN 185000€-250000€ BUILDING SI	-0.197	0.069
BETWEEN 250000€-350000€ BUILDING SI	-0.310	0.073
BETWEEN 350000€-535000€ BUILDING SI	-0.672	0.087
BETWEEN 535000€-1050000€ BUILDING SI	-1.136	0.125

ABOVE 1050000€ BUILDING SI	-3.078	0.704
BETWEEN 60000€-90000€ CONTENT SI	-0.464	0.071
BETWEEN 90000€-130000€ CONTENT SI	-0.455	0.070
BETWEEN 130000€-190000€ CONTENT SI	-0.582	0.071
BETWEEN 190000€-275000€ CONTENT SI	-0.816	0.079
BETWEEN 275000€-450000€ CONTENT SI	-0.993	0.088
BETWEEN 450000€-850000€ CONTENT SI	-1.089	0.105
ABOVE 850000€ CONTENT SI	-1.610	0.176

PEARSON RESIDUALS				
MIN	1^{ST} Q	MEDIAN	3^{RD} Q	MAX
-0.706	-0.314	-0.238	-0.148	68.523

Results for the POISSON, QUASI POISSON , NB, PIG and HURDLE models could be found on APENDIX A from Table 59 to Table 63.

To calculate the average model cost it is used continuous exponential distributions. The goodness measure is AIC as it is shown in Table 25. LOGNORMAL model is selected as it has the lower AIC figure.

Table 24: AIC for claim average cost.

AIC		
GAMMA	LOGNORMAL	INVERSE GAUSSIAN
199586	193491	212760

Table 26 shows the results of covariates parameter estimation using a LOGNORMAL. As it could be observed there are no many covariates that are statistically significant for this model. This also happened for the aggregate claim cost model. On this case only two variables are selected: Occupancy and Content Sum Insured. Moreover R-squared figure is about 0.02 showing that model could be improved. It is interesting to highlight that COMMERCIAL CENTRE has a negative sign, reverse of what was observed on frequency model. The sum insured follows the same behaviour as past models, it is riskier as greater is sum insured.

Table 25: LOGNORMAL parameter estimation for claim average cost.

PARAMETER ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT	6.015	0.022
AUTOMOTIVE OCCUPANCY	-0.162	0.043
COMMERCIAL CENTRE OCCUPANCY	-0.159	0.051
PAPER AND LEATHER OCCUPANCY	-0.230	0.073
BETWEEN 190000€ 275000€ CONTENT SI	0.174	0.044
BETWEEN 275000€ -450000€ CONTENT SI	0.210	0.041
BETWEEN 450000€-850000€ CONTENT SI	0.281	0.042
ABOVE 850000€ CONTENT SUM INSURED	0.514	0.036

DEVIANCE RESIDUAL				
MIN	1^{ST} Q	MEDIAN	3^{RD} Q	MAX
-10.902	-1.062	0.034	0.938	6.719

Results for the GAMMA and IG models could be found on APENDIX A Tables 64 and 65.

5. MODELLING CLAIM COST FOR BUILDING.

On this part of the work, it is done the same analysis that has been done in the previous part but focusing only on building or claims that affects building.

5.1. BINOMIAL AND COST MODELS.

Table 28 represents the statistically significant estimated parameters using a TOBIT model for claims that affects only building. Occupancy is one of the covariates with LEISURE AND RECREATIONAL occupancy and AUTOMOTIVE WITHOUT WORKSHOPS as riskiest ones. Moreover as more employees the SME have, riskier is the insured object. This is as expected because Employees is a control covariate as it has been said on previous models. It has to be pointed out that if the insurer covers building as a first loss it is riskier than cover the total sum insured for building. The presence of other perils like building theft, glasses and electrics are also present as covariates. Additionally their sum insured is also present as statistically significant. These sum insured covariates have the same behaviour of Building Sum Insured: as greater is the amount, more risky is the SME. Geographically covariate are also present as certain province and inhabitants figure. It should be pointed out that Canary islands have a discharge as it had in the all cost models. Finally, biannual frequency is more risky than the other ways of policy payment.

Table 26: TOBIT parameters estimation for building claim cost.

PARAMETERS ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT 1	-61060.00	1116.00
INTERCEPT 2	10.33	0.01
BIANNUAL FREQUENCY	1459.00	408.50
SALAMANCA PROVINCE	-9704.00	3349.00
GRAN CANARIA PROVINCE	-6426.00	1417.00
TENERIFE PROVINCE	-7375.00	1361.00
FIRST LOSS AS BUILDING SUM INSURED	2927.00	892.60
BETWEEN 15000€-50000€ BUILDING SI	3278.00	860.10
BETWEEN 50000€-90000€ BUILDING SI	3355.00	929.40
BETWEEN 90000€-130000€ BUILDING SI	4100.00	937.30
BETWEEN 130000€-185000€ BUILDING SI	5296.00	933.50

BETWEEN 185000€-250000€ BUILDING SI	6187.00	921.10
BETWEEN 250000€-350000€ BUILDING SI	7620.00	917.30
BETWEEN 350000€-535000€ BUILDING SI	11330.00	902.30
BETWEEN 535000€-1050000€ BUILDING SI	14270.00	901.40
ABOVE 1050000€ BUILDING SI	17960.00	915.70
AUTOMOTIVE OCCUPANCY	8567.00	629.50
AUTOMOTIVE WITH WORKSHOP OCCUPANCY	2387.00	495.90
FARM OCCUPANCY	7222.00	883.90
FOOD HANDLING OCCUPANCY	3861.00	738.20
FOOD WAREHOUSE OCCUPANCY	3320.00	753.80
LEISURE AND RECREATIONAL OCCUPANCY	10390.00	819.30
COMMERCIAL CENTRE	6518.00	752.10
STONE MANUFACTURING OCCPANCY	6737.00	1732.00
1 EMPLOYEE	-2146.00	573.20
BETWEEN 4 AND 9 EMPLOYEES	2897.00	494.10
ABOVE 9 EMPLOYEES	3085.00	503.80
BUILDING THEFT NOT COVERED	-2718.00	700.90
GLASS NOT COVERED	-5070.00	715.40
BETWEEN 750€ - 1500€ GLASS SI	1542.00	557.30
BETWEEN 1500€ - 3000€ GLASS SI	2926.00	570.80
ABOVE 3000€ GLASS SI	4358.00	582.80
ABOVE 9000€ ELECTRIC SI	2543.00	425.90
BETWEEN 7500 AND 30000 INHABITANTS	1434.00	475.60
BETWEEN 30000 AND 130000 INHABITANTS	2471.00	502.60
ABOVE 130000 INHABITANTS	1807.00	493.70

PERSON RESIDUALS				
MIN	1ST Q	MEDIAN	3RD Q	MAX
-13520.00	-0.16	-0.11	-0.08	10.11
-0.67	-0.25	-0.21	-0.17	5202.82

AIC
270502

Now the building cost is analysed under the random variable of having a claim or not. On Table 29 it is shown the AIC figure for binomial models. The two models are quite similar from AIC perspective but it is taken the lower one.

Table 27: AIC for binomial model on Building claim.

AIC	
PROBIT	LOGIT
64568	64558

Table 30 shows the estimated parameters of the covariates using a LOGIT model. Occupancy is one of the covariates present as in previous models having LEISURE AND RECREATIONAL and COMMERCIAL CENTRE on the riskiest activities. Employee covariate is also present and it has the same behaviour as the number of employees increase, riskier is the SME. First loss on building is also a rechargeable covariate as in TOBIT model. Other perils covariates are also present being statistically significant if those perils are covered and the sum insured related with them. If glasses, theft building and electric are not covered is less risky. The sum insured has a direct relation with the riskiness as in previous models. Geographical variables are also present with more granularity in the case of province covariate. Biannual as way of payment continues being a recharge on the model as it has been seen in TOBIT model.

Table 28: LOGIT parameter estimation for Building claims.

PARAMETERS ESTIMATION		
VARIABLE	ESTIMATE	SD
BIANNUAL FREQUENCY PAYMENT	0.052	0.013
SALAMANCA PROVINCE	-0.319	0.104
CUENCA PROVINCE	0.217	0.060
BALEARES PROVINCE	-0.163	0.031
GRAN CANARIA PROVINCE	-0.227	0.044
TENERIFE PROVINCE	-0.267	0.043
FIRST LOSS AS BUILDING SUM INSURED	0.106	0.028
AUTOMOTIVE OCCUPANCY	0.332	0.020
AUTOMOTIVE WITH WORKSHOP OCCUPANCY	0.094	0.016
FARM OCCUPANCY	0.244	0.028
FOOD HANDLING OCCUPANCY	0.146	0.023
FOOD WAREHOUSE OCCUPANCY	0.121	0.024
LEISURE AND RECREATIONAL OCCUPANCY	0.413	0.026
COMMERCIAL CENTRE OCCUPANCY	0.274	0.024
STONE MANUFACTURING OCCUPANCY	0.175	0.057

BETWEEN 15000€-50000€ BUILDING SI	0.110	0.027
BETWEEN 50000€-90000€ BUILDING SI	0.114	0.029
BETWEEN 90000€-130000€ BUILDING SI	0.145	0.029
BETWEEN 130000€-185000€ BUILDING SI	0.186	0.029
BETWEEN 185000€-250000€ BUILDING SI	0.218	0.029
BETWEEN 250000€-350000€ BUILDING SI	0.271	0.029
BETWEEN 350000€-535000€ BUILDING SI	0.389	0.028
BETWEEN 535000€-1050000€ BUILDING SI	0.469	0.028
ABOVE 1050000€ BUILDING SI	0.625	0.028
1 EMPLOYEE	-0.077	0.018
BETWEEN 4 AND 9 EMPLOYEES	0.106	0.016
ABOVE 9 EMPLOYEES	0.114	0.016
THEFT BUILDING NOT COVERED	-0.119	0.023
GLASS NOT COVERED	-0.158	0.021
BETWEEN 1500€ - 3000€ GLASS SI	0.064	0.015
ABOVE 3000€ GLASS SI	0.133	0.015
ELECTRIC NOT COVERED	-0.136	0.028
ABOVE 9000€ ELECTRIC SI	0.092	0.013
BETWEEN 7500 <= 30000 INHABITANTS	0.071	0.015
BETWEEN 30000 AND 130000 INHABITANTS	0.092	0.016
ABOVE 130000 INHABITANTS	0.082	0.016

DEVIANCE RESIDUALS				
MIN	1^{ST} Q	MEDIAN	3^{RD} Q	MAX
-0.927	-0.449	-0.358	-0.285	3.030

Results for PROBIT model could be found on Table 66 of Apendix A.

For Building claim cost a continuous exponential distribution model is used. On Table 31 it could be seen the different AIC figures of the models. As LOGNORMAL is the lowest one it is taken this model to construct the claim cost model.

Table 29: AIC figure for Building cost.

AIC		
GAMMA	LOGNORMAL	INVERSE GAUSSIAN
177233	171921	190111

Table 32 shows the parameter estimation the statistically significant covariates. The model has only two main covariates, Building Sum Insured and LEISURE AND RECREATIONAL OCCUPANCY. As there are only a few statistically significant covariates R-square is near 0.

Table 30: LOGNORMAL parameter estimation for Building claim cost.

PARAMETERS ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT	6.488	0.022
BETWEEN 350000€-535000€ BUILDING SI	0.291	0.050
BETWEEN 535000€-1050000€ BUILDING SI	0.488	0.047
ABOVE 1050000€ BUILDING SI	0.743	0.040
LEISURE AND RECREATIONAL OCCUPANCY	0.307	0.062

DEVIANCE RESIDUALS				
MIN	1ST Q	MEDIAN	3RD Q	MAX
-11.077	-1.178	0.008	1.047	7.566

Results for the GAMMA and IG model could be found on APENDIX A Tables 67 and 68.

5.2. FREQUENCY-SEVERITY MODELS.

As it has been done in the previous chapter, to adjust the best Tweedie model for our data, a dispersion parameter estimation must be chosen. If it is selected the null deviance as a parameter, the response is to use 1.7 as the value of the dispersion parameter.

Table 31: Dispersion parameter estimation for TWEEDIE model for Building claim cost.

P PARAMETER	NULL DEVIANCE
1.4	23932419
1.5	13842867
1.6	8472503
1.7	5604295

Table 35 shows which covariates are statistically significant. It seems that only Building Sum Insured and Glasses Sum Insured are relevant under this model.

Table 32: TWEEDIE parameter estimation for Building claim cost.

PARAMETERS ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT	4.463	0.088
BETWEEN 350000€ AND 535000€ BUILDING SI	1.135	0.149

BETWEEN 535000€ AND 1050000€ BUILDING SI	1.719	0.140
ABOVE 1050000€ BUILDING SI	1.846	0.140
BETWEEN 1500€ AND 3000€ GLASS SI	0.749	0.130
ABOVE 3000€ GLASS SI	0.784	0.133
BETWEEN 750€ AND 1500€ GLASS SI	0.588	0.127

DEVIANCE	AIC
5604295	301511

PERSON RESIDUALS				
MIN	1ST Q	MEDIAN	3RD Q	MAX
-7.482	-5.672	-5.508	-5.043	213.297

Table 36 shows the AIC values for different discrete exponential models. Also QUASI POISSON model has been tested in order to verify if the dispersion parameter is different to 1. As the test revealed that dispersion parameter is different to 1, models taking into account the dispersion parameter has been used. Furthermore, as in the previous model, it has been checked if data could be explained based in a Zero Inflated model. In this particular case seems that NB and PIG model suit better our data.

Table 33: AIC figure for Building frequency.

AIC				
POISSON	NB	PIG	HURDLE	ZINB
79404	75155	75058	81979	76246

Table 37 displays statistically significant covariates for building frequency using PIG model. Occupancy is a significant covariate, more granular than previous models. LEISURE AND RECREATIONAL and AUTOMOTIVE are the riskiest occupancies. It has to be pointed out that it appear METALLURGY for first with a discount factor. Employee covariate is also statistically significant for this SME model, as we expected in a control covariate. The more employers the riskier is the insured object. It appears Tennant as a surcharge covariate for first time. Also another covariates that where present in previous model are statistically significant in this model, like Renovation work or First loss covariate. Perils like Theft Building, Glasses and Electric are present on this model and their sum insured also calibrates the frequency too. Building Sum Insured has a direct relation with the frequency. Geographical covariates are also present on this model. Inhabitants have the same behaviour as Employees and Islands seem less risky on frequency under this model.

Table 34: PIG parameter estimation for Building frequency.

PARAMETERS ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT	-2.547	0.065

BIANNUAL FREQUENCY PAYMENT	0.089	0.026
GUIPUZCOA PROVINCE	-0.214	0.074
CUENCA PROVINCE	0.457	0.120
BALEARES PROVINCE	-0.391	0.064
GRAN CANARIA PROVINCE	-0.402	0.088
TENERIFE PROVINCE	-0.458	0.088
RENOVATION WORK ONLY	-0.137	0.050
FIRST LOSS AS BUILDING SUM INSURED	0.257	0.059
TENNANT	0.160	0.043
BETWEEN 15000€-50000€ BUILDING SI	0.256	0.058
BETWEEN 50000€-90000€ BUILDING SI	0.223	0.065
BETWEEN 90000€-130000€ BUILDING SI	0.330	0.066
BETWEEN 130000€-185000€ BUILDING SI	0.406	0.066
BETWEEN 185000€-250000€ BUILDING SI	0.476	0.066
BETWEEN 250000€-350000€ BUILDING SI	0.576	0.066
BETWEEN 350000€-535000€ BUILDING SI	0.812	0.065
BETWEEN 535000€-1050000€ BUILDING SI	1.025	0.065
OVER 1050000€ BUILDING SI	1.410	0.065
AUTOMOTIVE OCCUPANCY	0.626	0.040
AUTOMOTIVE WITH WORKSHOP OCCUPANCY	0.138	0.034
FARM OCCUPANCY	0.400	0.060
FOOD HANDLING OCCUPANCY	0.215	0.048
FOOD WAREHOUSE OCCUPANCY	0.208	0.049
LEISURE AND RECREATIONAL OCCUPANCY	0.814	0.048
COMMERCIAL CENTRE OCCUPANCY	0.602	0.045
METALLURGY OCCUPANCY	-0.162	0.036
1 EMPLOYEE	-0.181	0.038
BETWEEN 4 AND 9 EMPLOYEES	0.187	0.031
OVER 9 EMPLOYEES	0.213	0.032
THEFT BUILDING NOT COVERED	-0.263	0.049
OVER 6000€ THEFT BUILDING SI	0.071	0.026
ELECTRIC NOT COVERED	-0.304	0.061
OVER 9000€ ELECTRIC SI	0.238	0.027
BETWEEN 7500 AND 30000 INHABITANTS	0.169	0.030

BETWEEN 30000 AND 130000 INHABITANTS	0.217	0.032
OVER 130000 INHABITANT	0.183	0.031
GLASS NOT COVERED	-0.343	0.046
BETWEEN 1500€ AND 3000€ GLASS SI	0.152	0.029
OVER 3000€ GLASS SI	0.300	0.030

Results for the POISSON, QUASI POISSON, NB, HURDLE and ZINB and IG model could be found on APENDIX A Table 69 and 73.

Table 38 shows the different AIC values for average cost model using a continuous exponential distribution. It has been chosen LOGNORMAL model because it has the lowest AIC value.

Table 35: AIC for Building average cost.

AIC		
GAMMA	LOGNORMAL	INVERSE GAUSSIAN
157157.3	152206	168179

Table 39 shows the statistically significant covariates for average cost model using a LOGNORMAL model. As it could be appreciated there are only two covariates: Occupancy and Renovation work. The model as the previous ones has a low R square value indicating that it could be improved. As a remark for this model it could be perceived that the sign of Automotive occupancy is inverse in FREQUENCY and SEVERITY models indicating that this activity has more frequency of low claims cost than average policies.

Table 36: LOGNORMAL parameter estimation for Building average cost.

PARAMETERS ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT	5.943	0.016
RENOVATION WORK ONLY	-0.280	0.044
AUTOMOTIVE	-0.137	0.044
STONE MANAGING	0.705	0.138

DEVIANCE RESIDUALS				
MIN	1^{ST} Q	MEDIAN	3^{RD} Q	MAX
-10.548	-1.075	0.028	0.919	6.984

Results for the GAMMA and IG model could be found on APENDIX A Tables 74 and 75.

6. MODELLING CLAIM COST FOR CONTENT.

6.1. BINOMIAL AND COST MODELS.

Table 41 shows the statistically significant covariates under TOBIT model. Occupancy is present but on this model the riskiest Occupancy is FOOD HANDLING instead of LEISURE AND RECREATIONAL on Building models. Employees is a statistically significant variate on this model too as it has been said in previous models. Perils like Theft and Electrical with its highest sum insured are present on this model. Related with Theft is also presence or not of security box. Another variable that is present on the model is the risk location, nearer to urban core or towns are less risky than outback zones. Content Sum Insured has a direct relation with the riskiness of the model. As higher is the sum insured, riskier is the object insured. Content model has lost its province geographical variable but it is still present Inhabitants covariate for its highest value.

Table 37: TOBIT parameters estimation for aggregated Content claim cost.

PARAMETERS ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT 1	-136000.000	3650.000
INTERCEPT 2	11.200	0.013
BIANNUAL PAYMENT FREQUENCY	5880.000	1190.000
QUARTERLY PAYMENT FREQUENCY	6000.000	1410.000
UNIQUE PAYMENT FREQUENCY	27700.000	8600.000
URBAN CORE RISK LOCATION	-15300.000	2560.000
TOWN RISK LOCATION	-13200.000	2780.000
THEFT NOT COVERED	-27900.000	2930.000
NO SECURITY BOX	-5930.000	1120.000
AUTOMOTIVE OCCUPANCY	12600.000	1780.000
FARM OCCUPANCY	13400.000	2650.000
FOOD HANDLING OCCUPANCY	6740.000	2000.000
LEISURE AND RECREATIONAL OCCUPANCY	17200.000	2510.000
COMMERCIAL CENTRE OCCUPANCY	18100.000	2220.000
BETWEEN 3 AND 9 EMPLOYEES	6020.000	1310.000
MORE THAN 9 EMPLOYEES	6710.000	1380.000
ELECTRIC NOT COVERED	-10700.000	2300.000

MORE THAN 9000€ ELECTRICAL SI	9040.000	1220.000
BETWEEN 60000€ AND 90000€ CONTENT SI	11400.000	2290.000
BETWEEN 90000€ AND 130000€ CONTENT SI	12400.000	2250.000
BETWEEN 130000€ AND 190000€ CONTENT SI	15000.000	2200.000
BETWEEN 190000€ AND 275000€ CONTENT SI	19900.000	2200.000
BETWEEN 275000€ AND 450000€ CONTENT SI	22900.000	2180.000
BETWEEN 35000€ AND 60000€ CONTENT SI	7200.000	2330.000
BETWEEN 450000€ AND 850000€ CONTENT SI	27500.000	2250.000
ABOVE 850000€ CONTENT SI	32300.000	2280.000
ABOVE 130000 INHABITANTS	-4900.000	1200.000

PEARSON RESIDUALS				
MIN	1ST Q	MEDIAN	3RD Q	MAX
-4312.492	-0.096	-0.0669	-0.0473	16.22
-0.357	-0.197	-0.1629	-0.1335	1425.14

Focusing on binomial model, Table 42 shows the AIC value for the PROBIT and LOGIT model. It could be seen that AIC is basically the same but I select PROBIT model as it has lower AIC.

Table 38: AIC for Content claim binomial modeling.

AIC	
PROBIT	LOGIT
39633	39636

Table 43 presents the statistically significant covariates under PROBIT model. Occupancy is one of the significant covariates where LEISURE AND RECREATIONAL is the riskiest one. That is a difference from TOBIT model. In this case Employees has a similar behaviour if there are more than 3 employees. Covariates that are related with specific coverage are present like Electric, Glass and Theft. Glass and Electric with its sum insured too. To be tenant or not is important for this model as it is risk location. The geographical variable is statistically significant only in its province covariate and not in its inhabitants. Finally payment frequency covariate is a recharge if it is not annual or unique.

Table 39: PROBIT parameter estimation for Content binomial.

PARAMETERS ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT	-1.855	0.042
BIANNUAL PAYMENT FREQUENCY	0.084	0.016
QUARTERLY PAYMENT FREQUENCY	0.071	0.019
NAVARRA PROVINCE	-0.232	0.057
GRAN CANARIA PROVINCE	-0.346	0.063
TENERIFE PROVINCE	-0.304	0.058
URBAN CORE RISK LOCATION	-0.231	0.034
TOWN RISK LOCATION	-0.225	0.038
TENNANT	-0.051	0.017
THEFT NOT COVERED	-0.408	0.039
NON SECURITY BOX	-0.106	0.015
AUTOMOTIVE OCCUPANCY	0.213	0.024
FARM OCCUPANCY	0.212	0.036
FOOD HANDLING OCCUPANCY	0.116	0.028
FOOD WAREHOUSE OCCUPANCY	0.095	0.029
LEISURE AND RECREATIONAL OCCUPANCY	0.255	0.034
COMMERCIAL CENTRE	0.264	0.030
BETWEEN 4 AND 9 EMPLOYEES	0.098	0.018
ABOVE 9 EMPLOYEES	0.083	0.019
ABOVE 3000€ GLASS SI	0.064	0.017
ABOVE 9000€ ELECTRIC SI	0.147	0.017
BETWEEN 35000€-60000€ CONTENT SI	0.122	0.031
BETWEEN 60000€-90000€ CONTENT SI	0.184	0.031
BETWEEN 90000€-130000€ CONTENT SI	0.195	0.030
BETWEEN 130000€-190000€ CONTENT SI	0.231	0.030
BETWEEN 190000€-275000€ CONTENT SI	0.294	0.030
BETWEEN 275000€-450000€ CONTENT SI	0.334	0.029
BETWEEN 450000€-850000€ CONTENT SI	0.385	0.030
ABOVE 850000€ CONTENT SI	0.449	0.031

DEVIANCE RESIDUALS				
MIN	1^{ST} Q	MEDIAN	3^{RD} Q	MAX
-0.659	-0.327	-0.262	-0.212	3.322

DEVIANCE	AIC
41425	39633

Results for LOGIT model could be found on APENDIX A Table 76.

For content claim cost it is used a continuous exponential distribution model which AIC value could be observed on Table 44. Under AIC criteria LOGNORMAL is chosen as the best model.

Table 40: AIC for Content claim cost.

AIC		
GAMMA	LOGNORMAL	INVERSE GAUSSIAN
96863	87824	97412

Table 45 shows the result of fitting a LOGNORMAL model on content claim cost. There are few covariates for this model and R-squared value is low. It has to be pointed out that all the occupancies have an inverse relation with the claim cost in this case.

Table 41: LOGNORMAL parameter estimation for Content claims cost.

PARAMETERS ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT	7.429	0.031
ABOVE 850000€ CONTENT SI	0.545	0.063
AUTOMOTIVE OCCUPANCY	-0.413	0.082
LEISURE AND RECREATIONAL OCCUPANCY	-0.379	0.117
COMMERCIAL CENTRE OCCUPANCY	-0.509	0.102

DEVIANCE RESIDUALS				
MIN	1^{ST} Q	MEDIAN	3^{RD} Q	MAX
-7.213	-1.200	-0.024	1.114	7.268

Results for the GAMMA and IG model could be found on APENDIX A Table 77 and 78.

6.2. FREQUENCY-SEVERITY MODELS.

For TWEEDIE it has been selected a dispersion parameter following null deviance criteria. Table 47 displays the different values for the null deviance based on deviation parameter.

Table 42: Dispersion parameter estimation for TWEEDIE model for Content claim cost.

P PARAMETER	NULL DEVIANCE
1.4	31993813
1.5	17667258
1.6	10350933

Table 48 displays the statistically significant covariates under TWEEDIE model. As it could be pointed out Occupancy is not a selected covariate and only large number of employees are statistically significant. Also Content Sum Insured and Inhabitants are selected as covariates. Sum insured as in other models has a direct relation with the claim cost as it could be observed.

Table 43: TWEEDIE parameter estimation for aggregated Content cost.

PARAMETERS ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT	4.739	0.108
ABOVE 9 EMPLOYEES	0.481	0.123
ABOVE 130000 INHABITANTS	-0.446	0.130
BETWEEN 90000€ AND 130000€ CONTENT SI	0.727	0.202
BETWEEN 130000€ AND 190000€ CONTENT SI	0.988	0.195
BETWEEN 190000€ AND 275000€ CONTENT SI	1.455	0.189
BETWEEN 275000€ AND 450000€ CONTENT SI	1.52	0.185
BETWEEN 450000€ AND 850000€ CONTENT SI	2.014	0.185
ABOVE 850000€ CONTENT SI	2.242	0.180

PEARSON RESIDUALS				
MIN	1ST Q	MEDIAN	3RD Q	MAX
-7.909	-6.608	-5.862	-5.256	207.092

AIC
216856

AIC values for frequency and severity models could be seen on Table 49. It has to be pointed out that dispersion parameter in QUASSI POISSON is the lowest one of all two parts model studied. This also is reflected in the models AIC between POISSON and NB that its difference is not so high as in previous models. PIG model is selected as it has the lowest AIC.

Table 44: AIC for Content claims frequency.

AIC				
POISSON	NB	PIG	HURDLE	ZINB
42187	41518	41491	44663	42062

Table 50 shows the statistically significant covariates using PIG model. As it has been checked in other models Occupancy and Employee are a main covariate for this models. LEISURE AND RECREATIONAL and COMMERCIAL CENTRE are the riskiest activities on this model. Risk location it is still a significant covariate but tenant has fallen as covariate. Specific perils and their sum insured are present in this model too. As it was in previous models there is a direct relation between Continent Sum Insured and riskiness. Geographical component persist in the model but just province, inhabitants has disappeared.

Table 45: PIG parameter estimation for Content claim frequency.

PARAMETERS ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT	-2.458	0.090
BIANNUAL PAYMENT FREQUENCY	0.154	0.036
QUARTERLY PAYMENT FREQUENCY	0.213	0.041
UNIQUE PAYMENT FREQUENCY	1.080	0.241
NAVARRA PROVINCE	-0.438	0.123
ALBACETE PROVINCE	-0.435	0.161
GRAN CANARIA PROVINCE	-0.642	0.140
TENERIFE PROVINCE	-0.504	0.133
URBAN CORE RISK LOCATION	-0.480	0.072
TOWN RISK LOCATION	-0.478	0.080
THEFT NOT COVERED	-1.025	0.097
NON SECURITY BOX	-0.199	0.034
AUTOMOTIVE OCCUPANCY	0.480	0.051
FARM OCCUPANCY	0.497	0.078
FOOD HANDLING OCCUPANCY	0.220	0.059
LEISURE AND RECREATIONAL OCCUPANCY	0.585	0.070
COMMERCIAL CENTRE OCCUPANCY	0.594	0.063

1 EMPLOYEE	-0.169	0.052
BETWEEN 4 AND 9 EMPLOYEES	0.149	0.043
ABOVE 9 EMPLOYEES	0.126	0.045
ABOVE 3000€ GLASS SI	0.123	0.036
BETWEEN 6000€ - 9000€ ELECTRIC SI	0.122	0.042
ABOVE >9000€ ELECTRIC SI	0.384	0.039
BETWEEN 60000€-90000€ CONTENT SI	0.299	0.064
BETWEEN 90000€-130000€ CONTENT SI	0.299	0.063
BETWEEN130000€-190000€ CONTENT SI	0.395	0.061
BETWEEN 190000€-275000€ CONTENT SI	0.520	0.060
BETWEEN 275000€-450000€ CONTENT SI	0.638	0.058
BETWEEN 450000€-850000€ CONTENT SI	0.769	0.060
ABOVE 850000€ CONTENT SI	0.963	0.061

Results for POISSON, QUASI POISSON, NB, HURDLE and ZINB models could be found on APENDIX A from Table 81 to Table 85.

For content claim average cost a continuous exponential family distribution is used. Table 51 shows the different AIC for the models analysed. LOGNORMAL model is selected as it has the lower AIC.

Table 46: AIC for Content claim average cost.

AIC		
GAMMA	LOGNORMAL	INVERSE GAUSSIAN
88488	86646	89594

Table 52 explains the statistically significant covariates for content average cost using a LOGNORMAL distribution. On this model there are a few covariates considered and also R-square is low but it continue being Occupancy and Content Sum Insured as covariates. As it was pointed out before occupancies has an inverse relation with cost on this specific model.

Table 47: LOGNORMAL parameter estimation for Content claim average cost.

PARAMETERS ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT	6.736	0.029
AUTOMOTIVE OCCUPANCY	-0.444	0.076

LEISURE AND RECREATIONAL OCCUPANCY	-0.443	0.109
COMMERCIAL CENTRE OCCUPANCY	-0.456	0.095
ABOVE 850000€ CONTENT SI	0.421	0.059

DEVIANCE RESIDUALS				
MIN	1^{ST} Q	MEDIAN	3^{RD} Q	MAX
-7.157	-1.102	-0.042	1.036	6.667

AIC
86646

Results for the GAMMA and IG model could be found on APENDIX A Table 86 and 87.

7. CONCLUSIONS.

This piece of work has shown different types of models for SME line of business. Specifically perils has been regrouped based on if affects Building or Content instead of doing for each peril. So in this case it has been followed the approach from Frees et alt. (2010) to group the perils in two basic coverages Building and Content.

Under this assumption different models has been tested to obtain a model for analizing claims. Basically it has been analysed three variables: the possibility of having a claim, the number of claims and the claim cost.

The most important covariate on the approach used here is Occupancy. Other covariates are control variates related with the SME size like: Employee number, sum insured from Content and Building perils. Generally speaking the more sum insured, riskier. It could be theorized what happens with Extension peril. Under this peril there is also a weather component so it could be that geographical covariates are explaining Extension peril. Another significant factors is Payment frequency, showing that biannual is risky than others.

Focusing on Occupancies and models obtained, Table 55 shows which occupancies are statistically different (X) from the base level (B). The Occupancies that are in the Building and Content model are on the Aggregated model, but the other way around is false as it is expected. The major number of exposure in the Aggregated models makes that filthy weighted occupancies could be significant in some cases, so it could be useful to have a join model instead of splitting it in more perils. This could indicate as Veilleux (2007) and Frees et alt. (2010) exposed that you could only split into perils when you have a large number of claims and if not, it is better to do some kind of join models as it has been done taking into consideration if claim affects to Building or Content.

Moreover, as we said before Theft, Electricity and Glasses perils have their own covariates in the three models, so it is a good way to have a model when the exposure and number of claims is not enough to split into perils.

Table 48: Main group covariates for modelling

		WHOLE CLAIM AMOUNT		BUILDING CLAIM AMOUNT		CONTENT CLAIM AMOUNT	
OCCUPANCY	WEIGHT	BINOMIAL – COST	FREQUENCY – SEVERITY	BINOMIAL – COST	FREQUENCY – SEVERITY	BINOMIAL – COST	FREQUENCY – SEVERITY
WAREHOUSES WITHOUT FOOD	21%	B	B	B	B	B	B
AUTOMOTIVE WITH WORKSHOP	18%	X		X	X	X	
METALLURGY	15%		X		X		
AUTOMOTIVE WITHOUT WORKSHOP	8%	X	X	X	X	X	X
FARMS	6%	X	X	X	X	X	X
FOOD WAREHOUSES	5%	X	X	X	X	X	
COMMERCIAL CENTRE	5%	X	X	X	X	X	X
FOOD HANDLING	5%	X	X	X	X	X	X
LEISURE OR RECREATIONAL	3%	X	X	X	X	X	X
PAPER/LEATHER	3%	X					
WOOD MANUFACTURING	3%						
CLOTHES MANUFACTURING	2%						
CHEMICAL MANUFACTURING	1%						
ENERGY PRODUCTION	1%	X					
LABORATORIES	1%						

STONE MANUFACTURING	1%	X	X	X			
OTHER MACHINERY	0%	X	X				

Now, talking about the models selected. If binomial and cost models are used, the combination of logit-lognormal work quite good except for content. The difference between Probit or Logit in Content is very small so I could resume that a logit-lognormal model is good enough for the three risk grouping.

The second approach based on frequency-severity points out that Content and Building models are under PIG-lognormal distribution, so over dispersion has to be taken into account in this model, but if Content and Building is aggregated, the model that suits better is a ZINB-lognormal model. The number of zeros in the two grouped risks makes that when it is an aggregate model a Zero Inflated looks better. Perhaps a Zero Inflated with PIG could be a better model for the aggregated risk. Nevertheless, the aggregated model points out the presence of over dispersion and a big number of zeros.

Going further, it seems that models for aggregated risk are better than ones split into Building and Content. That suggest that there is an interdependency between these two models and it should be studied further in other to obtain an aggregated model from Content and Building or as a base to study the dependencies between these two.

8. APPENDIX A: MODEL TABLES AND SIGNIFICATIVE COVARIATES.

Table 49: PROBIT Parameter estimation for binomial.

PARAMETERS ESTIMATION		
VARABLE	ESTIMATE	SD
INTERCEPT	-1.765	0.041
BIANNUAL FREQUENCY	0.064	0.012
NAVARRA	-0.161	0.041
BALEARS PROVINCE	-0.119	0.029
GRAN CANARIA PROVINCE	-0.286	0.043
TENERIFE PROVINCE	-0.287	0.041
RENOVATION WORK ONLY	-0.046	0.018
URBAN CORE RISK LOCATION	-0.108	0.029
SUBURBAN RISK LOCATION	-0.107	0.032
AUTOMOTIVE OCCUPANCY	0.344	0.019
AUTOMOTIVE WITH WORKSHOP OCCUPANCY	0.062	0.015
ENERGY OCCUPANCY	0.226	0.062
FARM OCCUPANCY	0.254	0.030
FOOD HANDLING OCCUPANCY	0.169	0.022
FOOD WAREHOUSE OCCUPANCY	0.133	0.023
LEISURE AND RECREATIONAL OCCUPANCY	0.437	0.026
COMMERCIAL CENTRE OCCUPANCY	0.344	0.023
OTHER MACHINERY OCCUPANCY	0.262	0.099
PAPER OR LEATHER OCCUPANCY	0.113	0.029
STONE MANUFACTURING OCCUPANCY	0.216	0.052
BETWEEN 15000€ AND 130000€ BUILDING SI	0.061	0.021
BETWEEN 130000€ AND 185000€ BUILDING SI	0.079	0.026
BETWEEN 185000€ AND 250000€ BUILDING SI	0.105	0.026
BETWEEN 250000€ AND 350000€ BUILDING SI	0.144	0.026
BETWEEN 350000€ AND 535000€ BUILDING SI	0.234	0.026
BETWEEN 535000€ AND 1050000€ BUILDING SI	0.292	0.027
ABOVE 1050000€ BUILDING SI	0.436	0.028

1 EMPLOYEE	-0.075	0.017
BETWEEN 4 AND 9 EMPLOYEES	0.086	0.015
ABOVE 9 EMPLOYEES	0.085	0.015
GLASS NOT COVERED	-0.199	0.020
BETWEEN 750€ - 1500€ GLASS SI	0.045	0.016
BETWEEN 1500€ - 3000€ GLASS SI	0.091	0.017
ABOVE 3000€ GLASS SI	0.157	0.017
BETWEEN 7500 AND 30000 INHABITANTS	0.063	0.014
BETWEEN 30000 AND 130000 INHABITANTS	0.076	0.015
ABOVE 130000 INHABITANTS	0.054	0.015
BETWEEN 35000€ AND 60000€ CONTENT SI	0.084	0.023
BETWEEN 60000€ AND 90000€ CONTENT SI	0.175	0.023
BETWEEN 90000€ AND 130000€ CONTENT SI	0.168	0.023
BETWEEN 130000€ AND 190000€ CONTENT SI	0.206	0.022
BETWEEN 190000€ AND 275000€ CONTENT SI	0.266	0.023
BETWEEN 275000€ AND 450000€ CONTENT SI	0.292	0.023
BETWEEN 450000€ AND 850000€ CONTENT SI	0.308	0.024
ABOVE 850000€ CONTENT SI	0.357	0.025

DEVIANCE	AIC
78532	74073

RESIDUAL DEVIANCE				
MIN	1^{ST} Q	MEDIAN	3^{RD} Q	MAX
-1.029	-0.527	-0.391	-0.380	2.965

Table 50: GAMMA Parameter estimation for claim cost.

PARAMETERS ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT	8.455	0.077
AUTOMOTIVE OCCUPANCY	-0.525	0.158
BETWEEN 190000€ AND 275000€ CONTENT SI	0.531	0.164
BETWEEN 275000€ AND 450000€ CONTENT SI	0.466	0.152
BETWEEN 450000€ AND 850000€ CONTENT SI	0.927	0.154
ABOVE 850000€ CONTENT SI	0.975	0.135

DEVIANCE	AIC
43369	229029

RESIDUAL DEVIANCE				
MIN	1^{ST} Q	MEDIAN	3^{RD} Q	MAX
-4.945	-1.986	-1.306	-0.494	23.147

Table 51: IG parameter estimation for claim cost.

PARAMETERS ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT	8.389	0.058
BETWEEN 190000€ AND 275000€ CONTENT SI	0.562	0.163
BETWEEN 275000€ AND 450000€ CONTENT SI	0.474	0.144
BETWEEN 450000€ AND 850000€ CONTENT SI	0.955	0.181
ABOVE 850000€ CONTENT SI	1.035	0.158

DEVIANCE	AIC
172.96	243832

RESIDUAL DEVIANCE				
MIN	1^{ST} Q	MEDIAN	3^{RD} Q	MAX
-9.580	-0.049	-0.022	-0.006	0.180

Table 52: POISSON parameter estimation for frequency.

PARAMETER ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT	-1.865	0.063
BIANNUAL PAYMENT FREQUENCY	0.105	0.018
QUARTERLY PAYMENT FREQUENCY	0.089	0.021
UNIQUE PAYMENT FREQUENCY	0.510	0.117
GUIPUZCOA PROVINCE	-0.185	0.050
TERUEL PROVINCE	-0.294	0.095
CUENCA PROVINCE	0.401	0.083
ALBACETE PROVINCE	-0.257	0.077
BALEARS PROVINCE	-0.304	0.044
GRAN CANARIA PROVINCE	-0.394	0.063
TENERIFE PROVINCE	-0.472	0.064
WORK REFORM ONLY	-0.189	0.034
URBAN CORE RISK LOCATION	-0.189	0.040
SUBURBAN RISK LOCATION	-0.203	0.043

FIRST LOSS AS BUILDING SUM INSURED	0.165	0.041
TENNANT	0.121	0.030
FIRE DETECTOR PRESENT	-0.100	0.024
THEFT NOT COVERED	-0.400	0.043
PARTIAL VALUE 10% OF CONTENT SI AS THEFT SI	-0.202	0.044
PARTIAL VALUE 20% OF CONTENT SI AS THEFT SI	-0.116	0.035
2 LAYERS GLASS	0.118	0.023
3 LAYERS GLASS	0.244	0.048
SECURITY BOX NOT PRESENT	-0.112	0.017
AUTOMOTIVE OCCUPANCY	0.545	0.026
ENERGY OCCUPANCY	0.258	0.095
FARM OCCUPANCY	0.432	0.044
FOOD HANDLING OCCUPANCY	0.167	0.031
FOOD WAREHOUSE OCCUPANCY	0.180	0.032
LEISURE AND RECREATIONAL OCCUPANCY	0.778	0.030
COMMERCIAL CENTRE OCCUPANCY	0.723	0.029
METALLURGY OCCUPANCY	-0.130	0.024
OTHER MACHINERY OCCUPANCY	0.613	0.145
STONE MANUFACTURING OCCUPANCY	0.242	0.068
BETWEEN 15000€ AND 130000€ BUILDING SI	0.142	0.037
BETWEEN 130000€ AND 185000€ BUILDING SI	0.186	0.047
BETWEEN 185000€ AND 250000€ BUILDING SI	0.231	0.046
BETWEEN 250000€ AND 350000€ BUILDING SI	0.302	0.046
BETWEEN 350000€ AND 535000€ BUILDING SI	0.471	0.045
BETWEEN 535000€ AND 1050000€ BUILDING SI	0.600	0.046
ABOVE 1050000€ BUILDING SI	0.908	0.046
1 EMPLOYEE	-0.137	0.027
BETWEEN 4 AND 9 EMPLOYEES	0.119	0.022
ABOVE 9 EMPLOYEES	0.112	0.023
BUILDING THEFT NOT COVERED	-0.157	0.038
GLASS NOT COVERED	-0.209	0.036
BETWEEN 750€ - 1500€ GLASS SI	0.080	0.026
BETWEEN 1500€ AND 3000€ GLASS SI	0.157	0.026
ABOVE 3000€ GLASS SI	0.268	0.026

ELECTRIC NOT COVERED	-0.279	0.044
ABOVE 9000€ ELECTRIC SI	0.196	0.018
BETWEEN 7500 AND 30000 INHABITANTS	0.132	0.021
BETWEEN 30000 AND 130000 INHABITANTS	0.163	0.021
MORE 130000 INHABITANTS	0.144	0.021
BETWEEN 60000€ AND 90000€ CONTENT SI	0.239	0.033
BETWEEN 90000€ AND 130000€ CONTENT SI	0.188	0.032
BETWEEN 130000€ AND 190000€ CONTENT SI	0.303	0.031
BETWEEN 190000€ AND 275000€ CONTENT SI	0.363	0.031
BETWEEN 275000€ AND 450000€ CONTENT SI	0.400	0.030
BETWEEN 450000€ AND 850000€ CONTENT SI	0.486	0.032
ABOVE 850000€ CONTENT SI	0.618	0.033

DEVIANCE	AIC
86659	104000

RESIDUAL DEVIANCE				
MIN	**1ST Q**	**MEDIAN**	**3RD Q**	**MAX**
-2.918	-0.577	-0.399	-0.221	25.971

Table 53: QUASI POISSON parameter estimation for frequency.

PARAMETER ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT	-1.798	0.082
BIANNUAL PAYMENT FREQUENCY	0.103	0.024
QUARTERLY PAYMENT FREQUENCY	0.087	0.028
UNIQUE PAYMENT FREQUENCY	0.494	0.159
GUIPUZCOA PROVINCE	-0.181	0.068
CUENCA PROVINCE	0.401	0.112
BALEARS PROVINCE	-0.304	0.060
GRAN CANARIA PROVINCE	-0.399	0.086
TENERIFE PROVINCE	-0.470	0.087
RENOVATION WORKS ONLY	-0.190	0.047
URBAN CORE RISK LOCATION	-0.203	0.053
SUBURBAN RISK LOCATION	-0.216	0.058

FIRST LOSS AS BUILDING SI	0.163	0.056
TENNANT	0.122	0.040
FIRE DETECTOR PRESENT	-0.107	0.032
THEFT NOT COVERED	-0.397	0.058
PARTIAL VALUE 10% OF CONTENT SI AS THEFT SI	-0.186	0.059
2 LAYER GLASS	0.117	0.031
3 LAYER GLASS	0.244	0.065
NON SECURITY BOX	-0.114	0.023
AUTOMOTIVE OCCUPANCY	0.537	0.034
FARM OCCUPANCY	0.397	0.060
FOOD HANDLING OCCUPANCY	0.150	0.041
FOOD WAREHOUSE OCCUPANCY	0.168	0.044
LEISURE AND RECREATIONAL OCCUPANCY	0.773	0.041
LOCLA WITHOUT ACTIVITY OCCUPANCY	0.713	0.039
METALLURGY OCCUPANCY	-0.139	0.033
OTHER MACHINE OCCUPANCY	0.587	0.197
BETWEEN 15000€ AND 130000€ BUILDING SI	0.139	0.051
BETWEEN 130000€ AND 185000€ BUILDING SI	0.182	0.064
BETWEEN 185000€ AND 250000€ BUILDING SI	0.227	0.063
BETWEEN 250000€-350000€ BUILDING SI	0.297	0.063
BETWEEN 350000€ AND 535000€ BUILDING SI	0.466	0.062
BETWEEN 535000€ AND 1050000€ BUILDING SI	0.597	0.062
ABOVE 1050000€ BUILDING SI	0.905	0.063
1 EMPLOYEE	-0.142	0.037
BETWEEN 4 AND 9 EMPLOYEES	0.122	0.029
ABOVE 9 EMPLOYEES	0.116	0.031
THEFT BUILD NOT COVERED	-0.154	0.052
GLASS NOT COVERED	-0.242	0.044
BETWEEN 1500€ AND 3000€ GLASS SI	0.109	0.027
ABOVE 3000€ GLASS SI	0.217	0.027
ELECTRIC NOT COVERED	-0.290	0.060
ABOVE 9000€ ELECTRIC SI	0.201	0.024
BETWEEN 7500 AND 30000 INHABITANTS	0.132	0.028
BETWEEN 30000 AND 130000 INHABITANTS	0.161	0.029

ABOVE 130000 INHABITANTS	0.139	0.029
BETWEEN 60000€ AND 90000€ CONTENT SI	0.238	0.045
BETWEEN 90000€ AND 130000€ CONTENT SI	0.188	0.044
BETWEEN 130000€ AND 190000€ CONTENT SI	0.300	0.042
BETWEEN 190000€ AND 275000€ CONTENT SI	0.363	0.042
BETWEEN 275000€ AND 450000€ CONTENT SI	0.400	0.041
BETWEEN 450000€ AND 850000€ CONTENT SI	0.483	0.043
ABOVE 850000€ CONTENT SI	0.608	0.045

DEVIANCE
86659

RESIDUAL DEVIANCE				
MIN	1^{ST} Q	MEDIAN	3^{RD} Q	MAX
-2.882	-0.577	-0.400	-0.221	26.004

DISPERSION PARAMETER
1.845545

Table 54: NB parameter estimation for frequency.

PARAMETER ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT	-1.756	0.079
BIANNUAL PAYMENT FREQUENCY	0.124	0.024
QUARTERLY PAYMENT FREQUENCY	0.110	0.029
UNIQUE PAYMENT FREQUENCY	0.757	0.201
CUENCA PROVINCE	0.405	0.112
BALEAR PROVINCE	-0.255	0.057
GRAN CANARIA PROVINCE	-0.452	0.084
TENERIFE PROVINCE	-0.432	0.081
REFORM WORK ONLY	-0.151	0.045
URBAN CORE RISK LOCATION	-0.278	0.055
SUBURBAN RISK LOCATION	-0.282	0.060
FIRST LOSS AS BUILDING SI	0.165	0.052
TENNANT	0.111	0.040

PARTIAL VALUE 10% OF CONTENT SI AS THEFT SI	-0.259	0.064
THEFT NOT COVERED	-0.494	0.050
2 LAYER GLASS	0.096	0.034
3 LAYER GLASS	0.272	0.074
NON SECURITY BOX	-0.127	0.023
AUTOMOTIVE OCCUPANCY	0.542	0.036
FARM OCCUPANCY	0.457	0.057
FOOD HANDLING OCCUPANCY	0.180	0.043
FOOD WAREHOUSE OCCUPANCY	0.192	0.044
LEISURE AND RECREATIONAL OCCUPANCY	0.814	0.047
COMMERCIAL CENTRE OCCUPANCY	0.676	0.045
METALLURGY OCCUPANCY	-0.126	0.031
OTHER MACHINERY OCCUPANCY	0.693	0.177
STONE MANUFACTURING OCCUPANCY	0.283	0.100
BETWEEN 130000€ AND 185000€ BUILDING SI	0.186	0.058
BETWEEN 15000€ AND 130000€ BUILDING SI	0.141	0.045
BETWEEN 185000€ AND 250000€ BUILDING SI	0.235	0.057
BETWEEN 250000€ AND 350000€ BUILDING SI	0.313	0.058
BETWEEN 350000€ AND 535000€ BUILDING SI	0.493	0.058
BETWEEN 535000€ AND 1050000€ BUILDING SI	0.621	0.059
ABOVE 1050000€ BUILDING SI	0.927	0.061
BETWEEN 4 AND 9 EMPLOYEES	0.125	0.029
ABOVE 9 EMPLOYEES	0.115	0.031
1 EMPLOYEE	-0.132	0.034
BETWEEN 1500€ AND 3000€ GLASS SI	0.091	0.027
ABOVE 3000€ GLASS SI	0.192	0.028
GLASS NOT COVERED	-0.279	0.039
ABOVE 9000€ ELECTRIC SI	0.190	0.025
ELECTRIC NOT COVERED	-0.312	0.051
ABOVE 130000 INHABITANTS	0.129	0.029
BETWEEN 30000 AND 130000 INHABITANTS	0.174	0.029
BETWEEN 7500AND 30000 INHABITANTS	0.159	0.028
BETWEEN 130000€ AND 190000€ CONTENT SI	0.301	0.040
BETWEEN 190000€ AND 275000€ CONTENT SI	0.371	0.040

BETWEEN 275000€ AND 450000€ CONTENT SI	0.409	0.040
BETWEEN 450000€ AND 850000€ CONTENT SI	0.489	0.043
BETWEEN 60000€ AND 90000€ CONTENT SI	0.241	0.041
BETWEEN 90000€ AND 130000€ CONTENT SI	0.184	0.041
ABOVE 850000€ CONTENT SI	0.630	0.046

DEVIANCE	AIC
46894	94225

RESIDUAL DEVIANCE				
MIN	1^{ST} Q	MEDIAN	3^{RD} Q	MAX
-1.351	-0.518	-0.378	-0.217	7.322

Table 55: PIG parameter estimation for frequency.

PARAMETER ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT	-1.804	0.082
BIANNUAL PAYMENT FREQUENCY	0.132	0.025
QUARTERLY PAYMENT FREQUENCY	0.113	0.031
UNIQUE PAYMENT FREQUENCY	0.647	0.204
GUIPUZCOA PROVINCE	-0.215	0.071
CUENCA PROVINCE	0.404	0.117
BALEAR PROVINCE	-0.282	0.060
GRAN CANARIA PROVINCE	-0.501	0.087
TENERIFE PROVINCE	-0.471	0.085
WORK REFORM ONLY	-0.165	0.047
URBAN CORE RISK LOCATION	-0.245	0.057
SUBURBAN RISK LOCATION	-0.250	0.062
FIRST LOSS AS BUILDING SI	0.160	0.054
TENNANT	0.119	0.042
PARTIAL VALUE 10% OF CONTENT SI AS THEFT SI	-0.278	0.066
THEFT NOT COVERED	-0.524	0.052
3 LAYER GLASS	0.274	0.076
NON SECURITY BOX	-0.147	0.024

AUTOMOTIVE OCCUPANCY	0.571	0.038
FARM OCCUPANCY	0.459	0.059
FOOD HANDLING OCCUPANCY	0.212	0.045
FOOD WAREHOUSE OCCUPANCY	0.196	0.046
LEISURE AND RECREATIONAL OCCUPANCY	0.816	0.049
COMMERCIAL CENTRE OCCUPANCY	0.694	0.046
METALLURGY OCCUPANCY	-0.133	0.033
OTHER MACHINE OCCUPANCY	0.643	0.187
STONE MANUFACTURING OCCUPANCY	0.298	0.104
BETWEEN 15000€ AND 130000€ BUILDING SI	0.155	0.047
BETWEEN 130000€ AND 185000€ BUILDING SI	0.205	0.060
BETWEEN 185000€ AND 250000€ BUILDING SI	0.258	0.060
BETWEEN 250000€ AND 350000€ BUILDING SI	0.338	0.060
BETWEEN 350000€ AND 535000€ BUILDING SI	0.528	0.060
BETWEEN 535000€ AND 1050000€ BUILDING SI	0.672	0.061
ABOVE 1050000€ BUILDING SI	0.979	0.063
1 EMPLOYEE	-0.132	0.036
BETWEEN 4 AND 9 EMPLOYEES	0.141	0.030
ABOVE 9 EMPLOYEES	0.123	0.032
GLASS NOT COVERED	-0.295	0.041
BETWEEN 1500€ AND 3000€ GLASS SI	0.101	0.028
ABOVE 3000€ GLASS SI	0.225	0.029
ELECTRIC NOT COVERED	-0.317	0.054
ABOVE 9000€ ELECTRIC	0.212	0.026
BETWEEN 7500AND 30000 INHABITANTS	0.156	0.029
BETWEEN 30000 AND 130000 INHABITANTS	0.171	0.030
ABOVE 130000 INHABITANTS	0.120	0.030
BETWEEN 60000€ AND 90000€ CONTENT SI	0.254	0.043
BETWEEN 90000€ AND 130000€ CONTENT SI	0.203	0.043
BETWEEN 130000€ AND 190000€ CONTENT SI	0.307	0.042
BETWEEN 190000€ AND 275000€ CONTENT SI	0.396	0.042
BETWEEN 275000€ AND 450000€ CONTENT SI	0.435	0.042
BETWEEN 450000€ AND 850000€ CONTENT SI	0.512	0.044
ABOVE 850000€ CONTENT SI	0.641	0.047

DEVIANCE	AIC
N/A	94177

Table 56: HURDLE parameter estimation for frequency.

PARAMETER ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT	0.366	0.052
UNIQUE PAYMENT FREQUENCY	0.843	0.142
ALBACETE PROVINCE	-0.474	0.150
WORK REFORM ONLY	-0.144	0.039
URBAN CORE RISK LOCATION	-0.212	0.051
SUBURBAN RISK LOCATION	-0.232	0.057
FIRE DETECTOR PRESENT	-0.117	0.035
2 LAYER GLASS	0.138	0.033
AUTOMOTIVE OCCUPANCY	0.181	0.036
LEISURE AND RECREATIONAL OCCUPANCY	0.506	0.039
COMMERCIAL CENTRE OCCUPANCY	0.456	0.036
OTHER MACHINE OCCUPANCY	0.554	0.193
ABOVE 9 EMPLOYEES	0.132	0.025
THEFT BUILDING NOT COVERED	-0.390	0.059
ABOVE 3000€ GLASS SI	0.165	0.025
ABOVE 9000€ ELECTRIC	0.249	0.024
ABOVE 130000 INHABITANTS	0.119	0.027
BETWEEN 30000 AND 130000 INHABITANTS	0.082	0.028
ZERO MODEL		
INTERCEPT	0.044	0.040
BETWEEN 15000€ AND 130000€ BUILDING SI	0.173	0.041
BETWEEN 130000€ AND 185000€ BUILDING SI	0.256	0.050
BETWEEN 185000€ AND 250000€ BUILDING SI	0.342	0.049
BETWEEN 250000€ AND 350000€ BUILDING SI	0.439	0.048
BETWEEN 350000€ AND 535000€ BUILDING SI	0.634	0.047
BETWEEN 535000€ AND 1050000€ BUILDING SI	0.811	0.047
ABOVE 1050000€ BUILDING SI	1.209	0.047
BETWEEN 60000€ AND 90000€ CONTENT SI	0.331	0.041
BETWEEN 90000€ AND 130000€ CONTENT SI	0.338	0.040
BETWEEN 130000€ AND 190000€ CONTENT SI	0.449	0.038
BETWEEN 190000€ AND 275000€ CONTENT SI	0.572	0.038
BETWEEN 275000€ AND 450000€ CONTENT SI	0.633	0.037
BETWEEN 450000€ AND 850000€ CONTENT SI	0.624	0.038
ABOVE 850000€ CONTENT SI	0.656	0.039

DEVIANCE	AIC
N/A	102110.8

PEARSON RESIDUAL				
MIN	1^{ST} Q	MEDIAN	3^{RD} Q	MAX
-0.554	-0.327	-0.271	-0.223	100.93

Table 57: GAMMA paraemeter estimation for claim avarage cost.

PARAMETER ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT	7.134	0.061
AUTOMOTIVE OCCUPANCY	-0.563	0.122
COMMERCIAL CENTRE OCCUPANCY	-0.460	0.144
PAPER OR LEATHER OCCUPANCY	-0.601	0.207
BETWEEN 190000€ AND 275000€ CONTENT SI	0.392	0.126
BETWEEN 275000€ AND 450000€ CONTENT SI	0.337	0.117
BETWEEN 450000€ AND 850000€ CONTENT SI	0.802	0.118
ABOVE 850000€ CONTENT SI	0.797	0.104

DEVIANCE	AIC
31519	199586

DEVIANCE RESIDUAL				
MIN	1^{ST} Q	MEDIAN	3^{RD} Q	MAX
-4.808	-1.663	-0.994	-0.221	17.593

Table 58: IG parameter estimation for claim average cost.

PARAMETER ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT	7.138	0.050
AUTOMOTIVE OCCUPANCY	-0.522	0.089
COMMERCIAL CENTRE OCCUPANCY	-0.486	0.105
PAPER OR LEATHER OCCUPANCY	-0.528	0.156
BETWEEN 190000€ AND 275000€ CONTENT SI	0.367	0.111
BETWEEN 275000€ AND 450000€ CONTENT SI	0.312	0.099
BETWEEN 450000€ AND 850000€ CONTENT SI	0.782	0.125
ABOVE 850000€ CONTENT SI	0.786	0.109

DEVIANCE	AIC
263.15	212760

DEVIANCE RESIDUAL				
MIN	1^{ST} Q	MEDIAN	3^{RD} Q	MAX
-10	-0.0695	-0.030	-0.005	0.305

Table 59: PROBIT parameter estimation for Building binomial.

PARAMETERS ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT	-1.863	0.025
BIANNUAL PAYMENT FREQUENCY	0.052	0.013
SALAMANCA PROVINCE	-0.319	0.104
CUENCA PROVINCE	0.213	0.060
BALEAR PROVINCE	-0.161	0.031
GRAN CANARIA PROVINCE	-0.228	0.044
TENERIFE PROVINCE	-0.269	0.043
AUTOMOTIVE OCCUPANCY	0.332	0.020
AUTOMOTIVE WITH WORKSHOP OCCUPANCY	0.095	0.016
FARM OCCUPANCY	0.240	0.028
FOOD HANDLING OCCUPANCY	0.145	0.023
FOOD WAREHOUSE OCCUPANCY	0.121	0.024
LEISURE AND RECREATIONAL OCCUPANCY	0.415	0.026
COMMERCIAL CENTRE OCCUPANCY	0.276	0.024
STONE MANUFACTURING OCCUPANCY	0.175	0.057
BETWEEN 15000€ AND 50000€ BUILDING SI	0.082	0.026
BETWEEN 5000€ AND 90000€ BUILDING SI	0.072	0.027
BETWEEN 9000€ AND 130000€ BUILDING SI	0.100	0.027
BETWEEN 130000€ AND 185000€ BUILDING SI	0.140	0.026
BETWEEN 185000€ AND 250000€ BUILDING SI	0.172	0.026
BETWEEN 250000€ AND 350000€ BUILDING SI	0.225	0.026
BETWEEN 350000€ AND 535000€ BUILDING SI	0.343	0.025
BETWEEN 535000€ AND 1050000€ BUILDING SI	0.422	0.025
ABOVE 1050000€ BUILDING SI	0.579	0.025
1 EMPLOYEE	-0.079	0.018
BETWEEN 4 AND 9 EMPLOYEES	0.108	0.016
ABOVE 9 EMPLOYEES	0.113	0.016
THEFT BUILDING NOT COVERED	-0.119	0.023

GLASS NOT COVERED	-0.161	0.021
BETWEEN 1500€ AND 3000€ GLASS SI	0.064	0.015
ABOVE 3000€ GLASS SI	0.133	0.015
ELECTRIC NOT COVERED	-0.136	0.028
ABOVE 9000€ ELECTRIC	0.094	0.013
BETWEEN 7500AND 30000 INHABITANTS	0.071	0.015
BETWEEN 30000 AND 130000 INHABITANTS	0.093	0.016
ABOVE 130000 INHABITANTS	0.083	0.016
FIRST LOSS AS BUILDING SI	0.106	0.028

DEVIANCE RESIDUALS				
MIN	1ST Q	MEDIAN	3RD Q	MAX
-0.928	-0.449	-0.359	-0.284	3.051

DEVIANCE	AIC
68476	64568

Table 60: GAMMA parameter estimation for Building claim cost.

PARAMETERS ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT	7.706	0.065
BETWEEN 350000€ AND 535000€ BUILDING SI	0.574	0.148
BETWEEN 535000€ AND 1050000€ BUILDING SI	0.971	0.137
ABOVE 1050000€ BUILDING SI	0.924	0.118
THEFT BUILDING NOT COVERED	0.606	0.200

DEVIANCE RESIDUALS				
MIN	1ST Q	MEDIAN	3RD Q	MAX
-4.791	-1.794	-1.123	-0.297	19.293

DEVIANCE	AIC
29661	177233

Table 61: IG parameter estimation for Building claim cost.

PARAMETERS ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT	7.735	0.053
BETWEEN 350000€ AND 535000€ BUILDING SI	0.632	0.162
BETWEEN 535000€ AND 1050000€ BUILDING SI	1.068	0.181
ABOVE 1050000€ BUILDING SI	0.906	0.140

AIC
190111

Table 62: POISSON parameter estimation for Building frequency.

PARAMETERS ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT	-2.567	0.059
BIANNUAL PAYMENT FREQUENCY	0.075	0.021
UNIQUE PAYMENT FREQUENCY	0.416	0.147
GUIPUZCOA PROVINCE	-0.218	0.061
SALAMANCA PROVINCE	-0.514	0.184
TERUEL PROVINCE	-0.362	0.120
CUENCA PROVINCE	0.464	0.098
BALEAR PROVINCE	-0.386	0.053
GRAN CANARIA PROVINCE	-0.347	0.072
TENERIFE PROVINCE	-0.469	0.074
WORK REFORM ONLY	-0.152	0.042
FIRST LOSS AS BUILDING SI	0.262	0.051
TENNANT	0.148	0.035
BETWEEN 15000€ AND 50000€ BUILDING SI	0.251	0.051
BETWEEN 5000€ AND 90000€ BUILDING SI	0.209	0.058
BETWEEN 9000€ AND 130000€ BUILDING SI	0.307	0.058
BETWEEN 130000€ AND 185000€ BUILDING SI	0.389	0.058
BETWEEN 185000€ AND 250000€ BUILDING SI	0.453	0.057
BETWEEN 250000€ AND 350000€ BUILDING SI	0.544	0.057
BETWEEN 350000€ AND 535000€ BUILDING SI	0.761	0.056
BETWEEN 535000€ AND 1050000€ BUILDING SI	0.972	0.056
ABOVE 1050000€ BUILDING SI	1.362	0.055
AUTOMOTIVE OCCUPANCY	0.605	0.031
AUTOMOTIVE WITH WORKSHOP OCCUPANCY	0.145	0.028
FARM OCCUPANCY	0.378	0.051
FOOD HANDLING OCCUPANCY	0.202	0.038
FOOD WAREHOUSE OCCUPANCY	0.191	0.040

LEISURE AND RECREATIONAL OCCUPANCY	0.813	0.035
COMMERCIAL CENTRE OCCUPANCY	0.669	0.033
METALLURGY OCCUPANCY	-0.162	0.031
1 EMPLOYEE	-0.184	0.032
BETWEEN 4 AND 9 EMPLOYEES	0.155	0.026
ABOVE 9 EMPLOYEES	0.179	0.026
THEFT BUILDING NOT COVERED	-0.277	0.043
ABOVE 6000€ THEFT BUILDING SI	0.091	0.020
ELECTRIC NOT COVERED	-0.314	0.054
ABOVE 9000€ ELECTRIC	0.234	0.021
BETWEEN 7500AND 30000 INHABITANTS	0.157	0.025
BETWEEN 30000 AND 130000 INHABITANTS	0.217	0.026
ABOVE 130000 INHABITANTS	0.201	0.025
GLASS NOT COVERED	-0.282	0.044
BETWEEN 750€ AND 1500€ GLASS SI	0.087	0.031
BETWEEN 1500€ AND 3000€ GLASS SI	0.224	0.031
ABOVE 3000€ GLASS SI	0.352	0.031

DEVIANCE RESIDUALS				
MIN	1ST Q	MEDIAN	3RD Q	MAX
-2.122	-0.486	-0.337	-0.193	25.597

DEVIANCE	AIC
65998	79404

Table 63: QUASI POISSON parameter estimation for Building frequency.

PARAMETERS ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT	-2.526	0.071
BIANNUAL PAYMENT FREQUENCY	0.074	0.026
GUIPUZCOA PROVINCE	-0.213	0.076
CUENCA PROVINCE	0.465	0.122
BALEAR PROVINCE	-0.383	0.067
GRAN CANARIA PROVINCE	-0.351	0.089
TENERIFE PROVINCE	-0.467	0.093

WORK REFORM ONLY	-0.155	0.052
FIRST LOSS AS BUILDING SI	0.262	0.064
TENNANT	0.155	0.044
BETWEEN 15000€ AND 50000€ BUILDING SI	0.251	0.064
BETWEEN 50000€ AND 90000€ BUILDING SI	0.209	0.072
BETWEEN 9000€ AND 130000€ BUILDING SI	0.309	0.073
BETWEEN 130000€ AND 185000€ BUILDING SI	0.390	0.073
BETWEEN 185000€ AND 250000€ BUILDING SI	0.457	0.072
BETWEEN 250000€ AND 350000€ BUILDING SI	0.546	0.071
BETWEEN 350000€ AND 535000€ BUILDING SI	0.765	0.070
BETWEEN 535000€ AND 1050000€ BUILDING SI	0.977	0.069
ABOVE 1050000€ BUILDING SI	1.369	0.069
AUTOMOTIVE OCCUPANCY	0.605	0.039
AUTOMOTIVE WITH WORKSHOP OCCUPANCY	0.141	0.035
FARM OCCUPANCY	0.363	0.063
FOOD HANDLING OCCUPANCY	0.197	0.047
FOOD WAREHOUSE OCCUPANCY	0.192	0.050
LEISURE AND RECREATIONAL OCCUPANCY	0.814	0.043
COMMERCIAL CENTRE OCCUPANCY	0.672	0.041
METALLURGY OCCUPANCY	-0.162	0.038
1 EMPLOYEE	-0.191	0.040
BETWEEN 4 AND 9 EMPLOYEES	0.156	0.032
ABOVE 9 EMPLOYEES	0.182	0.032
THEFT BUILDING NOT COVERED	-0.278	0.053
ABOVE 6000€ THEFT BUILDING SI	0.091	0.025
ELECTRIC NOT COVERED	-0.322	0.068
ABOVE 9000€ ELECTRIC	0.236	0.026
BETWEEN 7500AND 30000 INHABITANTS	0.166	0.031
BETWEEN 30000 AND 130000 INHABITANTS	0.225	0.032
ABOVE 130000 INHABITANTS	0.209	0.031
GLASS NOT COVERED	-0.330	0.051
BETWEEN 1500€ AND 3000€ GLASS SI	0.168	0.030
ABOVE 3000€ GLASS SI	0.296	0.030

DEVIANCE RESIDUALS				
MIN	1ST Q	MEDIAN	3RD Q	MAX
-2.136	-0.486	-0.337	-0.194	25.575

DEVIANCE
65998

DISPERSION PARAMETER
1.554868

Table 64: NB parameter estimation for Building frequency.

PARAMETERS ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT	-2.545	0.064
BIANNUAL PAYMENT FREQUENCY	0.082	0.025
GUIPUZCOA PROVINCE	-0.201	0.073
CUENCA PROVINCE	0.459	0.118
BALEAR PROVINCE	-0.373	0.063
GRAN CANARIA PROVINCE	-0.389	0.086
TENERIFE PROVINCE	-0.453	0.086
WORK REFORM ONLY	-0.128	0.049
FIRST LOSS AS BUILDING SI	0.260	0.058
TENNANT	0.157	0.042
BETWEEN 15000€ AND 50000€ BUILDING SI	0.253	0.057
BETWEEN 50000€ AND 90000€ BUILDING SI	0.219	0.064
BETWEEN 130000€ AND 185000€ BUILDING SI	0.402	0.065
BETWEEN 185000€ AND 250000€ BUILDING SI	0.472	0.065
BETWEEN 250000€ AND 350000€ BUILDING SI	0.568	0.065
BETWEEN 350000€ AND 535000€ BUILDING SI	0.796	0.064
BETWEEN 535000€ AND 1050000€ BUILDING SI	1.006	0.064
BETWEEN 9000€ AND 130000€ BUILDING SI	0.326	0.065
ABOVE 1050000€ BUILDING SI	1.407	0.064
AUTOMOTIVE OCCUPANCY	0.624	0.039
AUTOMOTIVE WITH WORKSHOP OCCUPANCY	0.134	0.033
FARM OCCUPANCY	0.395	0.058

FOOD HANDLING OCCUPANCY	0.204	0.047
FOOD WAREHOUSE OCCUPANCY	0.203	0.048
LEISURE AND RECREATIONAL OCCUPANCY	0.841	0.047
COMMERCIAL CENTRE OCCUPANCY	0.614	0.045
METALLURGY OCCUPANCY	-0.163	0.036
1 EMPLOYEE	-0.178	0.037
BETWEEN 4 AND 9 EMPLOYEES	0.177	0.031
ABOVE 9 EMPLOYEES	0.206	0.031
THEFT BUILDING NOT COVERED	-0.259	0.048
ABOVE 6000€ THEFT BUILDING SI	0.080	0.026
ELECTRIC NOT COVERED	-0.308	0.060
ABOVE 9000€ ELECTRIC	0.234	0.026
BETWEEN 7500AND 30000 INHABITANTS	0.173	0.030
BETWEEN 30000 AND 130000 INHABITANTS	0.221	0.031
ABOVE 130000 INHABITANTS	0.192	0.031
GLASS NOT COVERED	-0.336	0.045
BETWEEN 1500€ AND 3000€ GLASS SI	0.158	0.029
ABOVE 3000€ GLASS SI	0.285	0.029

DEVIANCE RESIDUALS				
MIN	1^{ST} Q	MEDIAN	3^{RD} Q	MAX
-1.272	-0.459	-0.328	-0.192	9.019

DEVIANCE	AIC
43299	75155

Table 65: HURDLE parameter estimation for Building frequency.

PARAMETERS ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT	-0.789	0.047
WORK REFORM ONLY	-0.221	0.059
AUTOMOTIVE OCCUPANCY	0.357	0.051
LEISURE AND RECREATIONAL OCCUPANCY	0.765	0.049
COMMERCIAL CENTRE OCCUPANCY	0.765	0.046
METALLURGY OCCUPANCY	-0.179	0.060

ABOVE 9 EMPLOYEES	0.153	0.034
ABOVE 6000€ THEFT BUILDING SI	0.258	0.035
ABOVE 9000€ ELECTRIC	0.344	0.035
BETWEEN 7500AND 30000 INHABITANTS	0.152	0.048
BETWEEN 30000 AND 130000 INHABITANTS	0.270	0.048
ABOVE 130000 INHABITANTS	0.306	0.046
BETWEEN 1500€ AND 3000€ GLASS SI	0.272	0.045
ABOVE 3000€ GLASS SI	0.286	0.044
ZERO MODEL		
INTERCEPT	-3.041	0.040
BETWEEN 15000€ AND 90000€ BUILDING SI	0.177	0.049
BETWEEN 9000€ AND 130000€ BUILDING SI	0.242	0.057
BETWEEN 130000€ AND 185000€ BUILDING SI	0.372	0.056
BETWEEN 185000€ AND 250000€ BUILDING SI	0.490	0.054
BETWEEN 250000€ AND 350000€ BUILDING SI	0.639	0.053
BETWEEN 350000€ AND 535000€ BUILDING SI	0.910	0.051
BETWEEN 535000€ AND 1050000€ BUILDING SI	1.153	0.049
ABOVE 1050000€ BUILDING SI	1.607	0.047

PEARSON RESIDUALS				
MIN	1^{ST} Q	MEDIAN	3^{RD} Q	MAX
-0.487	-0.296	-0.241	-0.220	99.243

AIC
81979

Table 66: ZINB parameter estimation for Building frequency.

PARAMETERS ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT	-1.308	0.035
METALLURGY OCCUPANCY	-0.359	0.033
GUIPUZCOA PROVINCE	-0.253	0.073
CUENCA PROVINCE	0.429	0.125
BALEAR PROVINCE	-0.359	0.063
GRAN CANARIA PROVINCE	-0.325	0.086

TENERIFE PROVINCE	-0.407	0.086
ABOVE 9 EMPLOYEES	0.134	0.024
ABOVE 6000€ THEFT BUILDING SI	0.130	0.026
ABOVE 9000€ ELECTRIC	0.384	0.025
BETWEEN 7500AND 30000 INHABITANTS	0.188	0.030
BETWEEN 30000 AND 130000 INHABITANTS	0.261	0.031
ABOVE 130000 INHABITANTS	0.296	0.030
BETWEEN 1500€ AND 3000€ GLASS SI	0.355	0.028
ABOVE 3000€ GLASS SI	0.544	0.028
ZERO MODEL		
INTERCEPT	0.625	0.068
BETWEEN 15000€ AND 90000€ BUILDING SI	-0.240	0.073
BETWEEN 9000€ AND 130000€ BUILDING SI	-0.385	0.088
BETWEEN 130000€ AND 185000€ BUILDING SI	-0.564	0.090
BETWEEN 185000€ AND 250000€ BUILDING SI	-0.730	0.091
BETWEEN 250000€ AND 350000€ BUILDING SI	-0.961	0.096
BETWEEN 350000€ AND 535000€ BUILDING SI	-1.512	0.114
BETWEEN 535000€ AND 1050000€ BUILDING SI	-2.224	0.172
ABOVE 1050000€ BUILDING SI	-11.333	17.298

PEARSON RESIDUALS				
MIN	1^{ST} Q	MEDIAN	3^{RD} Q	MAX
-0.682	-0.289	-0.224	-0.141	111.274

AIC
76246

Table 67: GAMMA parameter estimation for Building average cost.

PARAMETERS ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT	7.103	0.058
WORK REFORM ONLY	-0.480	0.163
AUTOMOTIVE OCCUPANCY	-0.441	0.165
STONE MANUFACTURING OCCUPANCY	1.547	0.512

DEVIANCE RESIDUALS				
MIN	1^{ST} Q	MEDIAN	3^{RD} Q	MAX
-4.627	-1.610	-0.920	-0.173	25.756

AIC
157157.3

Table 68: IG parameter estimation for Building average cost.

PARAMETERS ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT	7.102	0.058
WORK REFORM ONLY	-0.473	0.132
AUTOMOTIVE OCCUPANCY	-0.436	0.135

DEVIANCE RESIDUALS				
MIN	1^{ST} Q	MEDIAN	3^{RD} Q	MAX
-9.999	-0.079	-0.03	-0.005	0.526

AIC
168179

Table 69: LOGIT parameter estimation for binomial Content.

PARAMETERS ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT	-3.489	0.092
BIANNUAL PAYMENT FREQUENCY	0.178	0.036
QUARTERLY PAYMENT FREQUENCY	0.148	0.041
NAVARRA PROVINCE	-0.501	0.132
GRAN CANARIA PROVINCE	-0.765	0.147
TENERIFE PROVINCE	-0.709	0.138
URBAN CORE RISK LOCATION	-0.457	0.072
SUBURBAN RISK LOCATION	-0.438	0.079
TENNANT	-0.117	0.039
THEFT NOT COVERED	-0.998	0.101
NON SECURITY BOX	-0.235	0.033
AUTOMOTIVE OCCUPANCY	0.468	0.051
FARM OCCUPANCY	0.491	0.079

FOOD HANDLING OCCUPANCY	0.244	0.058
FOOD WAREHOUSE OCCUPANCY	0.204	0.063
LEISURE AND RECREATIONAL OCCUPANCY	0.541	0.071
COMMERCIAL CENTRE OCCUPANCY	0.591	0.063
BETWEEN 4 AND 9 EMPLOYEES	0.218	0.039
ABOVE 9 EMPLOYEES	0.188	0.042
ABOVE 3000€ GLASS SI	0.133	0.035
ABOVE 9000€ ELECTRIC	0.310	0.035
BETWEEN 35000€ AND 60000€ CONTENT SI	0.283	0.074
BETWEEN 60000€ AND 90000€ CONTENT SI	0.426	0.072
BETWEEN 90000€ AND 130000€ CONTENT SI	0.460	0.071
BETWEEN 130000€ AND 190000€ CONTENT SI	0.539	0.068
BETWEEN 190000€ AND 275000€ CONTENT SI	0.677	0.068
BETWEEN 275000€ AND 450000€ CONTENT SI	0.760	0.067
BETWEEN 450000€ AND 850000€ CONTENT SI	0.868	0.068
ABOVE 850000€ CONTENT SI	0.996	0.069

DEVIANCE RESIDUALS				
MIN	1ST Q	MEDIAN	3RD Q	MAX
-0.6901	-0.324	-0.261	-0.214	3.256

DEVIANCE	AIC
41425	39636

Table 70: GAMMA parameters estimation for Content claim cost.

PARAMETERS ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT	9.247	0.086
AUTOMOTIVE OCCUPANCY	-0.960	0.189
FARM OCCUPANCY	-0.844	0.255
LEISURE AND RECREATIONAL OCCUPANCY	-0.975	0.268
COMMERCIAL CENTRE OCCUPANCY	-0.885	0.236
ABOVE 3000€ GLASS SI	-0.347	0.131
BETWEEN 450000€ AND 850000€ CONTENT SI	0.666	0.170
ABOVE 850000€ CONTENT SI	0.721	0.153

DEVIANCE RESIDUALS				
MIN	1^{ST} Q	MEDIAN	3^{RD} Q	MAX
-4.17	-1.974	-1.317	-0.47	15.097

DEVIANCE	AIC
18371	96863

Table 71: IG parameter estimation for claim Content cost.

PARAMETERS ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT	9.306	0.099
AUTOMOTIVE OCCUPANCY	0.470	0.220
LEISURE AND RECREATIONAL OCCUPANCY	-0.996	0.175
COMMERCIAL CENTRE OCCUPANCY	-1.081	0.232

AIC
97412

Table 72: POISSON parameter estimation for claim Content frequency.

PARAMETERS ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT	-2.468	0.081
BIANNUAL PAYMENT FREQUENCY	0.149	0.033
QUARTERLY PAYMENT FREQUENCY	0.204	0.037
UNIQUE PAYMENT FREQUENCY	0.979	0.191
NAVARRA PROVINCE	-0.404	0.114
ALBACETE PROVINCE	-0.438	0.152
GRAN CANARIA PROVINCE	-0.601	0.130
TENERIFE PROVINCE	-0.512	0.126
URBAN CORE RISK LOCATION	-0.430	0.064
SUBURBAN RISK LOCATION	-0.418	0.071
TENNANT	-0.104	0.036
THEFT NOT COVERED	-1.021	0.094
NON SECURITY BOX	-0.202	0.031

AUTOMOTIVE OCCUPANCY	0.478	0.045
FARM OCCUPANCY	0.477	0.072
FOOD HANDLING OCCUPANCY	0.197	0.052
LEISURE AND RECREATIONAL OCCUPANCY	0.557	0.062
COMMERCIAL CENTRE OCCUPANCY	0.588	0.056
1 EMPLOYEE	-0.170	0.049
BETWEEN 4 AND 9 EMPLOYEES	0.147	0.039
ABOVE 9 EMPLOYEES	0.127	0.041
ABOVE 3000€ GLASS SI	0.123	0.032
BETWEEN 6000€ AND 9000€ ELECTRIC SI	0.118	0.038
ABOVE 9000€ ELECTRIC	0.372	0.035
BETWEEN 130000€ AND 190000€ CONTENT SI	0.387	0.057
BETWEEN 190000€ AND 275000€ CONTENT SI	0.509	0.056
BETWEEN 275000€ AND 450000€ CONTENT SI	0.628	0.054
BETWEEN 450000€ AND 850000€ CONTENT SI	0.753	0.055
BETWEEN 60000€ AND 90000€ CONTENT SI	0.298	0.061
BETWEEN 90000€ AND 130000€ CONTENT SI	0.302	0.059
ABOVE 850000€ CONTENT SI	0.931	0.056

DEVIANCE RESIDUALS				
MIN	1ST Q	MEDIAN	3RD Q	MAX
-1.343	-0.343	-0.241	-0.147	8.277

DEVIANCE	AIC
34177	42187

Table 73: QUASI POISSON parameter estimation for claim Content frequency.

PARAMETERS ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT	-2.468	0.089
BIANNUAL PAYMENT FREQUENCY	0.149	0.036
QUARTERLY PAYMENT FREQUENCY	0.204	0.040
UNIQUE PAYMENT FREQUENCY	0.979	0.208
NAVARRA PROVINCE	-0.404	0.125
ALBACETE PROVINCE	-0.438	0.165

GRAN CANARIA PROVINCE	-0.601	0.142
TENERIFE PROVINCE	-0.512	0.137
URBAN CORE RISK LOCATION	-0.430	0.069
SUBURBAN RISK LOCATION	-0.418	0.077
TENNANT	-0.104	0.039
THEFT NOT COVERED	-1.021	0.103
NON SECURITY BOX	-0.202	0.033
AUTOMOTIVE OCCUPANCY	0.478	0.050
FARM OCCUPANCY	0.477	0.078
FOOD HANDLING OCCUPANCY	0.197	0.057
LEISURE AND RECREATIONAL OCCUPANCY	0.557	0.068
COMMERCIAL CENTRE OCCUPANCY	0.588	0.061
1 EMPLOYEE	-0.170	0.053
BETWEEN 4 AND 9 EMPLOYEES	0.147	0.042
ABOVE 9 EMPLOYEES	0.127	0.044
ABOVE 3000€ GLASS SI	0.123	0.035
BETWEEN 6000€ AND 9000€ ELECTRIC SI	0.118	0.042
ABOVE 9000€ ELECTRIC	0.372	0.038
BETWEEN 60000€ AND 90000€ CONTENT SI	0.298	0.066
BETWEEN 90000€ AND 130000€ CONTENT SI	0.302	0.064
BETWEEN 130000€ AND 190000€ CONTENT SI	0.387	0.062
BETWEEN 190000€ AND 275000€ CONTENT SI	0.509	0.061
BETWEEN 275000€ AND 450000€ CONTENT SI	0.628	0.059
BETWEEN 450000€ AND 850000€ CONTENT SI	0.753	0.060
ABOVE 850000€ CONTENT SI	0.931	0.061

DEVIANCE RESIDUALS				
MIN	1ST Q	MEDIAN	3RD Q	MAX
-1.3436	-0.3432	-0.2416	-0.1479	8.2773

DEVIANCE	AIC
34177	42187

DISPERSION
1.18

Table 74: NB parameter estimation for Content claim frequency.

PARAMETERS ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT	-2.463	0.089
BIANNUAL PAYMENT FREQUENCY	0.149	0.036
QUARTERLY PAYMENT FREQUENCY	0.212	0.041
UNIQUE PAYMENT FREQUENCY	1.068	0.238
NAVARRA PROVINCE	-0.418	0.123
ALBACETE PROVINCE	-0.435	0.160
GRAN CANARIA PROVINCE	-0.623	0.140
TENERIFE PROVINCE	-0.498	0.132
URBAN CORE RISK LOCATION	-0.474	0.071
SUBURBAN RISK LOCATION	-0.474	0.079
THEFT NOT COVERED	-1.022	0.097
NON SECURITY BOX	-0.193	0.033
AUTOMOTIVE OCCUPANCY	0.482	0.050
FARM OCCUPANCY	0.494	0.078
FOOD HANDLING OCCUPANCY	0.212	0.058
LEISURE AND RECREATIONAL OCCUPANCY	0.592	0.069
COMMERCIAL CENTRE OCCUPANCY	0.589	0.063
1 EMPLOYEE	-0.169	0.052
BETWEEN 4 AND 9 EMPLOYEES	0.148	0.042
ABOVE 9 EMPLOYEES	0.127	0.044
ABOVE 3000€ GLASS SI	0.120	0.036
BETWEEN 6000€ AND 9000€ ELECTRIC SI	0.121	0.041
ABOVE 9000€ ELECTRIC	0.379	0.038
BETWEEN 130000€ AND 190000€ CONTENT SI	0.395	0.060
BETWEEN 190000€ AND 275000€ CONTENT SI	0.518	0.060
BETWEEN 275000€ AND 450000€ CONTENT SI	0.639	0.058
BETWEEN 450000€ AND 850000€ CONTENT SI	0.766	0.060
BETWEEN 60000€ AND 90000€ CONTENT SI	0.300	0.064
BETWEEN 90000€ AND 130000€ CONTENT SI	0.298	0.063
ABOVE 850000€ CONTENT SI	0.963	0.060

DEVIANCE RESIDUALS				
MIN	1ST Q	MEDIAN	3RD Q	MAX
-1.0456	-0.3335	-0.238	-0.147	5.2007

DEVIANCE	AIC
26182	41518

Table 75: HURDLE parameters estimation for Content claim frequency.

PARAMETERS ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT	-1.100	0.055
UNIQUE PAYMENT FREQUENCY	1.180	0.361
NAVARRA PROVINCE	0.672	0.236
AUTOMOTIVE OCCUPANCY	0.425	0.102
LEISURE AND RECREATIONAL OCCUPANCY	0.522	0.142
ABOVE 9000€ ELECTRIC	0.360	0.075
ZERO MODEL		
INTERCEPT	-3.765	0.036
BETWEEN 60000€ AND 90000€ CONTENT SI	0.349	0.064
BETWEEN 90000€ AND 130000€ CONTENT SI	0.412	0.061
BETWEEN 130000€ AND 190000€ CONTENT SI	0.583	0.058
BETWEEN 190000€ AND 275000€ CONTENT SI	0.752	0.056
BETWEEN 275000€ AND 450000€ CONTENT SI	0.912	0.054
BETWEEN 450000€ AND 850000€ CONTENT SI	1.067	0.053
ABOVE 850000€ CONTENT SI	1.320	0.049

Table 76: ZINB parameters estimation for Content claim frequency.

PARAMETERS ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT	-1.865	0.038
UNIQUE PAYMENT FREQUENCY	1.118	0.229
NAVARRA PROVINCE	-0.482	0.122
AUTOMOTIVE OCCUPANCY	0.378	0.048
LEISURE AND RECREATIONAL OCCUPANCY	0.520	0.069

ABOVE 9000€ ELECTRIC	0.511	0.034
ZERO MODEL		
INTERCEPT	0.664	0.076
BETWEEN 60000€ AND 90000€ CONTENT SI	-0.536	0.110
BETWEEN 90000€ AND 130000€ CONTENT SI	-0.617	0.108
BETWEEN 130000€ AND 190000€ CONTENT SI	-0.893	0.114
BETWEEN 190000€ AND 275000€ CONTENT SI	-1.253	0.132
BETWEEN 275000€ AND 450000€ CONTENT SI	-1.632	0.157
BETWEEN 450000€ AND 850000€ CONTENT SI	-2.211	0.247
ABOVE 850000€ CONTENT SI	-15.244	366.326

Table 77: GAMMA parameter estimation for Content claim avarage cost.

PARAMETERS ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT	7.930	0.085
AUTOMOTIVE OCCUPANCY	-0.925	0.165
LEISURE AND RECREATIONAL OCCUPANCY	-0.871	0.235
COMMERCIAL CENTRE OCCUPANCY	-0.679	0.206
PAPER OR LEATHER OCCUPANCY	-0.764	0.271
BETWEEN 190000€ AND 275000€ CONTENT SI	0.505	0.168
BETWEEN 275000€ AND 450000€ CONTENT SI	0.463	0.156
BETWEEN 450000€ AND 850000€ CONTENT SI	0.913	0.155
ABOVE 850000€ CONTENT SI	0.879	0.138

DEVIANCE RESIDUALS				
MIN	1ST Q	MEDIAN	3RD Q	MAX
-3.960	-1.780	-1.143	-0.324	13.419

DEVIANCE	AIC
15647	88488

Table 78: IG parameter estimation for Content claim average cost.

PARAMETERS ESTIMATION		
VARIABLE	ESTIMATE	SD
INTERCEPT	8.437	0.077
AUTOMOTIVE OCCUPANCY	-1.135	0.140
LEISURE AND RECREATIONAL OCCUPANCY	-1.027	0.199
COMMERCIAL CENTRE OCCUPANCY	-0.748	0.199

DEVIANCE RESIDUALS				
MIN	1^{ST} Q	MEDIAN	3^{RD} Q	MAX
-0.999	-0.055	-0.027	-0.007	0.175

DEVIANCE	AIC
32490	89598

9. APPENDIX B: R CODE.

```
library(MASS)
library(gamlss)
library(pscl)
library(VGAM)
library(weights)
library(statmod)
library(tweedie)
data1<-read.table(file="datos_definitivo_mod_2.csv",header=TRUE,sep=";")
data2<-read.table(file="adjuntos.csv",header=TRUE,sep=";")
# discrete covariates non binary
summary(data1$PAYFREQ)
wpct(data1$PAYFREQ,data1$EXPOSURE)
summary(data1$PROVINCE)
wpct(data1$PROVINCE,data1$EXPOSURE)
summary(data1$INSBUTYPE)
wpct(data1$INSBUTYPE,data1$EXPOSURE)
summary(data1$BUTYPE)
wpct(data1$BUTYPE,data1$EXPOSURE)
summary(data1$RISKLOC)
wpct(data1$RISKLOC,data1$EXPOSURE)
summary(data1$GLASSTYPE)
wpct(data1$GLASSTYPE,data1$EXPOSURE)
summary(data1$SECBOX)
wpct(data1$SECBOX,data1$EXPOSURE)
summary(data1$DEELECT)
wpct(data1$DEELECT,data1$EXPOSURE)
summary(data1$DEEOW)
```

```
wpct(data1$DEEOW,data1$EXPOSURE)
summary(data1$YEAR)
wpct(data1$YEAR,data1$EXPOSURE)
summary(data1$OCCUPANCY)
wpct(data1$OCCUPANCY,data1$EXPOSURE)
summary(data1$EMPLOYEES)
wpct(data1$EMPLOYEES,data1$EXPOSURE)
summary(data1$CONTENTNUM)
wpct(data1$CONTENTNUM,data1$EXPOSURE)
summary(data1$BULDINGNUM)
wpct(data1$BULDINGNUM,data1$EXPOSURE)
summary(data1$CLAIMNUM)
wpct(data1$CLAIMNUM,data1$EXPOSURE)
summary(data1$INHABITANTS)
wpct(data1$INHABITANTS,data1$EXPOSURE)
summary(data1$BUILDYEAR)
wpct(data1$BUILDYEAR,data1$EXPOSURE)
# binary covariates
summary(data1$FIREEXT)
wpct(data1$FIREEXT,data1$EXPOSURE)
summary(data1$BUFORM)
wpct(data1$BUFORM,data1$EXPOSURE)
summary(data1$OWNER)
wpct(data1$OWNER,data1$EXPOSURE)
summary(data1$HYDRANT)
wpct(data1$HYDRANT,data1$EXPOSURE)
summary(data1$DETECTOR)
wpct(data1$DETECTOR,data1$EXPOSURE)
summary(data1$VIGILANCE)
wpct(data1$VIGILANCE,data1$EXPOSURE)
```

```
summary(data1$DEEXGA)
wpct(data1$DEEXGA,data1$EXPOSURE)
summary(data1$THEFTFORM)
wpct(data1$THEFTFORM,data1$EXPOSURE)
summary(data1$PHYSPROTECT)
wpct(data1$PHYSPROTECT,data1$EXPOSURE)
summary(data1$CONALARM)
wpct(data1$CONALARM,data1$EXPOSURE)
summary(data1$THEFTNORM)
wpct(data1$THEFTNORM,data1$EXPOSURE)
summary(data1$INDCONTENT)
wpct(data1$INDCONTENT,data1$EXPOSURE)
summary(data1$INDBUILDING)
wpct(data1$INDBUILDING,data1$EXPOSURE)
summary(data1$INDCLAIM)
wpct(data1$INDCLAIM,data1$EXPOSURE)
# Continuous variables
summary(data1$EXPOSURE)
summary(data1$CONTENTCOST)
summary(data1$BUILDINGCOST)
summary(data1$CLAIMCOST)
summary(data1$BUILDINGEXP)
summary(data1$CONTENTEXP)
summary(data1$BUILDINGSI)
summary(data1$CONTENTSI)
summary(data2$CONTENTSI)
hist(data2$capicdo,breaks=50000,xlim=c(0,1000000))
summary(data2$THEFTBUILDSI)
wpct(data1$THEFTBUILDSI,data1$EXPOSURE)
```

```
summary(data2$GLASSSI)

wpct(data1$GLASSSI,data1$EXPOSURE)

summary(data2$ELECTSI)

wpct(data1$ELECTSI,data1$EXPOSURE)

summary(data2$THEFTSI)

summary(data2$BUILDINGSI)

data14<-as.data.frame(cbind(data2$BUILDINGSI,data2$CONTENTSI))

names(data14)[names(data14)=="V1"] <- "BUILDING"

names(data14)[names(data14)=="V2"] <- "CONTENT"

boxplot(data14,outline=FALSE)

data24<-as.data.frame(cbind(data2$GLASSSI, data2$ELECTSI,
data2$THEFTBUILDSI))

names(data24)[names(data24)=="V1"] <- "GLASS"

names(data24)[names(data24)=="V2"] <- "ELECTRIC"

names(data24)[names(data24)=="V3"] <- "THEFT BUILDING"

boxplot(data24,outline=FALSE)

#TOTAL

#TOBIT

data3<-read.table(file="datos_definitivo_mod_2TOBIT.csv",header=TRUE,sep=";")

summary(m <- vglm(CLAIMCOSTEXP ~ PAYFREQ + PROVINCE + INSBUTYPE +
RISKLOC + FIREEXT + BUFORM + OWNER + HYDRANT + DETECTOR +
VIGILANCE + DEEXGA + THEFTFORM  + PHYSPROTECT + CONALARM +
GLASSTYPE + SECBOX + THEFTNORM + DEELECT + DEEOW + YEAR +
OCCUPANCY + BUILDINGSI + EMPLOYEES + THEFTBUILDSI + GLASSSI +
ELECTSI + INHABITANTS + BUILDYEAR + CONTENTSI , tobit(Lower = 0), data
= data1))

OCCUPANCY2 <-relevel(as.factor(data3$OCCUPANCY),ref="BASE")

PROVINCE2 <-relevel(as.factor(data3$PROVINCE),ref="BASE")

summary(m <- vglm(CLAIMCOSTEXP ~ PAYFREQ + PROVINCE2 + INSBUTYPE

 + OWNER + THEFTFORM + BUILDINGSI +EMPLOYEES + GLASSSI + ELECTSI
+ CONTENTSI + OCCUPANCY2, tobit(Lower = 0), data = data3,maxit=1500))

AIC(m)
```

```
out <- tweedie.profile( data1$CLAIMCOSTEXP ~ 1, p.vec=seq(1.5, 1.7,
by=0.01),maxit=1000)

out$p.max

out$ci

# Tested from 1.4 to 1.7.

summary(tweedie<-glm(CLAIMCOSTEXP ~ PAYFREQ + PROVINCE +
INSBUTYPE + RISKLOC + FIREEXT + BUFORM + OWNER + HYDRANT +
DETECTOR + VIGILANCE + DEEXGA + THEFTFORM + PHYSPROTECT +
CONALARM + GLASSTYPE + SECBOX + THEFTNORM + DEELECT + DEEOW
+ YEAR + OCCUPANCY + BUILDINGSI + EMPLOYEES + THEFTBUILDSI +
GLASSSI + ELECTSI + INHABITANTS + BUILDYEAR +
CONTENTSI ,family=tweedie(var.power=1.70,link.power=0),data=data1,maxit=1000))

data4<-
read.table(file="datos_definitivo_mod_2TWEEEDIE.csv",header=TRUE,sep=";")

PAYFREQ2 = relevel(as.factor(data4$PAYFREQ),ref="BASE")

BUILDINGSIM <-as.factor(data4$BUILDINGSI)

BUILDINGSI2 = relevel(BUILDINGSIM,ref="BASE")

CONTENTSIM <-as.factor(data4$CONTENTSI)

CONTENTSI2 = relevel(CONTENTSIM,ref="BASE")

summary(tweedie<-glm(CLAIMCOSTEXP ~ PAYFREQ2 + INSBUTYPE +

+ OWNER + DETECTOR + BUILDINGSI2 +
CONTENTSI2,family=tweedie(var.power=1.70,link.power=0),data=data4,maxit=1000)
)

AICtweedie(tweedie)

#probit

data5<-read.table(file="datos_definitivo_mod_2PROBIT.csv",header=TRUE,sep=";")

METALLURGY OCCUPANCY<-as.factor(data5$OCCUPANCY)

OCCUPANCY2 = relevel(METALLURGY OCCUPANCY,ref="BASE")

PAYFREQ2 = relevel(as.factor(data5$PAYFREQ),ref="BASE")

PROVINCE2 = relevel(as.factor(data5$PROVINCE),ref="BASE")

summary(probit<-glm(INDCLAIM ~ PAYFREQ2 + PROVINCE2 + INSBUTYPE +
RISKLOC + OCCUPANCY2 + BUILDINGSI + EMPLOYEES + GLASSSI +
INHABITANTS + CONTENTSI, family=binomial(link=probit), data=data5,
maxit=1000))

#logit
```

```
#It is used the same file because it matches with the groups done

data6<-read.table(file="datos_definitivo_mod_2LOGIT.csv",header=TRUE,sep=";")

METALLURGY OCCUPANCY<-as.factor(data6$OCCUPANCY)

OCCUPANCY2 = relevel(METALLURGY OCCUPANCY,ref="BASE")

PAYFREQ2 = relevel(as.factor(data6$PAYFREQ),ref="BASE")

PROVINCE2 = relevel(as.factor(data6$PROVINCE),ref="BASE")

SECBOX2 = relevel(as.factor(data6$SECBOX),ref="BASE")

summary(logit<-glm(INDCLAIM ~ PAYFREQ2 + PROVINCE2 + INSBUTYPE +

 + SECBOX2 + OCCUPANCY2 + BUILDINGSI + EMPLOYEES + GLASSSI +
ELECTSI + INHABITANTS +CONTENTSI, family=binomial(link=logit),
data=data6,maxit=1000))

hist(data7$CLAIMCOSTEXPLN)

#GAMMA

data7<-read.table(file="datos_definitivo_mod_2SEV.csv",header=TRUE,sep=";")

gamma<-glm(CLAIMCOSTEXP ~ PAYFREQ + PROVINCE + INSBUTYPE +
RISKLOC + FIREEXT + BUFORM + OWNER + HYDRANT + DETECTOR +
VIGILANCE + DEEXGA + THEFTFORM  + PHYSPROTECT + CONALARM +
GLASSTYPE + SECBOX + THEFTNORM + DEELECT + DEEOW +
OCCUPANCY + BUILDINGSI + EMPLOYEES + THEFTBUILDSI + GLASSSI +
ELECTSI + INHABITANTS + BUILDYEAR + CONTENTSI,
family=Gamma(link="log"), data=data7, maxit=1000)

data8<-read.table(file="datos_definitivo_mod_2SEVGAMMA.csv",
header=TRUE,sep=";")

OCCUPANCY2 = relevel(as.factor(data8$OCCUPANCY),ref="BASE")

PAYFREQ2 = relevel(as.factor(data8$PAYFREQ),ref="BASE")

summary(gamma<-glm(CLAIMCOSTEXP ~ OCCUPANCY2 + CONTENTSI

                ,family=Gamma(link="log"),data=data8,maxit=1000))

summary(gamma<-glm(CLAIMCOSTEXP ~ OCCUPANCY2 + CONTENTSI

                ,family=Gamma(link="inverse"),data=data8,maxit=1000))

summary(gamma<-glm(CLAIMCOSTEXP ~ OCCUPANCY2 + CONTENTSI

                ,family=Gamma(link="identity"),data=data8, start=c(8.4,-
0.4,0.4,0.4,0.9,9.6),maxit=1000))

#LOGNORMAL

data7<-read.table(file="datos_definitivo_mod_2SEV.csv",header=TRUE,sep=";")
```

```
RISKLOC2<-relevel(as.factor(data7$RISKLOC),ref="NU")

OCCUPANCY2<-relevel(as.factor(data7$OCCUPANCY),ref="W")

lognormal<-lm(CLAIMCOSTEXPLN ~ RISKLOC2 + OCCUPANCY2 +
INHABITANTS + CONTENTSI,data=data7)

summary(lognormal)

data9<-
read.table(file="datos_definitivo_mod_2SEVLOGNORMAL.csv",header=TRUE,sep="
;")

RISKLOC2<-relevel(as.factor(data9$RISKLOC),ref="BASE")

OCCUPANCY2<-relevel(as.factor(data9$OCCUPANCY),ref="BASE")

lognormal<-lm(CLAIMCOSTEXPLN ~ RISKLOC2 + OCCUPANCY2 +
INHABITANTS + CONTENTSI,data=data9)

summary(lognormal)

#INVERSE GAUSSIAN

data8<-read.table(file="datos_definitivo_mod_2SEVGAMMA.csv",
header=TRUE,sep=";")

summary(gamma<-glm(CLAIMCOSTEXP ~ OCCUPANCY2 + CONTENTSI

            ,family=Gamma(link="log"),data=data8,maxit=1000))

ig<-glm(CLAIMCOSTEXP ~ OCCUPANCY2 +
CONTENTSI ,family=inverse.gaussian(link="log"),data=data8,start=coefficients(gamm
a),maxit=1000)

summary(ig)

#POISSON

OCCUPANCY2<-relevel(as.factor(data7$OCCUPANCY),ref="W")

POISSON1 = glm(CLAIMNUM~PAYFREQ + PROVINCE + INSBUTYPE +
RISKLOC + FIREEXT + BUFORM + OWNER + HYDRANT + DETECTOR +
VIGILANCE + DEEXGA + THEFTFORM + PHYSPROTECT + CONALARM +
GLASSTYPE + SECBOX + THEFTNORM + DEELECT + DEEOW + OCCUPANCY
+ BUILDINGSI + EMPLOYEES + THEFTBUILDSI + GLASSSI + ELECTSI +
INHABITANTS + BUILDYEAR + CONTENTSI + offset(CLAIMEXPLN),
family=poisson(link=log),data=data1)

summary(POISSON1)

OCCUPANCY2<-relevel(as.factor(data1$OCCUPANCY),ref="W")

POISSON1 = glm(CLAIMNUM~PAYFREQ + PROVINCE + INSBUTYPE +
RISKLOC + BUFORM + OWNER + HYDRANT + DETECTOR + THEFTFORM +
GLASSTYPE + SECBOX + OCCUPANCY2 + BUILDINGSI + EMPLOYEES +
```

```
THEFTBUILDSI + GLASSSI + ELECTSI + INHABITANTS + CONTENTSI
+offset(CLAIMEXPLN), family=poisson(link=log),data=data1)

summary(POISSON1)

data10<-
read.table(file="datos_definitivo_mod_2POISSON.csv",header=TRUE,sep=";")

OCCUPANCY2<-relevel(as.factor(data10$OCCUPANCY),ref="BASE")

PROVINCE2<-relevel(as.factor(data10$PROVINCE),ref="BASE")

GLASSTYPE2<-relevel(as.factor(data10$GLASSTYPE),ref="BASE")

THEFTFORM2<-relevel(as.factor(data10$THEFTFORM),ref="BASE")

POISSON1 = glm(CLAIMNUM~PAYFREQ + PROVINCE2 + INSBUTYPE +
RISKLOC + BUFORM + OWNER +  DETECTOR + THEFTFORM2 +
GLASSTYPE2 + SECBOX + OCCUPANCY2 + BUILDINGSI + EMPLOYEES +
THEFTBUILDSI + GLASSSI + ELECTSI + INHABITANTS   + CONTENTSI
+offset(CLAIMEXPLN), family=poisson(link=log), data=data10)

summary(POISSON1)

#QPOISSON

data11<-
read.table(file="datos_definitivo_mod_2QPOISSON.csv",header=TRUE,sep=";")

OCCUPANCY2<-relevel(as.factor(data11$OCCUPANCY),ref="BASE")

PROVINCE2<-relevel(as.factor(data11$PROVINCE),ref="BASE")

GLASSTYPE2<-relevel(as.factor(data11$GLASSTYPE),ref="BASE")

THEFTFORM2<-relevel(as.factor(data11$THEFTFORM),ref="BASE")

QPOISSON1 = glm(CLAIMNUM~PAYFREQ + PROVINCE2 + INSBUTYPE +
RISKLOC + BUFORM + OWNER +  DETECTOR + THEFTFORM2 +
GLASSTYPE2 + SECBOX + OCCUPANCY2 + BUILDINGSI + EMPLOYEES +
THEFTBUILDSI + GLASSSI + ELECTSI + INHABITANTS + CONTENTSI

+ offset(CLAIMEXPLN),family=quasipoisson(link=log),data=data11)

summary(QPOISSON1)

data12<-read.table(file="datos_definitivo_mod_2NB.csv",header=TRUE,sep=";")

OCCUPANCY2<-relevel(as.factor(data12$OCCUPANCY),ref="BASE")

PROVINCE2<-relevel(as.factor(data12$PROVINCE),ref="BASE")

GLASSTYPE2<-relevel(as.factor(data12$GLASSTYPE),ref="BASE")

THEFTFORM2<-relevel(as.factor(data12$THEFTFORM),ref="BASE")

NB1 = glm.nb(CLAIMNUM~PAYFREQ + PROVINCE2 + INSBUTYPE + RISKLOC
```

```
+ BUFORM + OWNER +   THEFTFORM2 + GLASSTYPE2 + SECBOX

+ OCCUPANCY2 + BUILDINGSI + EMPLOYEES +  GLASSSI + ELECTSI +
INHABITANTS + CONTENTSI + offset(CLAIMEXPLN),link=log,data=data12)

summary(NB1)

data13<-read.table(file="datos_definitivo_mod_2PIG.csv",header=TRUE,sep=";")

OCCUPANCY2<-relevel(as.factor(data13$OCCUPANCY),ref="BASE")

PROVINCE2<-relevel(as.factor(data13$PROVINCE),ref="BASE")

GLASSTYPE2<-relevel(as.factor(data13$GLASSTYPE),ref="BASE")

THEFTFORM2<-relevel(as.factor(data13$THEFTFORM),ref="BASE")

PIG1 = gamlss(CLAIMNUM~PAYFREQ + PROVINCE2 + INSBUTYPE + RISKLOC

 + BUFORM + OWNER +  THEFTFORM2 + GLASSTYPE2 + SECBOX +
OCCUPANCY2 + BUILDINGSI + EMPLOYEES + GLASSSI + ELECTSI +
INHABITANTS + CONTENTSI + offset(CLAIMEXPLN),family=PIG,data=data13)

summary(PIG1)

data14<-
read.table(file="datos_definitivo_mod_2HURDLE.csv",header=TRUE,sep=";")

OCCUPANCY2<-relevel(as.factor(data14$OCCUPANCY),ref="BASE")

PROVINCE2<-relevel(as.factor(data14$PROVINCE),ref="BASE")

GLASSTYPE2<-relevel(as.factor(data14$GLASSTYPE),ref="BASE")

THEFTFORM2<-relevel(as.factor(data14$THEFTFORM),ref="BASE")

HURDLE1=hurdle(CLAIMNUM~PAYFREQ + PROVINCE2 + INSBUTYPE +
RISKLOC +  DETECTOR + GLASSTYPE2 + OCCUPANCY2 + EMPLOYEES +
THEFTBUILDSI + GLASSSI + ELECTSI + INHABITANTS
+offset(CLAIMEXPLN) | BUILDINGSI + CONTENTSI,dist="poisson",data=data14)

summary(HURDLE1)

data15<-read.table(file="datos_definitivo_mod_2ZNB.csv",header=TRUE,sep=";")

OCCUPANCY2<-relevel(as.factor(data15$OCCUPANCY),ref="BASE")

PROVINCE2<-relevel(as.factor(data15$PROVINCE),ref="BASE")

GLASSTYPE2<-relevel(as.factor(data15$GLASSTYPE),ref="BASE")

THEFTFORM2<-relevel(as.factor(data15$THEFTFORM),ref="BASE")

ZBN1=zeroinfl(CLAIMNUM~PAYFREQ + PROVINCE2 + INSBUTYPE +
RISKLOC + THEFTFORM2 + GLASSTYPE2 + OCCUPANCY2  + EMPLOYEES +
THEFTBUILDSI + GLASSSI + ELECTSI + INHABITANTS |BUILDINGSI +
CONTENTSI,offset=CLAIMEXPLN,dist="negbin",data=data15)
```

```
summary(ZBN1)

#ZPIG1=gamlss(CLAIMNUM~

#              PAYFREQ + PROVINCE2 + INSBUTYPE + RISKLOC

#            + THEFTFORM2

#           + GLASSTYPE2

##           + OCCUPANCY2

##            + EMPLOYEES + THEFTBUILDSI + GLASSSI + ELECTSI +
INHABITANTS

##             + offset(CLAIMEXPLN)

##            ,nu.formula = ~ BUILDINGSI + CONTENTSI,family = ZIPIG,data=data15)

#summary(ZPIG1)

#CME

data16<-
read.table(file="datos_definitivo_mod_2SEVCME.csv",header=TRUE,sep=";")

gamma<-glm(CLAIMCME ~ PROVINCE + INSBUTYPE + RISKLOC + FIREEXT

 + BUFORM + OWNER + HYDRANT + DETECTOR + VIGILANCE + DEEXGA +
THEFTFORM + PHYSPROTECT + CONALARM + GLASSTYPE + SECBOX +
THEFTNORM + DEELECT + DEEOW + OCCUPANCY + BUILDINGSI  +
EMPLOYEES + THEFTBUILDSI + GLASSSI + ELECTSI + INHABITANTS

+ BUILDYEAR +CONTENTSI , family=Gamma(link="log"), data=data16,
maxit=1000)

summary(gamma)

data17<-
read.table(file="datos_definitivo_mod_2SEVCMEGAMMA.csv",header=TRUE,sep=";
")

OCCUPANCY2<-relevel(as.factor(data17$OCCUPANCY),ref="BASE")

gamma<-glm(CLAIMCME ~ OCCUPANCY2
+CONTENTSI ,family=Gamma(link="log"),data=data17,maxit=1000)

summary(gamma)

lognormal<-lm(CLAIMCMELN ~ OCCUPANCY2
+CONTENTSI ,data=data17,maxit=1000)

summary(lognormal)
```

```
ig<-glm(CLAIMCME ~ OCCUPANCY2 +
CONTENTSI ,family=inverse.gaussian(link="log"),data=data17,start=coefficients(gam
ma),maxit=1000)

summary(ig)

#BUILDING

#TOBIT

data18<-
read.table(file="datos_definitivo_mod_2BUILDINGTOBIT.csv",header=TRUE,sep=";"
)

OCCUPANCY2<-relevel(as.factor(data18$OCCUPANCY),ref="BASE")

PAYFREQ2<-relevel(as.factor(data18$PAYFREQ),ref="BASE")

DEEOW2<-relevel(as.factor(data18$DEEOW),ref="BASE")

PROVINCE2<-relevel(as.factor(data18$PROVINCE),ref="BASE")

ELECTSI2<-relevel(as.factor(data18$ELECTSI),ref="BASE")

summary(m <- vglm(BUILDINCOSTEXPUES ~ PAYFREQ + PROVINCE2

+ BUFORM +  BUILDINGSI + OCCUPANCY2+ EMPLOYEES +THEFTBUILDSI +
GLASSSI + ELECTSI2 + INHABITANTS , tobit(Lower = 0), data = data1))


data19<-
read.table(file="datos_definitivo_mod_2BUILDINGTWEEDIE.csv",header=TRUE,sep
=";")

OCCUPANCY2<-relevel(as.factor(data19$OCCUPANCY),ref="BASE")

summary(tweedie<-glm(BUILDINCOSTEXPUES ~   BUILDINGSI+ GLASSSI

                ,family=tweedie(var.power=1.70,link.power=0),data=data19,maxit=1000))

data20<-
read.table(file="datos_definitivo_mod_2BUILDINGPROBIT.csv",header=TRUE,sep="
;")

OCCUPANCY2 = relevel(data20$OCCUPANCY,ref="BASE")

PROVINCE2<-relevel(as.factor(data20$PROVINCE),ref="BASE")

summary(probit<-glm(INDBUILDING ~ PAYFREQ + PROVINCE2 + BUFORM

 + OCCUPANCY2 + BUILDINGSI + EMPLOYEES + THEFTBUILDSI + GLASSSI
+ ELECTSI + INHABITANTS ,family=binomial(link=probit), data=data20,
maxit=1000))

summary(logit<-glm(INDBUILDING ~ PAYFREQ + PROVINCE2 + BUFORM
```

```
+ OCCUPANCY2 + BUILDINGSI  + EMPLOYEES + THEFTBUILDSI + GLASSSI
+ ELECTSI + INHABITANTS, family=binomial(link=logit),data=data20,maxit=1000))

data21<-
read.table(file="datos_definitivo_mod_2SEVBUILDING.csv",header=TRUE,sep=";")

OCCUPANCY2 = relevel(data21$OCCUPANCY,ref="BASE")

summary(gamma<-glm(BUILDINCOSTEXPUES ~  BUILDINGSI

+  THEFTBUILDSI,family=Gamma(link="log"),data=data21,maxit=10000))

lognormal<-lm(log(BUILDINCOSTEXPUES) ~ BUILDINGSI

            +  OCCUPANCY2,data=data21)

summary(lognormal)

ig<-glm(BUILDINCOSTEXPUES ~  BUILDINGSI  +  OCCUPANCY2 ,
family=inverse.gaussian(link="log"),data=data21,start=coefficients(lognormal),maxit=1
000)

summary(ig)

#TWO-PART

data22<-
read.table(file="datos_definitivo_mod_2BUILDINGPOISSON.csv",header=TRUE,sep
=";")

OCCUPANCY2 = relevel(data22$OCCUPANCY,ref="BASE")

PROVINCE2 =  relevel(data22$PROVINCE,ref="BASE")

POISSON1 = glm(BULDINGNUM~PAYFREQ+ PROVINCE2 + INSBUTYPE

 + BUFORM + OWNER + BUILDINGSI + OCCUPANCY2 + EMPLOYEES +
THEFTBUILDSI +  ELECTSI + INHABITANTS + GLASSSI
+offset(CLAIMEXPLN),family=poisson(link=log),data=data22)

summary(POISSON1)

data23<-
read.table(file="datos_definitivo_mod_2BUILDINGQPOISSON.csv",header=TRUE,se
p=";")

OCCUPANCY2 = relevel(data23$OCCUPANCY,ref="BASE")

PROVINCE2 =  relevel(data23$PROVINCE,ref="BASE")

QPOISSON1 = glm(BULDINGNUM~PAYFREQ+ PROVINCE2 + INSBUTYPE

 + BUFORM + OWNER + BUILDINGSI + OCCUPANCY2 + EMPLOYEES +
THEFTBUILDSI +  ELECTSI + INHABITANTS + GLASSSI
+offset(CLAIMEXPLN),family=quasipoisson(link=log),data=data23)

summary(QPOISSON1)
```

```
NB1 = glm.nb(BULDINGNUM~PAYFREQ+ PROVINCE2 + INSBUTYPE

+ BUFORM + OWNER + BUILDINGSI + OCCUPANCY2 + EMPLOYEES +
THEFTBUILDSI +  ELECTSI + INHABITANTS + GLASSSI
+offset(CLAIMEXPLN),link=log,data=data23)

summary(NB1)

PIG1 = gamlss(BULDINGNUM~PAYFREQ+ PROVINCE2 + INSBUTYPE

+ BUFORM + OWNER + BUILDINGSI + OCCUPANCY2 + EMPLOYEES +
THEFTBUILDSI +  ELECTSI + INHABITANTS + GLASSSI
+offset(CLAIMEXPLN),family=PIG,data=data23)

summary(PIG1)

data24<-
read.table(file="datos_definitivo_mod_2BUILDINGHURDLE.csv",header=TRUE,sep
=";")

OCCUPANCY2 = relevel(data24$OCCUPANCY,ref="BASE")

PROVINCE2 =  relevel(data24$PROVINCE,ref="BASE")

HURDLE1=hurdle(BULDINGNUM~ INSBUTYPE + OCCUPANCY2

+ EMPLOYEES + THEFTBUILDSI +  ELECTSI + INHABITANTS + GLASSSI

+offset(CLAIMEXPLN)|BUILDINGSI ,dist="poisson",data=data24)

summary(HURDLE1)

AIC(HURDLE1)

data25<-
read.table(file="datos_definitivo_mod_2BUILDINGHURDLE.csv",header=TRUE,sep
=";")

OCCUPANCY2 = relevel(data25$OCCUPANCY,ref="BASE")

PROVINCE2 =  relevel(data25$PROVINCE,ref="BASE")

BUILDINGF<-as.factor(data25$BUILDINGSI)

BUILDINGSI2 <-relevel(BUILDINGF,ref=">1050000€")

ZBN1=zeroinfl(BULDINGNUM~  OCCUPANCY + PROVINCE2 + EMPLOYEES +
THEFTBUILDSI +  ELECTSI + INHABITANTS + GLASSSI +
offset(CLAIMEXPLN) |BUILDINGSI,dist="negbin",data=data25)

summary(ZBN1)

AIC(ZBN1)
```

```
data21<-
read.table(file="datos_definitivo_mod_2SEVBUILDINGGAMMA.csv",header=TRUE,
sep=";")

OCCUPANCY2<-relevel(as.factor(data21$OCCUPANCY),ref="BASE")

gamma<-glm(BUILDINGCME ~ INSBUTYPE +  OCCUPANCY2

        ,family=Gamma(link="log"),data=data21,maxit=1000)

summary(gamma)

lognormal<-lm(log(BUILDINGCME) ~ INSBUTYPE +
OCCUPANCY2 ,data=data21,maxit=1000)

summary(lognormal)

AIC(lognormal)

ig<-glm(BUILDINGCME ~  INSBUTYPE +
OCCUPANCY2 ,family=inverse.gaussian(link="log"),data=data21,start=coefficients(lo
gnormal),maxit=1000)

summary(ig)

####CONTENT

data22<-
read.table(file="datos_definitivo_mod_2CONTENT.csv",header=TRUE,sep=";")

data23<-
read.table(file="datos_definitivo_mod_2CONTENTTOBIT.csv",header=TRUE,sep=";"
)

OCCUPANCY2<-relevel(as.factor(data23$OCCUPANCY),ref="BASE")

summary(m <- vglm(CONTENTCOSTEXP ~ PAYFREQ + RISKLOC +
+THEFTFORM + SECBOX + OCCUPANCY2 + EMPLOYEES  + ELECTSI
+CONTENTSI + INHABITANTS, tobit(Lower = 0), data = data23))

data24<-
read.table(file="datos_definitivo_mod_2CONTENTTWEEDIE.csv",header=TRUE,sep
=";")

OCCUPANCY2<-relevel(as.factor(data24$OCCUPANCY),ref="BASE")

summary(tweedie<-glm(CONTENTCOSTEXP ~ EMPLOYEES  + INHABITANTS +
CONTENTSI,family=tweedie(var.power=1.70,link.power=0),data=data24,maxit=1000)
)

data25<-
read.table(file="datos_definitivo_mod_2CONTENTPROBIT.csv",header=TRUE,sep="
;")

PROVINCE2<-relevel(as.factor(data25$PROVINCE),ref="BASE")
```

```
OCCUPANCY2<-relevel(as.factor(data25$OCCUPANCY),ref="BASE")

summary(probit<-glm(INDCONTENT ~ PAYFREQ + PROVINCE2 + RISKLOC

+ OWNER +  THEFTFORM + SECBOX + OCCUPANCY2 +  EMPLOYEES  +
GLASSSI + ELECTSI  + CONTENTSI,family=binomial(link=probit),data=data25,
maxit=1000))

summary(logit<-glm(INDCONTENT ~ PAYFREQ + PROVINCE2 + RISKLOC

+ OWNER +  THEFTFORM + SECBOX + OCCUPANCY2 + EMPLOYEES  +
GLASSSI + ELECTSI  +
CONTENTSI,family=binomial(link=logit),data=data25,maxit=1000))

data26<-
read.table(file="datos_definitivo_mod_2SEVCONTENT.csv",header=TRUE,sep=";")

OCCUPANCY2 = relevel(data26$OCCUPANCY,ref="W")

data27<-
read.table(file="datos_definitivo_mod_2SEVCONTENTGAMMA.csv",header=TRUE,
sep=";")

OCCUPANCY2 = relevel(data27$OCCUPANCY,ref="BASE")

summary(gamma<-glm(CONTENTCOSTEXP ~   OCCUPANCY2 + GLASSSI +
CONTENTSI,family=Gamma(link="log"),data=data27,maxit=10000))

data28<-
read.table(file="datos_definitivo_mod_2SEVCONTENTLOGNORMAL.csv",header=T
RUE,sep=";")

OCCUPANCY2 = relevel(data28$OCCUPANCY,ref="BASE")

lognormal<-lm(log(CONTENTCOSTEXP) ~ CONTENTSI  +OCCUPANCY2,
data=data28)

summary(lognormal)

ig<-glm(CONTENTCOSTEXP ~  CONTENTSI + OCCUPANCY2 ,
family=inverse.gaussian(link="log"),
data=data28,start=coefficients(lognormal),maxit=1000)

summary(ig)

#TWO-PART

data30<-
read.table(file="datos_definitivo_mod_2CONTENTPOISSON.csv",header=TRUE,sep=
";")

OCCUPANCY2 = relevel(data30$OCCUPANCY,ref="BASE")

PROVINCE2 = relevel(data30$PROVINCE,ref="BASE")
```

```
POISSON1 = glm(CONTENTNUM~PAYFREQ + PROVINCE2 + RISKLOC

+ OWNER + THEFTFORM + SECBOX +  OCCUPANCY2  + EMPLOYEES  + GLASSSI + ELECTSI + CONTENTSI +offset(CLAIMEXPLN), family=poisson(link=log),data=data30)

summary(POISSON1)

QPOISSON1 = glm(CONTENTNUM~PAYFREQ + PROVINCE2 + RISKLOC

+ OWNER + THEFTFORM + SECBOX +  OCCUPANCY2 + EMPLOYEES  + GLASSSI + ELECTSI + CONTENTSI +offset(CLAIMEXPLN), family=quasipoisson(link=log),data=data30)

summary(QPOISSON1)

NB1 = glm.nb(CONTENTNUM~PAYFREQ + PROVINCE2 + RISKLOC

+ THEFTFORM + SECBOX +  OCCUPANCY2 + EMPLOYEES  + GLASSSI + ELECTSI + CONTENTSI +offset(CLAIMEXPLN),link=log,data=data30)

summary(NB1)

AIC(NB1)

PIG1 = gamlss(CONTENTNUM~PAYFREQ + PROVINCE2 + RISKLOC

+ THEFTFORM + SECBOX +  OCCUPANCY2 + EMPLOYEES  + GLASSSI + ELECTSI + CONTENTSI+offset(CLAIMEXPLN),family=PIG,data=data30)

summary(PIG1)

data31<-
read.table(file="datos_definitivo_mod_2CONTENThurdle.csv",header=TRUE,sep=";")

OCCUPANCY2 = relevel(data31$OCCUPANCY,ref="BASE")

PROVINCE2 = relevel(data31$PROVINCE,ref="BASE")

HURDLE1=hurdle(CONTENTNUM~PAYFREQ +

+  OCCUPANCY2 + ELECTSI
+offset(CLAIMEXPLN)|CONTENTSI ,dist="poisson",data=data31)

summary(HURDLE1)

AIC(HURDLE1)

ZBN1=zeroinfl(CONTENTNUM~PAYFREQ +

+  OCCUPANCY2 + ELECTSI +offset(CLAIMEXPLN)|
CONTENTSI,dist="negbin",data=data31)

summary(ZBN1)

AIC(ZBN1)
```

```
data32<-read.table(file="datos_definitivo_mod_2SEVCONTENTCMEGAMMA.csv",header=TRUE,sep=";")

OCCUPANCY2<-relevel(as.factor(data32$OCCUPANCY),ref="BASE")

gamma<-glm(CONTENTCME ~   OCCUPANCY2 + CONTENTSI , family=Gamma(link="log"),data=data32,maxit=1000)

summary(gamma)

data33<-read.table(file="datos_definitivo_mod_2SEVCONTENTCMELOGNORMAL.csv",header=TRUE,sep=";")

OCCUPANCY2<-relevel(as.factor(data33$OCCUPANCY),ref="BASE")

lognormal<-lm(log(CONTENTCME)  ~   OCCUPANCY2 + CONTENTSI  ,data=data33)

summary(lognormal)

AIC(lognormal)

ig<-glm(CONTENTCME ~   OCCUPANCY2  + CONTENTSI  ,family=inverse.gaussian(link="log"),data=data33,start=coefficients(lognormal),maxit=1000)

summary(ig)
```

10. REFERENCES.

Bliss, C.I. (1935) The calculation of dosage-mortality curve. *Annuals of Applied Biology* 22, 134 – 167.

Berkson, J.(1944) Application of the Logistic function to bio-essay. *Journal of American Statistics* 39, 357 – 365.

Frees, E.W. (2010) *Regression Modelling with Actuarial and Financial Applications*. Cambridge University Press. New York. (USA).

Frees, E.W. Meyers G. and Cummings D. (2010) Dependent Multi-Peril Ratemaking models. *Astin Bulletin* Vol 40(2), 699 – 726.

ICEA (2015). Boletin_146-2_Inf_Coyuntura_4Trimestre_2015 http://www.icea.es/es-ES/Paginas/home.aspx (3rd of April 2016).

Jørgensen, P. and Paes de Souza M.C. (1994). Fitting Tweedies's Compound Poisson Model to Insurance Claim Data. *Scandinavian Actuarial Journal* 1, 69 – 93.

Jun Mo Lee, Kouros Nouri-Mahdavi, Esteban Morales, Abdelmonem Afifi, Fei Yu and Joseph Caprioli (2014). Comparison of regression models for serial visual field analysis. *The Official International Journal of the Japanese Ophtalmological Society*, 504 – 514.

Kuan-Chia Lin and Su-Fen Cheng (2011). Tobit Model for Outcome Variable Is Limited by Censoring in Nursing Research. *Nursing Research* September/October 2011 Vol. 60, No 5, 354 – 360.

McCullagh, P and Nelder J. A. (1989). *Generalized Linear Models*, Second Edition. Chapman and Hall, London (UK).

Nelder, J.A. and Wedderburn, R.W.M. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society*, Vol. 135 Nº3 pp. 370 – 384.

Panagiotis Ch. Anastopoulus, Shankar V.N., Haddok J.E. and Fred L. Mannering. (2012). A multivariate tobit analysis of highway accident-injury-severity rates. *Accident Analysis and Prevention* 45, 110 – 119.

Sun, Yunjie. (2011). *Micro-Econometric Modeling of Personal Lines Insurance*. Proquest LLC. University of Wisconsin – Madison (USA).

Tobin, J. (1958). Estimation of Relationships for Limied Dependent Variables. *Econometrica*, Vol. 26 No. 1, 24 - 36.

Tweedie, M.C.K. (1984). An index which distinguishes between some important exponential families. In Statistics: Applications and New Directions. *Proceedings of the Indian Statistical Institute Golden Jubilee International Conference*, 57 – 67.

Veilleux, G. (2007). Homeowners modelling. CAS Predictive Modeling Seminar. https://www.casact.org/education/specsem/f2007/handouts/veilleux.pdf (10th of May 2016).