# **Pricing using the address :** High dimensional data set

with heterogeneous quality

## Chatelain Pierre December, 2020 Laboratory of Actuarial and Financial Sciences & ADDACTIS France



## Introduction

To avoid adverse selection, insurance companies seek to have the most precised premium. In household insurance pricing, the models components stems from questionnaire with in average 20 questions. The more questions or the more complex are the questions, the more precise is the premium but the less quotation are done. Using the address to access computer vision and external data - type of roof, presence of windows on roof, house value, ... -, could household insurance be priced without questionnaire ?





lumber of floors	R+1		
iving surface	138	m²	
leight at the gutter	7,0	m	
idge height	8,8	m	

More variables and more complexe information available. Quicker underwriting process meaning more quotation finished.





(r^2)/3 Information on 0.5 the underwriter 0.0 Climat variables -0.5

selection constraints are optimised by coverage/risks.

Phase 3 & 4 : Pricing using GLM models (Log Gamma and Poisson) due to operational constraints.

#### 2 **Data selection per risk**



**Comparison between Traditional models and models using building's geolocalisation** 

**Traditional models** : Models based on questionnaire variables

Claims Cost models per risks (Ratio of Pseudo R)

## **Quick underwriting models** :

Models based only on external data given by the building's geolocalisation

**Performance models** : Models based on questionnaire variables and on external data

Frequency models per risks (Ratio of Pseudo R)

Coverage	Distributions	Glm - Traditional	Glm – Quick underwriting	Glm -Performance
Water damage	20%	100%	104%	115%
Theft	42%	100%	107%	118%
Fire damage	24%	100%	115%	130%
Electric damage	6%	100%	110%	131%
Glass damage	7%	100%	114%	123%
Total	100 %	100 %	108%	120%

\* Log-Gamma models done on a french attritional claims

### Results

Adding new variables thanks to geolocalised building, the Performance model is significantly improving.

#### **Difference Gini indexes for the** aggregate pure premium **GLM-Quick** GLM -**GLM-Traditional** Performance Underwriting + 1 pts 0 pts 0 pts

Coverage	Distributions	Glm - Traditional	Glm – Quick underwriting	Glm - Performance		
Water damage	31%	100%	102%	128%		
Theft	13%	100%	118%	139%		
Fire damage	20%	100%	103%	119%		
Electric damage	16%	100%	117%	132%		
Glass damage	20%	100%	87%	108%		
Total	100 %	100 %	104%	124%		
Lag Deissen medels dens en a franch attritional deime						

<sup>c</sup> Log-Poisson models done on a french attritional claims

Without questionnaire, Pricing models are as performing as the Traditional models.

DATA

INITIATIVE 🔍 🌢 🍡

SCIENCE



46 bis Chemin du Vieux Moulin, 69160 TASSIN LA DEMI LUNE- Tél : +33 (0)4 72 18 58 58 SAS au capital de 100 000 Euros Immatriculée au RCS de Lyon sous le numéro 413 611 344 Numéro de TVA Intracommunautaire : FR 53 413611344 Organisme de formation agréé sous le numéro : 82 69 06035 69



Laboratory of Actuarial and Financial Sciences (SAF) Université Claude Bernard - Lyon 1 Domaine de Gerland 50 Avenue Tony Garnier 69366 Lyon Cedex 07 Tél : 33 (0)4 37 28 74 40. The LSAF (research group n° 2429) has around fifteen lecturersresearchers working on the sciences of management, applied mathematics (probability and statistics) and economics.





