

# L'assurance : mutualisation des risques, segmentation, classification et compétition

Arthur Charpentier (UQAM)

Séminaire du CIMMUL, Janvier 2022

January 26, 2022



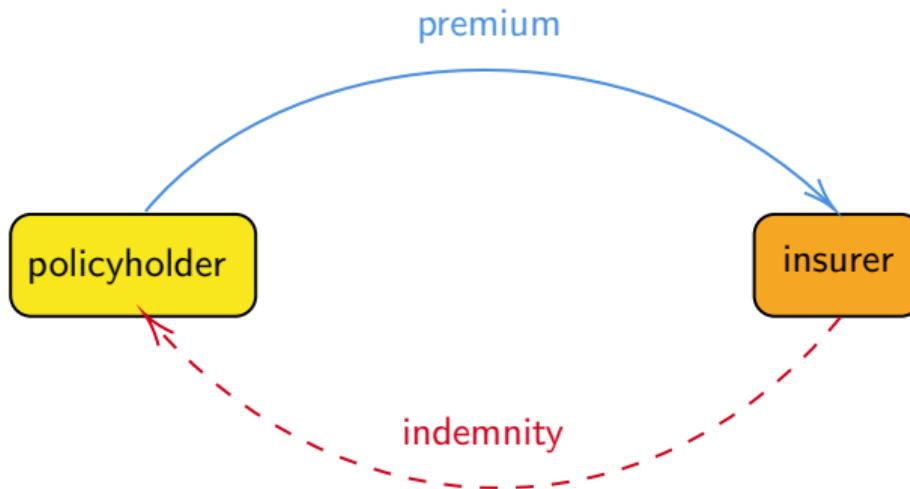
# Arthur Charpentier

- Professor (Mathematics Department) UQAM
- Researcher at Quantact (& CRM)
- Blog editor <http://freakonometrics.hypotheses.org/>
- Joint Research Initiative (P.I.) AXA Research Fund  
Unusual Data for Insurance <https://jridata.github.io/>
- Insurance pricing game(s) <https://pricing-game.com/>
- see <https://freakonometrics.github.io/> for more details

actuarial models; algorithms; climate risk; data; decision theory; discrimination; extreme value; insurance; fairness; machine learning; predictive modeling; uncertainty; reinforcement learning;



# Insurance & Actuarial Science

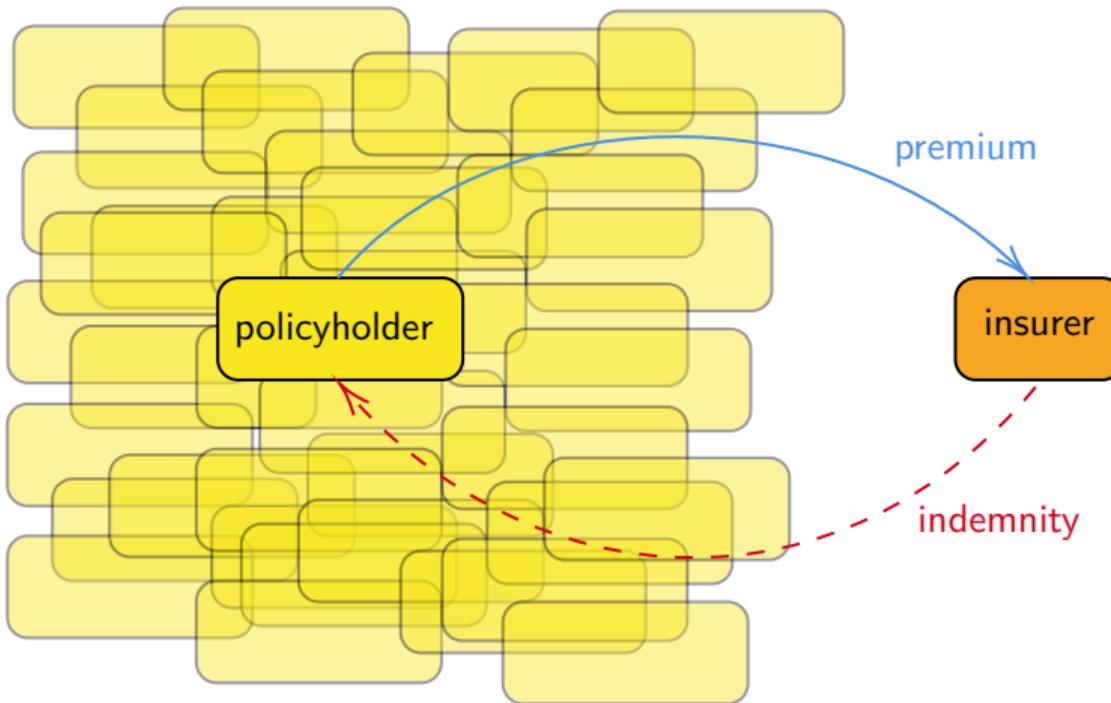


What would be a “fair premium” ? see O’Neill (1997)

- **pure actuarial fairness** contributions for individual policyholders should perfectly reflect their predicted risk levels → **predictive modeling**
- **choice-sensitive fairness** contributions should take into account only risks that result from choices - **luck-egalitarianism** (Cohen (1989) or Arneson (2011))

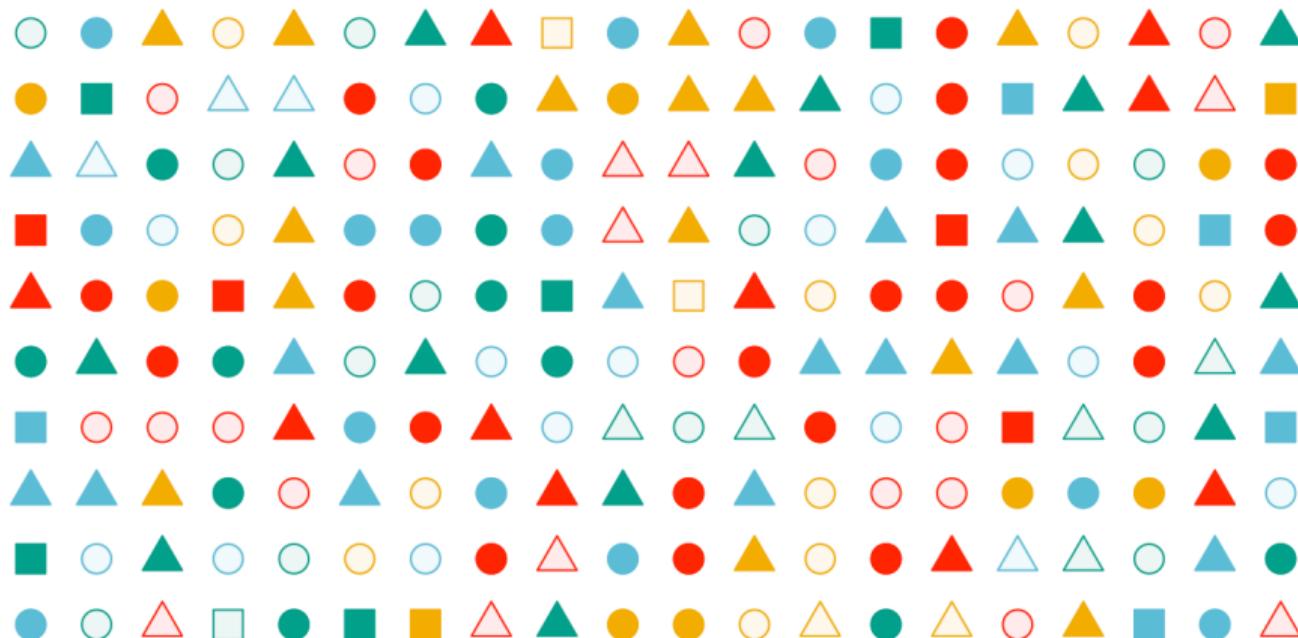
# Insurance & Actuarial Science

*“Insurance is the contribution of the many to the misfortune of the few”*



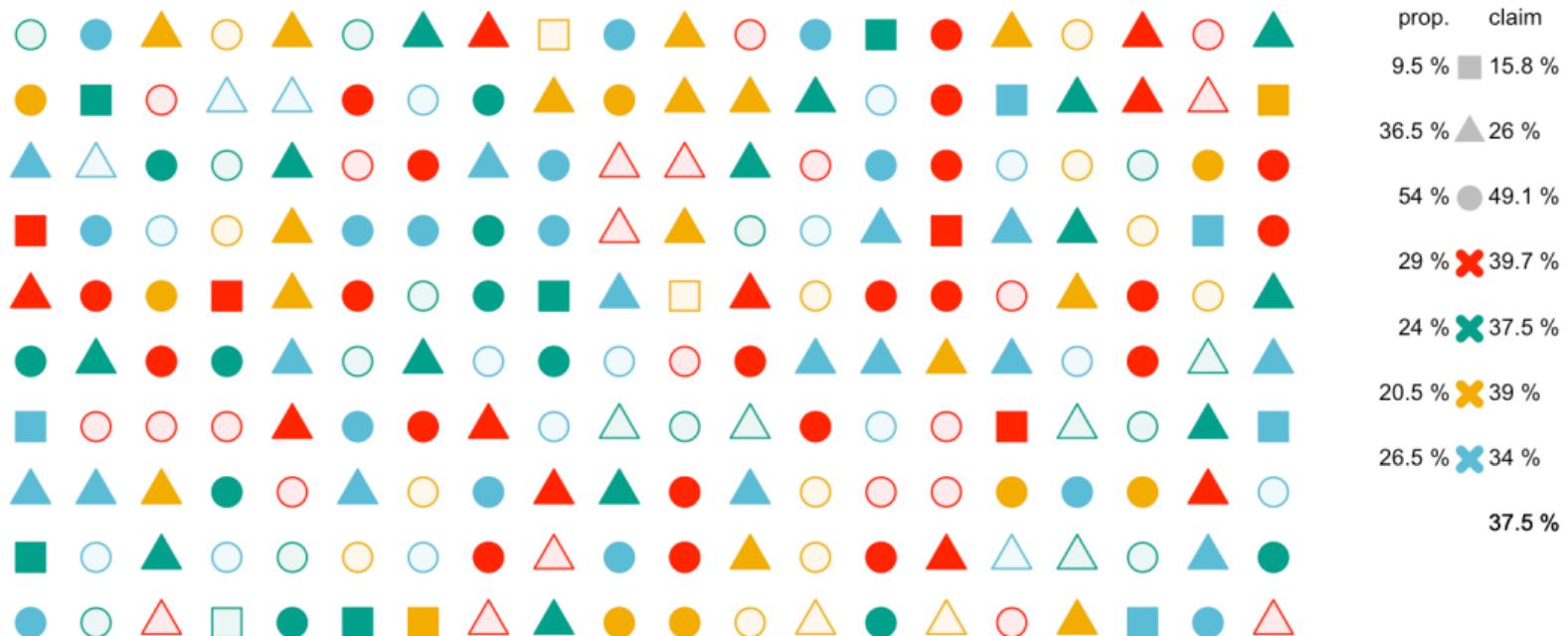
## Insurance & Actuarial Science

Consider two features,  $x_1 \in \{ \text{□} \text{△} \text{○} \}$  and  $x_2 \in \{ \text{■} \text{■} \text{■} \text{■} \}$



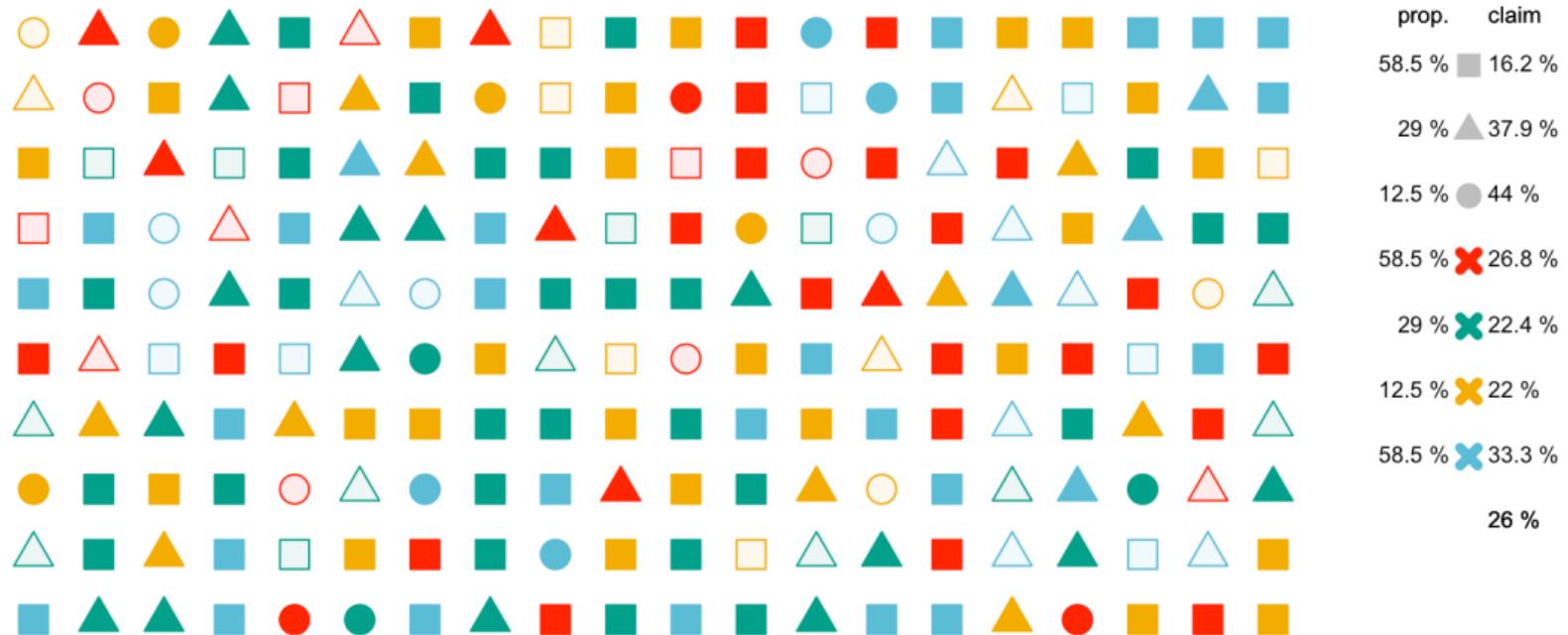
# Insurance & Actuarial Science

Consider two features,  $x_1 \in \{ \square, \triangle, \circ \}$  and  $x_2 \in \{ \blacksquare, \blacksquare, \blacksquare, \blacksquare \}$



# Insurance & Actuarial Science

Consider two features,  $x_1 \in \{ \text{■} \triangle \circ \}$  and  $x_2 \in \{ \text{■} \textcolor{blue}{\square} \textcolor{red}{\square} \textcolor{teal}{\square} \}$



## Insurance & Actuarial Science



38	49	34	45
38	26	34	23
38	16	34	13
38	49	39	52
38	26	39	28
38	16	39	17
38	49	38	51
38	26	38	27
38	16	38	17
38	49	40	49
38	26	40	26
38	16	40	16

26	44	33	53
26	38	33	49
26	16	33	22
26	44	22	37
26	38	22	34
26	16	22	13
26	44	22	37
26	38	22	34
26	16	22	13
26	44	27	44
26	38	27	40
26	16	27	16

## Insurance Pricing

“*Insurance is the contribution of the many to the misfortune of the few* ”

Insurance pricing is (1) **data-driven** and (2) **model-driven**

**mutualization** (of risk) : *spread of risk among several parties* see **pooling, sharing**

$$\pi = \mathbb{E}_{\mathbb{P}}[S]$$

(market) **segmentation** : *division of a market into identifiable groups* see **differentiation, customization**, i.e. for some (unobservable) risk factor  $\Omega$

$$\pi(\omega) = \mathbb{E}_{\mathbb{P}}[S|\Omega = \omega]$$

Use of features (covariates)  $x$  as a proxy of (unobservable) risk factor  $\Omega$

$$\pi(x) = \mathbb{E}_{\mathbb{P}}[S|X = x] = \mathbb{E}_{\mathbb{P}_x}[S]$$

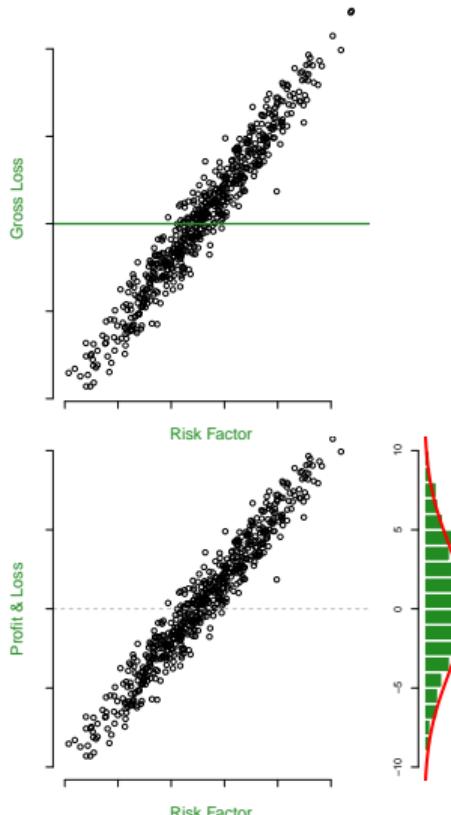
# Risk Transfert without Segmentation

	Insured	Insurer
Loss	$\mathbb{E}[S]$	$S - \mathbb{E}[S]$
Average Loss	$\mathbb{E}[S]$	0
Variance	0	$\text{Var}[S]$

All the risk -  $\text{Var}[S]$  - is kept by the insurance company.

**Remark:** interpretations are discussed in

Denuit & Charpentier (2004).



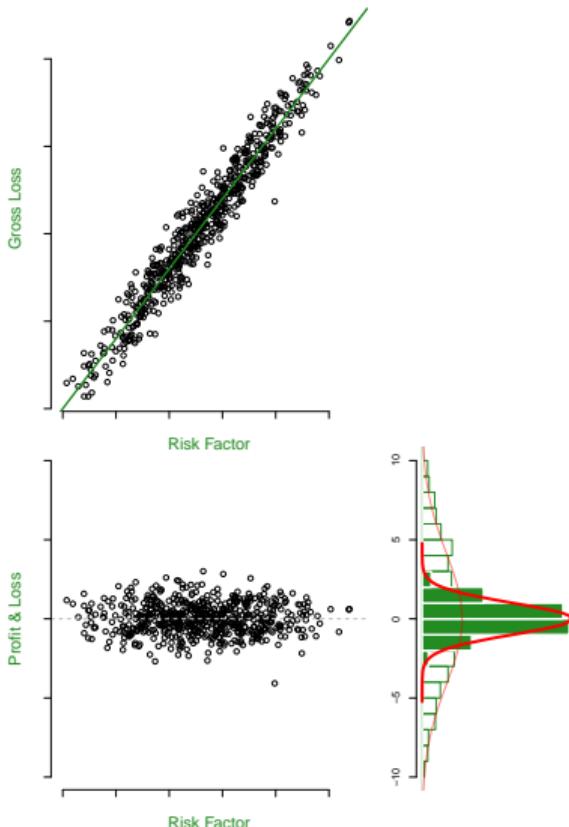
# Risk Transfert with Segmentation and Perfect Information

Assume that information  $\Omega$  is observable,

	Insured	Insurer
Loss	$E[S \Omega]$	$S - E[S \Omega]$
Average Loss	$E[S]$	0
Variance	$Var[E[S \Omega]]$	$Var[S - E[S \Omega]]$

Observe that  $Var[S - E[S|\Omega]] = E[Var[S|\Omega]]$ , so that

$$Var[S] = \underbrace{E[Var[S|\Omega]]}_{\rightarrow \text{insurer}} + \underbrace{Var[E[S|\Omega]]}_{\rightarrow \text{insured}}.$$



# Segmentation and Imperfect Information

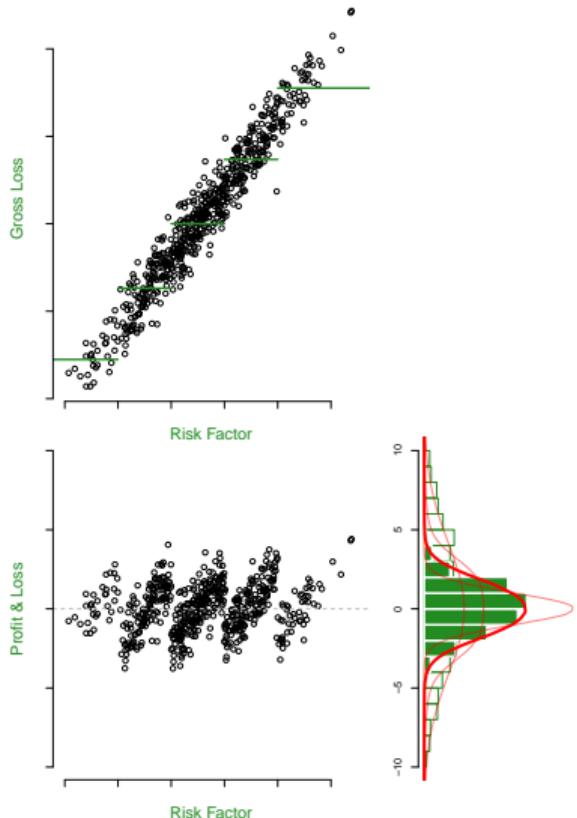
Assume that  $\mathbf{X} \subset \Omega$  is observable

	Insured	Insurer
Loss	$\mathbb{E}[S \mathbf{X}]$	$S - \mathbb{E}[S \mathbf{X}]$
Average Loss	$\mathbb{E}[S]$	0
Variance	$\text{Var}[\mathbb{E}[S \mathbf{X}]]$	$\mathbb{E}[\text{Var}[S \mathbf{X}]]$

Now

$$\begin{aligned}\mathbb{E}[\text{Var}[S|\mathbf{X}]] &= \mathbb{E}[\mathbb{E}[\text{Var}[S|\Omega]|\mathbf{X}]] + \mathbb{E}[\text{Var}[\mathbb{E}[S|\Omega]|\mathbf{X}]] \\ &= \underbrace{\mathbb{E}[\text{Var}[S|\Omega]]}_{\text{perfect pricing}} + \underbrace{\mathbb{E}\{\text{Var}[\mathbb{E}[S|\Omega]|\mathbf{X}]\}}_{\text{misfit}}.\end{aligned}$$

spiral of segmentation...



# Segmentation and Imperfect Information

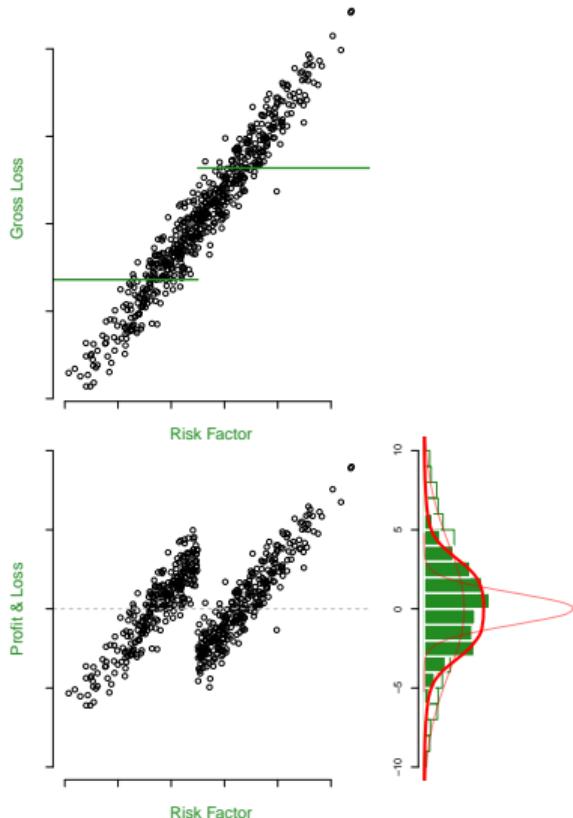
Assume that  $\mathbf{X} \subset \Omega$  is observable

	Insured	Insurer
Loss	$\mathbb{E}[S \mathbf{X}]$	$S - \mathbb{E}[S \mathbf{X}]$
Average Loss	$\mathbb{E}[S]$	0
Variance	$\text{Var}[\mathbb{E}[S \mathbf{X}]]$	$\mathbb{E}[\text{Var}[S \mathbf{X}]]$

Now

$$\begin{aligned}\mathbb{E}[\text{Var}[S|\mathbf{X}]] &= \mathbb{E}[\mathbb{E}[\text{Var}[S|\Omega]|\mathbf{X}]] + \mathbb{E}[\text{Var}[\mathbb{E}[S|\Omega]|\mathbf{X}]] \\ &= \underbrace{\mathbb{E}[\text{Var}[S|\Omega]]}_{\text{perfect pricing}} + \underbrace{\mathbb{E}\{\text{Var}[\mathbb{E}[S|\Omega]|\mathbf{X}]\}}_{\text{misfit}}.\end{aligned}$$

spiral of segmentation...



## Actuarial Pricing Model

Premium is  $\mathbb{E}[S|\mathbf{X} = \mathbf{x}] = \mathbb{E}\left[\sum_{i=1}^N Y_i \middle| \mathbf{X} = \mathbf{x}\right] = \mathbb{E}[N|\mathbf{X} = \mathbf{x}] \cdot \mathbb{E}[Y|\mathbf{X} = \mathbf{x}]$

Statistical and modeling issues to approximate based on some training datasets, with claims frequency  $\{n_i, \mathbf{x}_i\}$  and individual losses  $\{y_i, \mathbf{x}_i\}$ .

*“Most firms [...] rely on traditional generalised linear models (GLMs)”, FCA (2016)*

Use GLM to approximate  $\mathbb{E}[N|\mathbf{X} = \mathbf{x}]$  and  $\mathbb{E}[Y|\mathbf{X} = \mathbf{x}]$ , i.e. Poisson model for claims frequency,  $N|\mathbf{X} \sim \mathcal{P}(\lambda_{\mathbf{x}})$  and Gamma regression for claims severity,  $Y|\mathbf{X} \sim \mathcal{G}(\mu_{\mathbf{x}}, \varphi)$

$$\hat{\pi}_j(\mathbf{x}) = \widehat{\mathbb{E}}[N_1|\mathbf{X} = \mathbf{x}] \cdot \widehat{\mathbb{E}}[Y|\mathbf{X} = \mathbf{x}] = \underbrace{\exp(\hat{\alpha}^T \mathbf{x})}_{\text{Poisson } \mathcal{P}(\lambda_{\mathbf{x}})} \cdot \underbrace{\exp(\hat{\beta}^T \mathbf{x})}_{\text{Gamma } \mathcal{G}(\mu_{\mathbf{x}}, \varphi)}$$

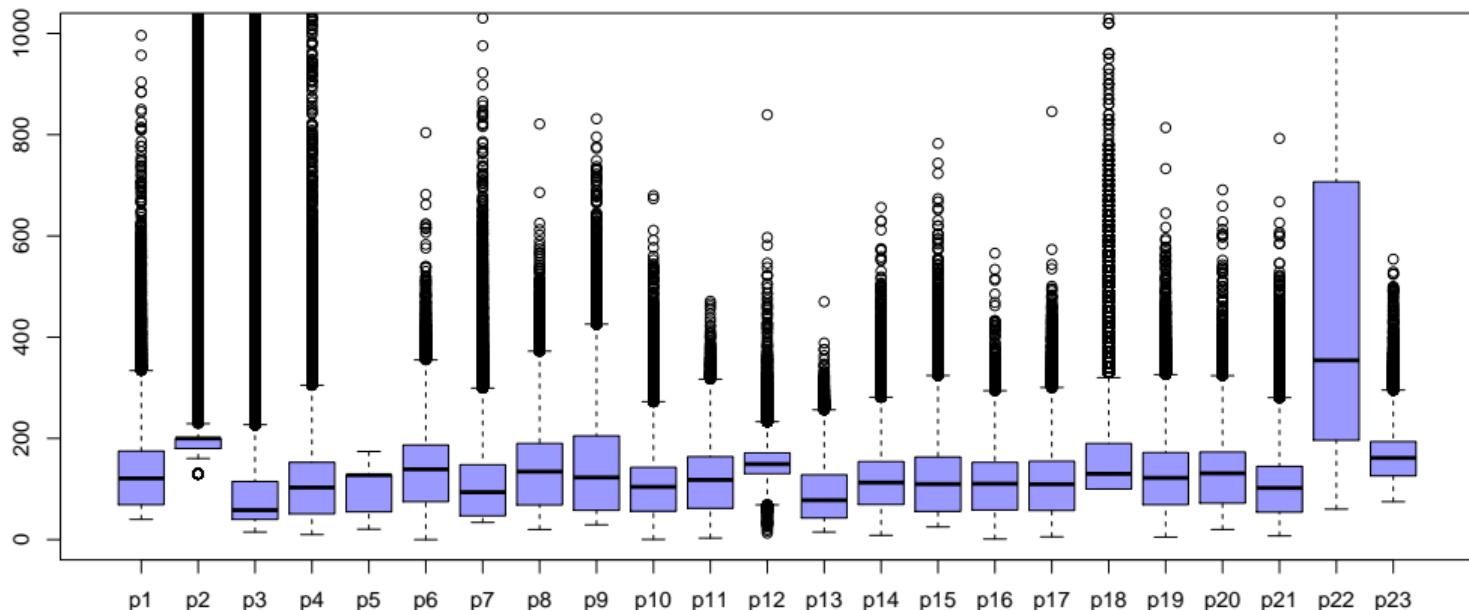
(see Charpentier & Denuit (2005) or Kaas et al. (2008)) or any other statistical model

$$\hat{\pi}_j(\mathbf{x}) \text{ where } \hat{\pi}_j \in \operatorname{argmin}_{m \in \mathcal{F}_j: \mathcal{X}_j \rightarrow \mathbb{R}} \left\{ \sum_{i=1}^n \ell(s_i, m(\mathbf{x}_i)) + \lambda \cdot \text{penalty}(m) \right\}$$

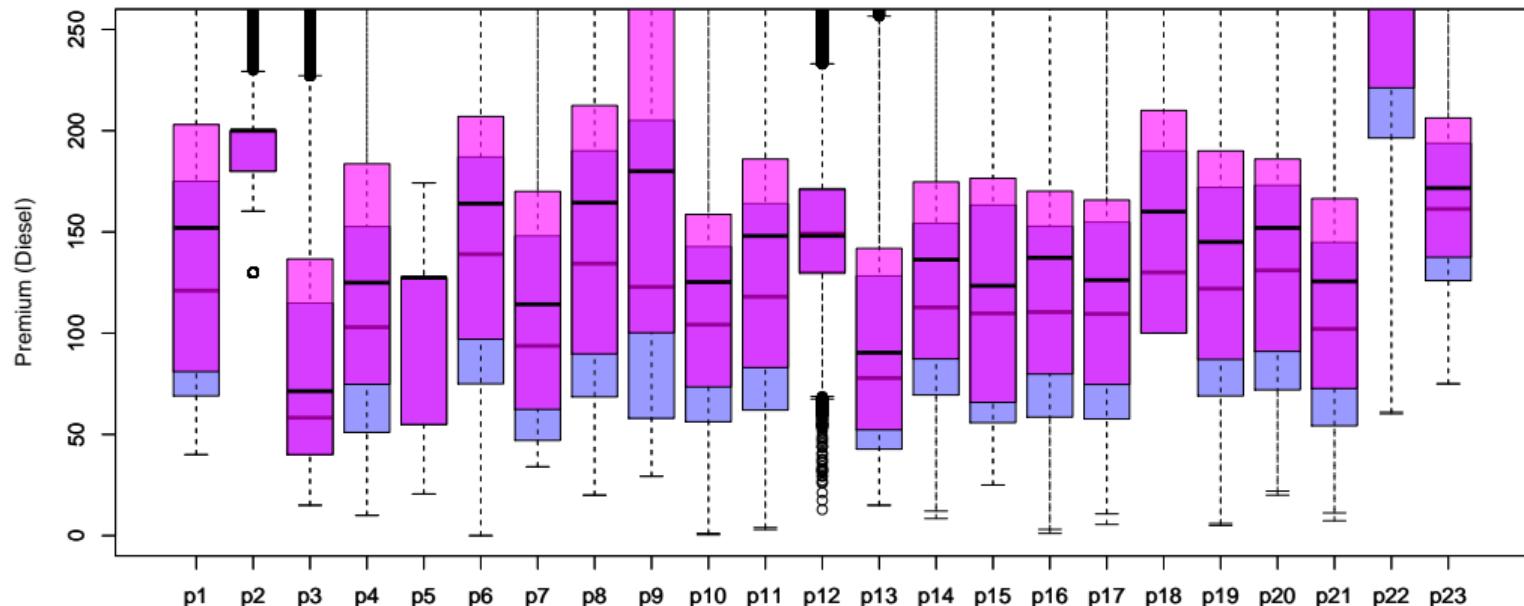
for some loss function  $\ell : \mathbb{R}^2 \rightarrow \mathbb{R}_+$  (usually an  $L_2$  based loss,  $\ell(s, y) = (s - y)^2$ ).

freakonometrics.hypotheses.org

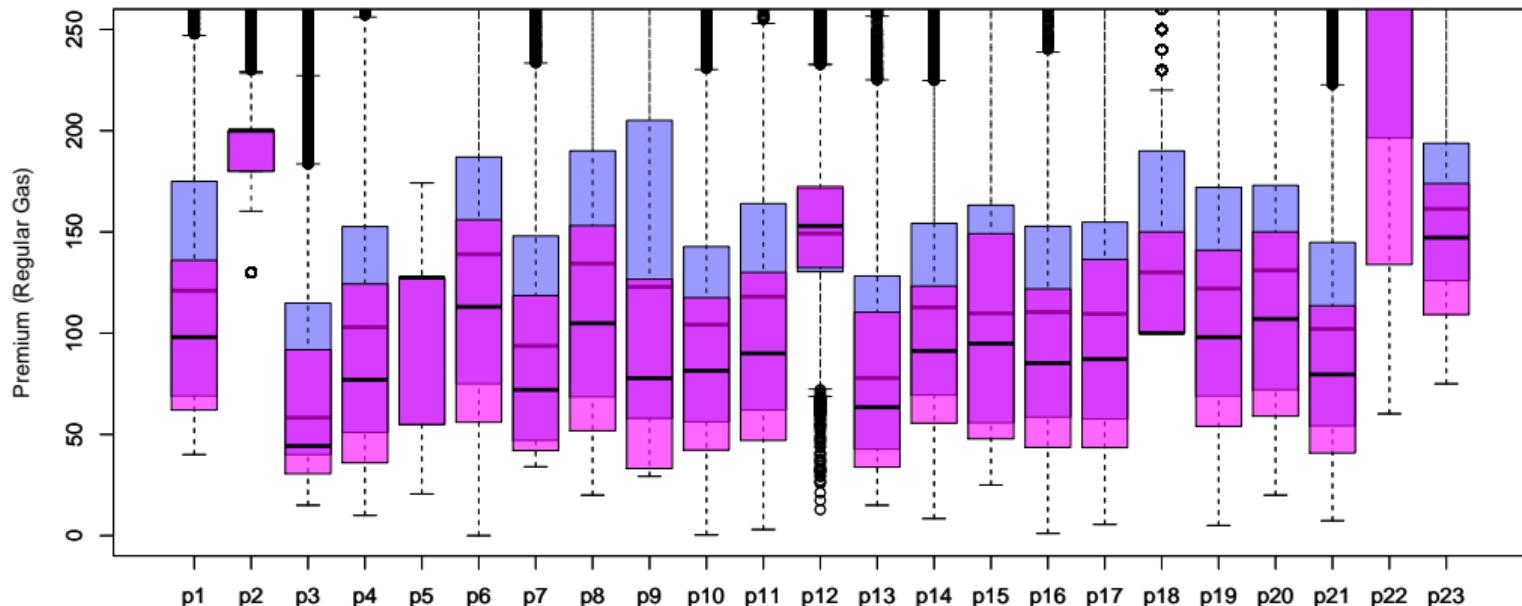
## Comparing models



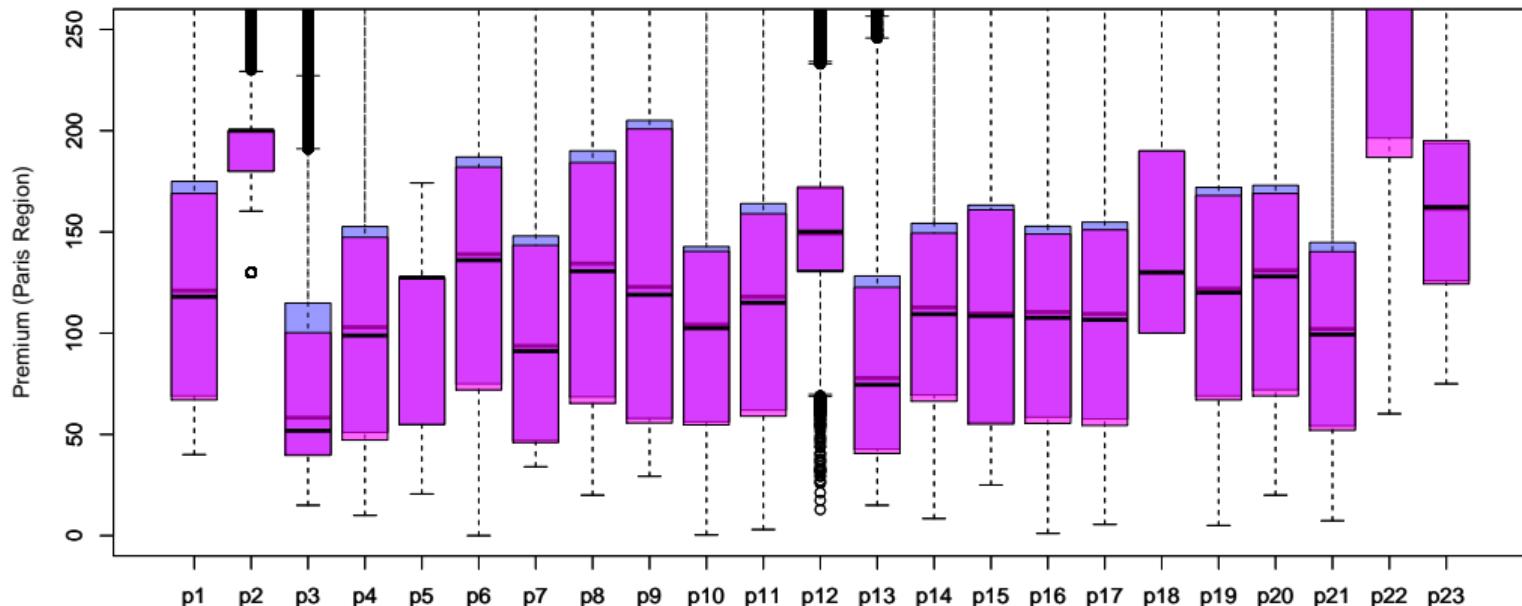
## Comparing models ( Gas Type Diesel)



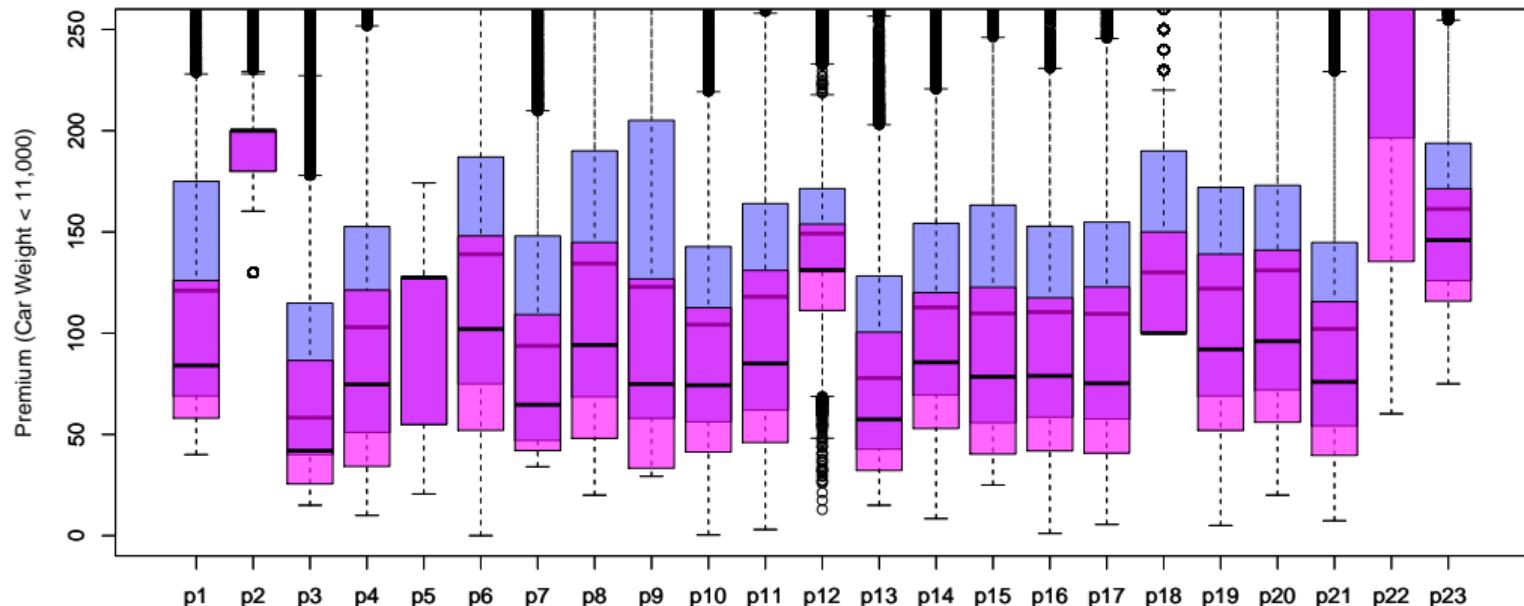
## Comparing models ( Gas Type Regular)



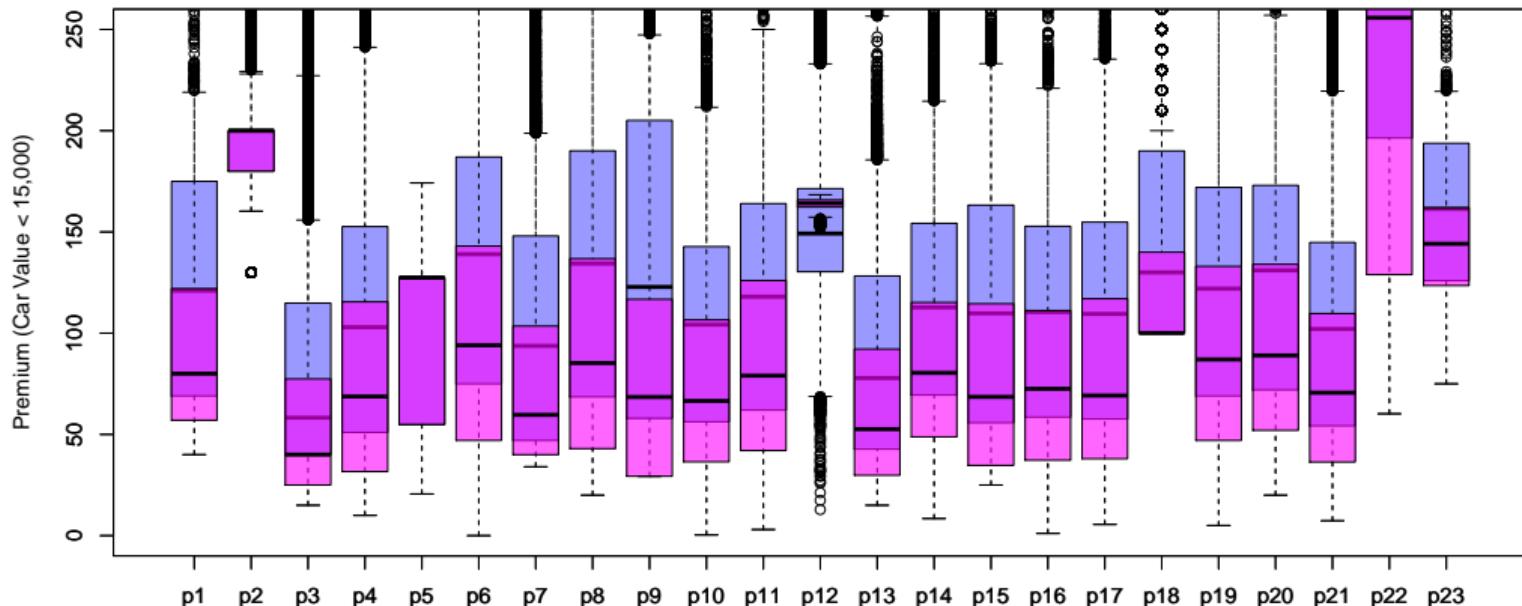
## Comparing models ( Paris Region)



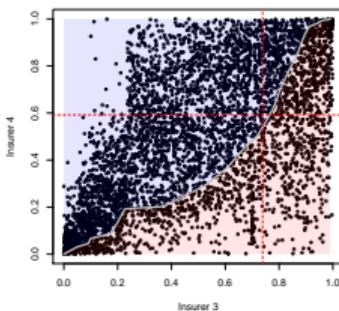
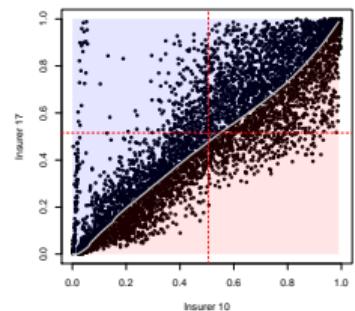
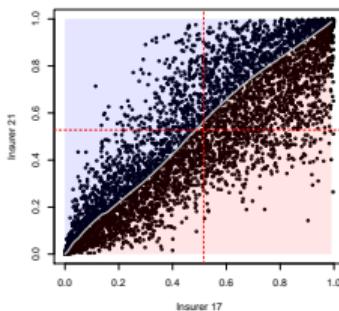
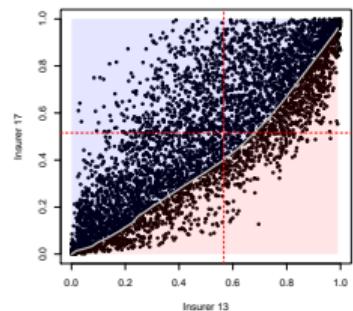
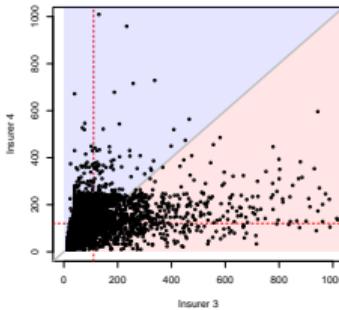
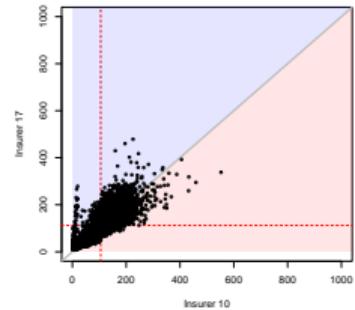
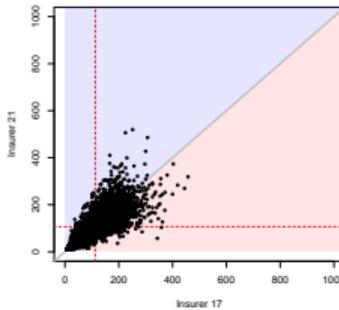
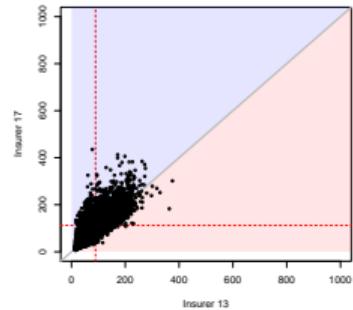
# Comparing models ( Car Weight)



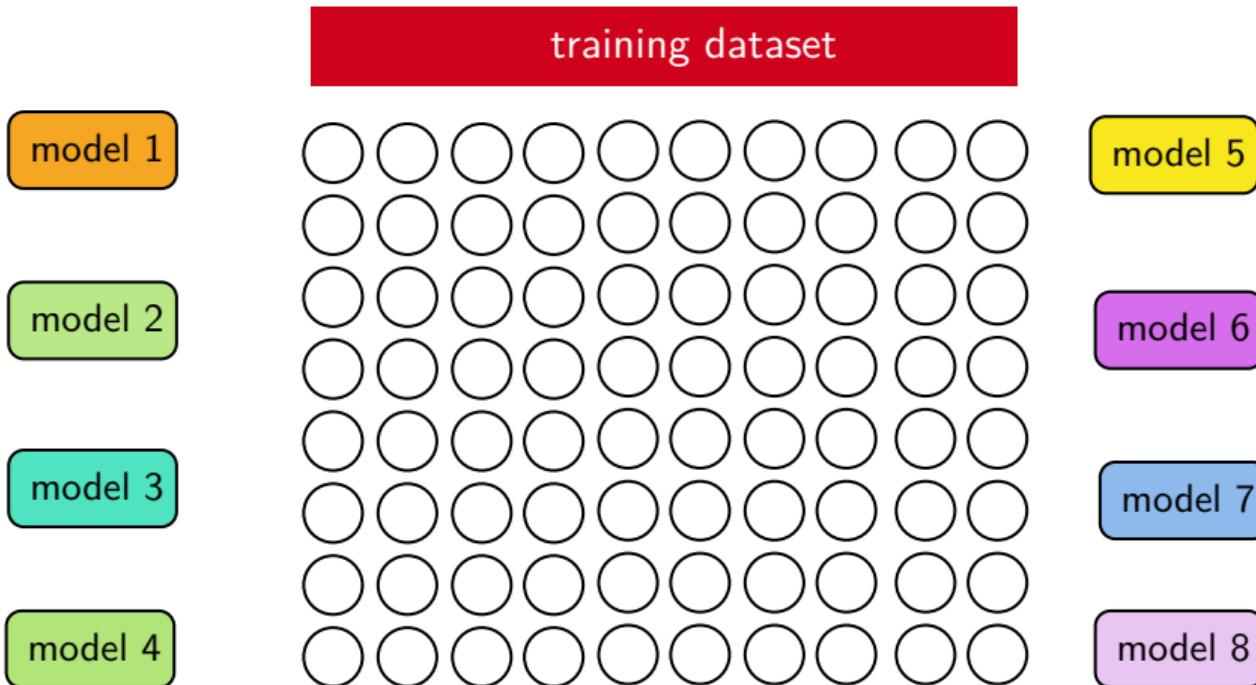
## Comparing models ( Car Value)



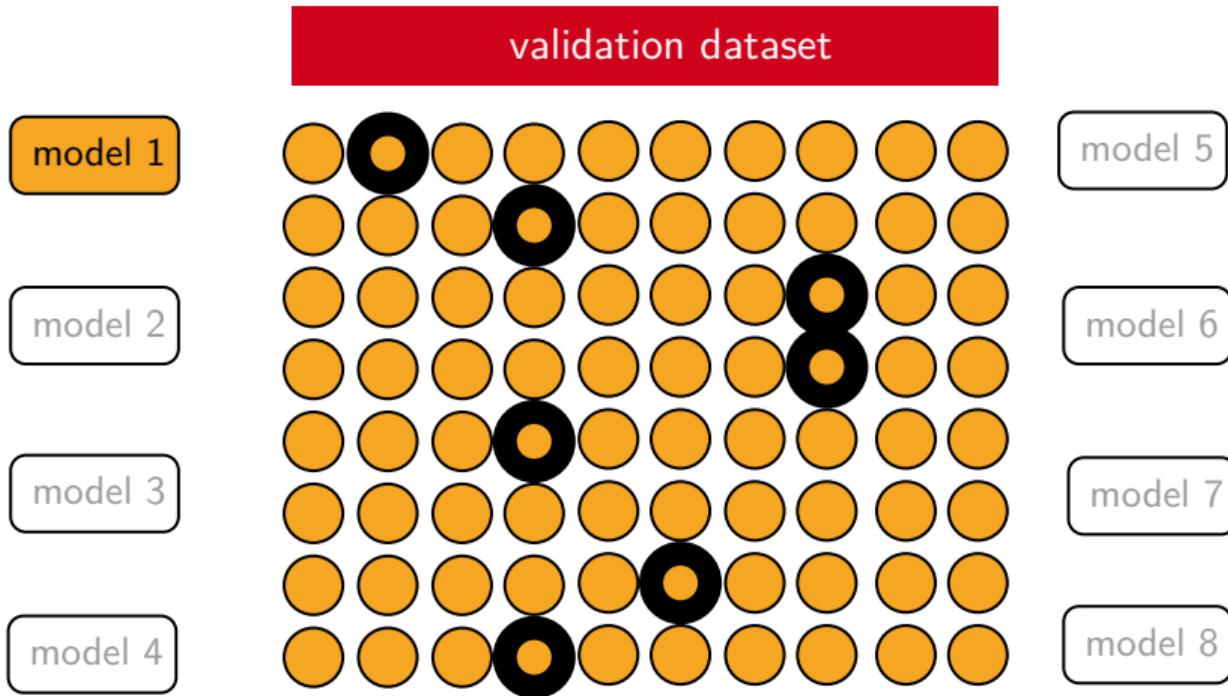
# Comparing models ( Correlation)



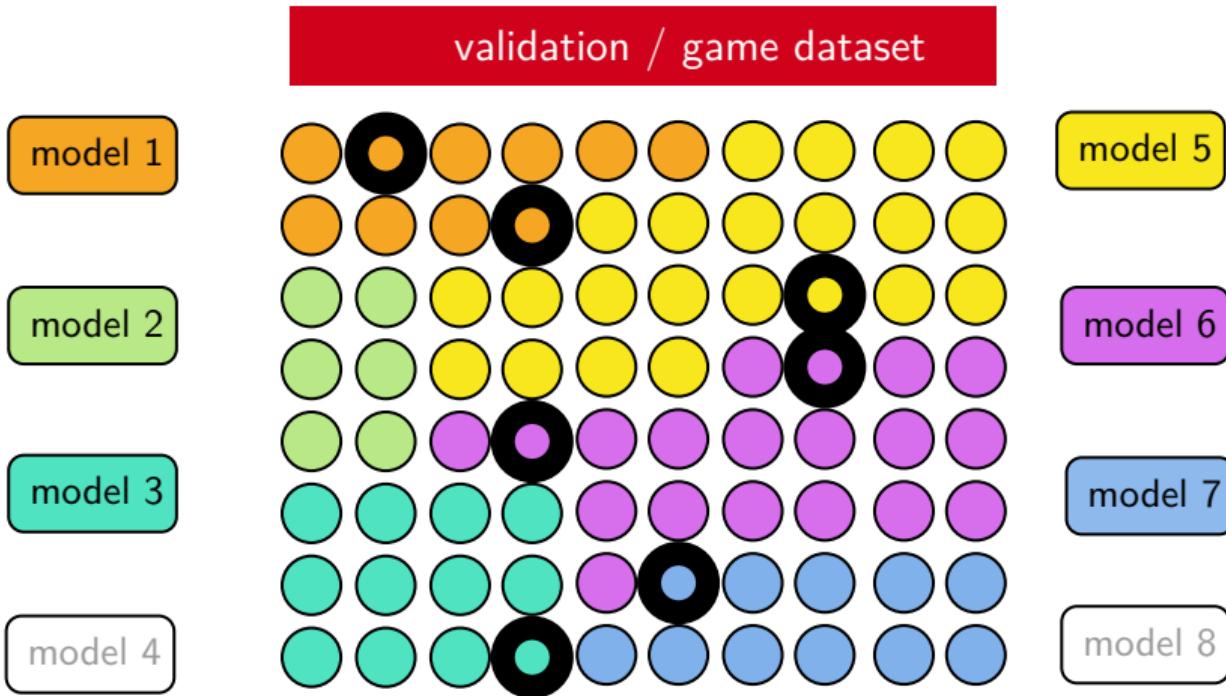
# Insurance & Actuarial Science



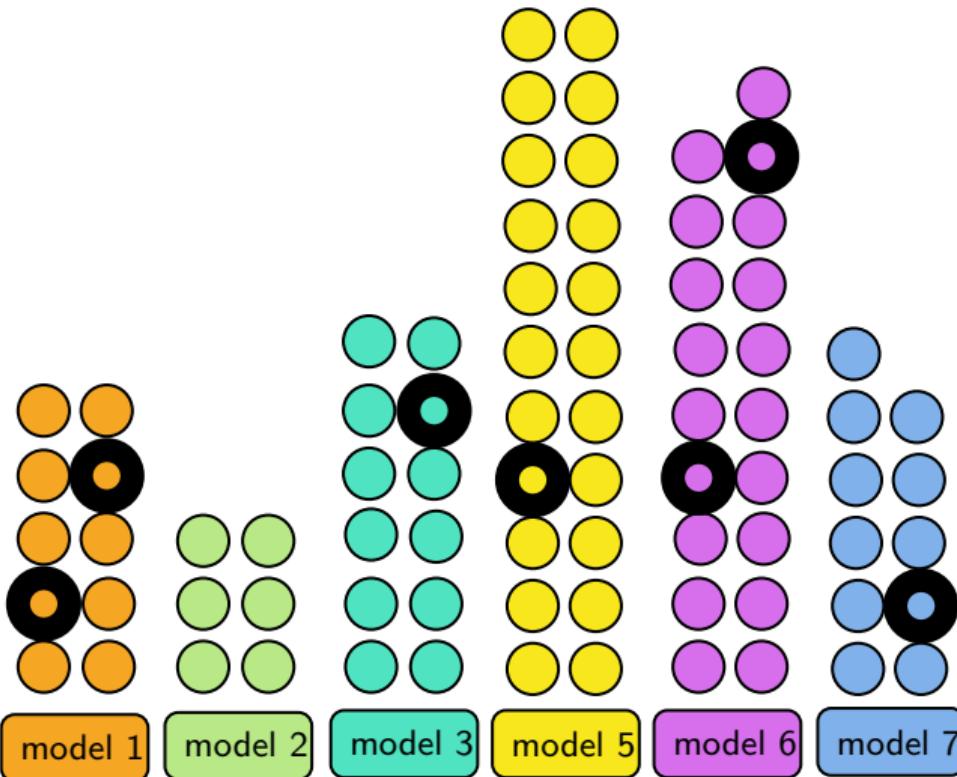
# Insurance & Actuarial Science



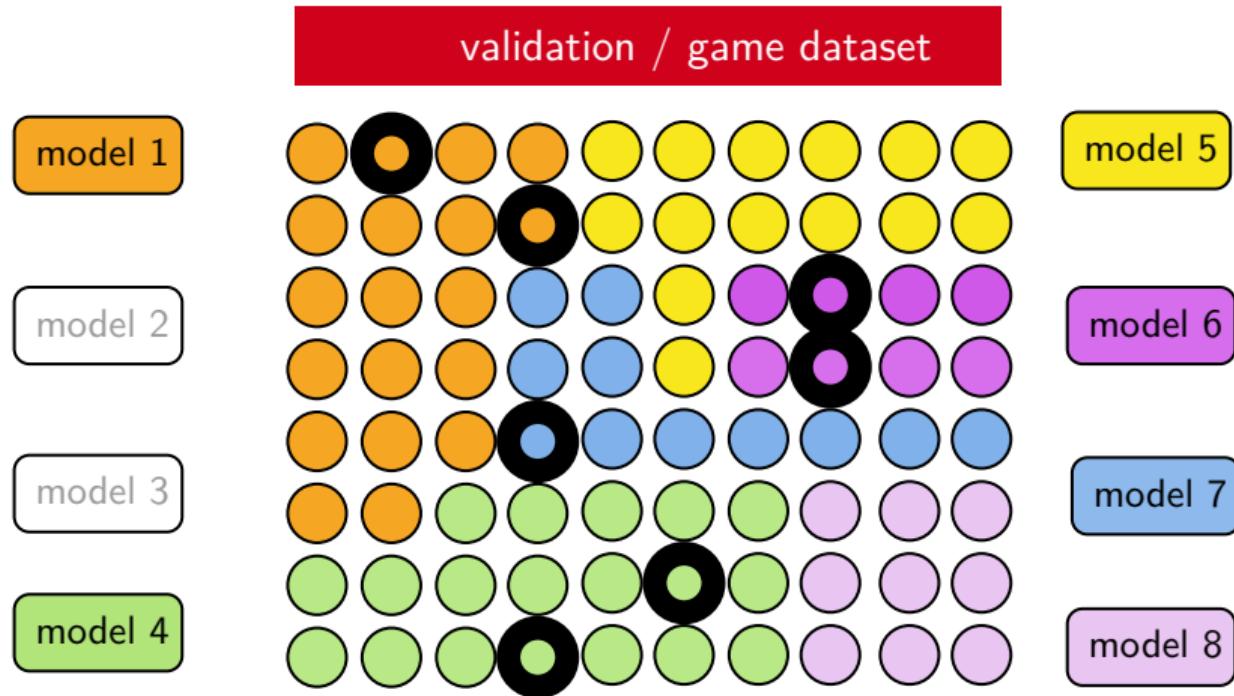
# Insurance & Actuarial Science

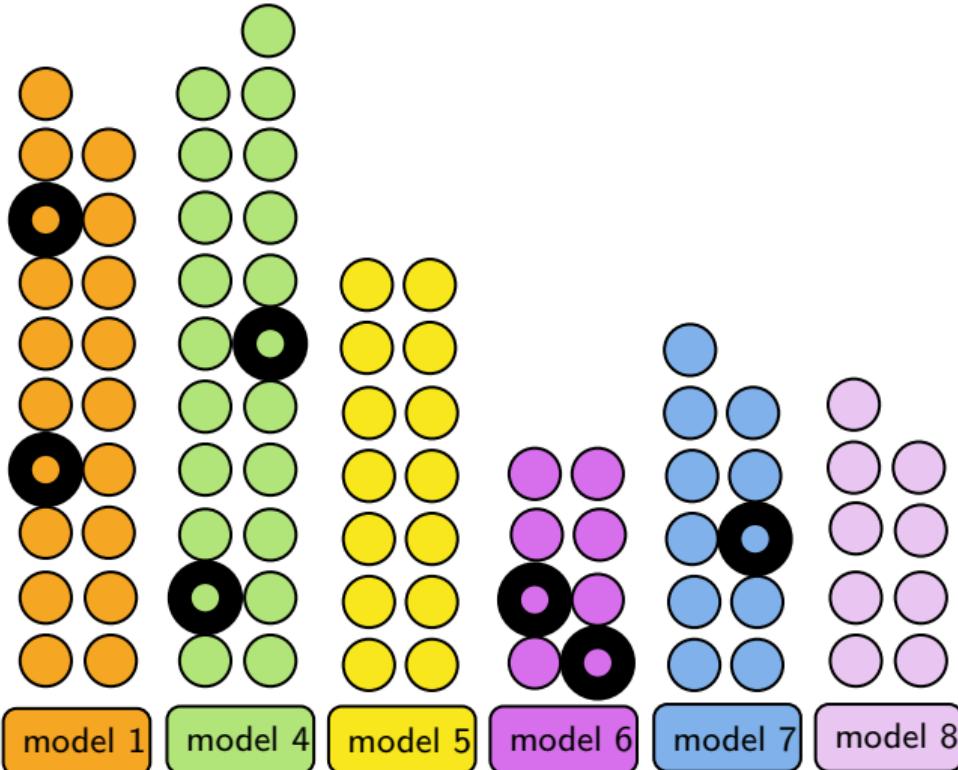


# Insurance & Actuarial Science



# Insurance & Actuarial Science





<https://www.aicrowd.com/challenges/insurance-pricing-game>

Market simulation competition : Completed   Bonus round: Ageless dataset: 8 days left   #supervised\_learning   #educational   #insurance

# AIcrowd Motor insurance market simulation

Prizes \$12,000 (USD)

By Imperial CPG   53.2k   1847   215   10.3k   87   Follow

Overview Leaderboard Notebooks Resources Submissions Pick submission for profit leaderboard Rules More ▾   [Participate](#)

Announcing our community engagement prizes!

- Join the office hours on Discord! (Wednesday 2PM CET)
- Have a question? Visit the discussion forum
- Python Starter Notebook
- R Starter Notebook
- >/> Code based starter kit

CHAT 93 ONLINE

PARTICIPANTS

Participate

Overview Leaderboard Notebooks Resources Submissions Pick submission for profit leaderboard Rules More ▾ Participate

look like:

FrankSchmid failed

jyotish graded

Overview

Cheapest-wins Market

Leaderboards

Evaluation Metric

Market Rules

Weekly Market Feedback

Dataset

How To Submit

Prizes

Timeline

Research Sponsors

Contact

Claim amount

Company 1 Company 2

Policy 1 120 90

Policy 2 30 25

Policy 3 10 15

Policy 4 5 10

100  
10  
0  
0

Total revenue → 15 115

0 110

Most profit

Total loss ←

Total profit → 15 5

So, once the claims are taken into account in the cheapest-wins market, we can see that Company 1 wins as it has the most competitive profit.

	Company 1	Company 2
Policy 1	120	90
Policy 2	30	25
Policy 3	10	15
Policy 4	5	10

	100	10	0	0
--	-----	----	---	---

	15	115
Total revenue →	15	115

	0	110
	0	110

	15	5
Total profit →	15	5

✓ Overview

✗ Cheapest-wins Market

★ Leaderboards

✖ Evaluation Metric

⌚ Market Rules

📊 Weekly Market Feedback

💾 Dataset

✉️ How To Submit

🏆 Prizes

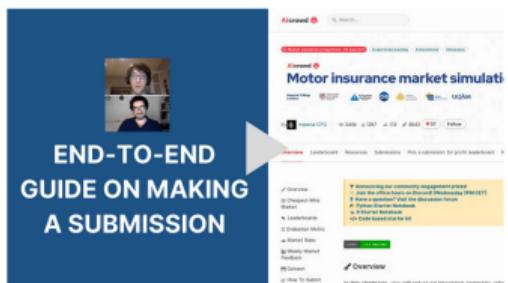
📅 Timeline

🌐 Research Sponsors

📞 Contact

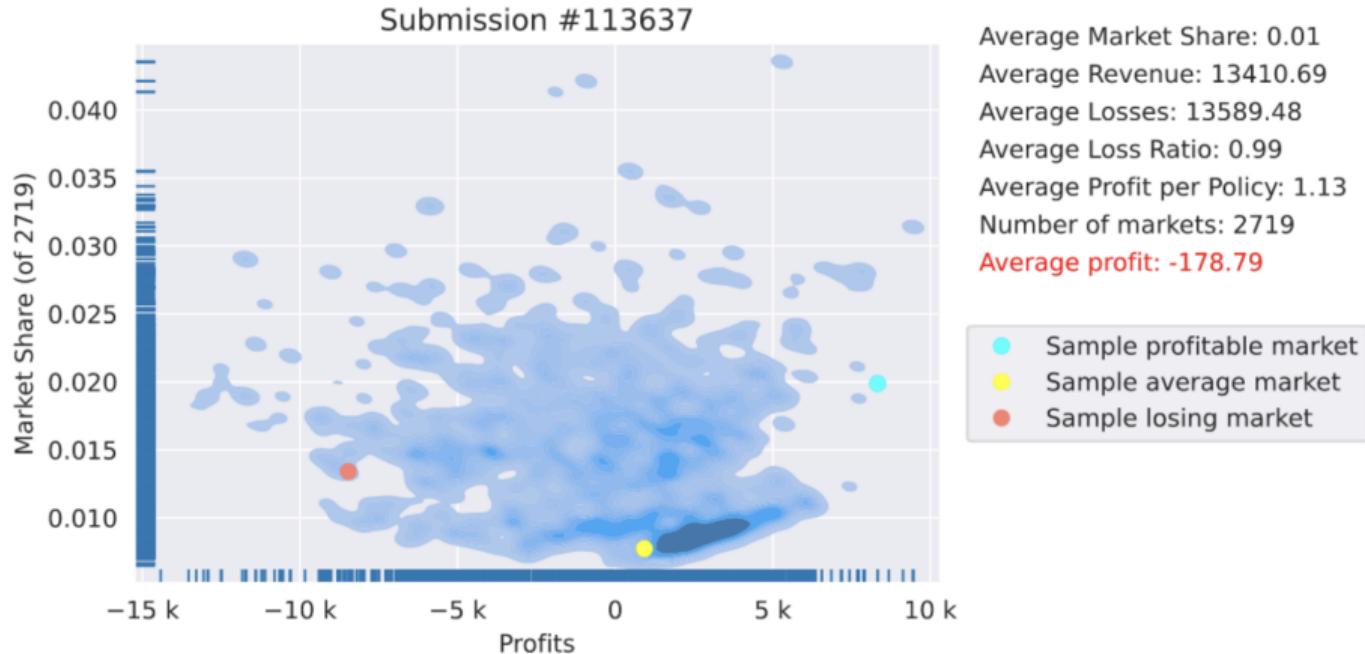
1. Submission through Google Colaboratory notebooks.
2. Submission from a `.zip` file.

#### SUBMISSION VIA COLABORATORY NOTEBOOKS



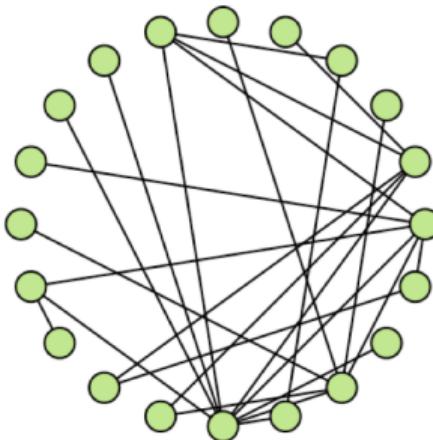
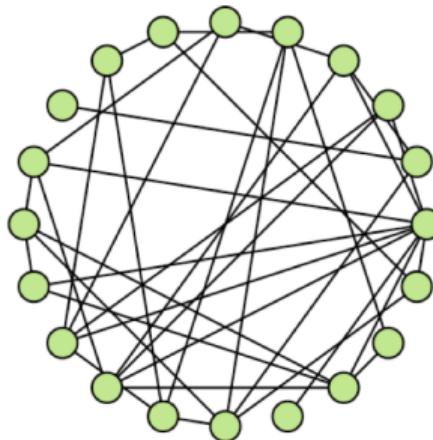
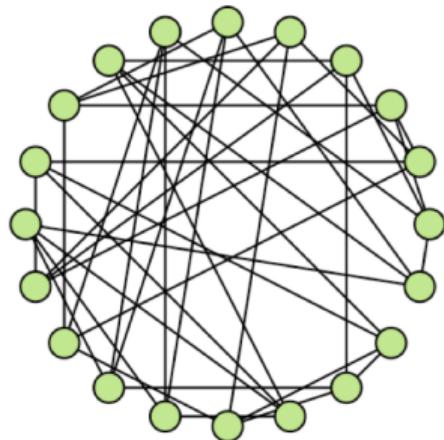
1. **Python.** Visit the [starter notebook here](#).
2. **R.** Visit the [starter notebook here](#). Here is an [end-to-end walkthrough for a simple R model](#).

The notebooks are self-contained and submissions are made through the notebooks themselves.



# Networks & P2P Insurance

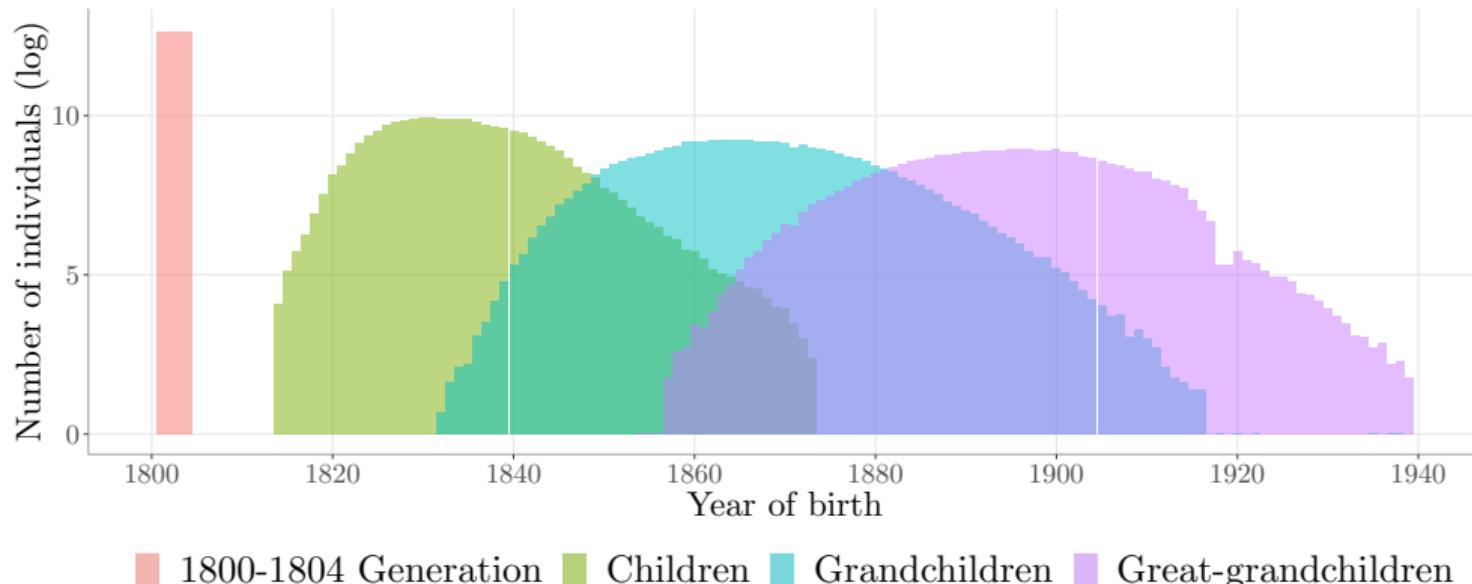
Charpentier *et al.* (2021) Collaborative Insurance Sustainability and Network Structure



# Family Trees (Genealogy) & Life Insurance

See [Charpentier et al. \(2020\)](#) on Modeling Joint Lives within Families

Initial starting generation ■(born in [1800, 1805]), children ■(born  $\sim$  [1815, 1870]),  
grand-children ■(born  $\sim$  [1830, 1915]), grand-grand-children ■(born  $\sim$  [1850, 1940])



## Wrap-Up

- work on **predictive models** in this model competition setting
  - hard to assess how good a prediction model is (with insurance losses)
  - hard to assess how a model will perform when competing against another one
  - ex: what to do if  $\hat{\pi}_{\text{your model}}(\mathbf{x}) > \hat{\pi}_{\text{competitor}}(\mathbf{x})$  ?  
importance of game theory (and sequential learning)
- work on **networks** (families, friends, hierarchies)
- work on **climate related risks** (flood, drought/subsidence, wildfire)
- work on **bias, discrimination and fairness**, see **Charpentier (2021)**
  - how to insure that a pricing model does not discriminate according to some sensitive information (gender, age, etc) ?
  - how to legitimate actuarial discrimination in that context ?
  - how to get a valid counterfactual ?