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

Arthur Charpentier (UQAM)

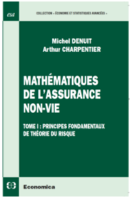
Séminaire du CIMMUL, Février 2022

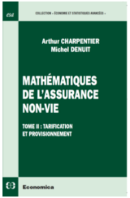
February 4, 2022

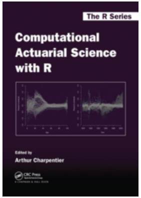
# Arthur Charpentier

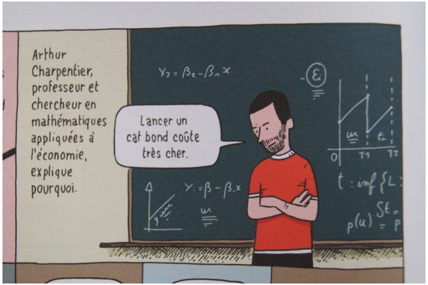
- Professor (Mathematics Dpt) UQAM
- Previously professor (Economics Dpt) Rennes, France
- Researcher at Quantact (& CRM)
- Blog editor http://freakonometrics.hypotheses.org/
- Joint Research Initiative (P.I.) AXA Research Fund
  Unusual Data for Insurance https://jridata.github.io/
- Insurance pricing 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;



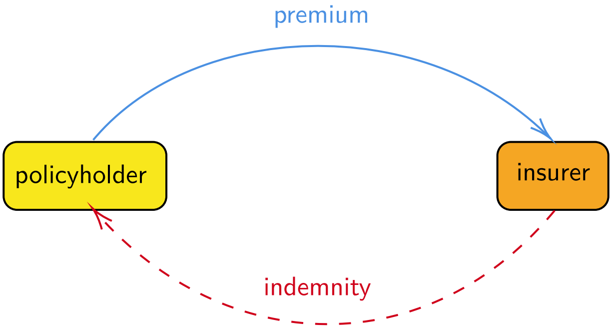




Joint work with Philipp Ratz (PhD), Yves-Alexandre de Montjoye (Imperial College), Ewen Gallic (AMSE), Laurence Barry (PARI), Lariosse Kouakou (MSc), Matthias Löwe (Universität Münster), Franck Vermet (Brest), Rawanda Matar (MSc), Michel Denuit (Louvain), Julien Trufin (Bruxelles), etc

# Insurance & Actuarial Science

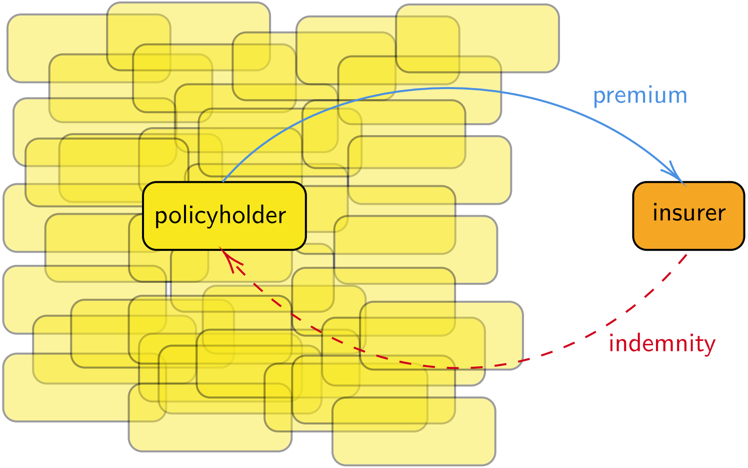



What would be a "fair premium" ? see O'Neill (1997)

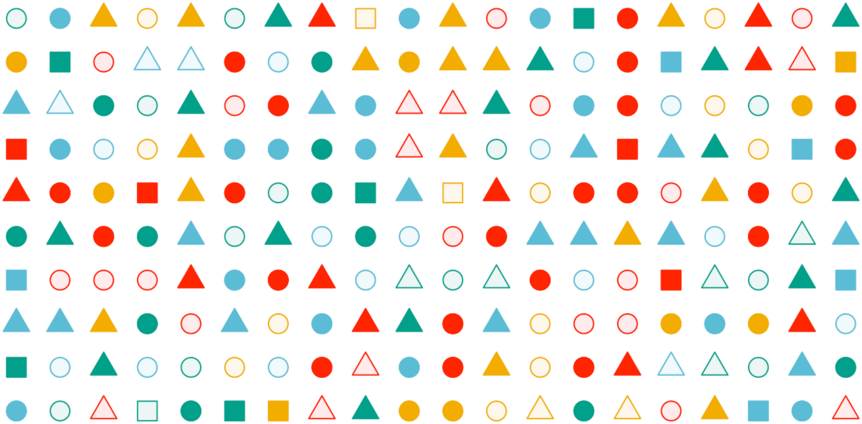
- pure actuarial fairness contributions for individual policyholders should perfectly reflect their predicted risk levels $\rightarrow$ predictive modeling
- choice-sensitive fairness contributions should take into account only risks that result from choices - luck-egalitarianism (Cohen (1989) or Arneson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

# Insurance & Actuarial Science

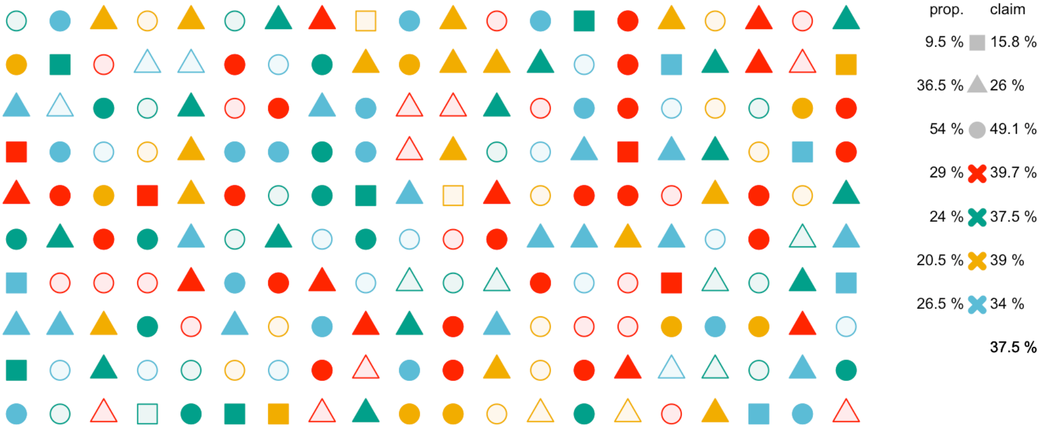
"*Insurance is the contribution of the many to the misfortune of the few*"



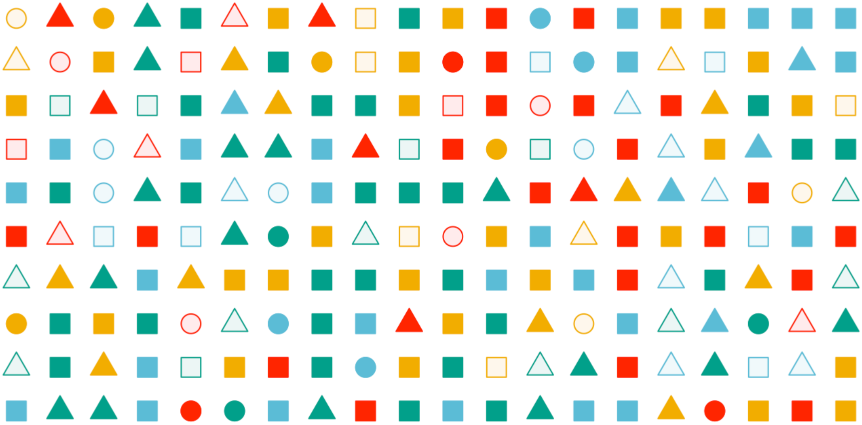
Consider two features, $x_1 \in \{$ ■ ▲ ● $\}$ and $x_2 \in \{$ ■ ■ ■ ■ $\}$



Consider two features, $x_1 \in$ { ■ ▲ ● } and $x_2 \in$ { ■ ■ ■ ■ }



Consider two features, $x_1 \in \{$ ■ ▲ ● $\}$ and $x_2 \in \{$ ■ ■ ■ ■ $\}$



| prop. | | claim |
|---|---|---|
| 58.5 % | ■ | 16.2 % |
| 29 % | ▲ | 37.9 % |
| 12.5 % | ● | 44 % |
| 58.5 % | ✖ | 26.8 % |
| 29 % | ✖ | 22.4 % |
| 12.5 % | ✖ | 22 % |
| 58.5 % | ✖ | 33.3 % |
| | | 26 % |

| | | | |
|---|---|---|---|
| 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 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 $\Omega$

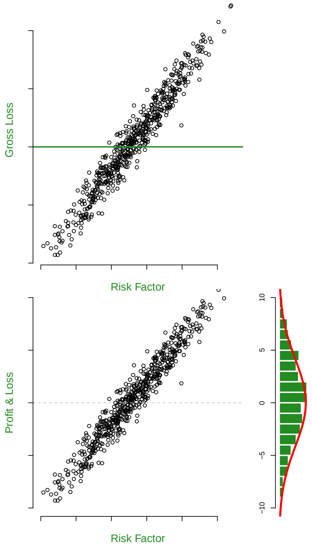
$$\pi(\omega) = \mathbb{E}_{\mathbb{P}}[S|\Omega = \omega]$$

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

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

# 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).

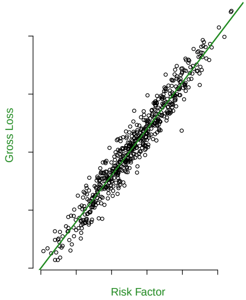
# Risk Transfert with Segmentation and Perfect Information

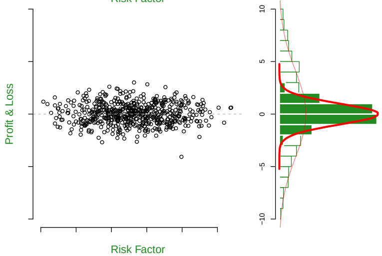
Assume that information $\mathbf{\Omega}$ is observable,

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

Observe that $\text{Var}\Big[S-\mathbb{E}[S|\mathbf{\Omega}]\Big]=\mathbb{E}\Big[\text{Var}[S|\mathbf{\Omega}]\Big]$, so that

$$\text{Var}[S]=\underbrace{\mathbb{E}\Big[\text{Var}[S|\mathbf{\Omega}]\Big]}_{\to\text{ insurer}}+\underbrace{\text{Var}\Big[\mathbb{E}[S|\mathbf{\Omega}]\Big]}_{\to\text{ insured}}.$$



# Segmentation and Imperfect Information

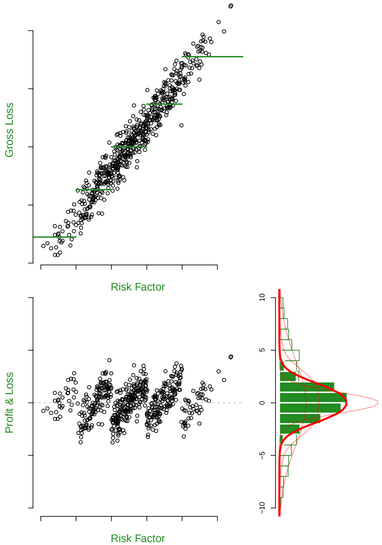
Assume that $\boldsymbol{X} \subset \boldsymbol{\Omega}$ is observable

| | Insured | Insurer |
|---|---|---|
| Loss | $\mathbb{E}[S\vert\boldsymbol{X}]$ | $S - \mathbb{E}[S\vert\boldsymbol{X}]$ |
| Average Loss | $\mathbb{E}[S]$ | $0$ |
| Variance | $\text{Var}\Big[\mathbb{E}[S\vert\boldsymbol{X}]\Big]$ | $\mathbb{E}\Big[\text{Var}[S\vert\boldsymbol{X}]\Big]$ |

Now

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

spiral of segmentation...



# Segmentation and Imperfect Information

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

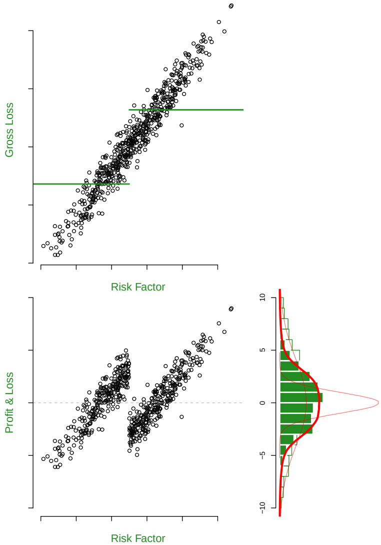
| | Insured | Insurer |
|---|---|---|
| Loss | $\mathbb{E}[S\|\boldsymbol{X}]$ | $S - \mathbb{E}[S\|\boldsymbol{X}]$ |
| Average Loss | $\mathbb{E}[S]$ | $0$ |
| Variance | $\text{Var}\big[\mathbb{E}[S\|\boldsymbol{X}]\big]$ | $\mathbb{E}\big[\text{Var}[S\|\boldsymbol{X}]\big]$ |

Now

$$\mathbb{E}\Big[\text{Var}[S|\boldsymbol{X}]\Big] = \mathbb{E}\Big[\mathbb{E}\Big[\text{Var}[S|\boldsymbol{\Omega}]\Big|\boldsymbol{X}\Big]\Big] + \mathbb{E}\Big[\text{Var}\Big[\mathbb{E}[S|\boldsymbol{\Omega}]\Big|\boldsymbol{X}\Big]\Big]$$

$$= \underbrace{\mathbb{E}\Big[\text{Var}[S|\boldsymbol{\Omega}]\Big]}_{\text{perfect pricing}} + \underbrace{\mathbb{E}\Big\{\text{Var}\Big[\mathbb{E}[S|\boldsymbol{\Omega}]\Big|\boldsymbol{X}\Big]\Big\}}_{\text{misfit}}.$$

spiral of segmentation...



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

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

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

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

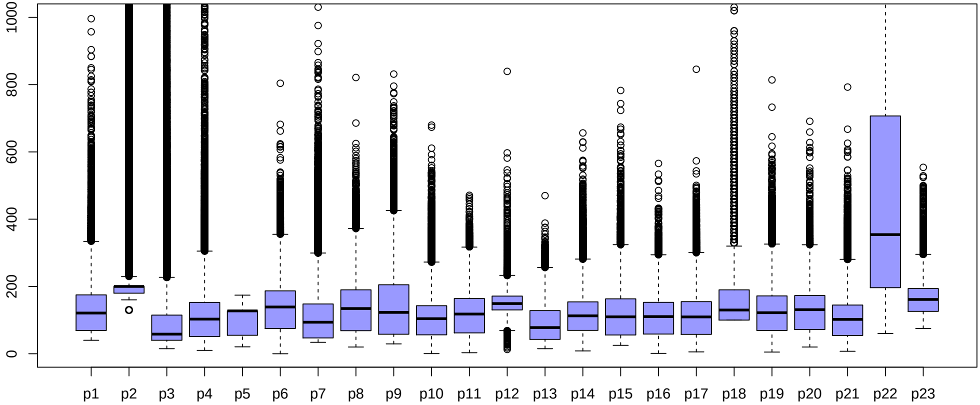
$$\widehat{\pi}_j(\boldsymbol{x})=\widehat{\mathbb{E}}[N_1|\boldsymbol{X}=\boldsymbol{x}]\cdot\widehat{\mathbb{E}}[Y|\boldsymbol{X}=\boldsymbol{x}] \;=\; \underbrace{\exp(\widehat{\alpha}^\top\boldsymbol{x})}_{\text{Poisson }\mathcal{P}(\lambda_{\boldsymbol{x}})}\cdot\underbrace{\exp(\widehat{\beta}^\top\boldsymbol{x})}_{\text{Gamma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

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

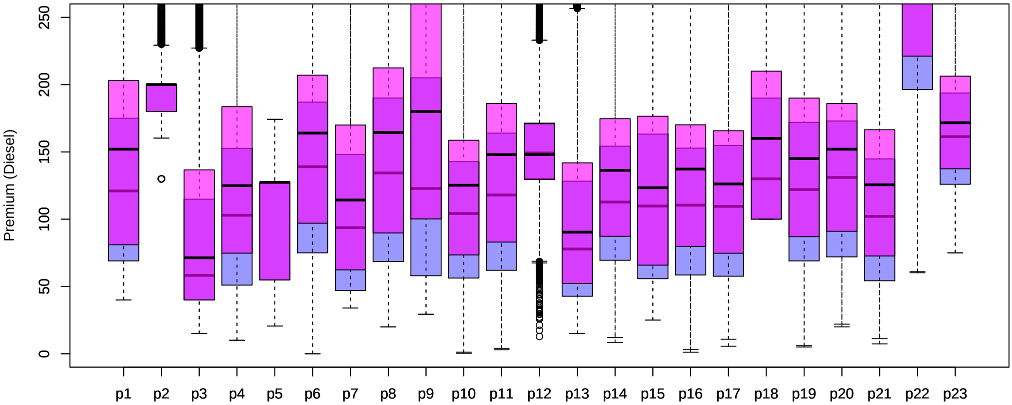
$$\widehat{\pi}_j(\boldsymbol{x})\text{ where }\widehat{\pi}_j\in\underset{m\in\mathcal{F}_j:\mathcal{X}_j\to\mathbb{R}}{\text{argmin}}\left\{\sum_{i=1}^{n}\ell(s_i,m(\boldsymbol{x}_i))+\lambda\cdot\text{penalty}(m)\right\}$$

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

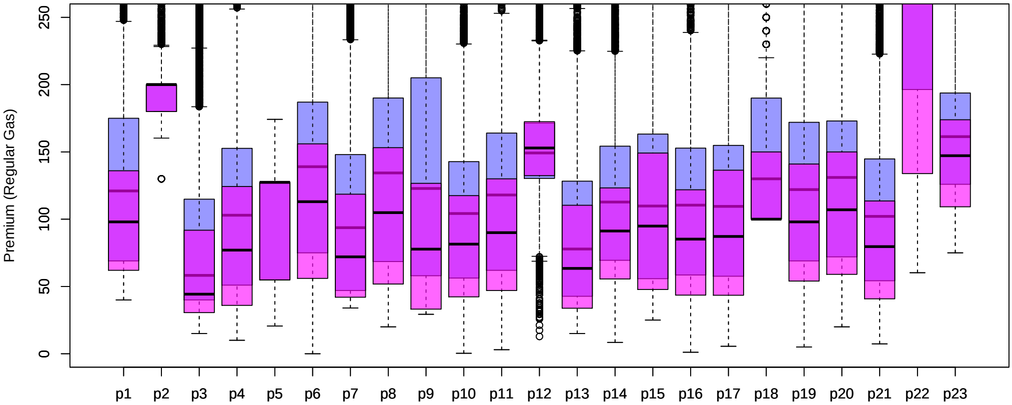
# Comparing models



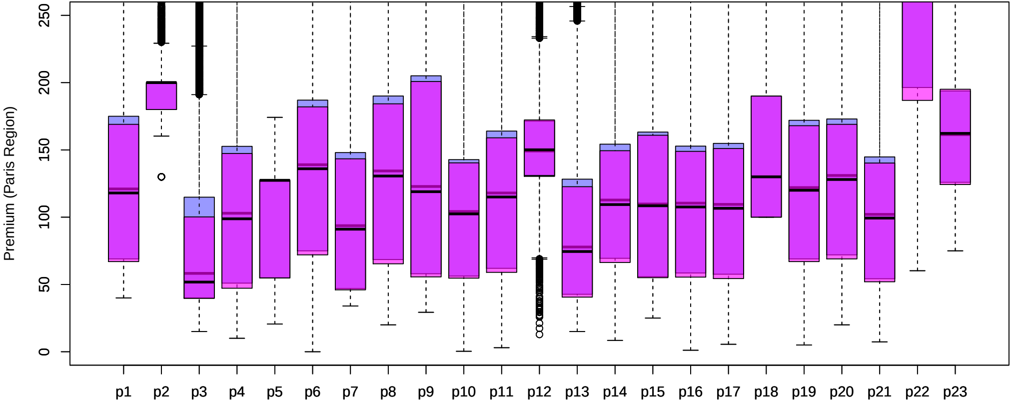
# Comparing models ( Gas Type Diesel)



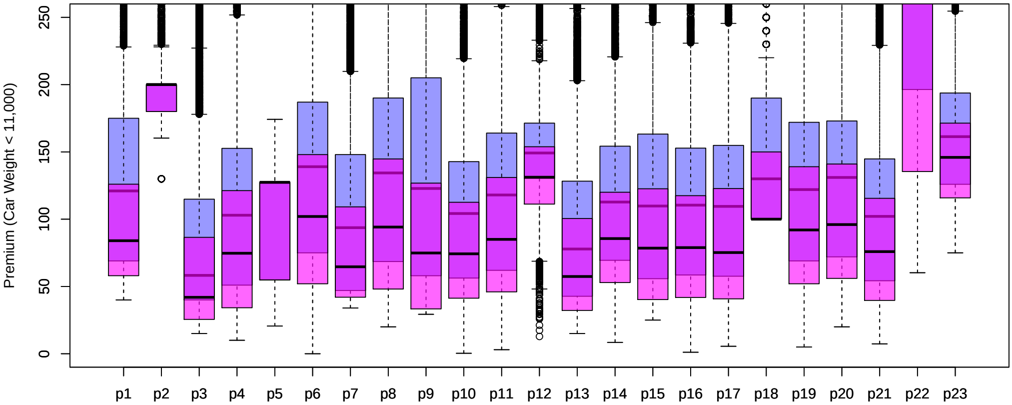
# Comparing models ( Gas Type Regular)



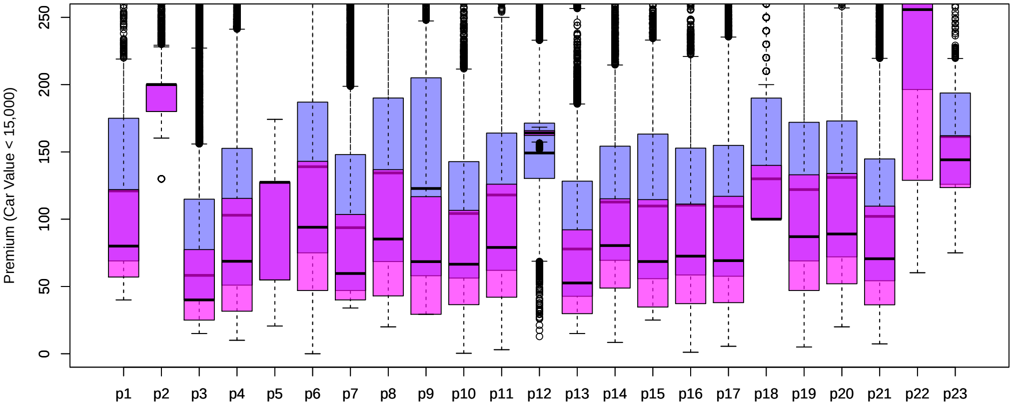
# Comparing models ( Paris Region)



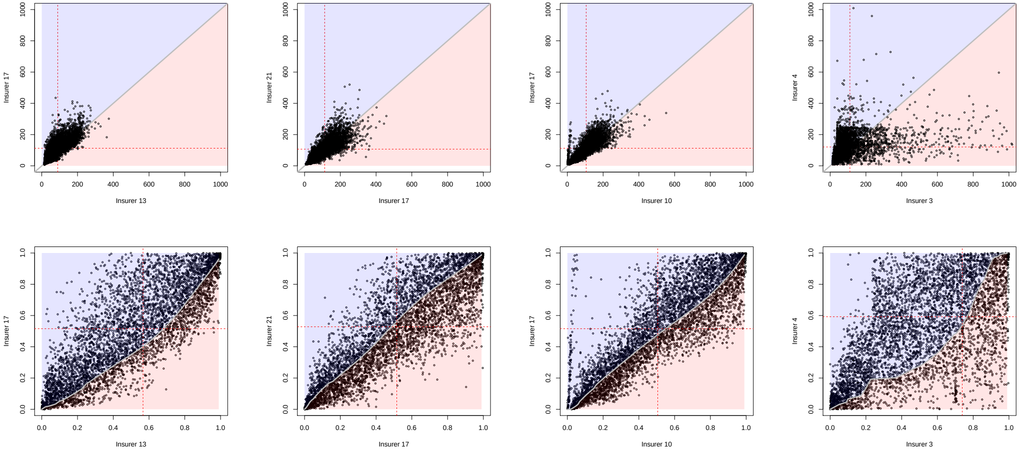
# Comparing models ( Car Weight)



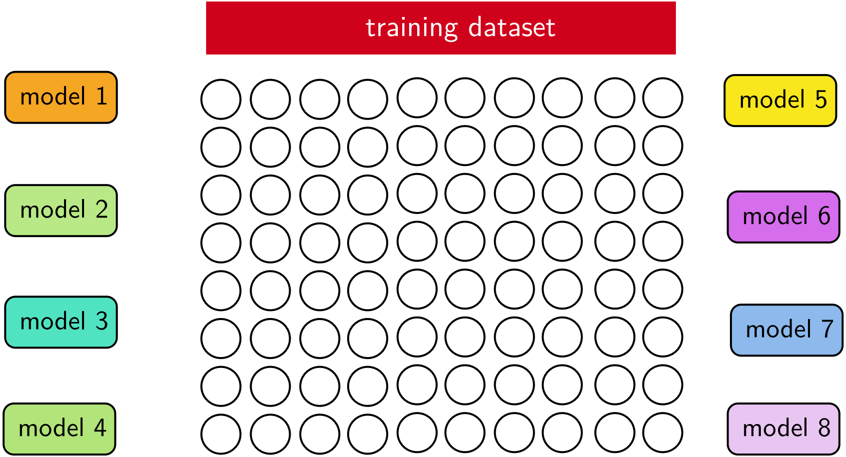
# Comparing models ( Car Value)

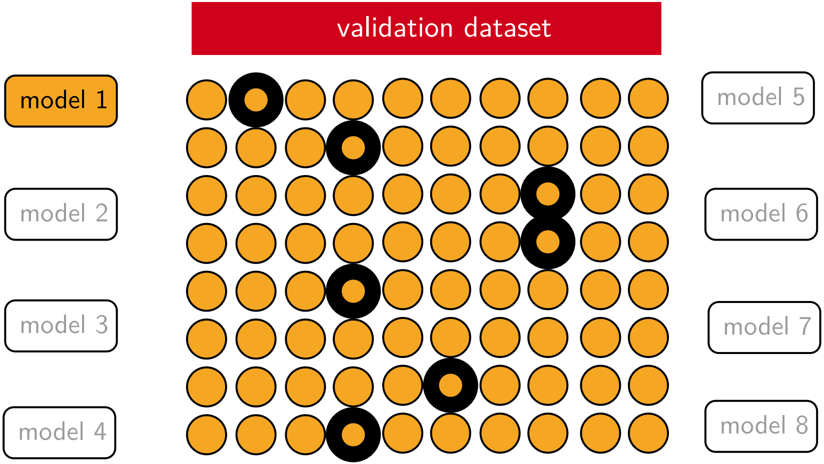


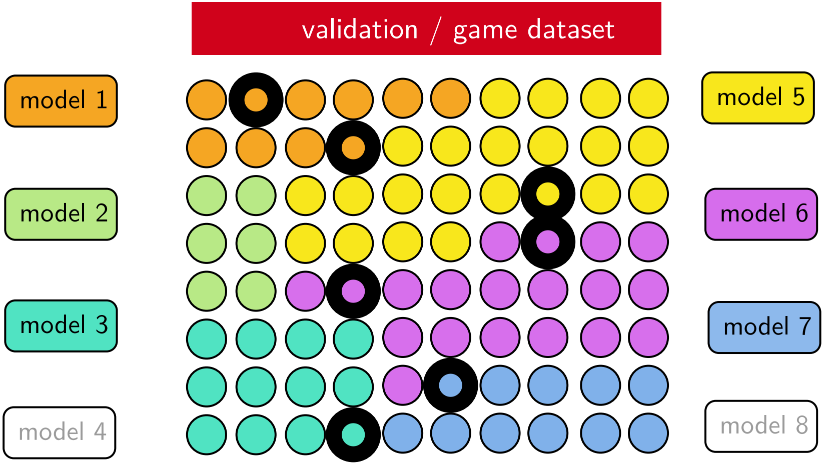
# Comparing models ( Correlation)

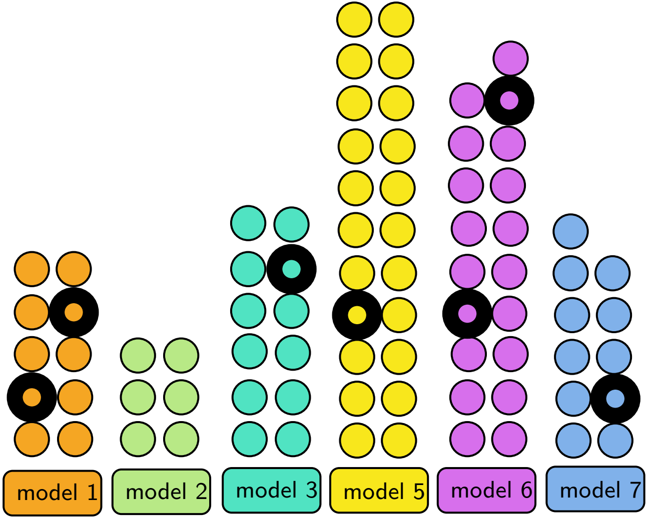


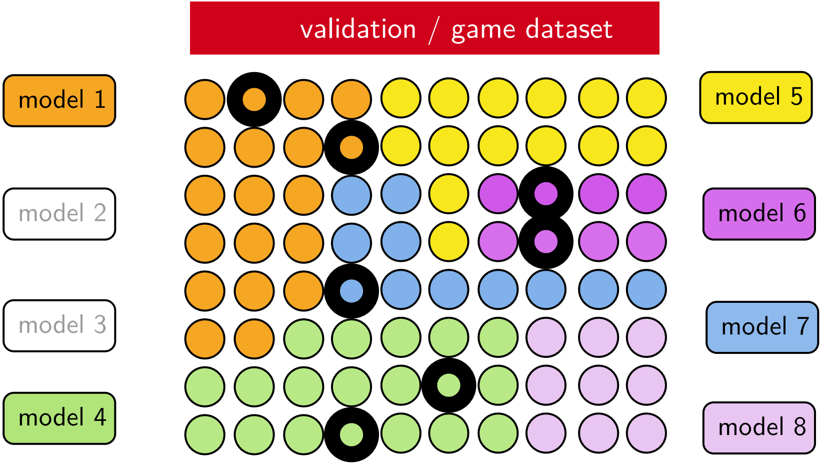
# Insurance & Actuarial Science

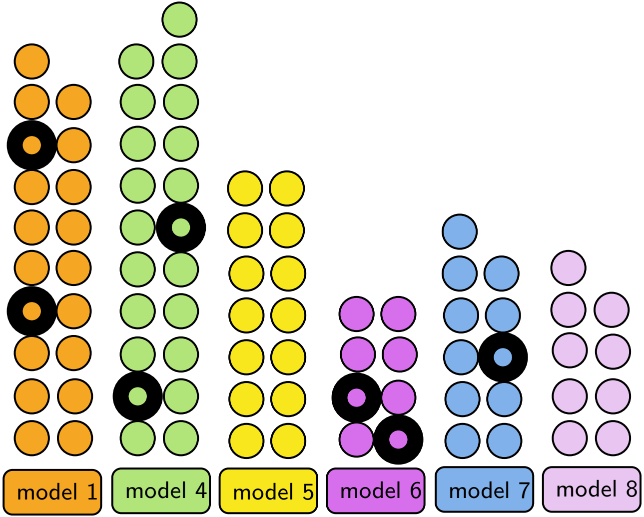



validation dataset
model 1
model 2
model 3
model 4
model 5
model 6
model 7
model 8

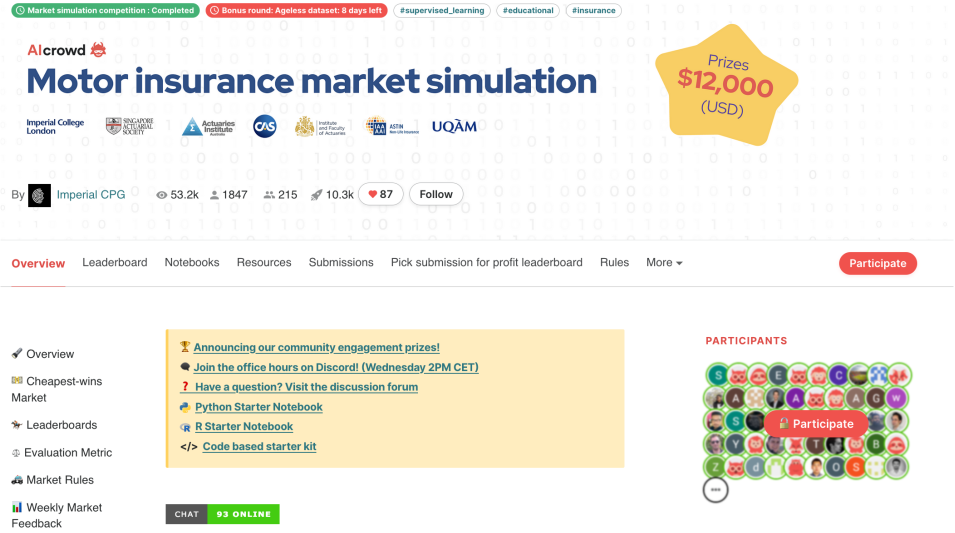

validation / game dataset
model 1
model 2
model 3
model 4
model 5
model 6
model 7
model 8


model 1
model 2
model 3
model 5
model 6
model 7

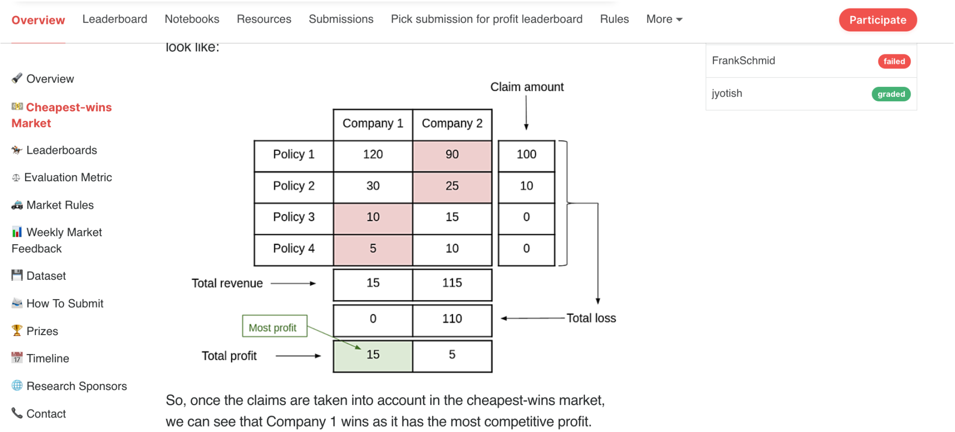

validation / game dataset
model 1
model 2
model 3
model 4
model 5
model 6
model 7
model 8


model 1
model 4
model 5
model 6
model 7
model 8

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



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



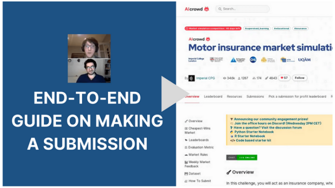
https://www.aicrowd.com/challenges/insurance-pricing-game

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

- 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.

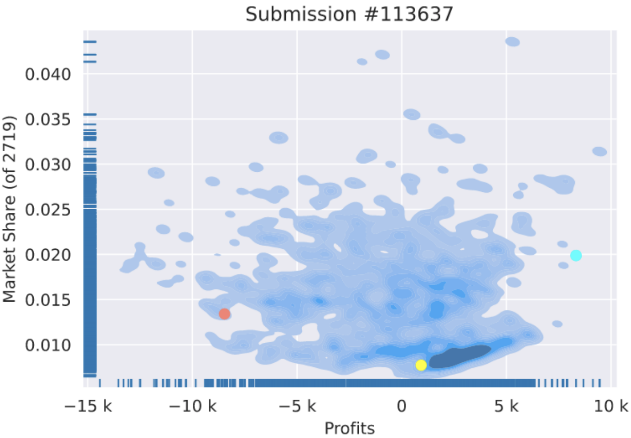
**SUBMISSION VIA COLABORATORY NOTEBOOKS**



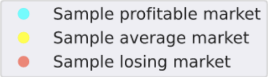

1. **Python.** Visit the starter notebook here.
2. **R.** Visit the starter notebook here. Here is an end-to-end walkthrough for a simple R model.

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

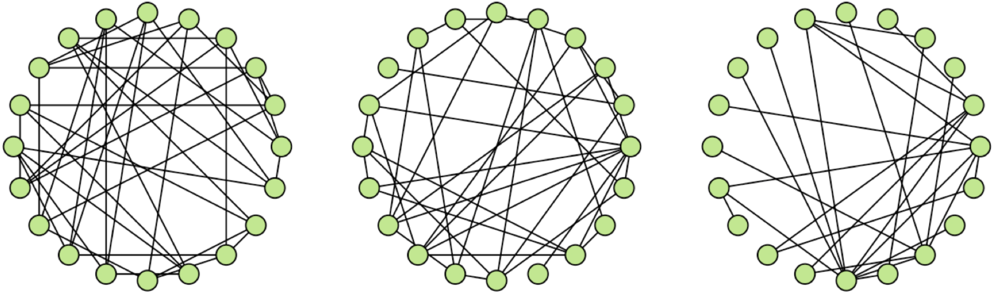
https://www.aicrowd.com/challenges/insurance-pricing-game




Average Market Share: 0.01
Average Revenue: 13410.69
Average Losses: 13589.48
Average Loss Ratio: 0.99
Average Profit per Policy: 1.13
Number of markets: 2719
Average profit: -178.79



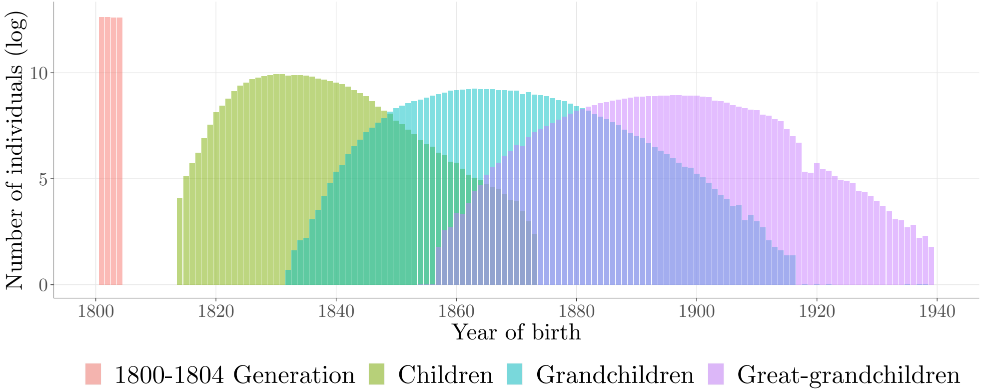
Charpentier *et al.* (2021) Collaborative Insurance Sustainability and Network Structure



# 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 competiting against another one

  ex: what to do if $\widehat{\pi}_{\text{your model}}(\boldsymbol{x}) > \widehat{\pi}_{\text{competitor}}(\boldsymbol{x})$ ?

  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 ?