La mutualisation et l'inclusion à l'épreuve de la segmentation

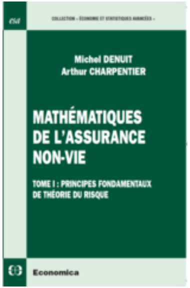
Arthur Charpentier (Montréal, Canada)

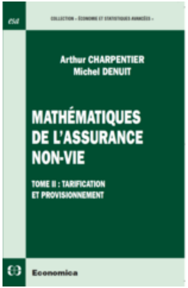
3^{e} Colloque International de l'Actuariat Francophone

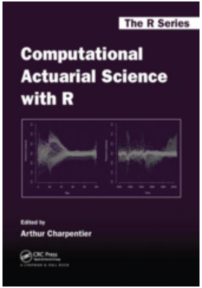
bio (succinte)

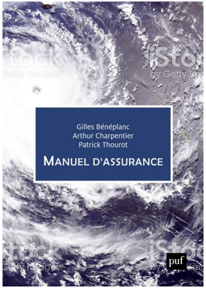
Professeur à l'Université du Québec à Montréal (https://freakonometrics.github.io/)

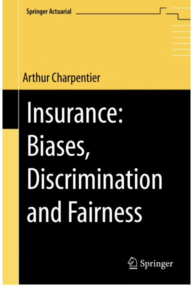
- Denuit and C. (2004, 2005) Mathématiques de l'Assurance Non-Vie,
- C. (2014) Computational Actuarial Science with R,
- Bénéplanc et al. (2022) Manuel d'Assurance,
- C. (2024) Insurance: Biases, Discrimination and Fairness.







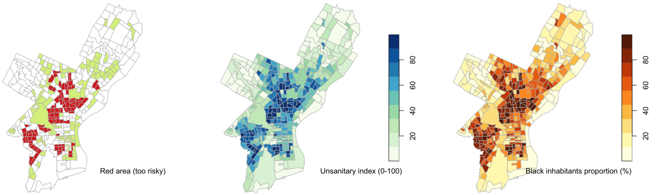




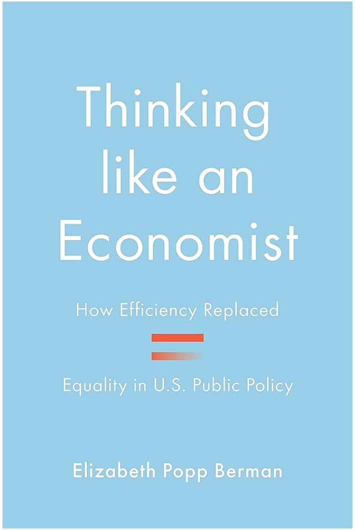
justification économique de la segmentation

- notion de de rationalité et d'actuarial fairness,
- nécessaire de tenir compte de l'hétérogénéité
- cf efficacité en économie, Berman (2022)

- quelle hétérogénéité ? est-ce juste ?



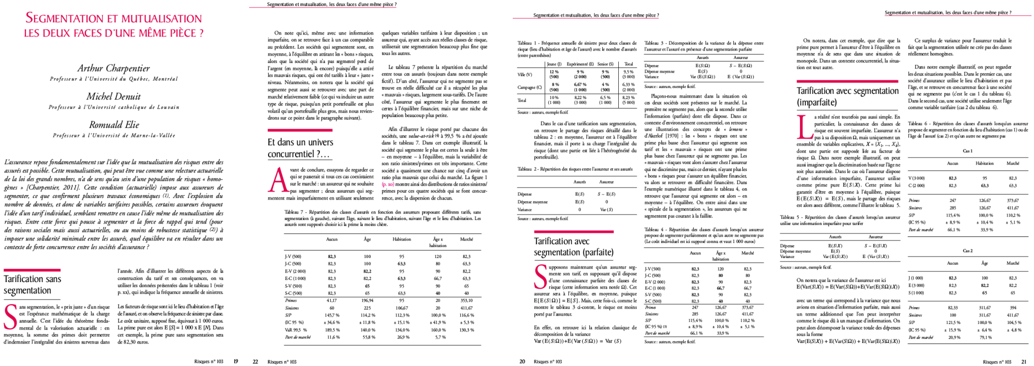
- cf "redlining" (assurance ou crédit)
  la prime d'assurance est-elle plus chère
  - parce que les maisons sont vétustes ?
  - parce que la population est noire ?



de la spirale de la segmentation

- "*Insurance is the contribution of the many to the misfortune of the few*"
- Mathématiques de la segmentation (De Wit and Van Eeghen (1984), C. et al. (2014))

$$\mathrm{Var}[Y] = \underbrace{\mathrm{Var}[\mathbb{E}[Y|\boldsymbol{X}]]}_{\to \text{ assuré}} + \underbrace{\mathbb{E}[\mathrm{Var}[Y|\Theta]]}_{\to \text{ modèle "parfait"}} + \underbrace{\mathbb{E}\{\mathrm{Var}[\mathbb{E}[Y|\Theta]|\boldsymbol{X}]\}}_{\to \text{ imperfection}}.$$



de la théorie aux expériences, les premiers "pricing games"

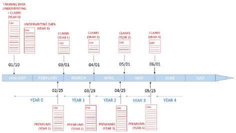
- **2015-2017, "pricing games"** (en lien avec l'institut des actuaires)

- Trois compétitions (2015, 2016 et 2017)
- De 10 à 25 participants (qui donnaient des prix, pas des modèles)

THIRD ACTUARIAL PRICING GAME

From January 2017 till June 2017, we organize the **Third Actuarial Pricing Game**, as part of a research project conducted by **Arthur Charpentier**, Université de Rennes 1 (France) & Quantact (Montréal, Canada), with the support of the ACTINFO chair of the **Institut Louis Bachelier**, and the **Institut des Actuaires**, the French Institute of Actuaries.

1. THE RULES : AGENDA



On **January 10th**, datasets with 100,000 insured is provided to all (potential) players. There are two datasets, for Year 0 :
— an underwriting dataset (U_0), with information about insurance policies, insured drivers and their cars (see page 5 for description of the variables)
— an claims dataset (C_0), with all claims occured during year 0 to all policyholders
An underwriting dataset (U_1) with (the same) 100,000 policies (drivers willing to purchase insurance for Year 1) is provided. Players must provide prices for those 100,000 insured.
A **player** is a dataset we receive. It can be sent by individuals, or groups, practitionners, students or academics.

For **February 25th**, players must send a dataset with 100,000 prices to pricing-game@univ-rennes1.fr in a csv file, with two columns : the id_policy, and the premium. Let $p_{i,j,1}$ denote the annual premium for policy i, offered by player j (for year 1). Let $\ell_{i,1}$ denote the losses for policy i (unknown by players when submitting their premiums).

1

lightgray **Num**	**Family Type**	**Name**	**Label**	**Format**
1	ID	id_client	ID- Client ID	string
2	ID	id_vehicle	ID- Vehicle ID	string
3	ID	id_year	ID- Year	string
4	Claims	id_claim	Claims- Claim ID	string
5	Claims	claim_nb	Claims- Number of Claims	int
6	Claims	claim_amount	Claims- Total Claims Amount	int

Variables List : Claims database

lightgray **Num**	**Family Type**	**Name**	**Label**	**Format**
1	ID	id_client	ID- Client ID	string
2	ID	id_vehicle	ID- Vehicle ID	string
3	ID	id_policy	ID- Policy ID	string
4	ID	id_year	ID- Year	string
5	Policy	pol_bonus	Policy- Bonus Coefficient	float
6	Policy	pol_coverage	Policy- Coverage	string
7	Policy	pol_duration	Policy- Duration	int
8	Policy	pol_sit_duration	Policy- Current Endorsment Duration	int
9	Policy	pol_pay_freq	Policy- Payment Frequency	string
10	Policy	pol_payd	Policy- Payd Indicator	string
11	Policy	pol_usage	Policy- Usage	string
12	Policy	pol_insee_code	Policy- Insee Town Code	string
13	Drivers	drv_drv2	Drivers- Secondary Driver Presence Indicator	string
14	Drivers	drv_age1	Drivers- First Driver Age	int
15	Drivers	drv_age2	Drivers- Secondary Driver Age	int
16	Drivers	drv_sex1	Drivers- First Driver Gender	string
17	Drivers	drv_sex2	Drivers- Secondary Driver Gender	string
18	Drivers	drv_age_lic1	Drivers- First Driver Licence Age	int
19	Drivers	drv_age_lic2	Drivers- Secondary Driver Licence Age	int
20	Vehicle	vh_age	Vehicle- Vehicle Age	int
21	Vehicle	vh_cyl	Vehicle- Engine Capacity	int
22	Vehicle	vh_din	Vehicle- Din Power	int
23	Vehicle	vh_fuel	Vehicle- Fuel Type	string
24	Vehicle	vh_make	Vehicle- Make	string
25	Vehicle	vh_model	Vehicle- Model	string
26	Vehicle	vh_sale_begin	Vehicle- Sales Date Beginning	int
27	Vehicle	vh_sale_end	Vehicle- Sales Date End	int
28	Vehicle	vh_speed	Vehicle- Max Speed	int
29	Vehicle	vh_type	Vehicle- Type	string
30	Vehicle	vh_value	Vehicle- Value	int
31	Vehicle	vh_weight	Vehicle- Weight	int

Variables List : Underwriting database

2 Instructions, partie 1 (15 Juillet - 15 Septembre)

Le but est de proposer une prime pure pour les 37 772 contrats de la base pricing. Sont attendus

- une base (un fichier csv) constituée de deux colonnes (seulement): la première nu_affa sera le numéro de police et la seconde Premium contiendra la prime pure proposée. La base sera constituée de 37 772 lignes (il est impératif de proposer une prime pour *tous* les contrats).
- un rapide descriptif de la méthodologie utilisée, décrivant les variables retenues.

Merci d'indiquer un nom déquipe pour chaque modèle proposé, et d'envoyer le fichier csv avec un rapide descriptif à PricingGame@institutdesactuaires.com, avant le 15 Septembre.

3 Instructions, partie 2 (20 Septembre - 20 Octobre)

Le 20 Septembre, tous les participants recevront les primes de 2 de leurs compétiteur au moins.

Les participants auront alors la possibilité de réviser (ou pas) la prime qu'ils proposent. Un fichier csv révisé - et un rapide descriptif de la méthodologie adoptée - devront être envoyés à PricingGame@institutdesactuaires.com pour le 20 Octobre, au plus tard.

4 Règles du Jeu

Une fois les primes collectées, nous fonctionnerons comme un agrégateur de prix: chacun des xxx assurés se verra affecter une compagnie selon une m 'ethode qui sera basée sur une sélection aléatoire parmi les primes les moins chères.
Les organisateurs normaliseront les prix pour éviter les stratégies de dumping. En l'occurence, la somme des primes proposées par un assureur sera égalisée à 2,5 millions d'euros (par un facteur d'inflation appliqué uniformément).

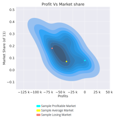
Les résultats seront présentés lors de la journée 100% Data Science organisée à Paris (voir également sur http://freakonometrics.hypotheses.org).

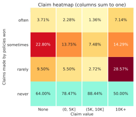
Pour plus d'information, PricingGame@institutdesactuaires.com ou @freakonometrics sur Twitter.

retour sur la compétition de 2021

▶ 2021, "IA Crowd pricing game"

- 11 semaines $\approx 2,000$ participants, $10,000$ modèles
- retours sur les performances en compétition

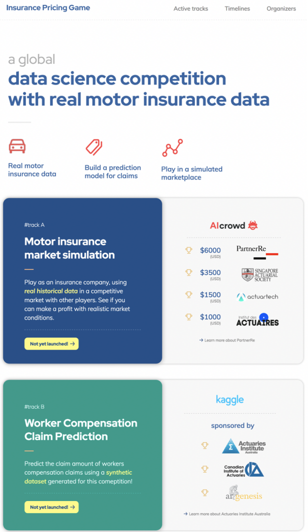




Driver Summaries

Policies won	drv_sex1	drv_age1	drv_age_lic1	drv_drv2	drv_sex2	drv_age2	drv_age_lic2
often: 54.5 - 100.0%	M (63%)	61.57	38.70	No (73%)	F (66%)	50.00	27.74
sometimes: 18.2 - 54.5%	M (65%)	59.19	36.87	No (72%)	F (63%)	51.05	29.14
rarely: 9.1 - 18.2%	M (63%)	60.92	38.35	No (71%)	F (62%)	51.09	29.12
never:	M (58%)	58.17	36.08	No (64%)	F (60%)	51.16	29.31
in sample profitable market	M (64%)	61.05	38.33	No (74%)	F (64%)	49.29	27.25
in sample average market	M (65%)	61.73	38.99	No (75%)	F (65%)	50.43	28.29
in sample losing market	M (63%)	59.89	37.37	No (70%)	F (64%)	50.70	28.69

- compétition avec un prix à gagner (donc une métrique cible de succès)

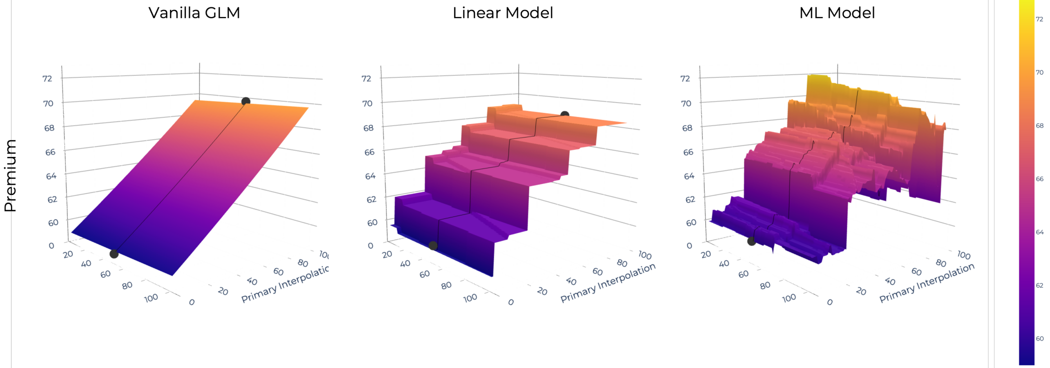


compétition entre modèles, quel équilibre ?

- modèles d'apprentissage machine souvent mal calibrés, cf Denuit et al. (2021)

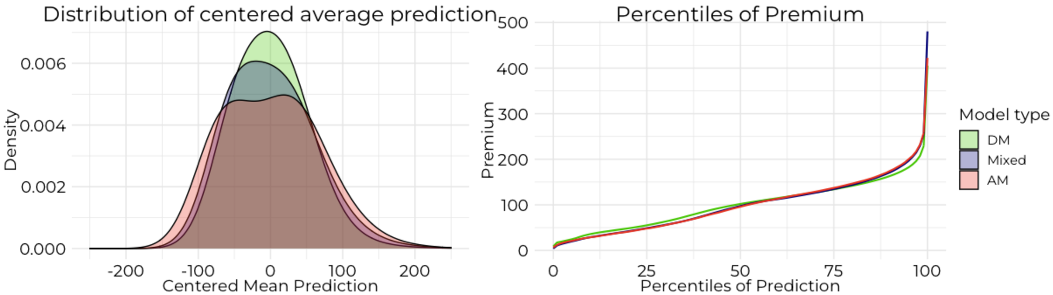
$$\mathbb{E}[Y|\widehat{Y}=y]=y,\ \forall y$$

- apprentissage machine = plus de variabilité (avant la compétition)



compétition entre modèles, quel équilibre ?

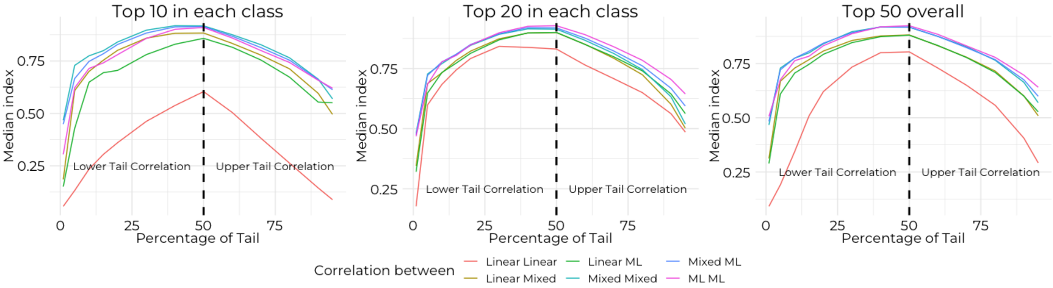
- apprentissage machine = plus de variabilité (avant la compétition)



(via Ratz et al. (2023))

compétition entre modèles, quel équilibre ?

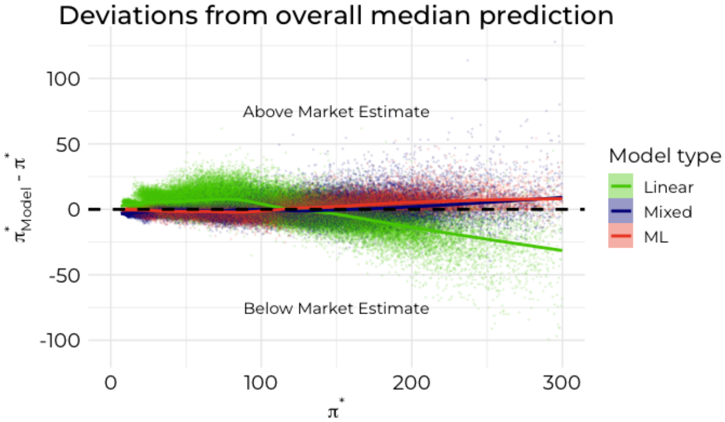
- peu de consensus entre modèles d'apprentissage machine
- ici corrélation (moyenne) de queues entre les modèles
  - à gauche, $u \to \mathbb{P}[X_i \leq F_i^{-1}(u) | X_j \leq F_j^{-1}(u)]$
  - à droite, $u \to \mathbb{P}[X_i > F_i^{-1}(u) | X_j > F_j^{-1}(u)]$



(via Ratz et al. (2023))

compétition entre modèles, quel équilibre ?

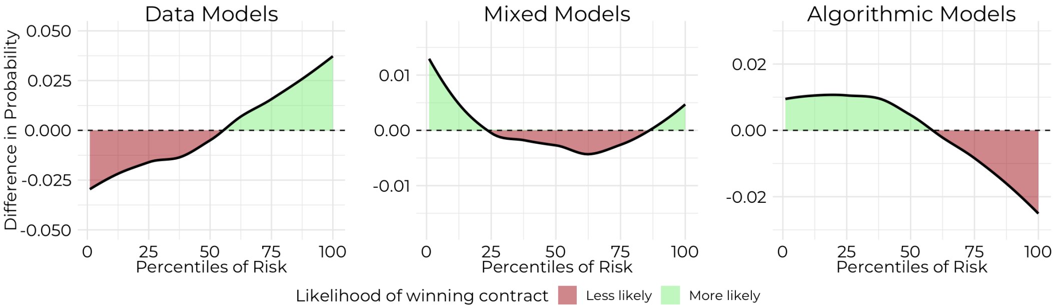
- les modèles linéaires sont chers sur les bas risques, pas chers sur les hauts risques



(via Ratz et al. (2023))

compétition entre modèles, quel équilibre ?

- les modèles linéaires sont chers sur les bas risques, pas chers sur les hauts risques



(via Ratz et al. (2023))