



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

Arthur Charpentier (Montréal, Canada)

3^e Colloque International de l'Actuariat Francophone



International Actuarial Association
Association Actuarielle Internationale



International Actuarial Association
Let's build the future!
Savoir, mutualiser
Institut des Actuairens en Belgique
Institut van de Actuarissen in België
Institute of Actuaries in Belgium

IA|BE



Etablissement des Actuairens
du Cameroun



Institut
canadien
des actuairens
Canadian
Institute
of Actuaries



Institut des
ACTUAIRES



LEBANESE
ASSOCIATION
OF ACTUARIES
الجمعية اللبنانية للتأمين



Association
Marocaine
des Actuairens



INAS
Institut National
des Actuairens
au Sénégal



SAV
ASA
ASA
Schweizerische
Aktuarvereinigung
Association Suisse
des Actuairens
Associazione Svizzera
degli Attuari

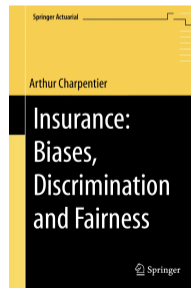
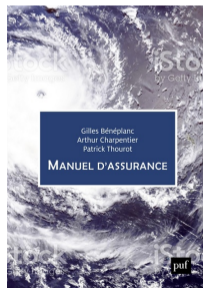
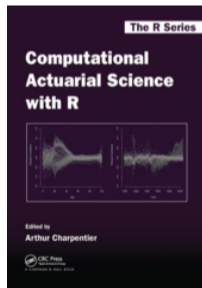
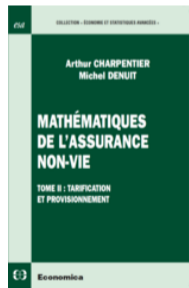
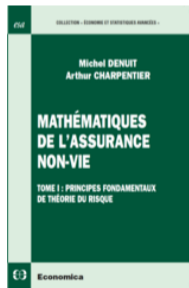


Association Tunisienne des Actuairens

bio (succinte)

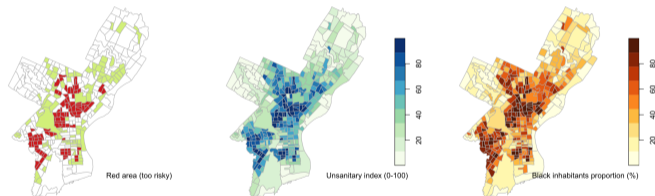
Professeur à l'Université du Québec à Montréal (<https://freakonometrics.github.io/>)

- 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.

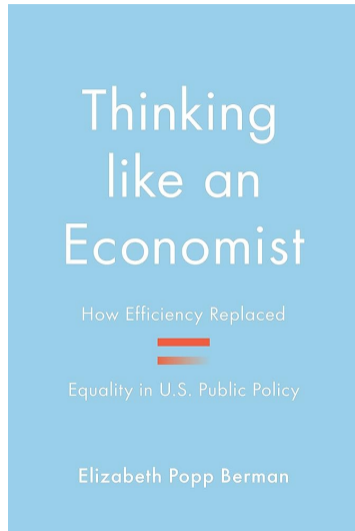


justification économique de la segmentation

- 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))

$$\text{Var}[Y] = \underbrace{\text{Var}[\mathbb{E}[Y|X]]}_{\rightarrow \text{assuré}} + \underbrace{\mathbb{E}[\text{Var}[Y|\Theta]]}_{\rightarrow \text{modèle "parfait"}} + \underbrace{\mathbb{E}\{\text{Var}[\mathbb{E}[Y|\Theta]|X]\}}_{\rightarrow \text{imperfecton}}$$

SEGMENTATION ET MUTUALISATION LES DEUX FACES D'UNE MÊME PIÈCE ?

Arthur Charpentier
Professeur à l'Université du Québec, Montréal

Michel Demut
Professeur à l'Université catholique de Louvain

Romuald Elie
Professeur à l'Université de Moncton-La-Valley

Assurance repose fondamentalement sur l'idée que la mutualisation des risques entre des assurés est possible. Cette mutualisation, qui peut être vue comme une sélection actuarielle de la loi des grands nombres, n'a de sens qu'en tant que celle d'une population de risque « homogène » (Charpentier, 2011). Cette sélection (actuarielle) impose aux assurés de segmenter, ce qui confère plusieurs avantages homogènes (1). Avec l'évolution des modes de données, et donc de variables disponibles, certaines assurances dépassent l'idée d'un tarif individuel, semblant remettre en cause l'idée même de mutualisation des risques. Entre ceux ceux qui peuvent segmenter et la force de regrouper qui impose des risques sociaux mais aussi actuariels, on se interroge sur le statut (ou) à imposer aux individus mutualisant entre les assurés, qui équilibrent ou en résultent dans un contexte de forte concurrence entre les assurés d'assurance ?

Tarifcation sans segmentation

Si une segmentation, le « prix pour » d'un risque est l'espérance mathématique de la charge annuelle. C'est l'idée de la loi des grands nombres de la volatilité actuarielle : on suppose, la somme des primes doit permettre d'équilibrer l'espérance des sinistres sur une durée de 30-50 ans.

La fonction de risque unitaire à la fois d'habitation et d'âge de l'assuré, on se interroge sur la façon de décrire ce risque dans un contexte de forte concurrence. Les coûts sociaux supportés des sinistres à 100 ans. Le premier pas est d'être (1) à 1000 € (2) Dans ce contexte, la somme des primes doit permettre d'équilibrer l'espérance des sinistres sur une durée de 30-50 ans.

Segmentation et mutualisation, les deux faces d'une même pièce ?

Ce sera qu'il, même avec une information imparfaite, on se interroge sur la façon de décrire ce risque dans un contexte de forte concurrence. Les coûts sociaux supportés des sinistres à 100 ans. Le premier pas est d'être (1) à 1000 € (2) Dans ce contexte, la somme des primes doit permettre d'équilibrer l'espérance des sinistres sur une durée de 30-50 ans.

Le tableau 7 présente la répartition du marché entre tous les assurés habitation dans notre exemple fictif. Il est clair, l'assuré qui ne segmente pas se trouve en réalité défavorisé car il est réparti sur le plus « mauvais » risque, logiquement sous-évalué. De ce fait, l'assuré qui segmente le plus bénéficie en tant qu'il équilibre l'assurance, mais son acte est-elle en ce point dans le paysage social.

Et dans un univers concurrentiel ? ...

Assurance repose fondamentalement sur l'idée que la mutualisation des risques entre des assurés est possible. Cette mutualisation, qui peut être vue comme une sélection actuarielle de la loi des grands nombres, n'a de sens qu'en tant que celle d'une population de risque « homogène » (Charpentier, 2011). Cette sélection (actuarielle) impose aux assurés de segmenter, ce qui confère plusieurs avantages homogènes (1). Avec l'évolution des modes de données, et donc de variables disponibles, certaines assurances dépassent l'idée d'un tarif individuel, semblant remettre en cause l'idée même de mutualisation des risques. Entre ceux ceux qui peuvent segmenter et la force de regrouper qui impose des risques sociaux mais aussi actuariels, on se interroge sur le statut (ou) à imposer aux individus mutualisant entre les assurés, qui équilibrent ou en résultent dans un contexte de forte concurrence entre les assurés d'assurance ?

Tableau 7 - Répartition des classes d'assurés en fonction des sinistres supportés différents coûts, une segmentation à forfait, selon la loi d'habitation, selon l'âge et la loi d'habitation. Le marché nous suppose donc à la prime le même prix.

Année	Age	Habitation	Age + Habitation	Marché
1/1/2000	82,3	39,9	91	120
1/1/2000	82,3	39,9	82,3	80
1/1/2000	82,3	82,3	91	80
1/1/2000	82,3	82,3	82,3	80
1/1/2000	82,3	40	91	80
1/1/2000	82,3	40	82,3	80
Total	112,7	198,8	91	28
Assurés	80	215	1062,7	20
SP	142,7	114,7	112,3	108,6
IC (100)	6,844%	4,137%	1,112%	4,53%
IC (100)	10,5%	140,7%	108,6%	108,6%
Année de marché	114,6%	518,8%	30,9%	5,7%

Segmentation et mutualisation, les deux faces d'une même pièce ?

Ce sera qu'il, même avec une information imparfaite, on se interroge sur la façon de décrire ce risque dans un contexte de forte concurrence. Les coûts sociaux supportés des sinistres à 100 ans. Le premier pas est d'être (1) à 1000 € (2) Dans ce contexte, la somme des primes doit permettre d'équilibrer l'espérance des sinistres sur une durée de 30-50 ans.

Le tableau 8 présente la répartition du marché entre tous les assurés habitation dans notre exemple fictif. Il est clair, l'assuré qui ne segmente pas se trouve en réalité défavorisé car il est réparti sur le plus « mauvais » risque, logiquement sous-évalué. De ce fait, l'assuré qui segmente le plus bénéficie en tant qu'il équilibre l'assurance, mais son acte est-elle en ce point dans le paysage social.

Tarifcation avec segmentation (parfaite)

Assurance repose fondamentalement sur l'idée que la mutualisation des risques entre des assurés est possible. Cette mutualisation, qui peut être vue comme une sélection actuarielle de la loi des grands nombres, n'a de sens qu'en tant que celle d'une population de risque « homogène » (Charpentier, 2011). Cette sélection (actuarielle) impose aux assurés de segmenter, ce qui confère plusieurs avantages homogènes (1). Avec l'évolution des modes de données, et donc de variables disponibles, certaines assurances dépassent l'idée d'un tarif individuel, semblant remettre en cause l'idée même de mutualisation des risques. Entre ceux ceux qui peuvent segmenter et la force de regrouper qui impose des risques sociaux mais aussi actuariels, on se interroge sur le statut (ou) à imposer aux individus mutualisant entre les assurés, qui équilibrent ou en résultent dans un contexte de forte concurrence entre les assurés d'assurance ?

Tableau 8 - Répartition des classes d'assurés en fonction des sinistres supportés différents coûts, une segmentation à forfait, selon la loi d'habitation, selon l'âge et la loi d'habitation. Le marché nous suppose donc à la prime le même prix.

Année	Age	Habitation	Age + Habitation	Marché
1/1/2000	82,3	39,9	91	120
1/1/2000	82,3	39,9	82,3	80
1/1/2000	82,3	82,3	91	80
1/1/2000	82,3	82,3	82,3	80
1/1/2000	82,3	40	91	80
1/1/2000	82,3	40	82,3	80
Total	112,7	198,8	91	28
Assurés	80	215	1062,7	20
SP	142,7	114,7	112,3	108,6
IC (100)	6,844%	4,137%	1,112%	4,53%
IC (100)	10,5%	140,7%	108,6%	108,6%
Année de marché	114,6%	518,8%	30,9%	5,7%

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 Brest (France) & Quantax (Montreal, Canada), with the support of the ACTREC 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)
- a claims dataset (C_0), with all claims occurred during year 0 to all policyholders

An underwriting dataset (U_i) with (the same) 100,000 policies (drivers willing to purchase insurance for Year i) is provided. Players must provide prices for these 100,000 insured. A player i a dataset was month i . It can be sent by individuals, or groups, practitioners, students or academics.

For February 25th, players must send a dataset with 100,000 prices to pricinggame@institutesactuaires.com. It is a csv file, with two columns : the id, policy, and the premium. Let $P_{i,j}$ denote the annual premium for policy j , offered by player i (for year i). Let $\ell_{i,j}$ denote the losses for policy j (unknown by players when submitting their premiums).

lightgray Num	Family Type	Name	Label	Format
1	ID	id_client	ID-Client ID	string
2	ID	id_vehicule	ID-Vehicule 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_vehicule	ID-Vehicule 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_at_duration	Policy-Current Endorsment Duration	int
9	Policy	pol_pay_freq	Policy-Payment Frequency	string
10	Policy	pol_payd	Policy-Paid Indentance	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_lc1	Drivers-First Driver License Age	int
19	Drivers	drv_age_lc2	Drivers-Secondary Driver License 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_sak_begin	Vehicle-Sales Date Beginning	int
27	Vehicle	vh_sak_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 m'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 désigné pour chaque modèle proposé, et d'envoyer le fichier csv avec un rapide descriptif à PricingGame@institutesactuaires.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 mois.

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@institutesactuaires.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 de vos assurés se verra affecter une compagnie selon une méthode 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'occurrence, la somme des primes proposées par un assureur sera égale à 2,5 millions d'euros (par un facteur d'inflation appliqué uniformément).

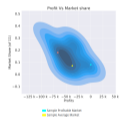
Les résultats seront présentés lors de la journée 100% Data Science organisée à Paris (voir également sur <http://freakonometrics.hypotheses.org>).

Pour plus d'information, PricingGame@institutesactuaires.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_lik1	drv_dmv2	drv_sex2	drv_age2	drv_age_lik2
often: 54.5 - 100.0%	M (93%)	61.57	35.70	No (73%)	F (66%)	50.00	27.34
sometimes: 18.2 - 54.5%	M (92%)	59.19	35.87	No (72%)	F (63%)	51.85	25.14
rarely: 0.1 - 18.2%	M (93%)	60.32	38.35	No (71%)	F (62%)	51.89	26.12
never:	M (88%)	58.17	36.08	No (64%)	F (60%)	51.16	26.31
in sample profitable market	M (94%)	61.05	36.93	No (74%)	F (64%)	49.25	27.25
in sample average market	M (92%)	61.73	36.39	No (73%)	F (63%)	50.41	26.23
in sample losing market	M (93%)	59.89	37.27	No (75%)	F (64%)	50.70	26.69

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

a global data science competition with real motor insurance data

- Real motor insurance data
- Build a prediction model for claims
- Play in a simulated marketplace

#Track A

Motor Insurance market simulation

Play as an insurance company, using real historical data in a competitive market with other players. See if you can make a profit with realistic market conditions.

[Not yet launched! →](#)

Aicrowd

- \$6000 0.0% PartnerRe
- \$3500 0.0%
- \$1500 0.0%
- \$1000 0.0%

[Learn more about PartnerRe](#)

#Track B

Worker Compensation Claim Prediction

Predict the claim amount of workers compensation claims using a synthetic dataset generated for this competition!

[Not yet launched! →](#)

kaggle

sponsored by

-
-
-

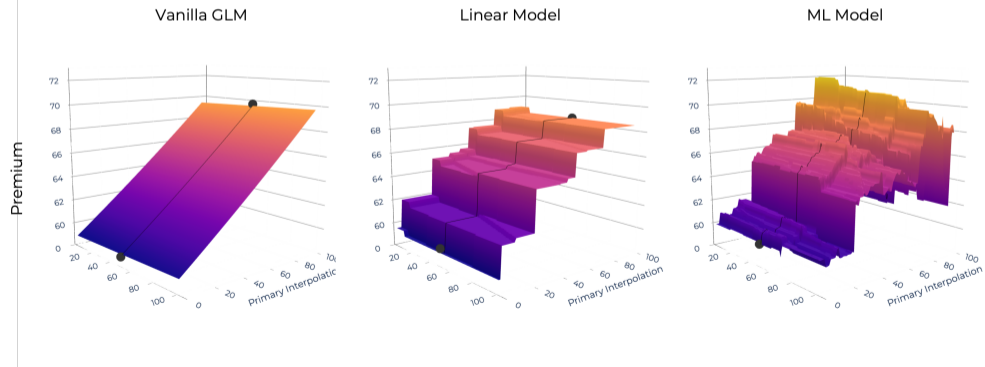
[Learn more about Actuarium Institute Australia](#)

compétition entre modèles, quel équilibre ?

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

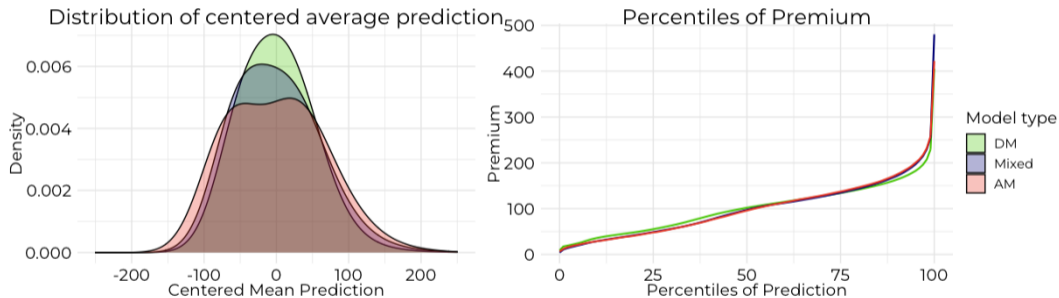
$$\mathbb{E}[Y|\hat{Y} = y] = y, \forall y$$

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



compétition entre modèles, quel équilibre ?

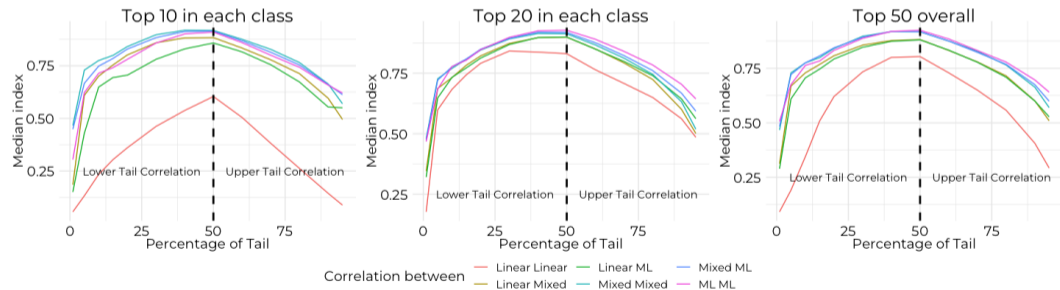
- apprentissage machine = plus de **variabilité** (avant la compétition)



(via [Ratz et al. \(2023\)](#))

compétition entre modèles, quel équilibre ?

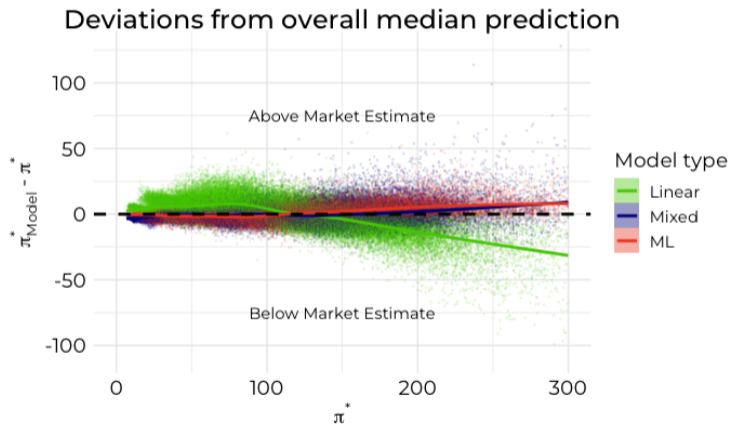
- peu de consensus entre modèles d'apprentissage machine
- ici corrélation (moyenne) de queues entre les modèles
 - à gauche, $u \rightarrow \mathbb{P}[X_i \leq F_i^{-1}(u) | X_j \leq F_j^{-1}(u)]$
 - à droite, $u \rightarrow \mathbb{P}[X_i > F_i^{-1}(u) | X_j > F_j^{-1}(u)]$



(via Ratz et al. (2023))

compétition entre modèles, quel équilibre ?

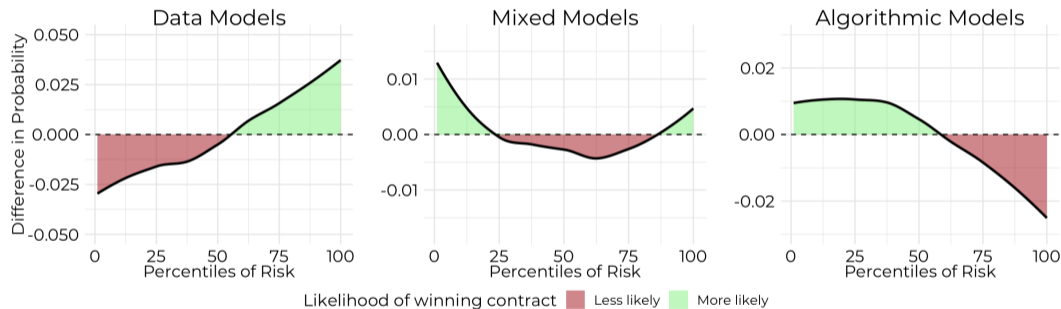
- les modèles linéaires sont chers sur les bas risques, pas chers sur les hauts risques



(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. \(2023\)](#))