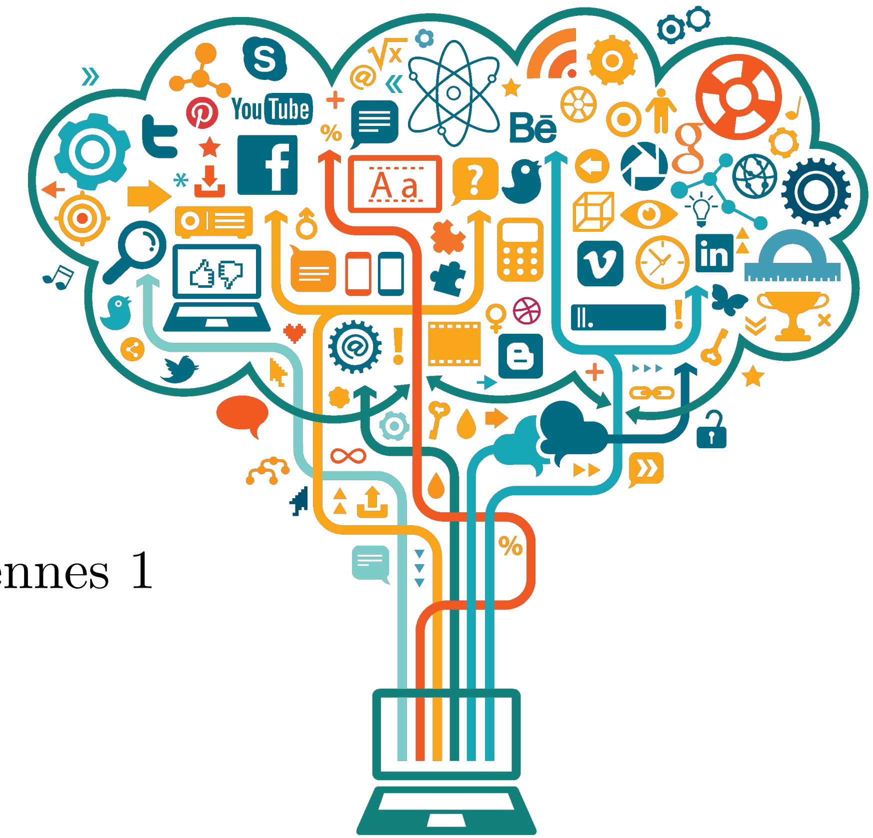
Tarification en assurance dans un contexte concurrentiel

A. Charpentier (Université Rennes 1 & CREST)

avec R. Élie (Université Paris-Est & CREST)



Chaire ACTINFO

Covéa / CREST / Univ. Paris Est / Univ. Rennes 1

Institut Louis Bachelier, Paris, Juin 2016.

http://freakonometrics.hypotheses.org

http://actinfo.hypotheses.org

manière des caractéristiques $\mathbf{\Omega}$ de l'assuré, et lui réclame donc une prime pure de montant $\mathbb{E}[S]$, la même que celle qu'il réclame à tous les assurés du portefeuille. Dans ce cas, la situation est telle que présentée au Tableau 3.7.

	Assurés	Assureur
Dépense	$\mathbb{E}[S]$	$S-\mathbb{E}[S]$
Dépense moyenne	$\mathbb{E}[S]$	0
Variance	0	$\mathbb{V}[S]$

TAB. 3.7 – *Situation des assurés et de l'assureur en l'absence de segmentation.*

L'assureur prend donc l'entièreté de la variance des sinistres $\mathbb{V}[S]$ à sa charge, que celle-ci soit due à l'hétérogénéité du portefeuille, ou à la variabilité intrinsèque des montants des sinistres.

Transfert de risque en information complète

A l'autre extrême, supposons que l'assureur incorpore toute l'information $\mathbf{\Omega}$ dans la tarification. On serait alors dans la situation décrite au Tableau 3.8.

	Assurés	Assureur
Dépense	$\mathbb{E}[S\vert\mathbf{\Omega}]$	$S-\mathbb{E}[S\vert\mathbf{\Omega}]$
Dépense moyenne	$\mathbb{E}[S]$	0
Variance	$\mathbb{V}\Big[\mathbb{E}[S\vert\mathbf{\Omega}]\Big]$	$\mathbb{V}\Big[S-\mathbb{E}[S\vert\mathbf{\Omega}]\Big]$

TAB. 3.8 – *Situation des assurés et de l'assureur dans le cas où la segmentation est opérée sur base de $\mathbf{\Omega}$.*

Contrairement au cas précédent, la prime payée par un assuré prélevé au hasard dans le portefeuille est à présent une variable aléatoire: $\mathbb{E}[S|\mathbf{\Omega}]$ dépend des caractéristiques $\mathbf{\Omega}$ de cet assuré. Comme la variable aléatoire $S-\mathbb{E}[S|\mathbf{\Omega}]$ est centrée, le risque assumé par l'assureur la variance du résultat financier de l'opération d'assurance, i.e.

$$\mathbb{V}\Big[S-\mathbb{E}[S|\mathbf{\Omega}]\Big] = \mathbb{E}\Big[\big(S-\mathbb{E}[S|\mathbf{\Omega}]\big)^2\Big]$$

Aucune segmentation

	Assuré	Assureur
Perte	$\mathbb{E}[S]$	$S-\mathbb{E}[S]$
Perte moyenne	$\mathbb{E}[S]$	0
Variance	0	$\text{Var}[S]$

Information parfaite: $\mathbf{\Omega}$ observable

	Assuré	Assureur
Perte	$\mathbb{E}[S\vert\mathbf{\Omega}]$	$S-\mathbb{E}[S\vert\mathbf{\Omega}]$
Perte moyenne	$\mathbb{E}[S]$	0
Variance	$\text{Var}\Big[\mathbb{E}[S\vert\mathbf{\Omega}]\Big]$	$\text{Var}\Big[S-\mathbb{E}[S\vert\mathbf{\Omega}]\Big]$

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

3.8. La prime pure en univers segmenté 177

On assiste dans ce cas à un partage de la variance totale de S (c'est-à-dire du risque) entre les assurés et l'assureur, matérialisé par la formule

$$\mathbb{V}[S] = \underbrace{\mathbb{E}\Big[\mathbb{V}[S|\boldsymbol{\Omega}]\Big]}_{\to\text{assureur}} + \underbrace{\mathbb{V}\Big[\mathbb{E}[S|\boldsymbol{\Omega}]\Big]}_{\to\text{assurés}}.$$

Ainsi, lorsque toutes les variables pertinentes $\boldsymbol{\Omega}$ ont été prises en compte, l'intervention de l'assureur se limite à la part des sinistres due exclusivement au hasard; en effet, $\mathbb{V}[S|\boldsymbol{\Omega}]$ représente les fluctuations de S dues au seul hasard. Dans cette situation idéale, l'assureur mutualise le risque et il n'y a donc aucune solidarité induite entre les assurés du portefeuille: chacun paie en fonction de son propre risque.

Transfert des risques en information partielle

Bien entendu, la situation décrite au paragraphe précédent est purement théorique puisque parmi les variables explicatives $\boldsymbol{\Omega}$ nombreuses sont celles qui ne peuvent pas être observées par l'assureur. En assurance automobile par exemple, l'assureur ne peut pas observer la vitesse à laquelle roule l'assuré, son agressivité au volant, ni le nombre de kilomètres qu'il parcourt chaque année[2]. Dès lors, l'assureur ne peut utiliser qu'un sous-ensemble $\boldsymbol{X}$ des variables explicatives contenues dans $\boldsymbol{\Omega}$, i.e. $\boldsymbol{X} \subset \boldsymbol{\Omega}$. La situation est alors semblable à celle décrite au Tableau 3.9.

	Assuré	Assureur
Dépense	$\mathbb{E}[S\|\boldsymbol{X}]$	$S-\mathbb{E}[S\|\boldsymbol{X}]$
Dépense moyenne	$\mathbb{E}[S]$	0
Variance	$\mathbb{V}\Big[\mathbb{E}[S\|\boldsymbol{X}]\Big]$	$\mathbb{E}\Big[\mathbb{V}[S\|\boldsymbol{X}]\Big]$

Tab. 3.9 – *Situation de l'assuré et de l'assureur dans le cas où la segmentation est opérée sur base de $\boldsymbol{X} \subset \boldsymbol{\Omega}$.*

Il est intéressant de constater que

$$\begin{aligned}\mathbb{E}\Big[\mathbb{V}[S|\boldsymbol{X}]\Big] &= \mathbb{E}\Big[\mathbb{E}\Big[\mathbb{V}[S|\boldsymbol{\Omega}]\Big|\boldsymbol{X}\Big]\Big] + \mathbb{E}\Big[\mathbb{V}\Big[\mathbb{E}[S|\boldsymbol{\Omega}]\Big|\boldsymbol{X}\Big]\Big] \\ &= \underbrace{\mathbb{E}\Big[\mathbb{V}[S|\boldsymbol{\Omega}]\Big]}_{\text{mutualisation}} + \underbrace{\mathbb{E}\Big\{\mathbb{V}\Big[\mathbb{E}[S|\boldsymbol{\Omega}]\Big|\boldsymbol{X}\Big]\Big\}}_{\text{solidarité}}. \qquad (3.22)\end{aligned}$$

Information imparfaite: $\boldsymbol{X} \subset \boldsymbol{\Omega}$ est observable

	Assuré	Assureur
Perte	$\mathbb{E}[S\|\boldsymbol{X}]$	$S-\mathbb{E}[S\|\boldsymbol{X}]$
Perte moyenne	$\mathbb{E}[S]$	0
Variance	$\text{Var}\Big[\mathbb{E}[S\|\boldsymbol{X}]\Big]$	$\mathbb{E}\Big[\text{Var}[S\|\boldsymbol{X}]\Big]$

$$\begin{aligned}\text{Var}[S] &= \mathbb{E}\Big[\text{Var}[S|\boldsymbol{X}]\Big] + \text{Var}\Big[\mathbb{E}[S|\boldsymbol{X}]\Big] \\ &= \underbrace{\underbrace{\mathbb{E}\Big[\text{Var}[S|\boldsymbol{\Omega}]\Big]}_{\text{mutualisation}} + \underbrace{\mathbb{E}\Big[\text{Var}\Big[\mathbb{E}[S|\boldsymbol{\Omega}]\Big|\boldsymbol{X}\Big]\Big]}_{\text{solidarité}}}_{\to\text{assureur}} \\ &\quad + \underbrace{\text{Var}\Big[\mathbb{E}[S|\boldsymbol{X}]\Big]}_{\to\text{ assuré}}.\end{aligned}$$

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

Arthur Charpentier
Professeur à l'Université du Québec, Montréal

Michel Denuit
Professeur à l'Université catholique de Louvain

Romuald Elie
Professeur à l'Université de Marne-la-Vallée

L'assurance repose fondamentalement sur l'idée que la mutualisation des risques entre des assurés est possible. Cette mutualisation, qui peut être vue comme une relecture actuarielle de la loi des grands nombres, n'a de sens qu'au sein d'une population de risques « homogènes » [Charpentier, 2011]. Cette condition (actuarielle) impose aux assureurs de segmenter, ce que confirment plusieurs travaux économiques (1). Avec l'explosion du nombre de données, et donc de variables tarifaires possibles, certains assureurs évoquent l'idée d'un tarif individuel, semblant remettre en cause l'idée même de mutualisation des risques. Entre cette force qui pousse à segmenter et la force de rappel qui tend (pour des raisons sociales mais aussi actuarielles, ou au moins de robustesse statistique (2)) à imposer une solidarité minimale entre les assurés, quel équilibre va en résulter dans un contexte de forte concurrence entre les sociétés d'assurance ?

Tarification sans segmentation

Sans segmentation, le « prix juste » d'un risque est l'espérance mathématique de la charge annuelle. C'est l'idée du théorème fondamental de la valorisation actuarielle : en moyenne, la somme des primes doit permettre d'indemniser l'intégralité des sinistres survenus dans l'année. Afin d'illustrer les différents aspects de la construction du tarif et ses conséquences, on va utiliser les données présentées dans le tableau 1 (voir p. xx), qui indique la fréquence annuelle de sinistres.

Les facteurs de risque sont ici le lieu d'habitation et l'âge de l'assuré, et on observe la fréquence de sinistre par classe. Le coût unitaire, supposé fixe, équivaut à 1 000 euros. La prime pure est alors E [S] = 1 000 x E [N]. Dans cet exemple, la prime pure sans segmentation sera de 82,30 euros.

Modèle simple, $\boldsymbol{\Omega}=\{\boldsymbol{X}_1,\boldsymbol{X}_2\}$.

Quatre modèles

$$\begin{cases}\widehat{m}_0(\boldsymbol{x}_1,\boldsymbol{x}_2)=\mathbb{E}[S]\\ \widehat{m}_1(\boldsymbol{x}_1,\boldsymbol{x}_2)=\mathbb{E}[S|\boldsymbol{X}_1=\boldsymbol{x}_1]\\ \widehat{m}_2(\boldsymbol{x}_1,\boldsymbol{x}_2)=\mathbb{E}[S|\boldsymbol{X}_2=\boldsymbol{x}_2]\\ \widehat{m}_{12}(\boldsymbol{x}_1,\boldsymbol{x}_2)=\mathbb{E}[S|\boldsymbol{X}_1=\boldsymbol{x}_1,\boldsymbol{X}_2=\boldsymbol{x}_2]\end{cases}$$

PolNum	CalYear	Gender	Type	Category	Occupation	Age	Group1	Bonus	Poldur	Value	Adind	SubGroup2	Group2	Density
200285786	2010	Male	E	Large	Employed	48	14	40	0	32345	1	O31	O	35.43401501
200285787	2010	Male	B	Medium	Employed	30	8	-30	9	8995	0	Q29	Q	239.4551701
200285788	2010	Female	B	Large	Housewife	47	2	-50	2	9145	1	U21	U	88.29014956
200285789	2010	Female	D	Large	Self-employed	48	13	-30	15	22075	1	R21	R	275.2822626
200285790	2010	Male	C	Medium	Housewife	57	12	-50	1	24985	1	Q5	Q	99.6400095
200285791	2010	Male	D	Medium	Self-employed	21	15	50	1	12100	1	R11	R	259.0040603
200285792	2010	Male	B	Small	Employed	44	5	-40	15	9820	1	Q10	Q	169.7885554
200285793	2010	Male	F	Small	Self-employed	37	17	-50	5	28680	1	Q5	Q	99.6400095
200285794	2010	Female	C	Large	Retired	49	3	20	4	28470	0	L94	L	84.22903844
200285795	2010	Female	A	Medium	Unemployed	35	5	20	5	8590	0	L112	L	66.06668352
200285796	2010	Male	E	Large	Self-employed	50	10	-30	3	20490	1	Q10	Q	169.7885554
200285797	2010	Female	B	Medium	Housewife	31	8	140	1	8385	1	P28	P	41.2451199
200285798	2010	Female	E	Medium	Self-employed	41	11	90	3	6410	1	L47	L	66.76541883
200285799	2010	Female	A	Medium	Housewife	44	10	-30	8	8485	0	P29	P	20.86448407
200285800	2010	Male	B	Large	Retired	69	8	-40	11	9380	1	U14	U	123.0152076
200285801	2010	Male	F	Medium	Housewife	45	11	30	0	19700	0	L40	L	76.05272599
200285802	2010	Male	E	Large	Retired	53	8	-30	6	10980	1	U19	U	61.79475865
200285803	2010	Male	C	Small	Employed	47	10	-10	9	21980	0	L96	L	45.66982293
200285804	2010	Female	D	Large	Retired	46	7	-50	1	28925	1	U12	U	54.93181221
200285805	2010	Female	C	Large	Retired	67	17	-50	9	14525	1	L52	L	73.25249905

Numtppd	Numtpbi	Indtppd	Indtpbi
0	1	0	1056.0334927
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
3	1	5800.0189068	16.507641942
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0

$X_{1,i}$ $X_{k,i}$ Y_i

PolNum	CalYear	Gender	Type	Category	Occupation	Age	Group1	Bonus	Poldur	Value	Adind	SubGroup2	Group2	Density
200375666	2011	Female	A	Large	Employed	46	11	50	0	42975	0	L18	L	58.91132801
200375667	2011	Male	B	Large	Unemployed	31	8	80	11	14835	0	U8	U	125.1320458
200375668	2011	Female	D	Medium	Employed	27	7	-40	13	19000	1	R30	R	296.4319078
200375670	2011	Male	B	Small	Self-employed	22	7	-10	14	33305	0	Q33	Q	129.6690079
200375672	2011	Male	B	Small	Employed	21	17	-20	14	25995	0	T25	T	28.51184808
200375674	2011	Male	C	Medium	Employed	45	19	-50	0	8320	1	N21	N	71.18027901
200375675	2011	Male	C	Medium	Housewife	51	19	30	3	8445	0	L110	L	83.90453994
200375676	2011	Male	E	Large	Self-employed	49	16	-50	3	19545	0	L58	L	64.53563007
200375677	2011	Male	C	Small	Housewife	31	11	-20	5	5030	1	Q7	Q	83.76263662
200375678	2011	Female	A	Medium	Housewife	31	9	-50	14	15480	1	P7	P	25.62227499
200375679	2011	Male	B	Large	Housewife	69	13	-50	7	29580	0	Q23	Q	205.4307964
200375682	2011	Male	A	Medium	Self-employed	43	13	140	3	3735	0	U16	U	91.54176264
200375683	2011	Female	A	Medium	Self-employed	64	18	-20	6	13670	1	O35	O	21.45273029
200375685	2011	Male	E	Large	Employed	25	8	-10	6	17315	0	O22	O	32.18545326
200375688	2011	Male	B	Small	Retired	55	7	-40	3	19410	1	R49	R	208.8164363
200375689	2011	Female	F	Medium	Self-employed	54	9	-40	14	4165	0	U12	U	54.93181221
200375690	2011	Male	D	Large	Housewife	42	9	80	0	11970	1	L125	L	44.16537902
200375692	2011	Male	E	Large	Employed	36	12	-20	7	28415	0	L48	L	71.62174491
200375693	2011	Male	F	Medium	Self-employed	26	10	-30	6	4300	0	L97	L	63.82886936
200375694	2011	Female	B	Small	Unemployed	24	6	-40	7	24005	0	M17	M	201.6569069

Marché en Concurence

Règle de choix: les assurés choisissent la prime la moins chère,

	A	B	C	D	E	F
	787.93	706.97	1032.62	907.64	822.58	**603.83**
	**170.04**	197.81	285.99	212.71	177.87	265.13
	473.15	447.58	**343.64**	410.76	414.23	425.23
	337.98	**336.20**	468.45	339.33	383.55	672.91

Marché en Concurence

Règle de choix: les assurés choisissent (au hasard) parmi les trois primes les moins chères

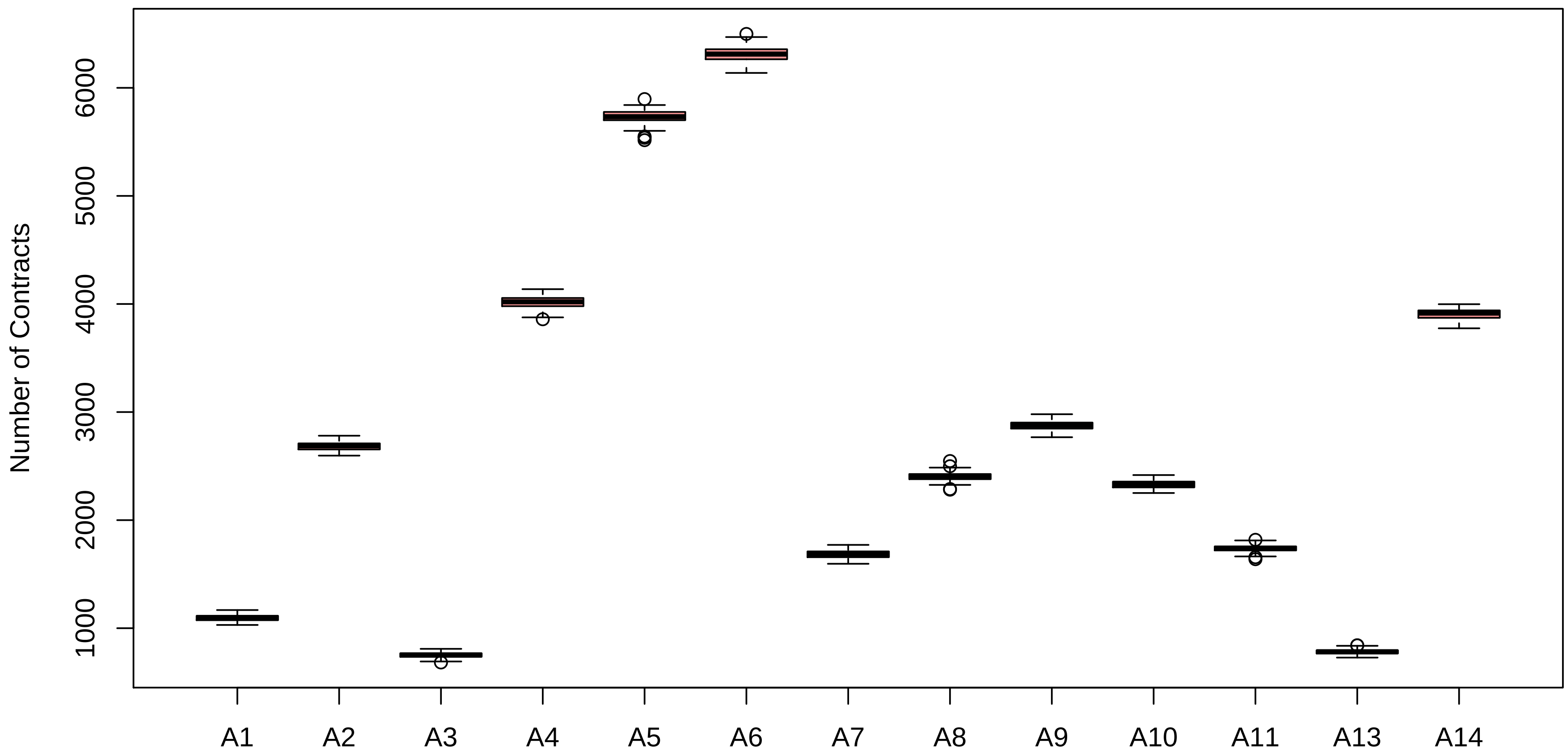
	A	B	C	D	E	F
	787.93	706.97	1032.62	907.64	822.58	603.83
	170.04	197.81	285.99	212.71	177.87	265.13
	473.15	447.58	343.64	410.76	414.23	425.23
	337.98	336.20	468.45	339.33	383.55	672.91

Marché en Concurence

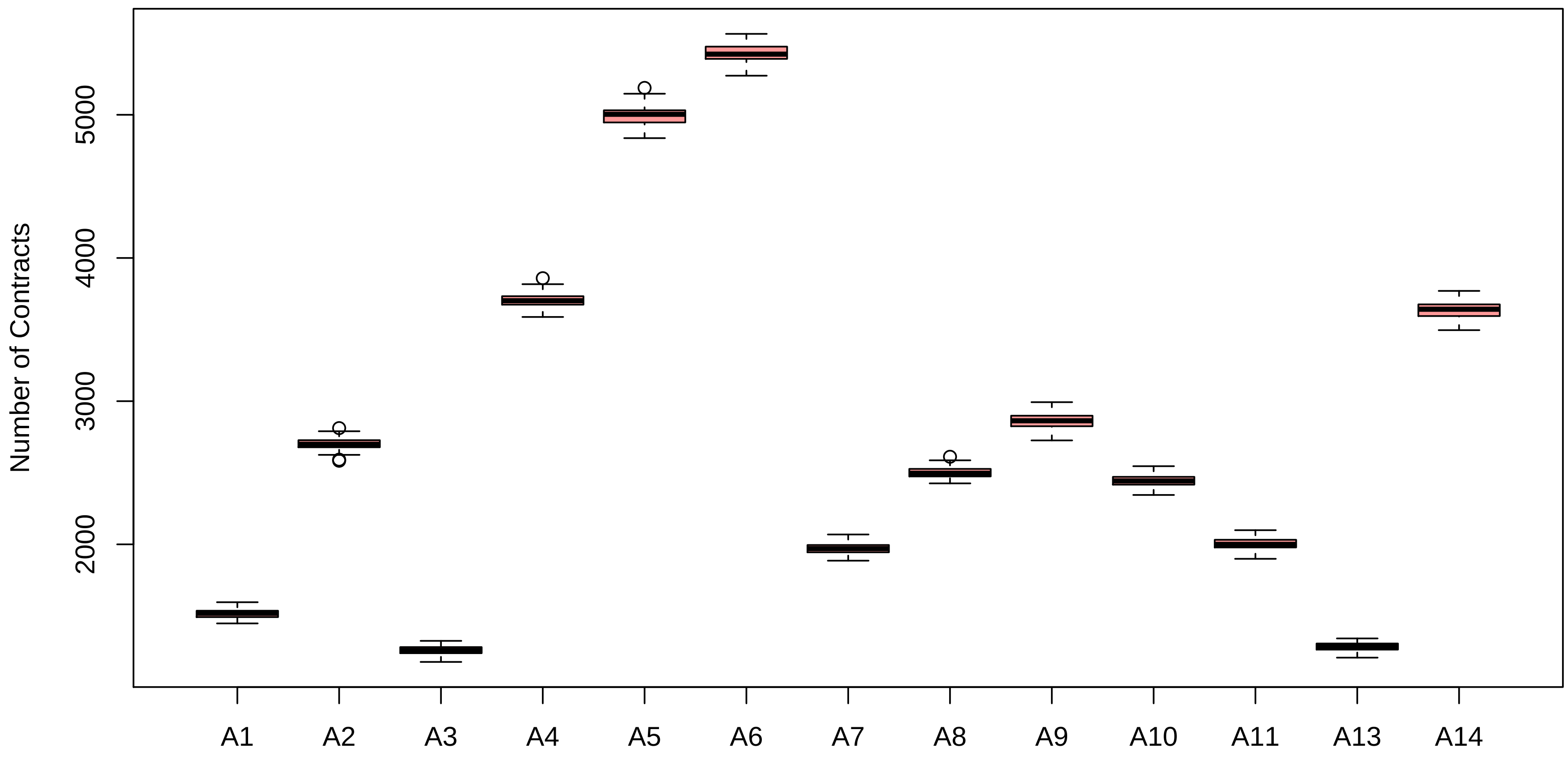
Règle de choix: les assurés se voient $\boxed{\text{attribuer}}$ (au hasard) un assureur pour l'année $n-1$. L'année n, si leur assureur est parmi les 3 moins chers, il est retenu, sinon choix au hasard parmi 4.

	A	B	C	D	E	F
	787.93	706.97	1032.62	$\boxed{907.64}$	822.58	603.83
	170.04	197.81	285.99	212.71	$\boxed{\mathbf{177.87}}$	265.13
	473.15	447.58	343.64	410.76	414.23	$\boxed{425.23}$
	$\boxed{\mathbf{337.98}}$	336.20	468.45	339.33	383.55	672.91

Parts de marché, règle 2

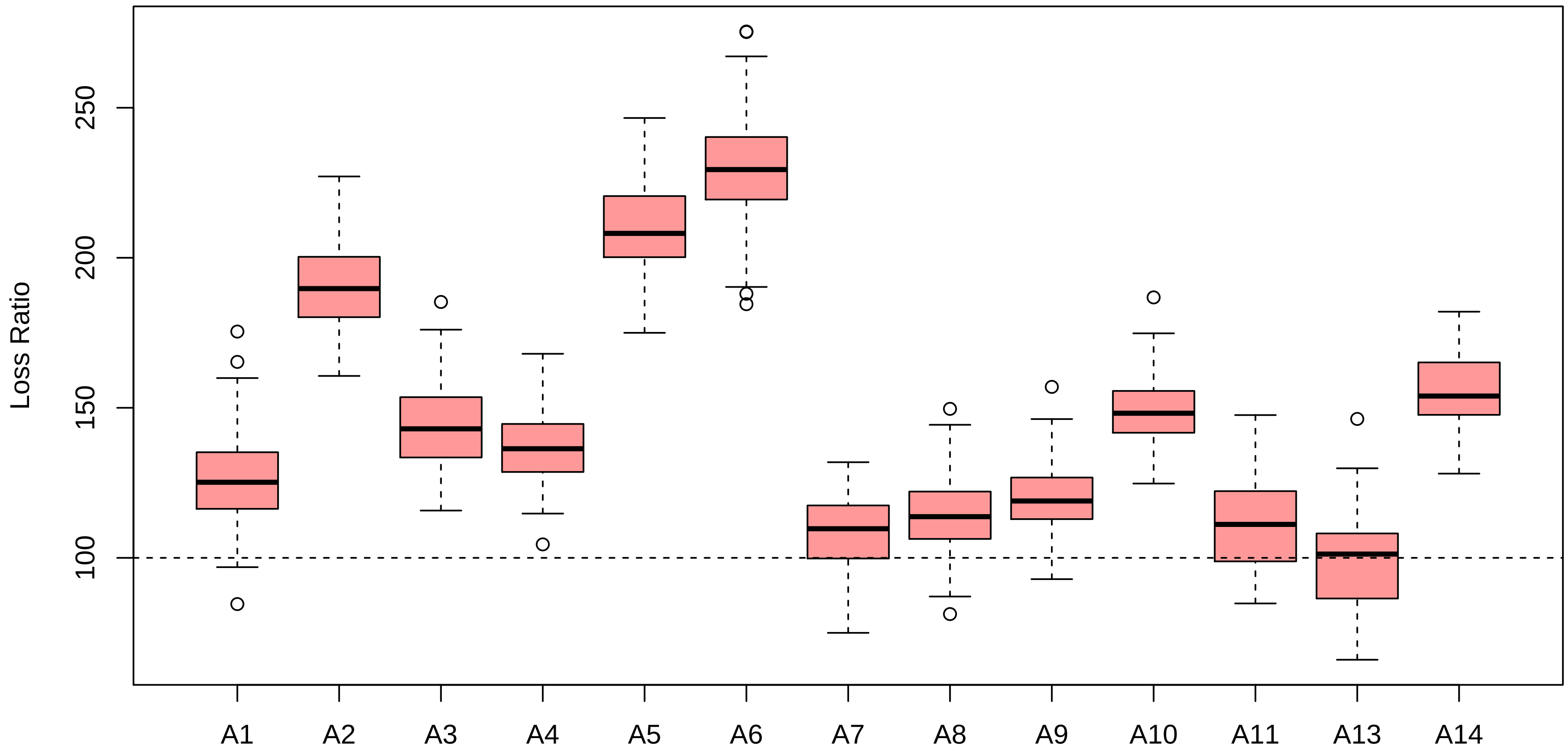


Parts de marché, règle 3



Ratio S/P (Sinistres / Primes), règle 2

Ratio S/P marché $\sim 154\%$.

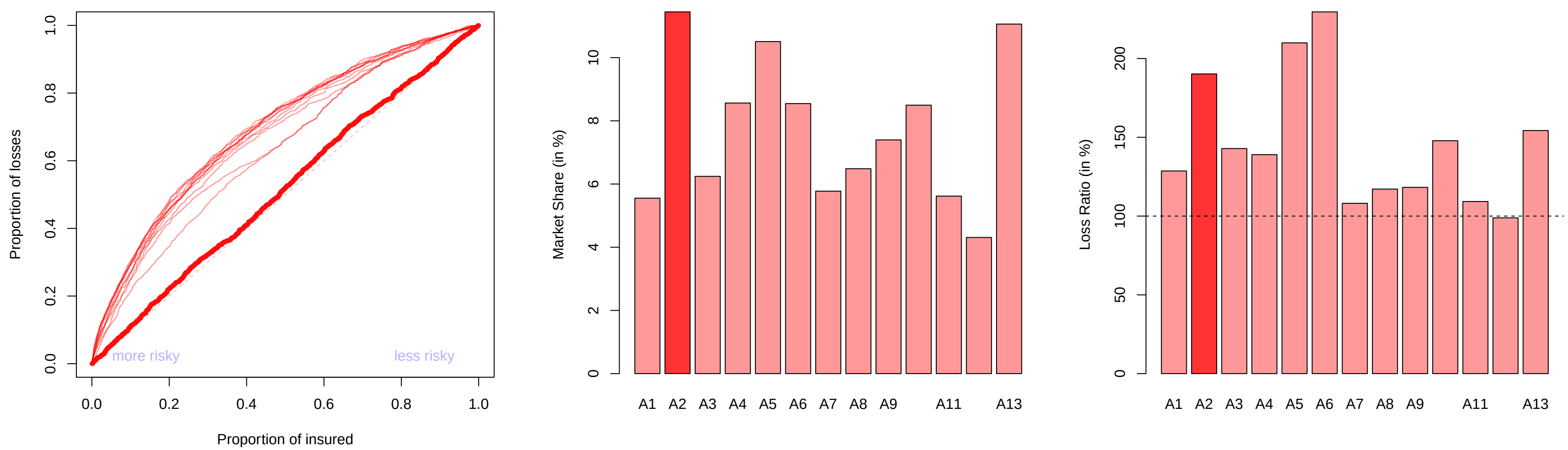


Assureur A2

Pas de segmentation, prime unique

Remarque tous les prix ont été normalisés,

$$\pi_2 = m_2(\boldsymbol{x}_i) = \frac{1}{n}\sum_{i=1}^{n} m_j(\boldsymbol{x}_i)\ \forall j$$



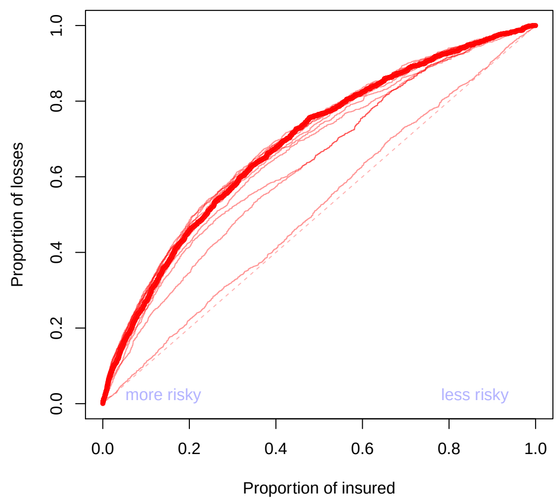
Assureur A1

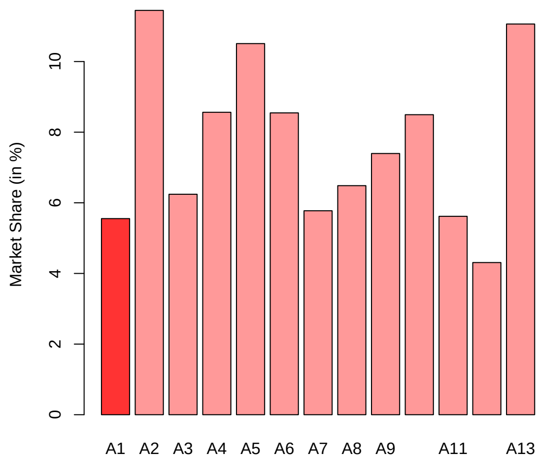
Modèles GLM, fréquence RC matériel/corporel et coûts matériels

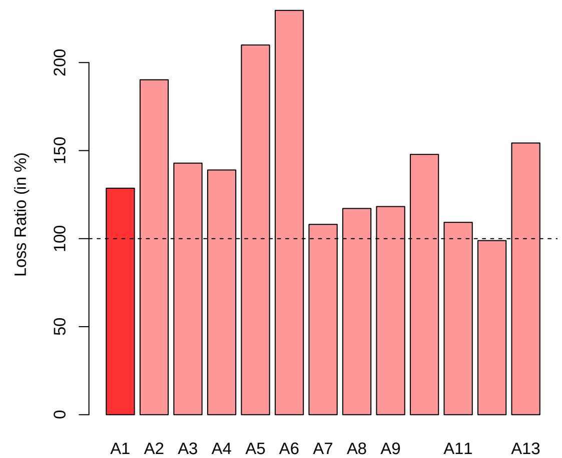
Age coupé en classe [18-30], [30-45], [45-60] et [60+], effets croisés avec occupation

Lissage manuel, SAS et Excel

Actuaire dans une mutuelle (en France)







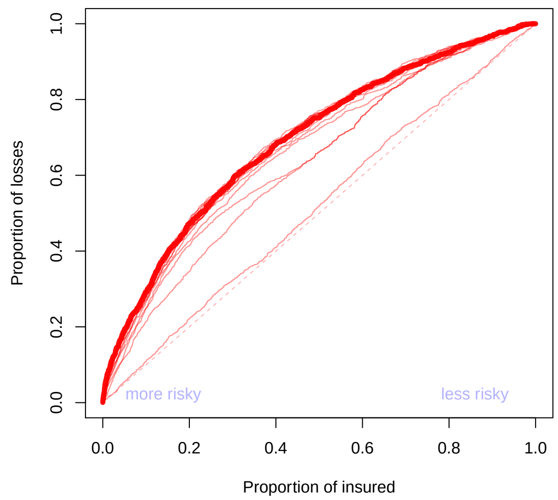
Assureur A8/A9

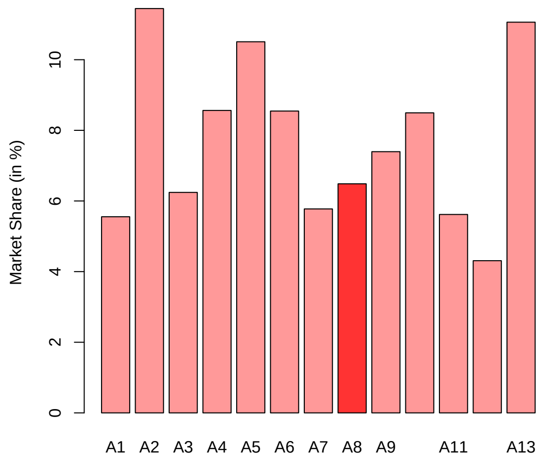
Modèles GLM, fréquence et coûts, suppression des graves (>15k)

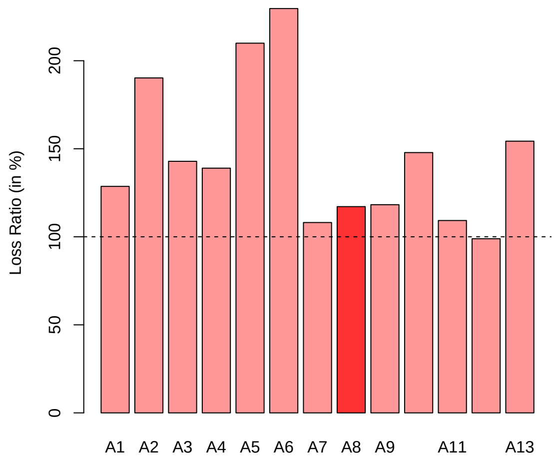
Interaction âge-genre

Utilisation d'un logiciel commercial de pricing (développé par Actuaris)

Actuaire dans une mutuelle (en France)





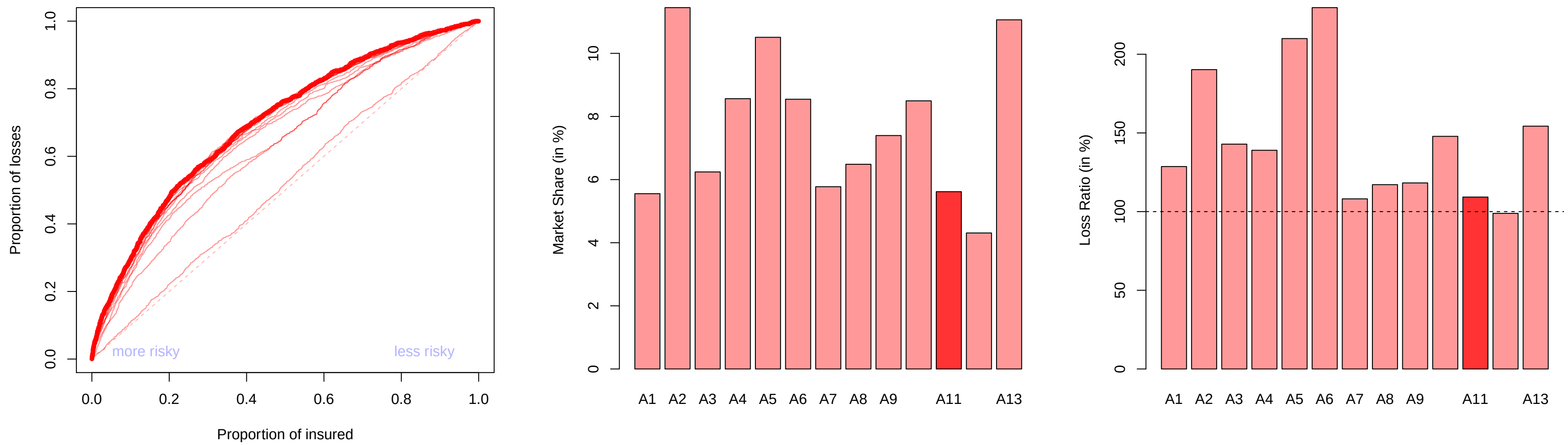


Assureur A11

Toutes les variables sauf une, utilisation de deux modèles XGBoost (gradient boosting)

Correction pour primes négatives (plafonnées)

Programmé en Python par un actuaire dans une compagnie d'assurance (inscrit à la formation ADS).

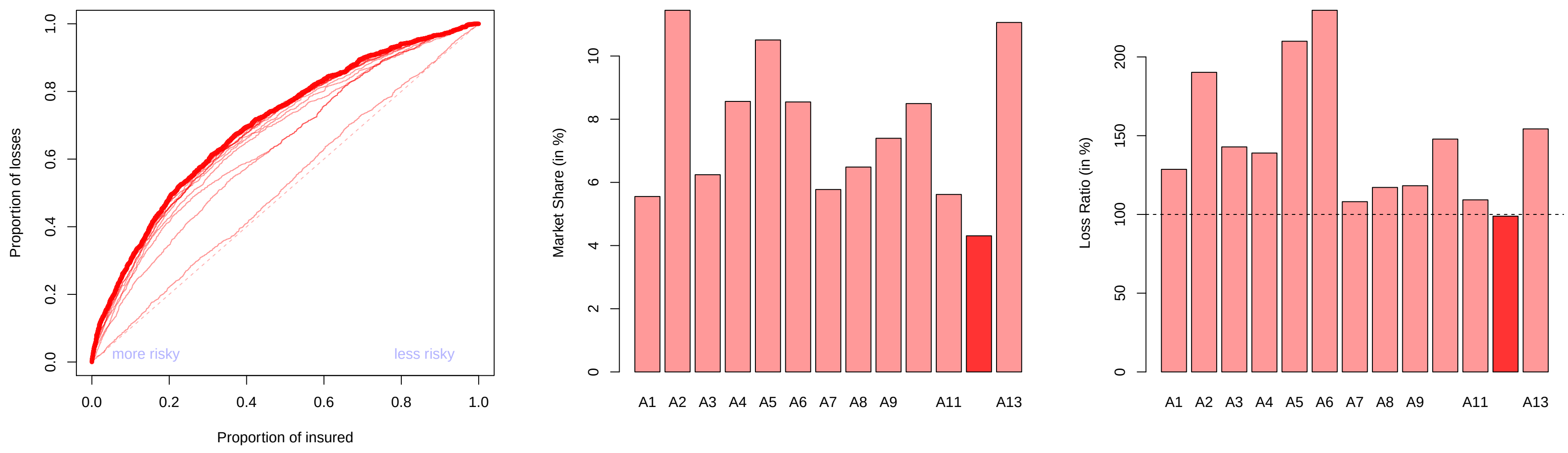


Assureur A12

Toutes les variables, utilisation de deux modèles XGBoost (gradient boosting)

Correction pour primes négatives (plafonnées)

Programmé en R par un actuaire dans une compagnie d'assurance en Europe.



Conclusion

- début de réflexions sur la mise en concurrence de modèles statistiques/économétriques/actuariels
- comportement des modèles en concurrence difficile à prévoir
- pour l'instant rien sur la convergence des prix, et la prise en compte d'information sur les prix des concurrents