Actuarial Pricing Game

A. Charpentier
(UQAM & Université de Rennes 1)

with R. Élie & J. Jakubowicz

Paris, 100% Actuaires, November 2015.

http://freakonometrics.hypotheses.org

Pricing Game, *Charpentier, Denuit & Élie* (2015)

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 $E[S] = 1\,000 \times E[N]$. Dans cet exemple, la prime pure sans segmentation sera de 82,30 euros.

Tableau 1 - Fréquence annuelle de sinistre pour deux classes de risque (lieu d'habitation et âge de l'assuré) avec le nombre d'assurés (entre parenthèses)

	Jeune (J)	Expérimenté (E)	Senior (S)	Total
Ville (V)	**12 % (500)**	**9 % (2 000)**	**9 % (500)**	9,5 % (3 000)
Campagne (C)	**8 % (500)**	**6,67 % (1 000)**	**4 % (500)**	6,33 % (2 000)
Total	10 % (1 000)	8,22 % (3 000)	6,5 % (1 000)	8,23 % (5 000)

Source : auteurs, exemple fictif.

Dans le cas d'une tarification sans segmentation, on retrouve le partage des risques détaillé dans le tableau 2 : en moyenne, l'assureur est à l'équilibre financier, mais il porte à sa charge l'intégralité du risque (dont une partie est liée à l'hétérogénéité du portefeuille).

Tableau 2 - Répartition des risques entre l'assureur et ses assurés

	Assurés	Assureur
Dépense	$E(S)$	$S - E(S)$
Dépense moyenne	$E(S)$	0
Variance	0	$Var(S)$

Source : auteurs, exemple fictif.

Tarification avec segmentation (parfaite)

Supposons maintenant qu'un assureur segmente son tarif, en supposant qu'il dispose d'une connaissance parfaite des classes de risque (cette information sera notée Ω). Cet assureur sera à l'équilibre, en moyenne, puisque $E[E(S|\Omega)] = E[S]$. Mais, cette fois-ci, comme le montre le tableau 3 ci-contre, le risque est moins porté par l'assureur.

En effet, on retrouve ici la relation classique de décomposition de la variance
$Var(E(S|\Omega)) + E(Var(S|\Omega)) = Var(S)$

Tableau 3 - Décomposition de la variance de la dépense entre l'assureur et l'assuré en présence d'une segmentation parfaite

	Assurés	Assureur
Dépense	$E(S\|\Omega)$	$S - E(S\|\Omega)$
Dépense moyenne	$E(S)$	0
Variance	$Var(E(S\|\Omega))$	$E(Var(S\|\Omega))$

Source : auteurs, exemple fictif.

Plaçons-nous maintenant dans la situation où ces deux sociétés sont présentes sur le marché. La première ne segmente pas, alors que la seconde utilise l'information (parfaite) dont elle dispose. Dans ce contexte d'environnement concurrentiel, on retrouve une illustration des concepts de « *lemons* » d'Akerlof [1970] : les « bons » risques ont une prime plus basse chez l'assureur qui segmente son tarif et les « mauvais » risques ont une prime plus basse chez l'assureur qui ne segmente pas. Les « mauvais » risques vont alors s'assurer chez l'assureur qui ne discrimine pas, mais ce dernier, n'ayant plus les « bons » risques pour s'assurer un équilibre financier, va alors se retrouver en difficulté financière. Dans l'exemple numérique illustré dans le tableau 4, on retrouve que l'assureur qui segmente est alors – en moyenne – à l'équilibre. On entre ainsi dans une « spirale de la segmentation », les assureurs qui ne segmentent pas courant à la faillite.

Tableau 4 - Répartition des classes d'assurés lorsqu'un assureur propose de segmenter parfaitement et qu'un autre ne segmente pas (Le coût individuel est ici supposé connu et vaut 1 000 euros)

	Aucun	Âge x habitation	Marché
J-V (500)	**82,3**	120	82,3
J-C (500)	82,3	**80**	80
E-V (2 000)	**82,3**	90	82,3
E-C (1 000)	82,3	**66,7**	66,7
S-V (500)	**82,3**	90	82,3
S-C (500)	82,3	**40**	40
Primes	247	126,67	373,67
Sinistres	285	126,67	411,67
S/P	115,4 %	100,0 %	110,2 %
(IC 95 %) (3)	± 8,9 %	± 10,4 %	± 5,1 %
Part de marché	66,1 %	33,9 %	

Source : auteurs, exemple fictif.

Pricing Game, *Charpentier, Denuit & Élie* (2015)

On notera, dans cet exemple, que dire que la prime pure permet à l'assureur d'être à l'équilibre en moyenne n'a de sens que dans une situation de monopole. Dans un contexte concurrentiel, la situation est tout autre.

Tarification avec segmentation (imparfaite)

La réalité n'est toutefois pas aussi simple. En particulier, la connaissance des classes de risque est souvent imparfaite. L'assureur n'a pas à sa disposition Ω, mais uniquement un ensemble de variables explicatives, $X = \{X_1, ..., X_k\}$, dont une partie est supposée liée au facteur de risque Ω. Dans notre exemple illustratif, on peut aussi imaginer que la discrimination basée sur l'âge ne soit plus autorisée. Dans le cas où l'assureur dispose d'une information imparfaite, l'assureur utilise comme prime pure $\mathrm{E}(S|X)$. Cette prime lui garantit d'être en moyenne à l'équilibre, puisque $\mathrm{E}(\mathrm{E}(S|X)) = \mathrm{E}(S)$, mais le partage des risques est alors assez différent, comme l'illustre le tableau 5.

Tableau 5 - Répartition des classes d'assurés lorsqu'un assureur utilise une information imparfaite pour tarifer

	Assurés	Assureur
Dépense	$\mathrm{E}(S\|X)$	$S - \mathrm{E}(S\|X)$
Dépense moyenne	$\mathrm{E}(S)$	0
Variance	$\mathrm{Var}(\mathrm{E}(S\|X))$	$\mathrm{E}(\mathrm{Var}(S\|X))$

Source : auteurs, exemple fictif.

On notera que la variance de l'assureur est ici
$\mathrm{E}(\mathrm{Var}(S|X)) = \mathrm{E}(\mathrm{Var}(S|\Omega))+\mathrm{E}(\mathrm{Var}(\mathrm{E}(S|\Omega)|X))$

avec un terme qui correspond à la variance que nous avions en situation d'information parfaite, mais aussi un terme additionnel que l'on peut interpréter comme le risque dû à un manque d'information. On peut alors décomposer la variance totale des dépenses sous la forme
$\mathrm{Var}(\mathrm{E}(S|X)) + \mathrm{E}(\mathrm{Var}(S|\Omega)) + \mathrm{E}(\mathrm{Var}(\mathrm{E}(S|\Omega)|X))$

Ce surplus de variance pour l'assureur traduit le fait que la segmentation utilisée ne crée pas des classes réellement homogènes.

Dans notre exemple illustratif, on peut regarder les deux situations possibles. Dans le premier cas, une société d'assurance utilise le lieu d'habitation et pas l'âge, et se retrouve en concurrence face à une société qui ne segmente pas (c'est le cas 1 du tableau 6). Dans le second cas, une société utilise seulement l'âge comme variable tarifaire (cas 2 du tableau 6).

Tableau 6 - Répartition des classes d'assurés lorsqu'un assureur propose de segmenter en fonction du lieu d'habitation (cas 1) ou de l'âge de l'assuré (cas 2) et qu'un autre ne segmente pas

Cas 1

	Aucun	Habitation	Marché
V (3 000)	**82,3**	95	82,3
C (2 000)	82,3	**63,3**	63,3
Primes	247	126,67	373,67
Sinistres	285	126,67	411,67
S/P	115,4 %	100,0 %	110,2 %
(IC 95 %)	± 8,9 %	± 10,4 %	± 5,1 %
Part de marché	66,1 %	33,9 %	

Cas 2

	Aucun	Âge	Marché
J (1 000)	**82,3**	100	82,3
E (3 000)	82,3	**82,2**	82,2
S (1 000)	82,3	**65**	65
Primes	82,33	311,67	394
Sinistres	100	311,67	411,67
S/P	121,5 %	100,0 %	104,5 %
(IC 95 %)	± 15,9 %	± 6,4 %	± 4,8 %
Part de marché	20,9 %	79,1 %	

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

On note qu'ici, même avec une information imparfaite, on se retrouve face à un cas comparable au précédent. Les sociétés qui segmentent sont, en moyenne, à l'équilibre en attirant les « bons » risques, alors que la société qui n'a pas segmenté perd de l'argent (en moyenne, là encore) puisqu'elle a attiré les mauvais risques, qui ont été tarifés à leur « juste » niveau. Néanmoins, on notera que la société qui segmente peut aussi se retrouver avec une part de marché relativement faible (ce qui va induire un autre type de risque, puisqu'un petit portefeuille est plus volatil qu'un portefeuille plus gros, mais nous reviendrons sur ce point dans le paragraphe suivant).

Et dans un univers concurrentiel ?…

Avant de conclure, essayons de regarder ce qui se passerait si tous ces cas coexistaient sur le marché : un assureur qui ne souhaite pas segmenter ; deux assureurs qui segmentent mais imparfaitement en utilisant seulement quelques variables tarifaires à leur disposition ; un assureur qui, ayant accès aux réelles classes de risque, utiliserait une segmentation beaucoup plus fine que tous les autres.

Le tableau 7 présente la répartition du marché entre tous ces assurés (toujours dans notre exemple fictif). D'un côté, l'assureur qui ne segmente pas se trouve en réelle difficulté car il a récupéré les plus « mauvais » risques, largement sous-tarifés. De l'autre côté, l'assureur qui segmente le plus finement est certes à l'équilibre financier, mais sur une niche de population beaucoup plus petite.

Afin d'illustrer le risque porté par chacune des sociétés, une *value-at-risk* [4] à 99,5 % a été ajoutée dans le tableau 7. Dans cet exemple illustratif, la société qui segmente le plus est certes la seule à être – en moyenne – à l'équilibre, mais la variabilité de son ratio sinistres/primes est très importante. Cette société a quasiment une chance sur cinq d'avoir un ratio plus mauvais que celui du marché. La figure 1 (p. xo) montre ainsi des distributions de ratios sinistres/primes pour ces quatre sociétés qui se font concurrence, avec la dispersion de chacun.

Tableau 7 - Répartition des classes d'assurés en fonction des assureurs proposant différents tarifs, sans segmentation (à gauche), suivant l'âge, suivant le lieu d'habitation, suivant l'âge et le lieu d'habitation. Les assurés sont supposés choisir ici la prime la moins chère.

	Aucun	Âge	Habitation	Âge x habitation	Marché
J-V (500)	**82,3**	100	95	120	82,3
J-C (500)	82,3	100	**63,3**	80	63,3
E-V (2 000)	82,3	**82,2**	95	90	82,2
E-C (1 000)	82,3	82,2	**63,3**	66,7	63,3
S-V (500)	82,3	**65**	95	90	65
S-C (500)	82,3	65	63,3	**40**	40
Primes	41,17	196,94	95	20	353,10
Sinistres	60	225	106,67	20	411,67
S/P	145,7 %	114,2 %	112,3 %	100,0 %	116,6 %
(IC 95 %)	± 34,6 %	± 11,8 %	± 15,1 %	± 41,9 %	± 5,3 %
VaR 99,5 %	189,5 %	140,0 %	134,0 %	160,0 %	130,3 %
Part de marché	11,6 %	55,8 %	26,9 %	5,7 %	

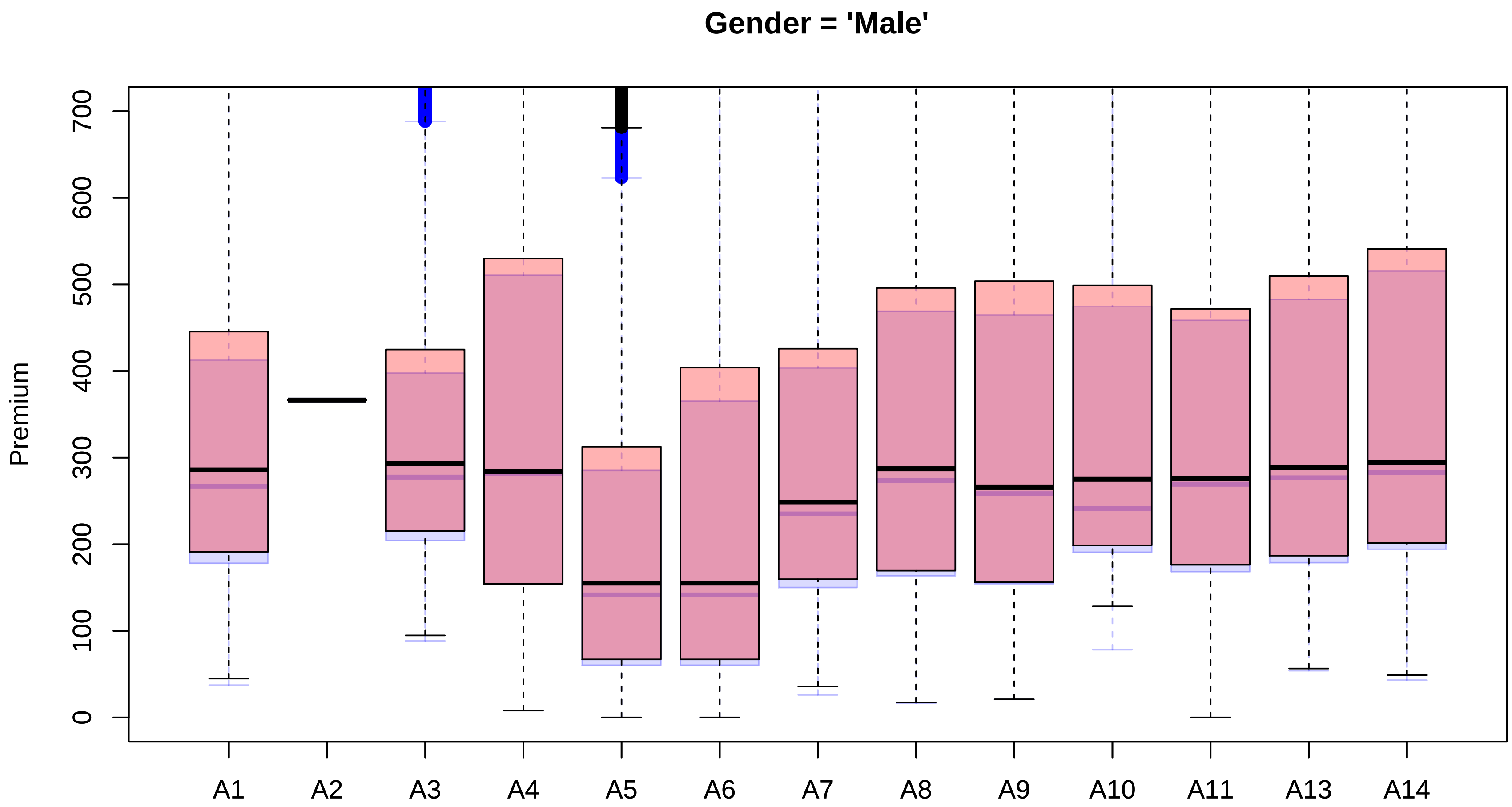
Pricing Game, with a Toy Dataset

	PolNum	CalYear	Gender	Type	Category	Occupation	Age	Group1	Bonus	Poldur	Value	Adind	SubGroup2	Exppdays	Numad	Numtppd	Numtpbi	Incad	Indtppd	Indtpbi	Group2
80	200114978	2009	Male	C	Large	Employed	25	18	90	3	15080	0	L46	365	0	1	0	0.0000	0	0	L
81	200114994	2009	Male	E	Large	Employed	20	11	30	2	22370	1	O38	365	1	1	0	862.0558	0	0	O
82	200115001	2009	Female	E	Large	Unemployed	42	11	150	0	39650	0	Q28	365	0	2	0	0.0000	0	0	Q
83	200115011	2009	Female	C	Medium	Housewife	21	5	0	0	12600	1	L6	365	0	1	0	0.0000	0	0	L
84	200115015	2009	Female	D	Large	Employed	33	12	30	10	9065	0	N4	365	0	2	0	0.0000	0	0	N
85	200115016	2009	Female	D	Small	Employed	26	13	40	7	27335	1	N16	365	0	1	0	0.0000	0	0	N
86	200115023	2009	Female	C	Small	Unemployed	20	7	80	13	7710	0	Q65	365	0	1	0	0.0000	0	0	Q
87	200115043	2009	Female	B	Medium	Employed	29	3	-20	12	8965	0	R19	365	0	1	0	0.0000	0	0	R
88	200115048	2009	Male	E	Medium	Unemployed	31	3	-40	10	21030	1	R9	355	0	1	0	0.0000	0	0	R
89	200115063	2009	Male	D	Medium	Employed	35	7	120	1	19995	1	Q22	365	0	1	0	0.0000	0	0	Q
90	200115068	2009	Male	C	Small	Employed	27	11	30	0	18395	0	Q13	365	0	1	0	0.0000	0	0	Q
91	200115070	2009	Female	A	Large	Housewife	65	9	-30	15	11880	0	R35	365	0	1	0	0.0000	0	0	R

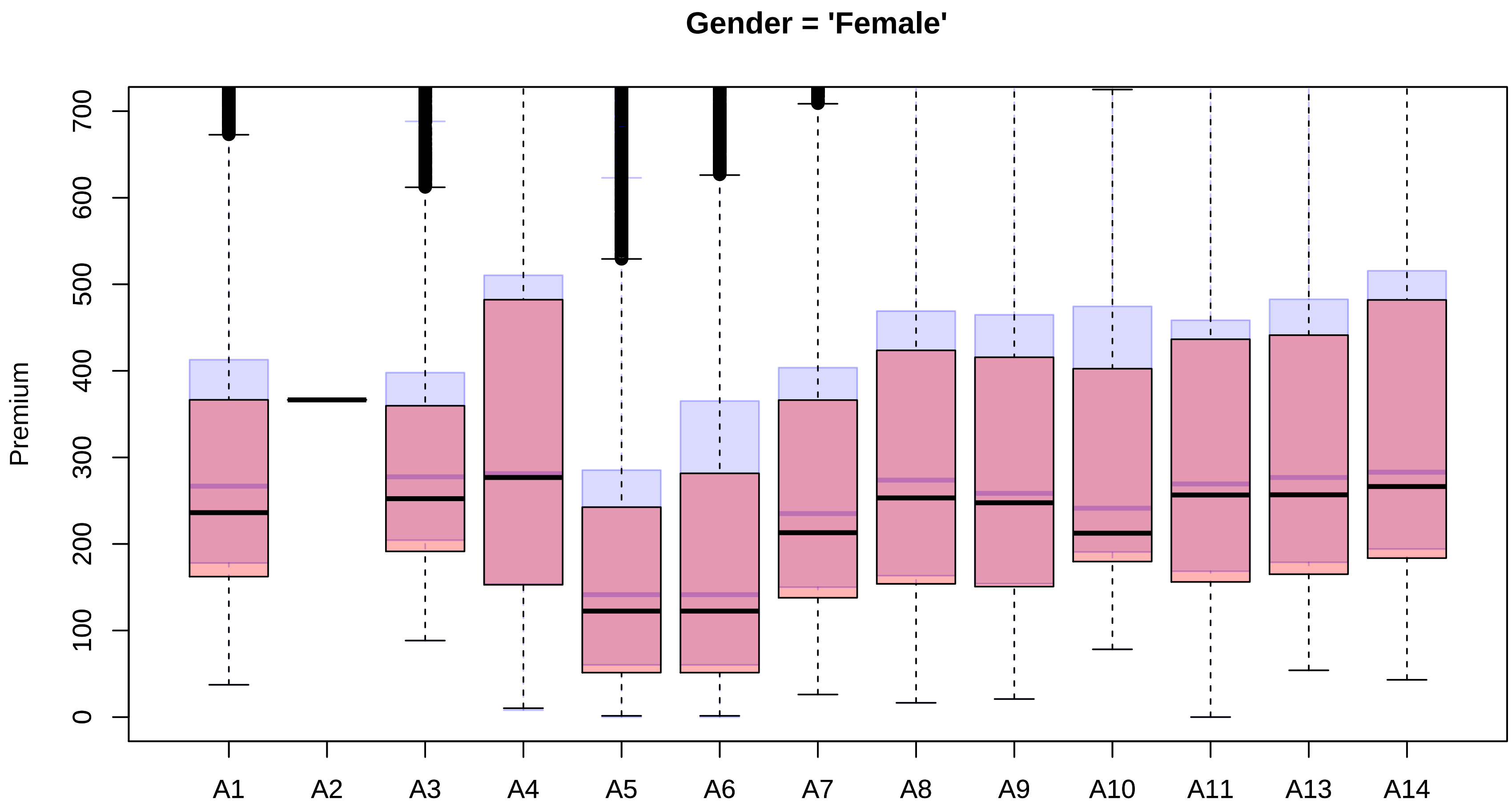
	PolNum	Prime
1	200325695	787.93
2	200325696	170.04
3	200325697	473.15
4	200325698	107.35
5	200325699	102.02
6	200325700	337.98
7	200325701	326.92
8	200325702	586.35
9	200325703	455.05
10	200325704	195.54
11	200325705	309.8
12	200325706	194.38
13	200325707	195.67
14	200325710	202.35
15	200325711	198.02
16	200325712	336.64
17	200325713	611.04
18	200325716	195.19
19	200325718	73.97

	PolNum	Prime
1	200325695	706.97
2	200325696	197.81
3	200325697	447.58
4	200325698	146.15
5	200325699	141.75
6	200325700	336.2
7	200325701	327.08
8	200325702	540.86
9	200325703	432.66
10	200325704	218.82
11	200325705	312.98
12	200325706	217.86
13	200325707	218.92
14	200325710	224.43
15	200325711	220.86
16	200325712	335.09
17	200325713	561.21
18	200325716	218.53
19	200325718	118.64

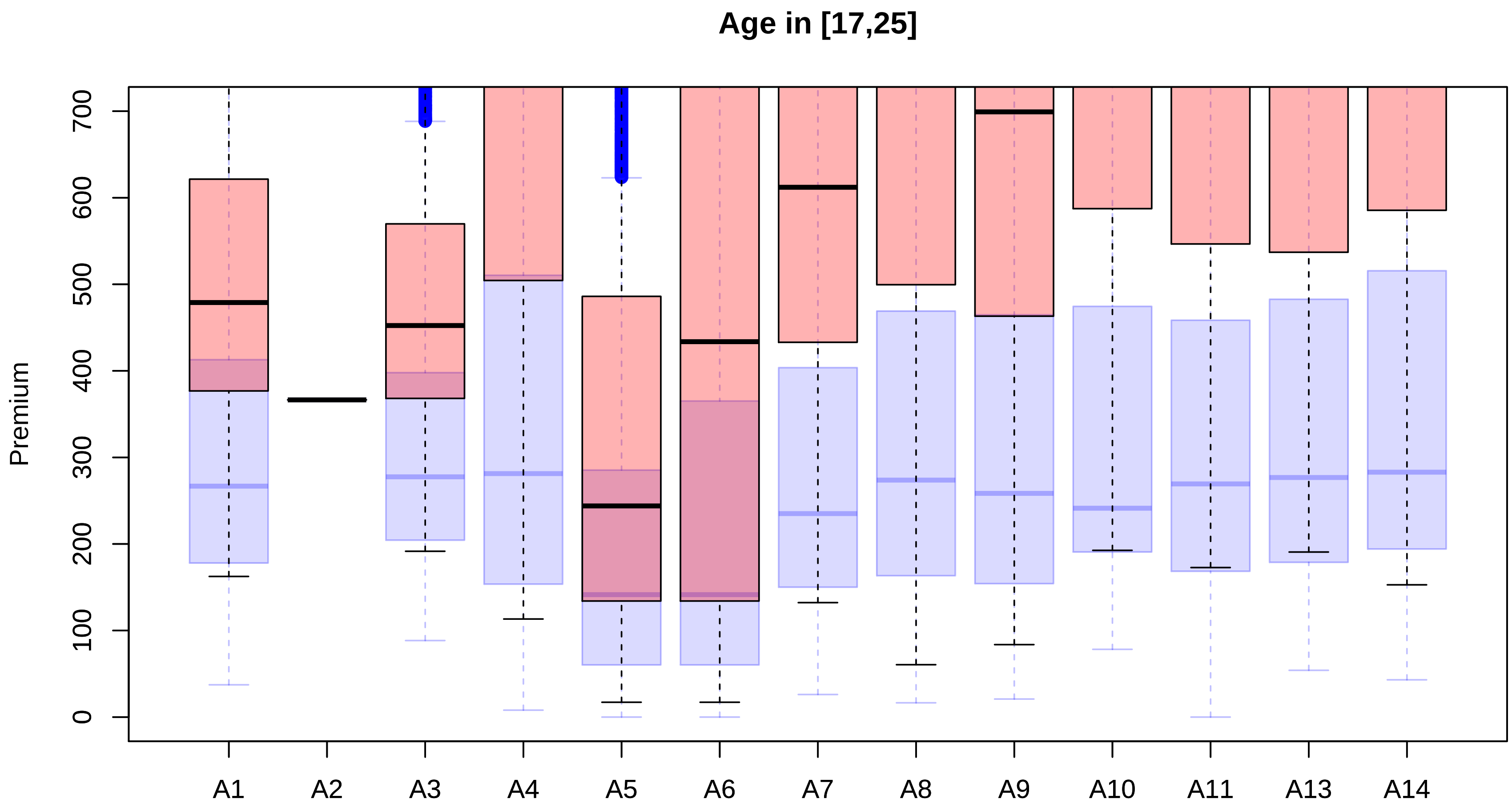
The Premiums, Brief Summary



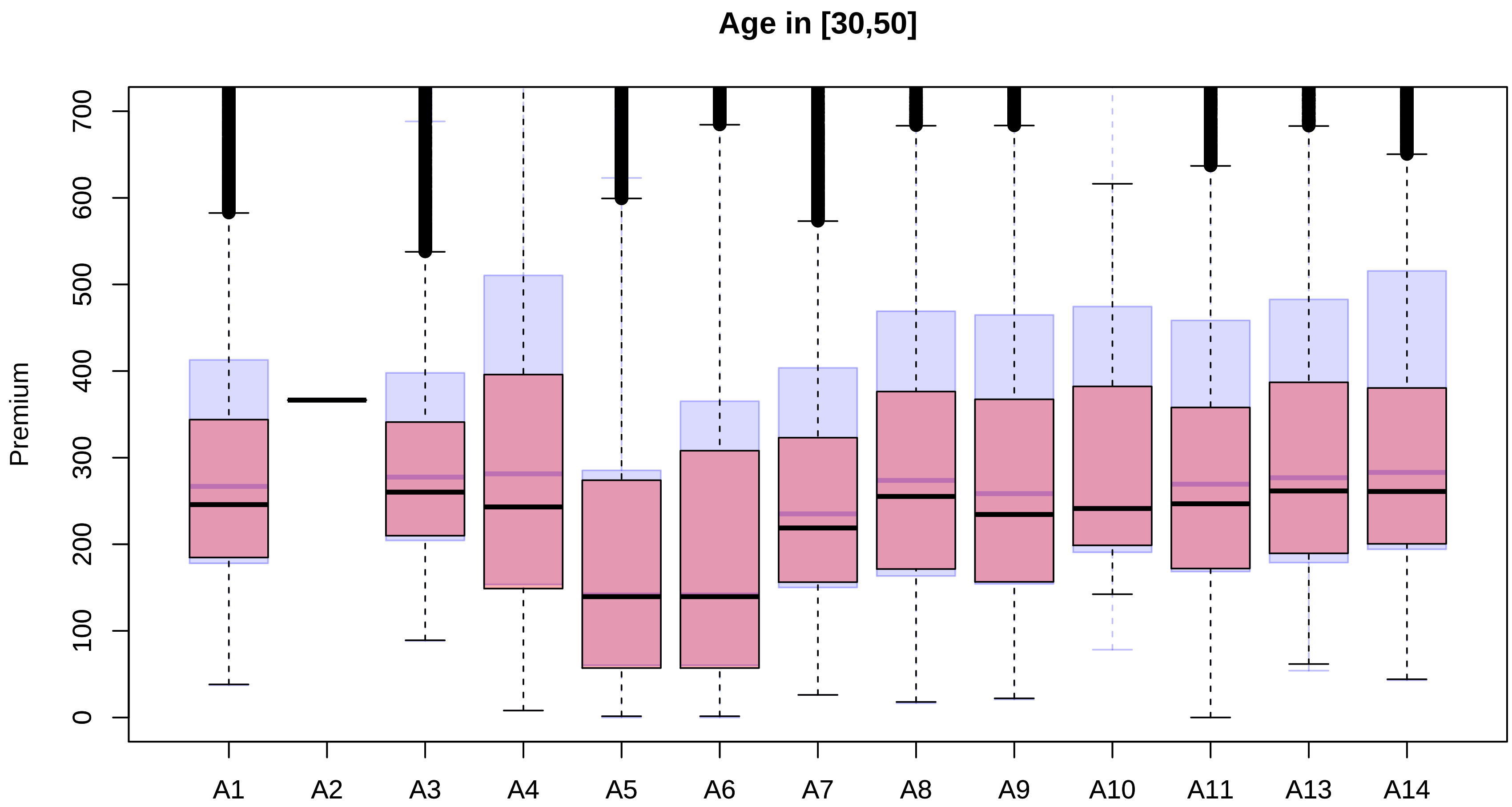
The Premiums, Brief Summary



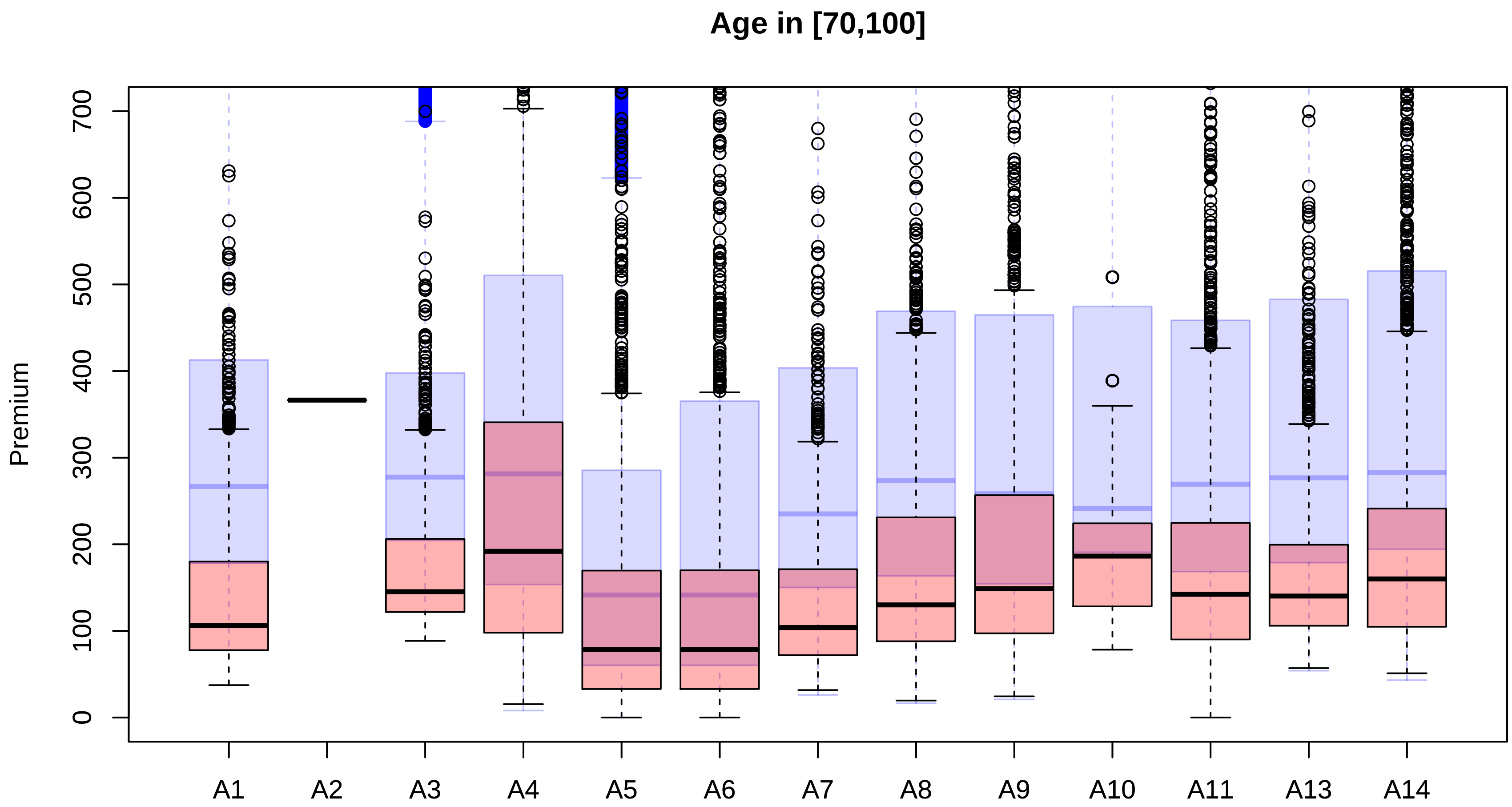
The Premiums, Brief Summary



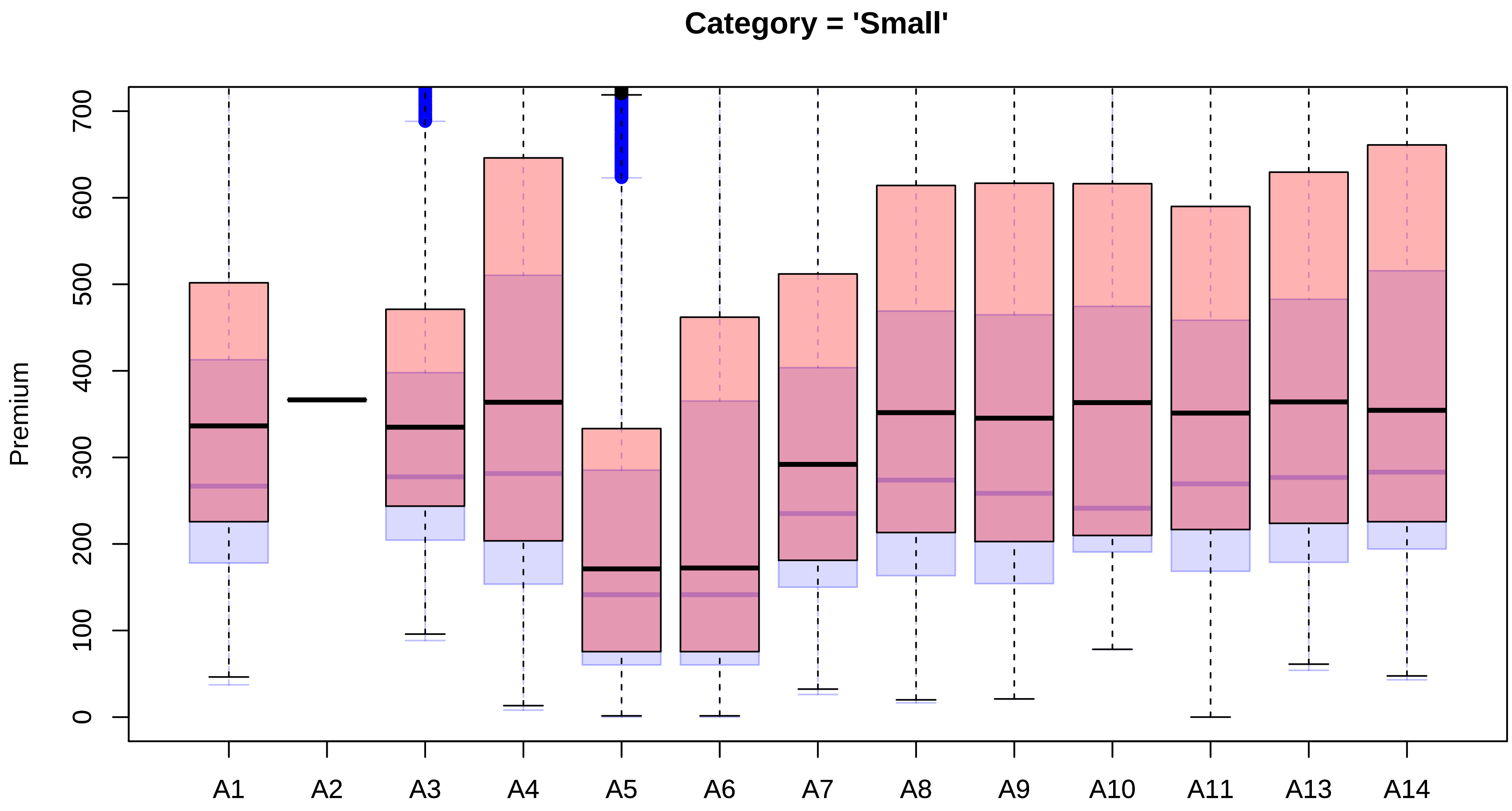
The Premiums, Brief Summary



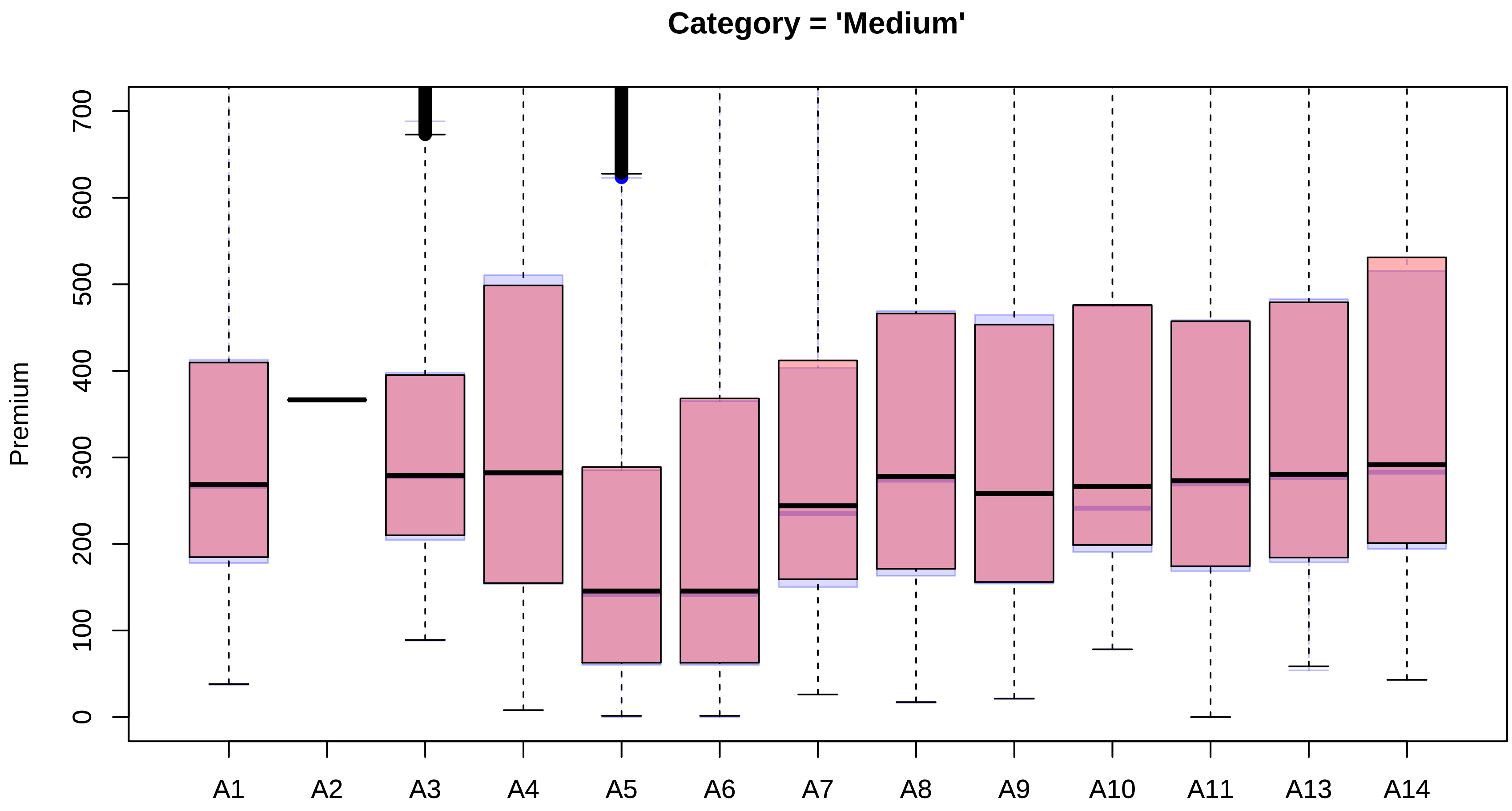
The Premiums, Brief Summary



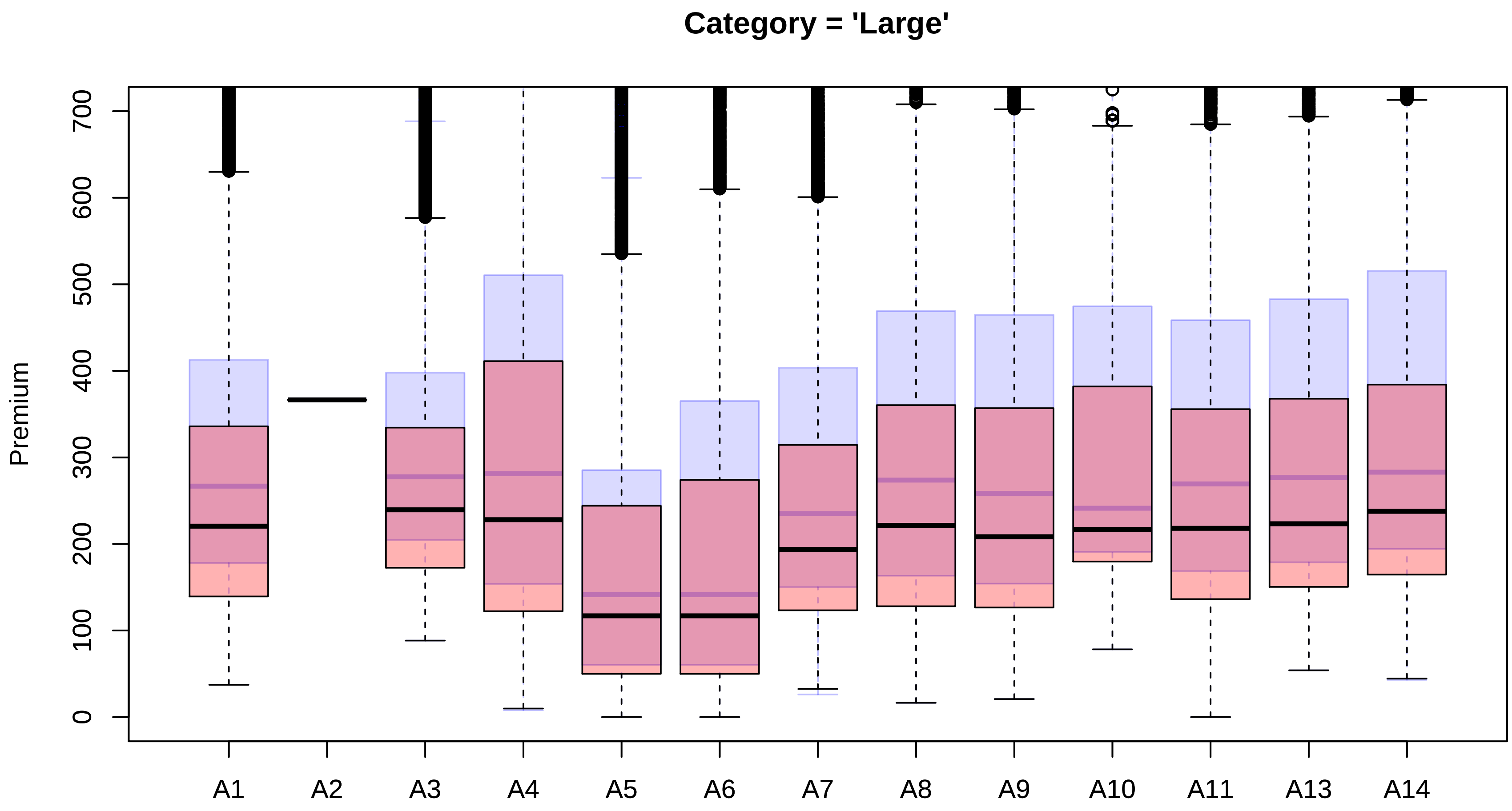
The Premiums, Brief Summary



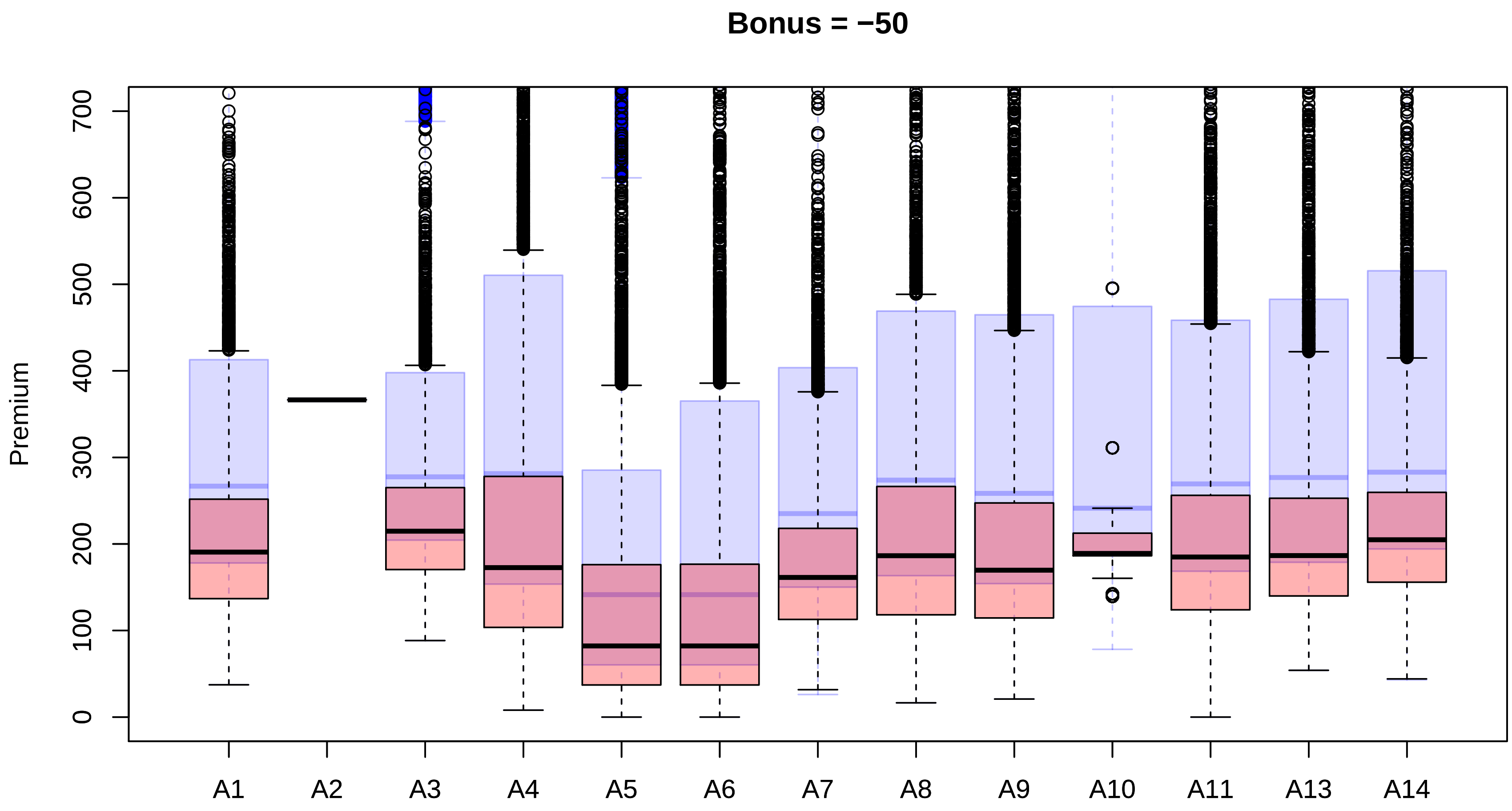
The Premiums, Brief Summary



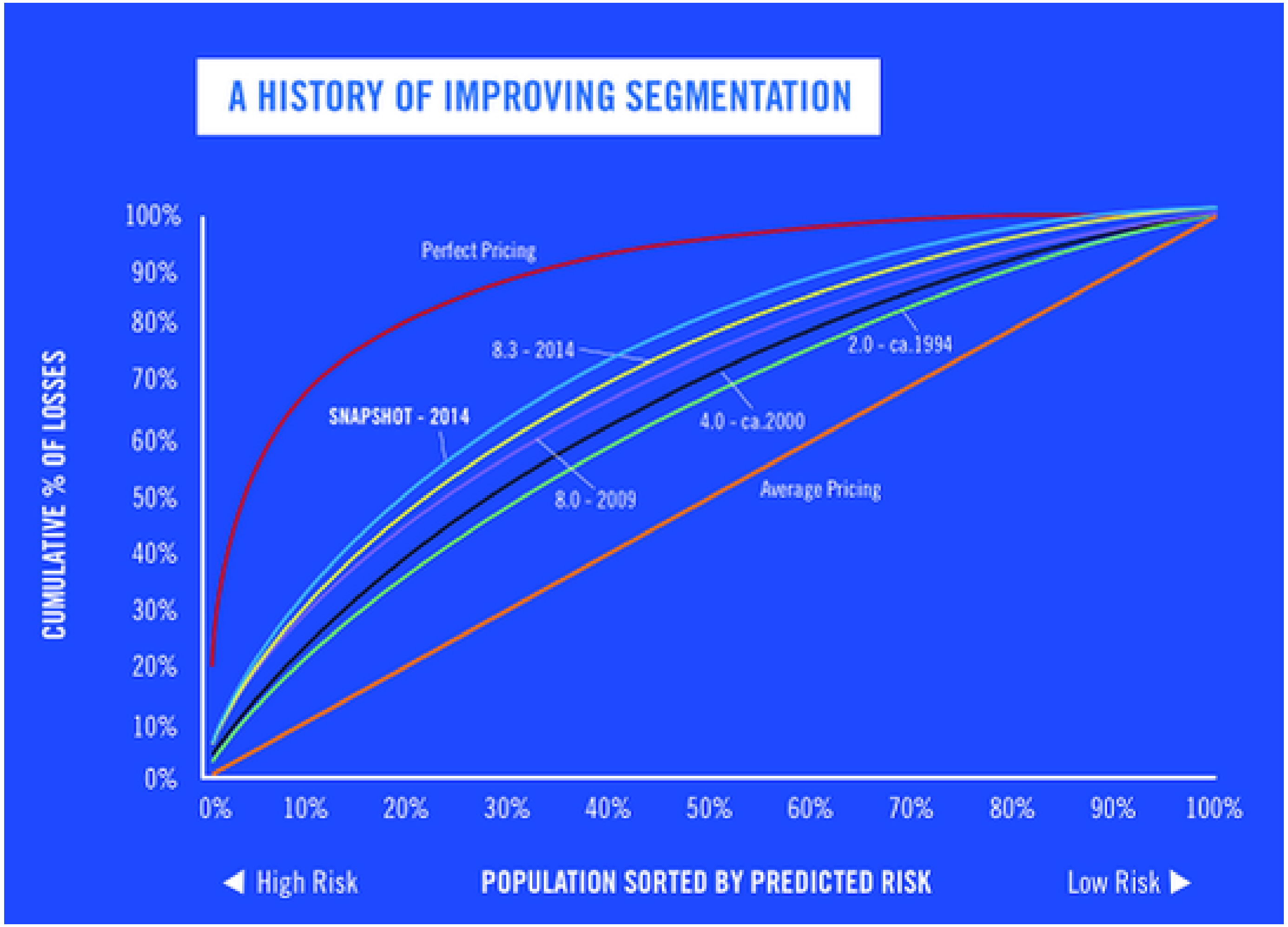
The Premiums, Brief Summary



The Premiums, Brief Summary



Comparing Models

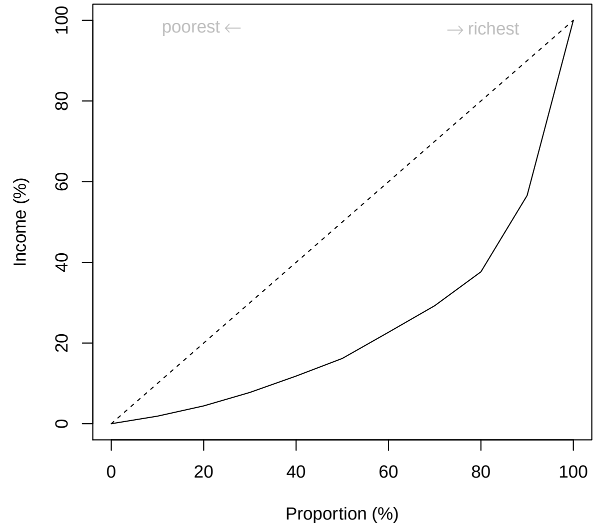


Source: *unkown, so far...*

Comparing Models

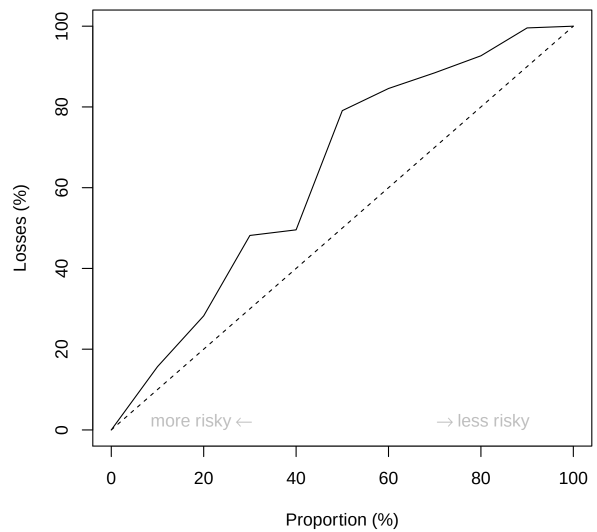
Consider an ordered sample $\{y_1, \cdots, y_n\}$ of incomes, with $y_1 \leq y_2 \leq \cdots \leq y_n$, then Lorenz curve is

$$\{F_i, L_i\} \text{ with } F_i = \frac{i}{n} \text{ and } L_i = \frac{\sum_{j=1}^{i} y_j}{\sum_{j=1}^{n} y_j}$$

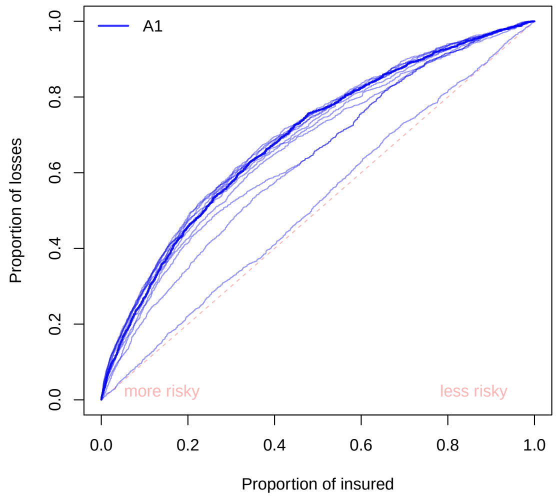


We have observed losses y_i and premiums $\widehat{\pi}(\boldsymbol{x}_i)$. Consider an ordered sample by the model, see Frees, Meyers & Cummins (2014), $\widehat{\pi}(\boldsymbol{x}_1) \geq \widehat{\pi}(\boldsymbol{x}_2) \geq \cdots \geq \widehat{\pi}(\boldsymbol{x}_n)$, then plot

$$\{F_i, L_i\} \text{ with } F_i = \frac{i}{n} \text{ and } L_i = \frac{\sum_{j=1}^{i} y_j}{\sum_{j=1}^{n} y_j}$$



The Models, Brief Summary



Market Competition

Decision Rule: the insured selects the cheapeast premium,

	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

Market Competition

Decision Rule: the insured selects randomly from the three cheapeast premium

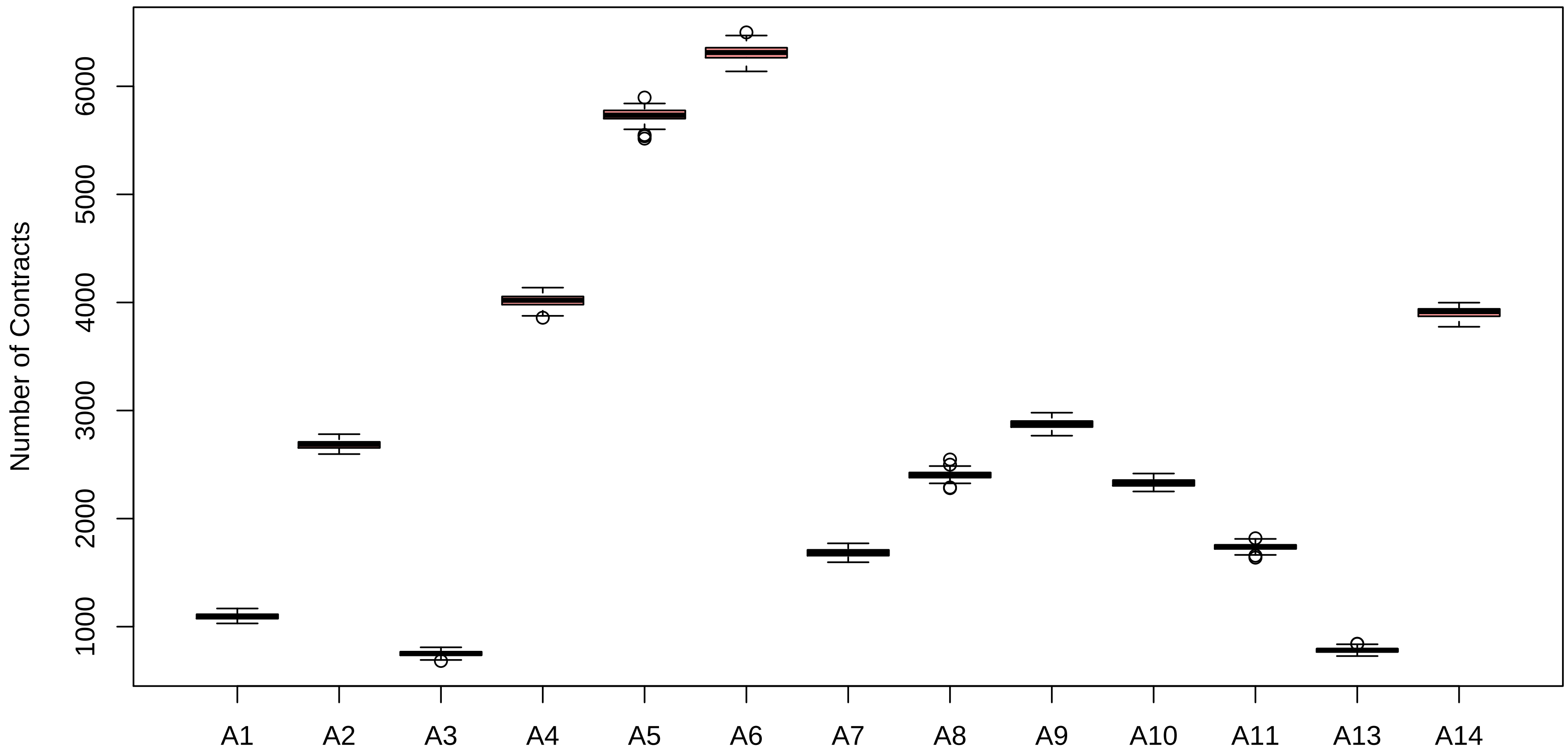
	A	B	C	D	E	F
	787.93	706.97	1032.62	907.64	822.58	603.83
	170.04	197.81	285.99	212.71	177.87	265.13
	473.15	447.58	343.64	410.76	414.23	425.23
	337.98	336.20	468.45	339.33	383.55	672.91

Market Competition

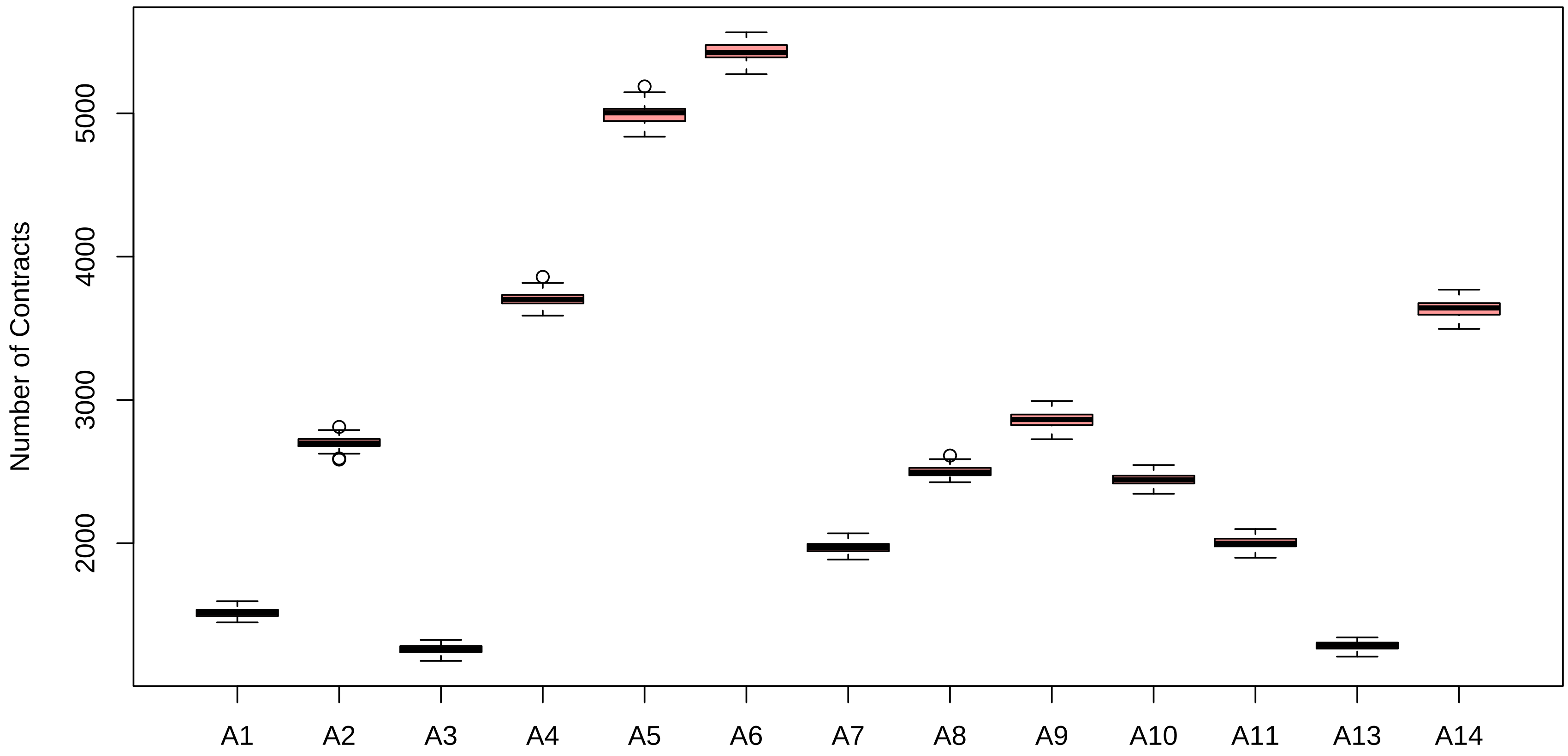
Decision Rule: the insured were assigned randomly to some insurance company for year $n-1$. For year n, they stay with their company if the premium is one of the three cheapeast premium, if not, random choice among the four

	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

Financial Results, Market Share, Method 2

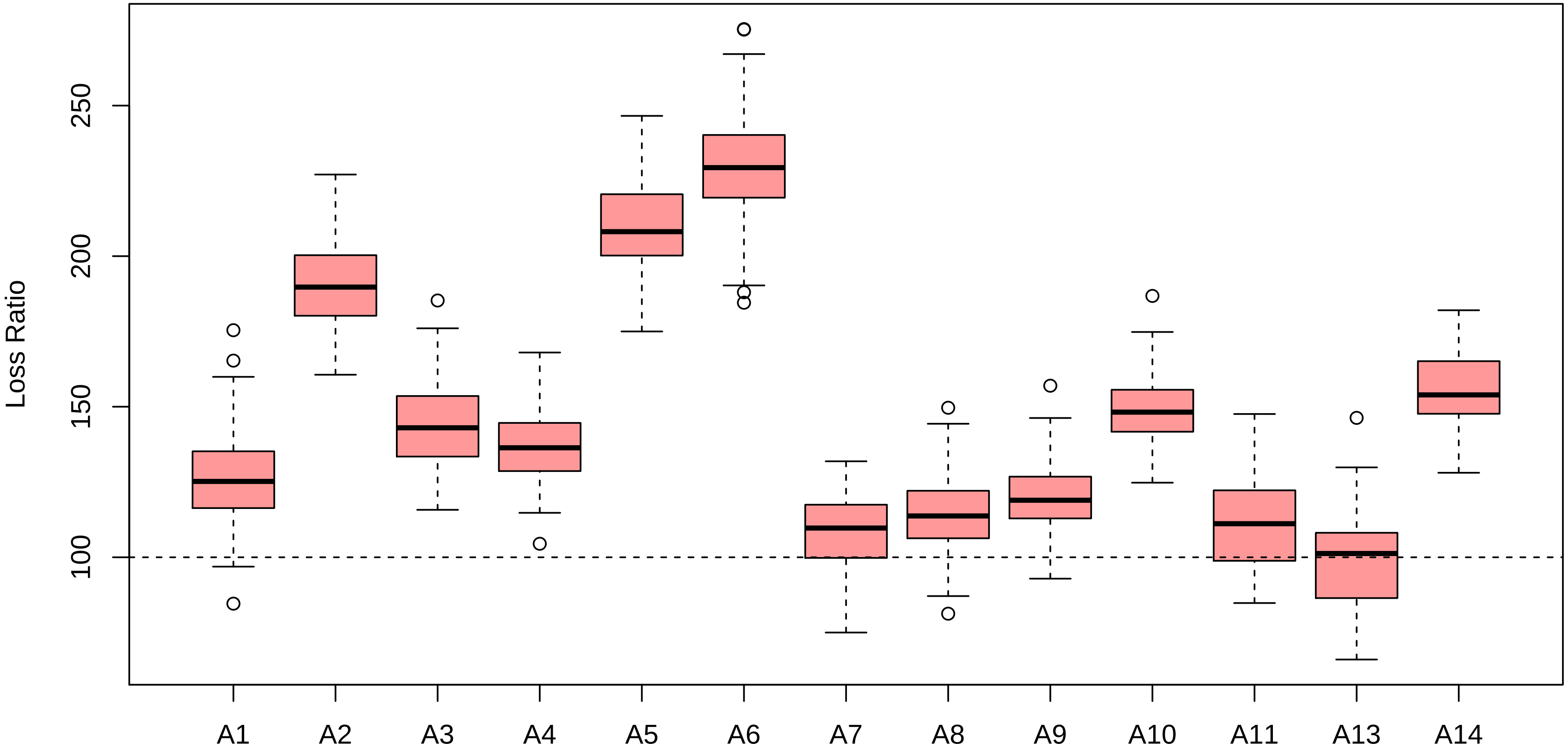


Financial Results, Market Share, Method 3

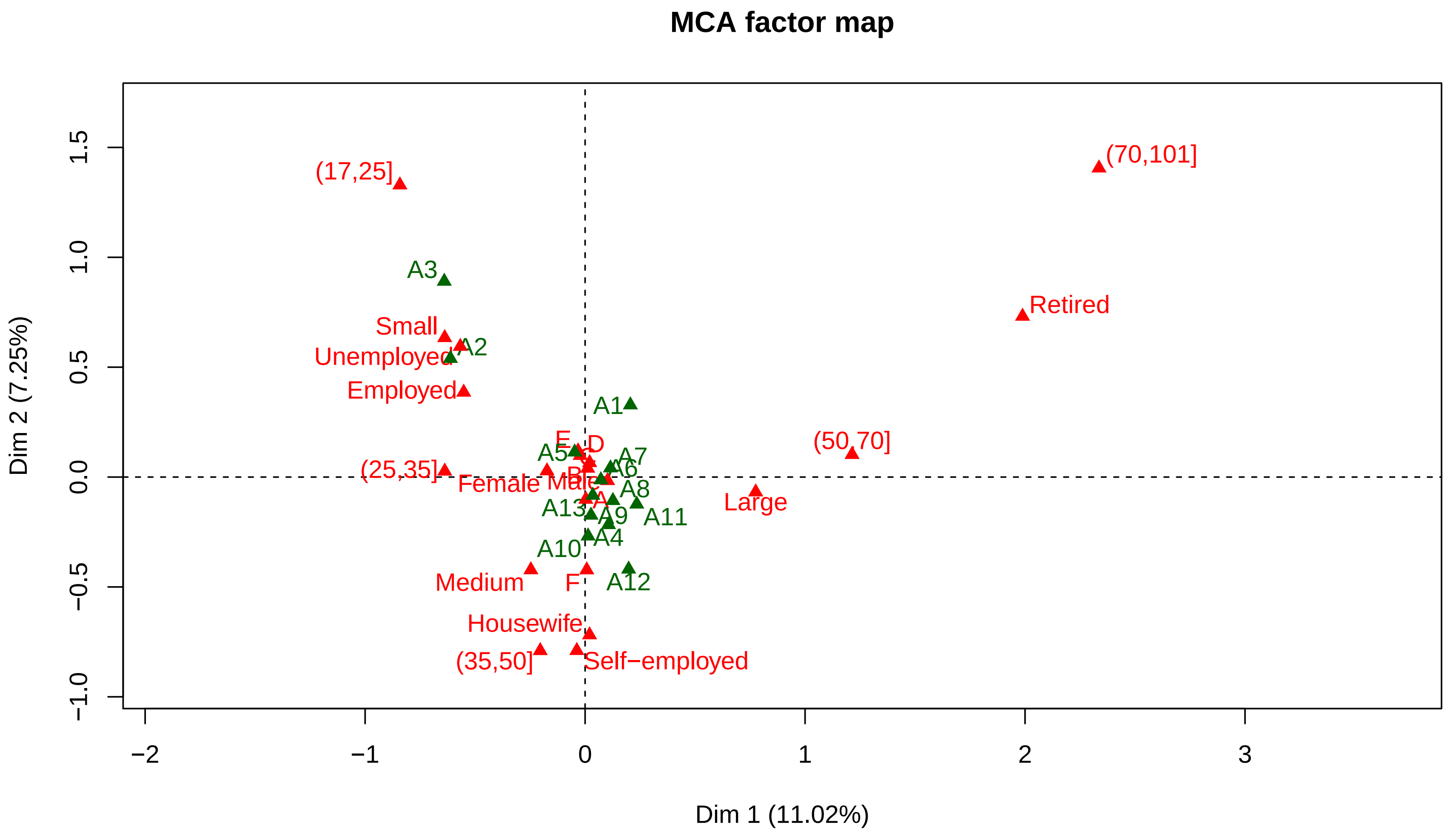


Financial Results, Loss Ratio, Method 2

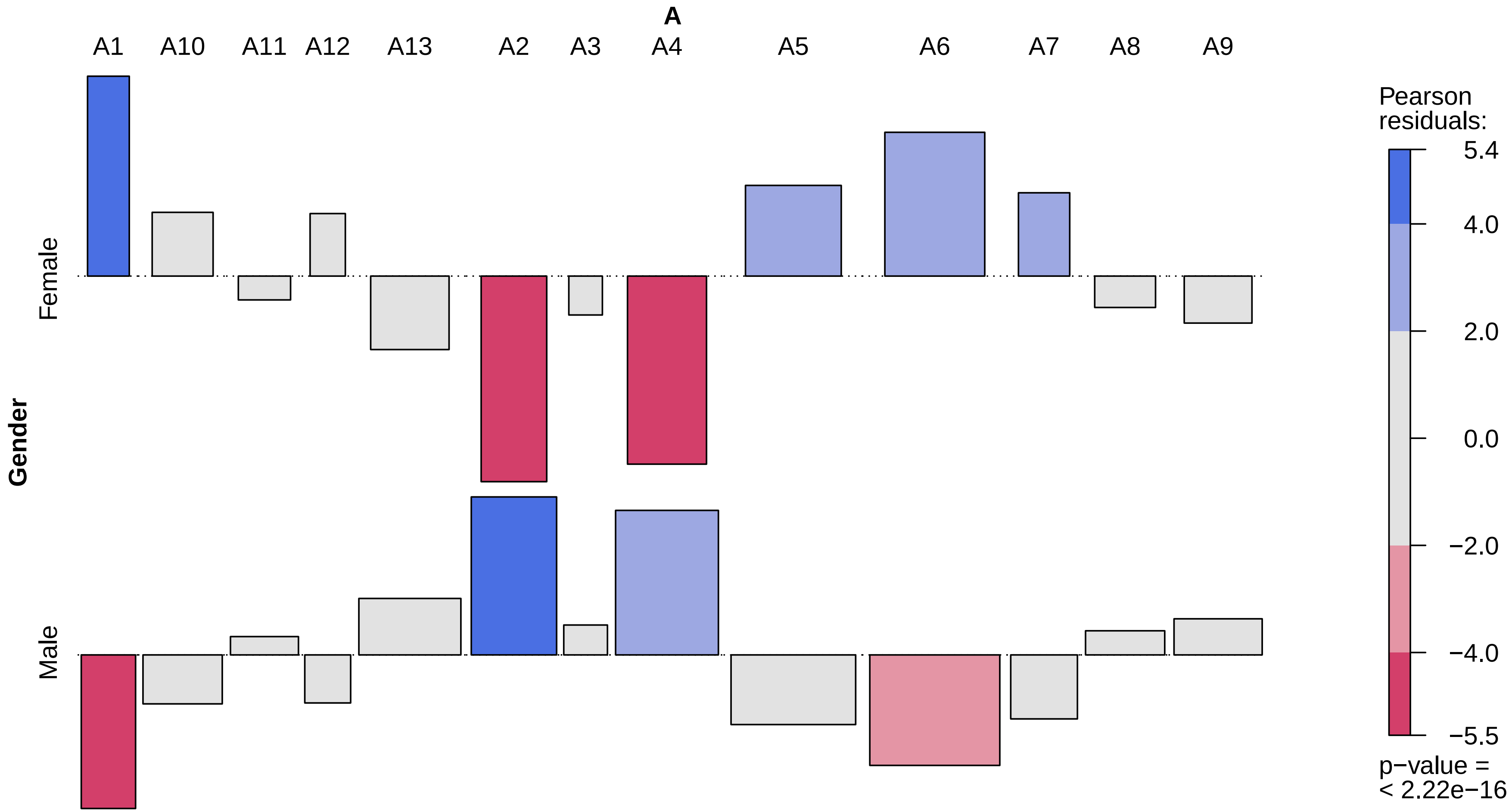
Market Loss Ratio $\sim 154\%$.



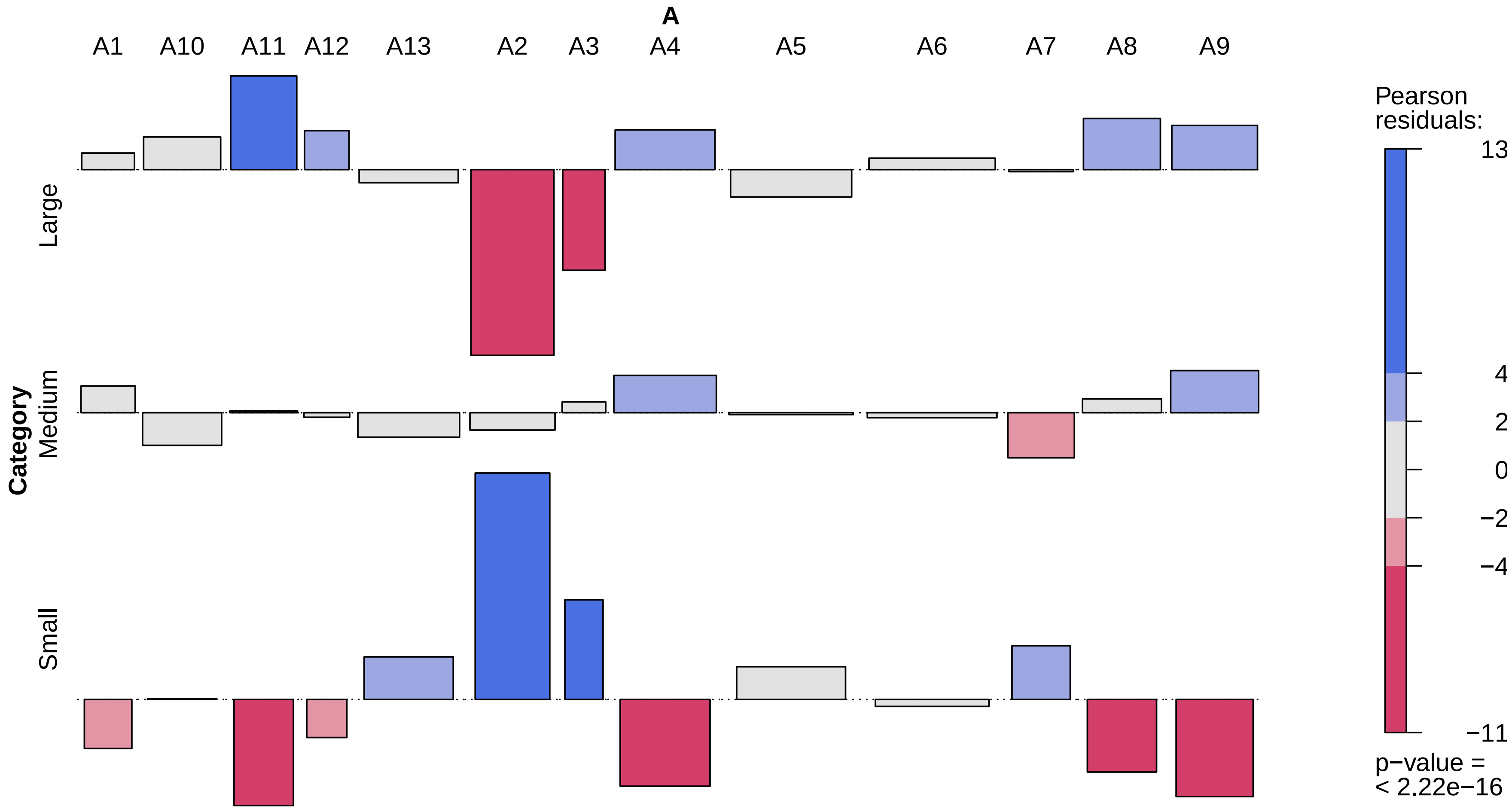
Insurers and Insurance Companies



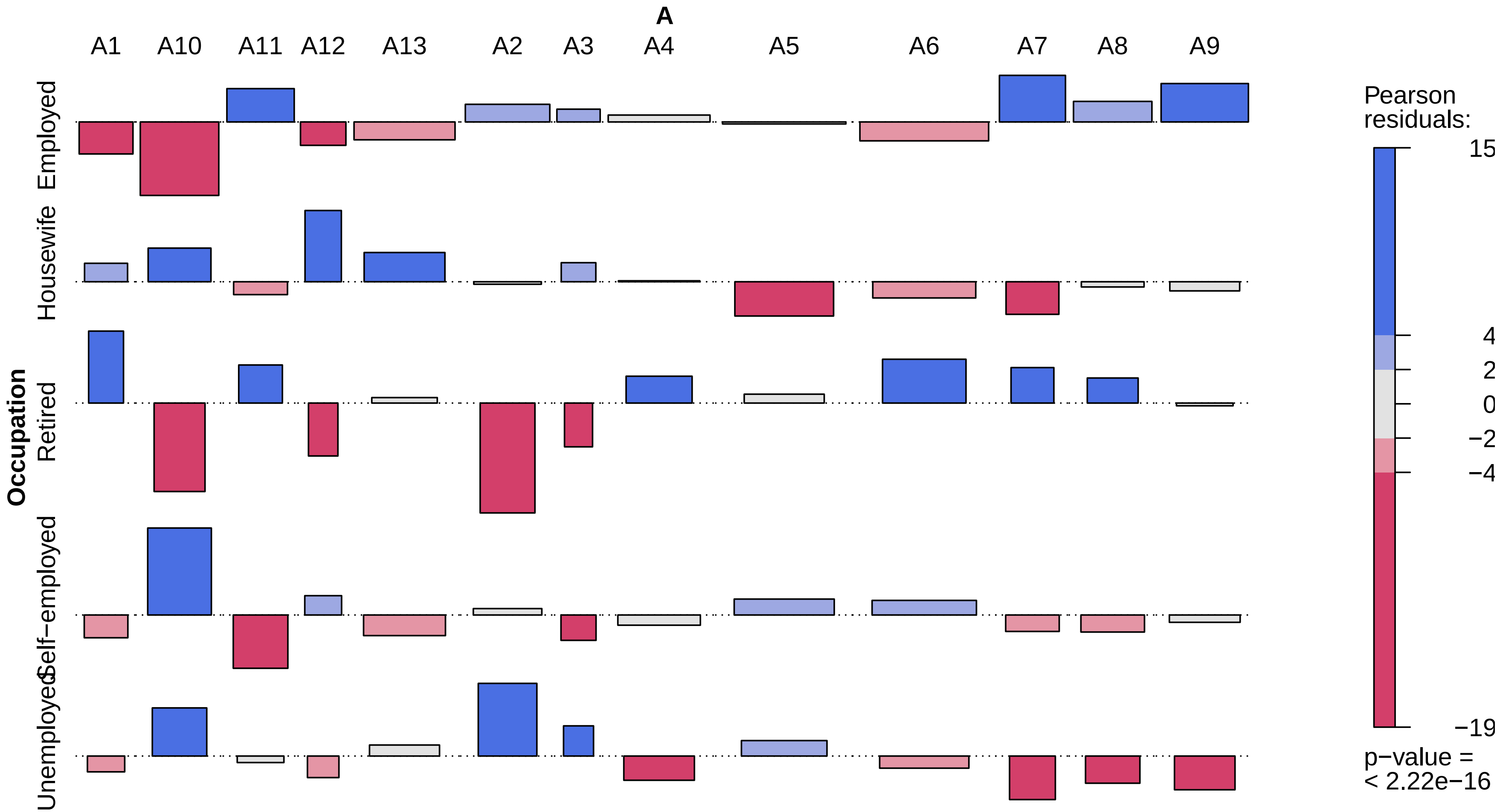
Insurers and Insurance Companies



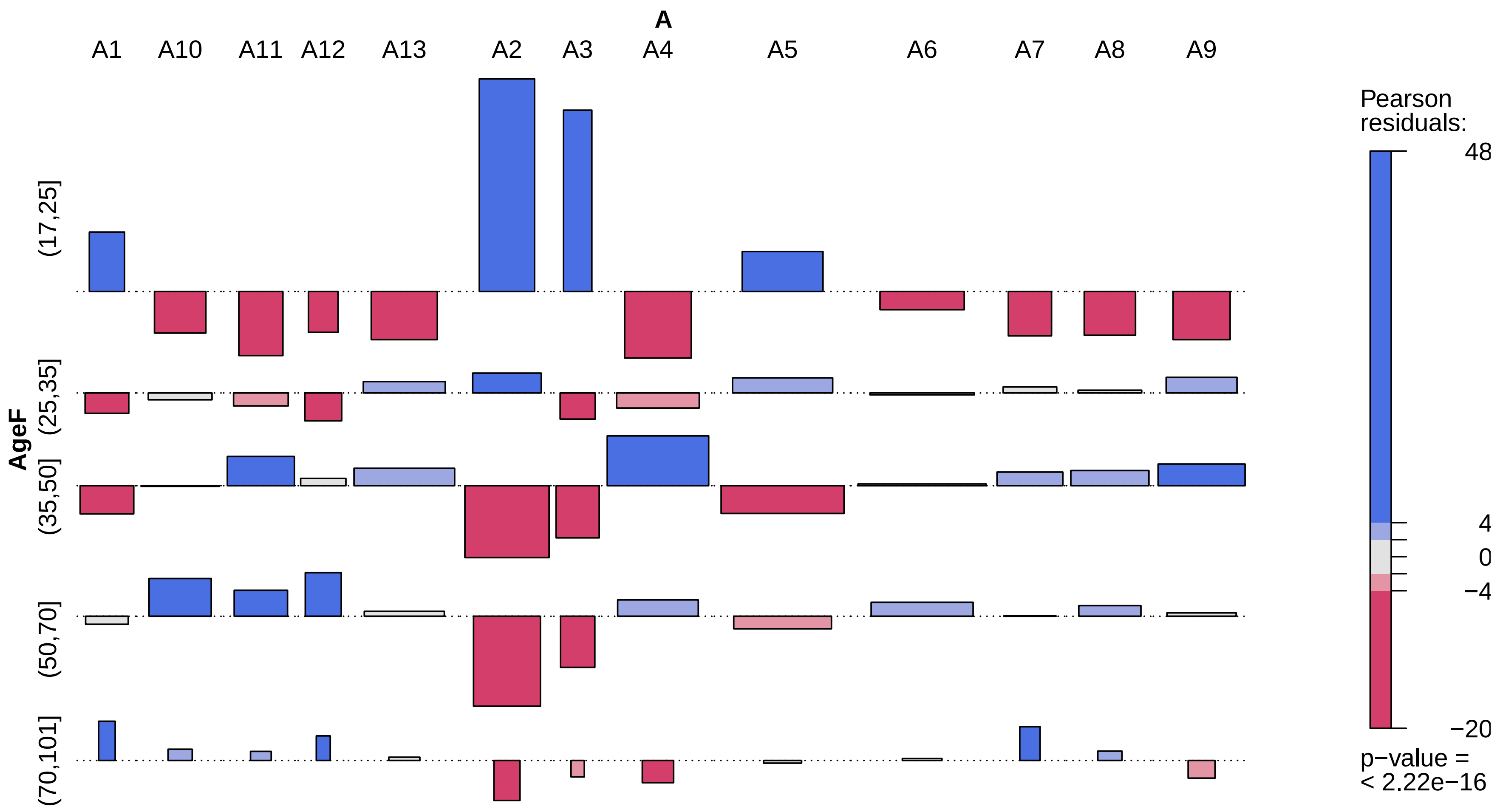
Insurers and Insurance Companies



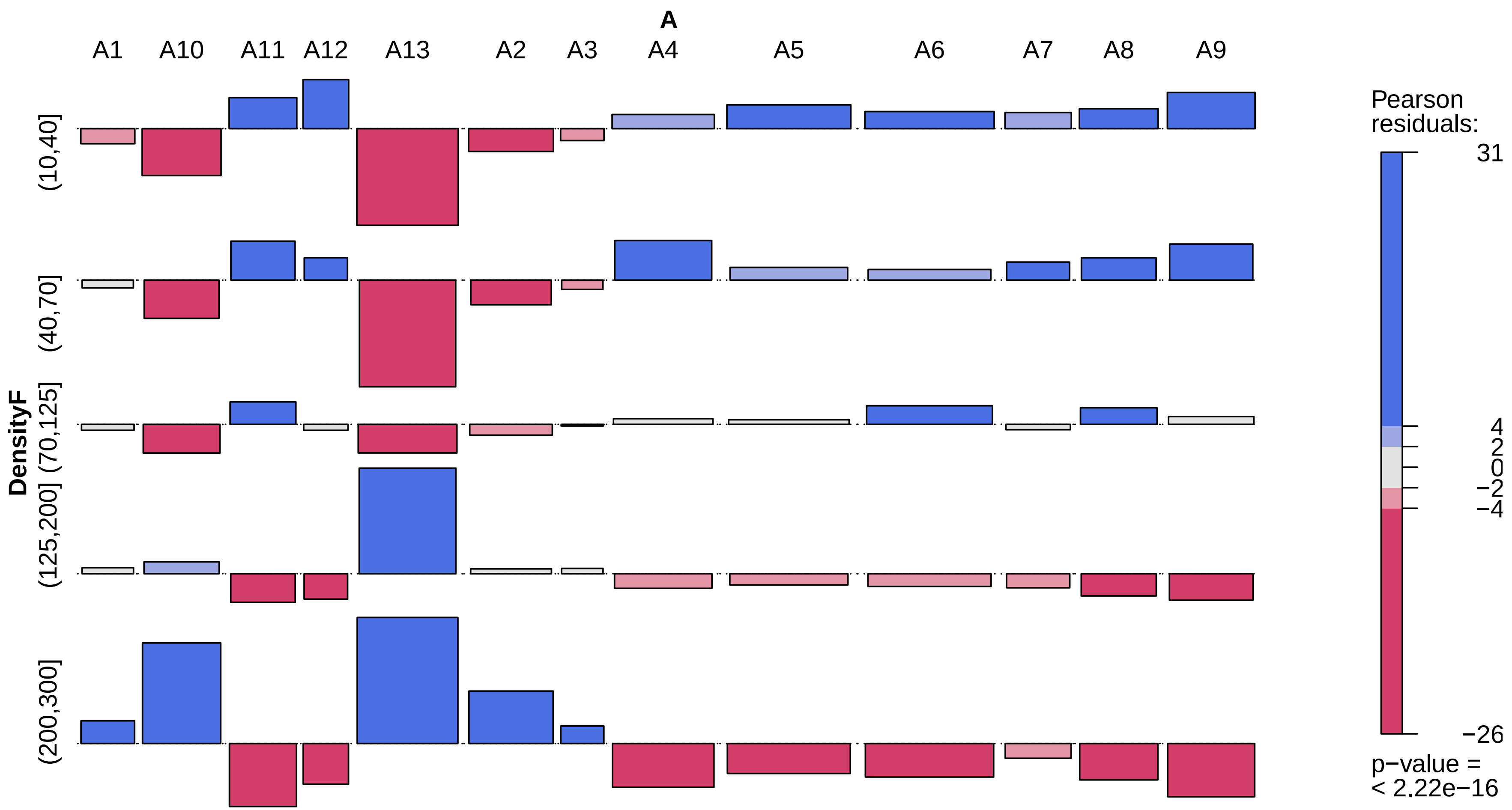
Insurers and Insurance Companies



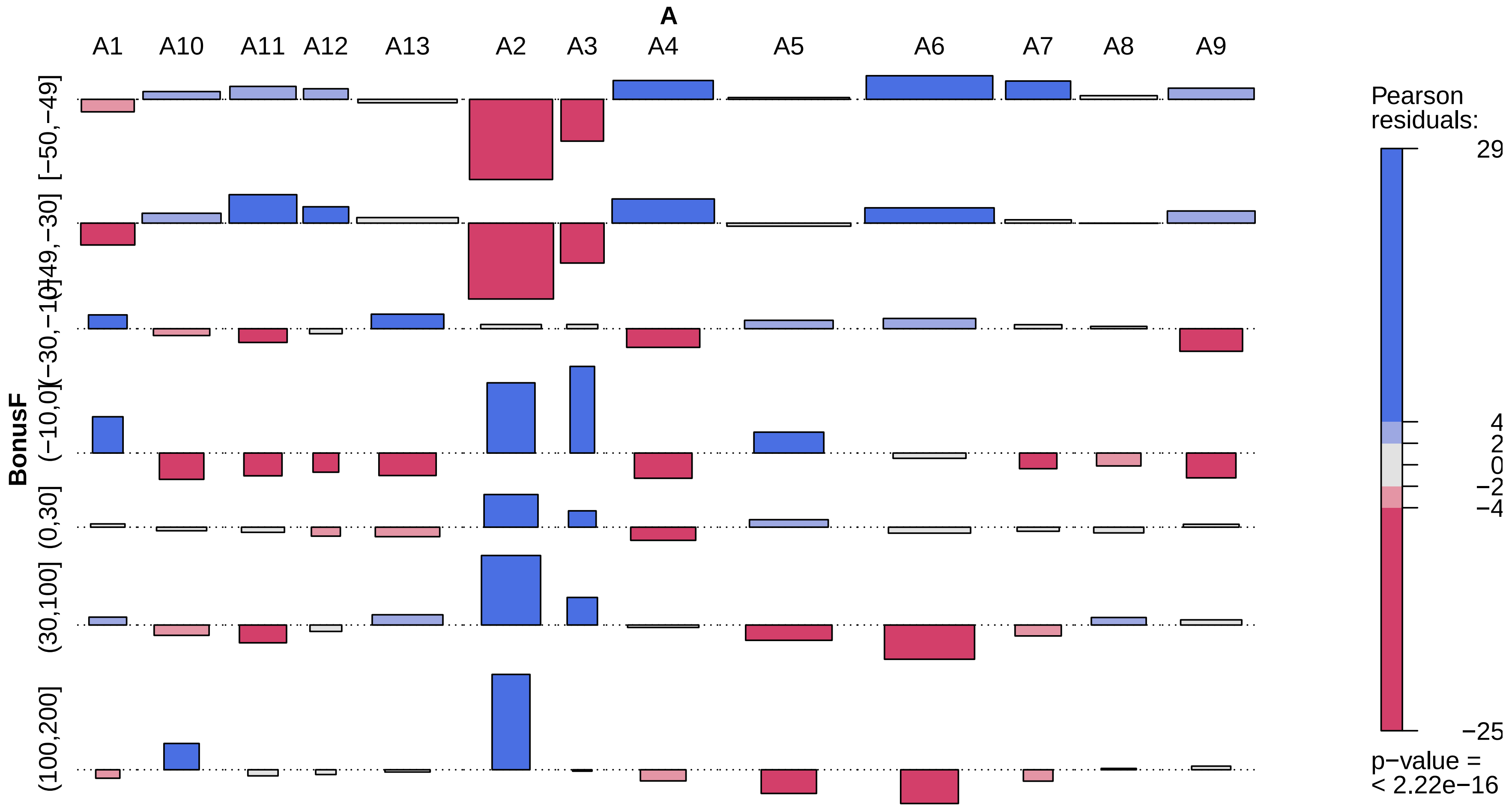
Insurers and Insurance Companies



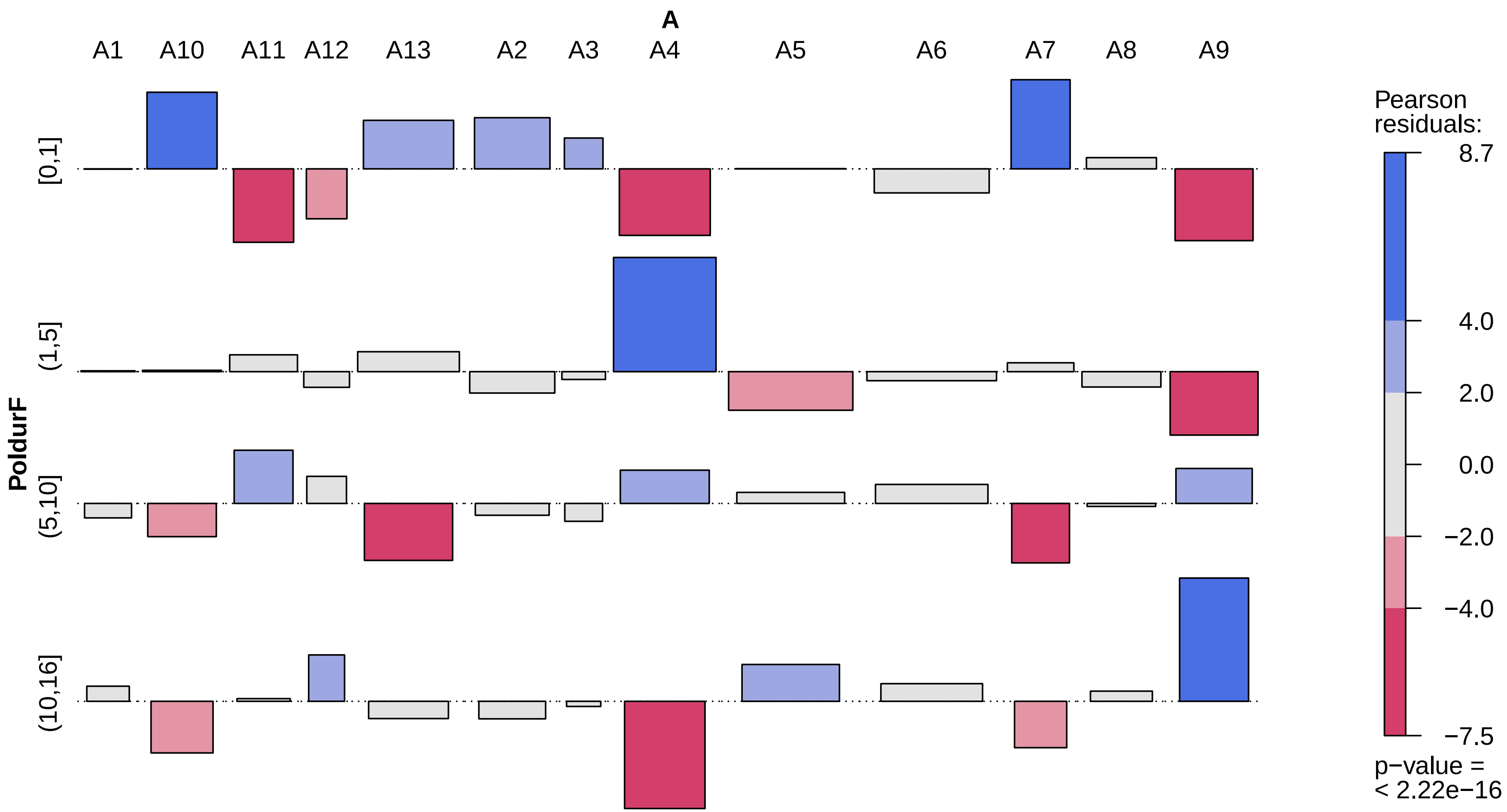
Insurers and Insurance Companies



Insurers and Insurance Companies

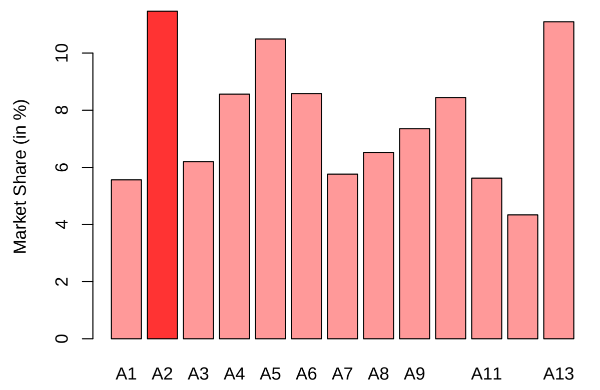


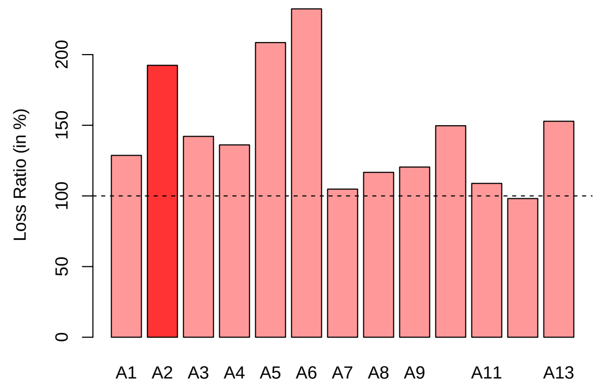
Insurers and Insurance Companies



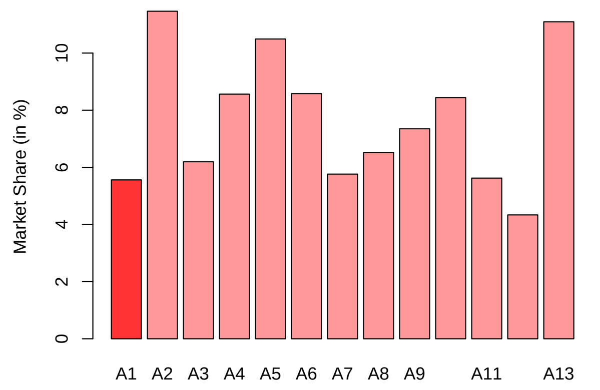
Market Competition, A2

No segmentation, unique price





Market Competition, A1

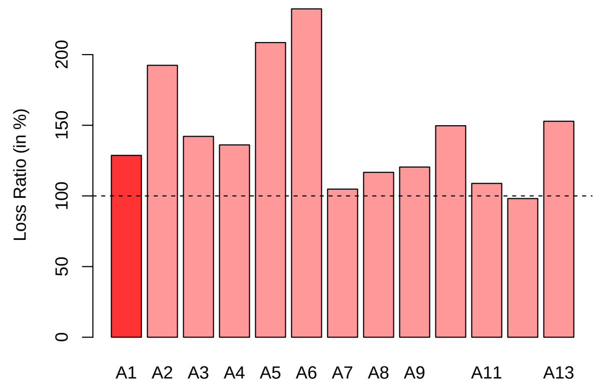


Three models (GLM) frequency bodily injury - frequency material damage - loss material

Age in 4 categories, $30-$, $30-45$, $45-60$ and $60+$, an interaction with occupation.

Manual Smoothing of parameters

Done with SAS and Excel.

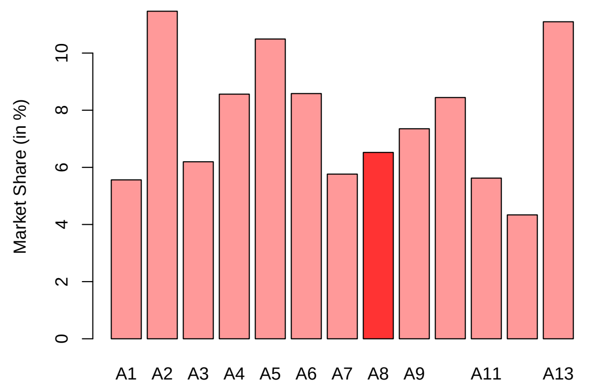


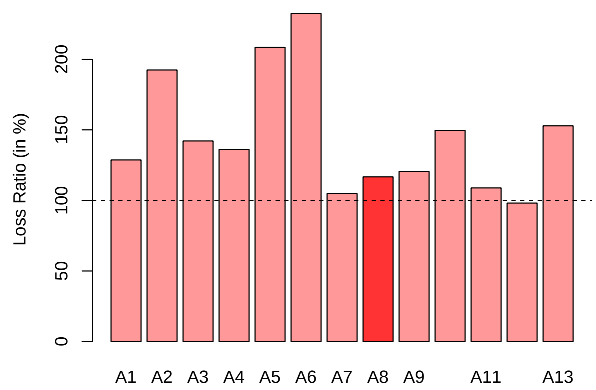
Sent by an actuary working for a *mutuelle*

Market Competition, A8-A9

GLM for frequency and standard cost (large claimes were removed, above 15k)
Interaction Age and Gender
All variables but density and group2.
Done with a software developped by Actuaris.

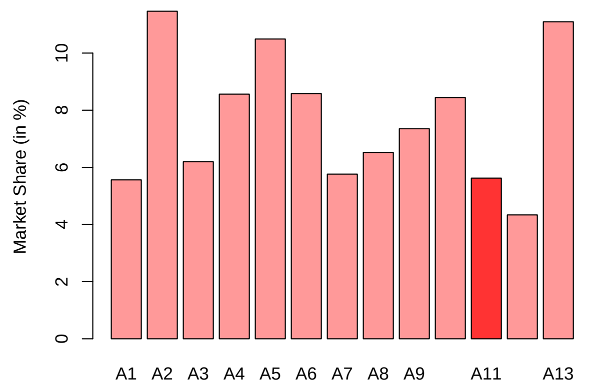
Sent by an actuary working for a *mutuelle*



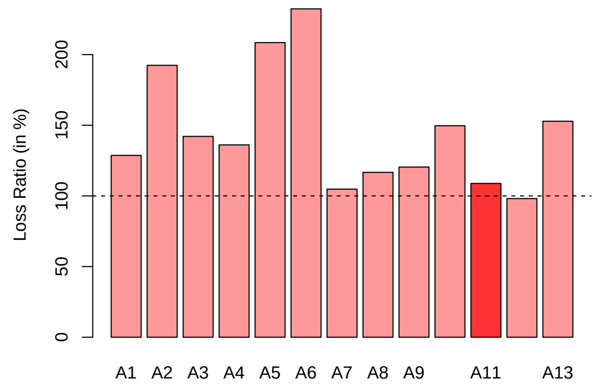


Market Competition, A11

Use of all variables, but Subgroup2,
Use of two XGBoost models (bodily injury and material),
Correction for negative premiums
Use of Python



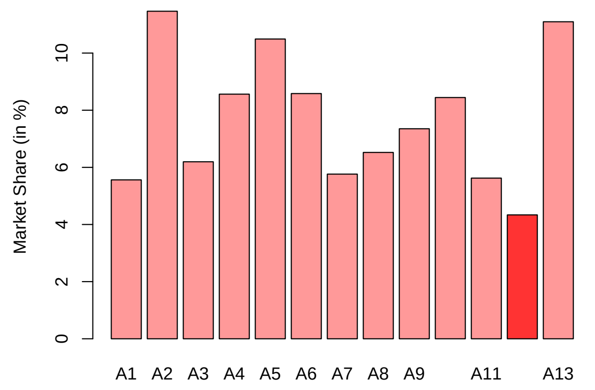
Sent by an actuary working for a private insurance company (sudent at the *Data Science for Actuaries* program)

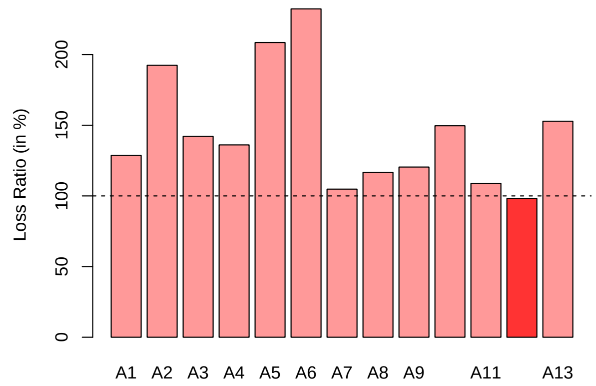


Market Competition, A12

Use of two xgboost models (bodily injury and material),
Correction for negative premiums
Use of R

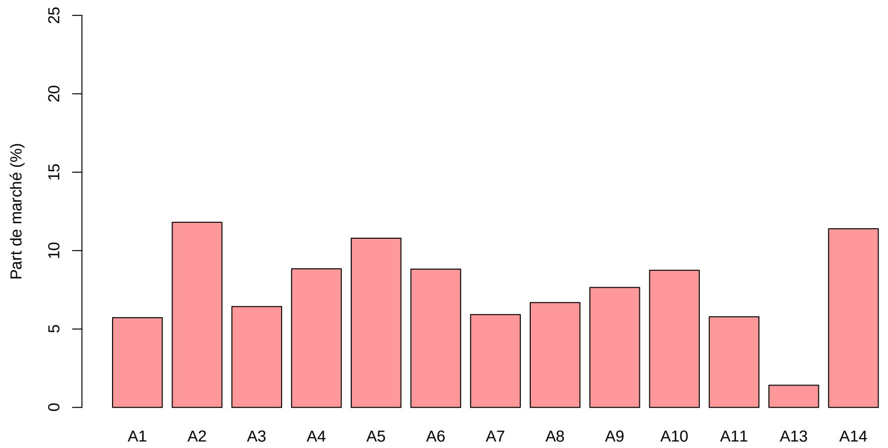
Sent by an actuary working for a private insurance company, in Europe, but not in France.

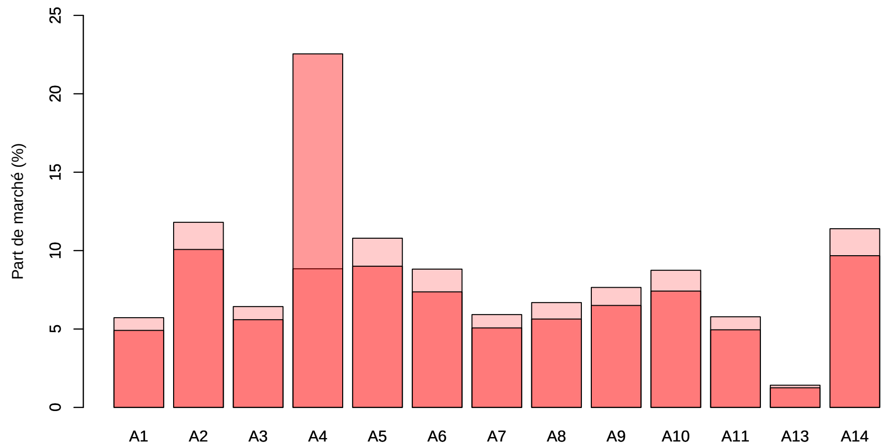




Distorting the Market: the Price Aggregator as a Market Player

Partnership between A4 and the price aggregator: if A4 is either the 4^{th} or 5^{th}, it returns the 3^{rd} price $(-\varepsilon)$

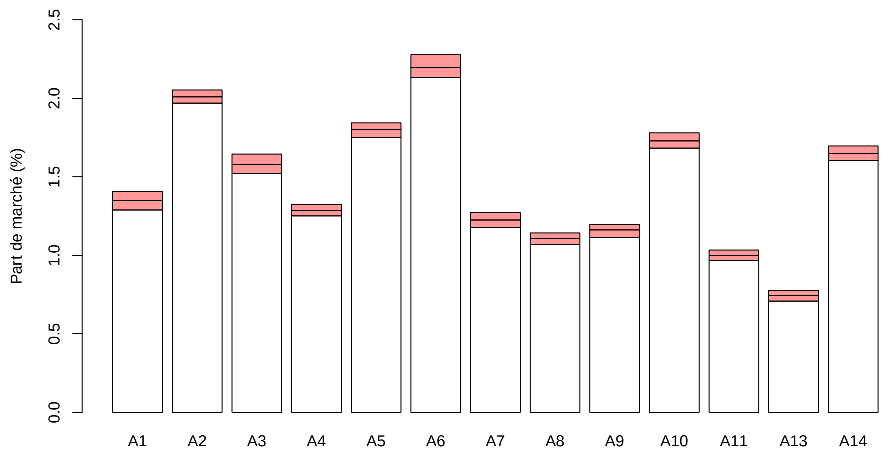


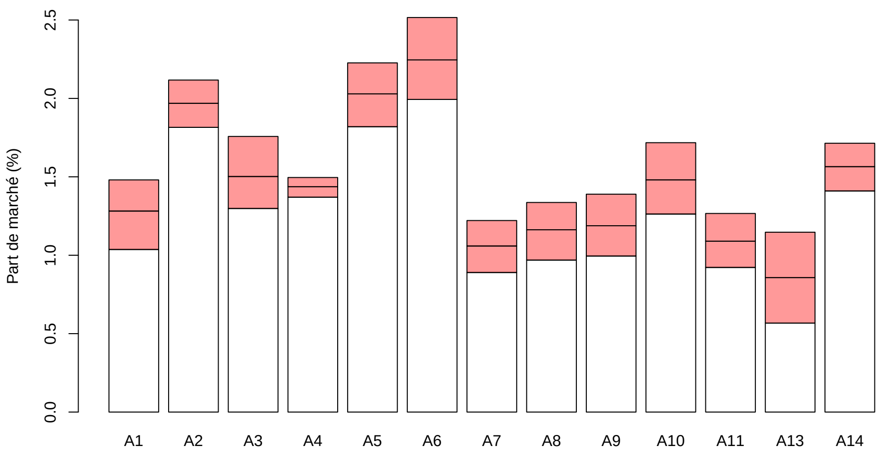


Happens for 33% of the prices. Market share 8.8% up to 22.5%

Distorting the Market: the Price Aggregator as a Market Player

Increase of loss ratio of A4 from 128% to 143%





Much higher volatility on other companies...

Take-Home Message

- different but consistent models, from different actuaries, different backgrounds, different softwares, different languages, different models
- hard to predict model's behavior in a competitive market from standard tools (lift curves)
- the choice of the insurance company has no big impact on market results, but has a big impact on how to use information on competitors