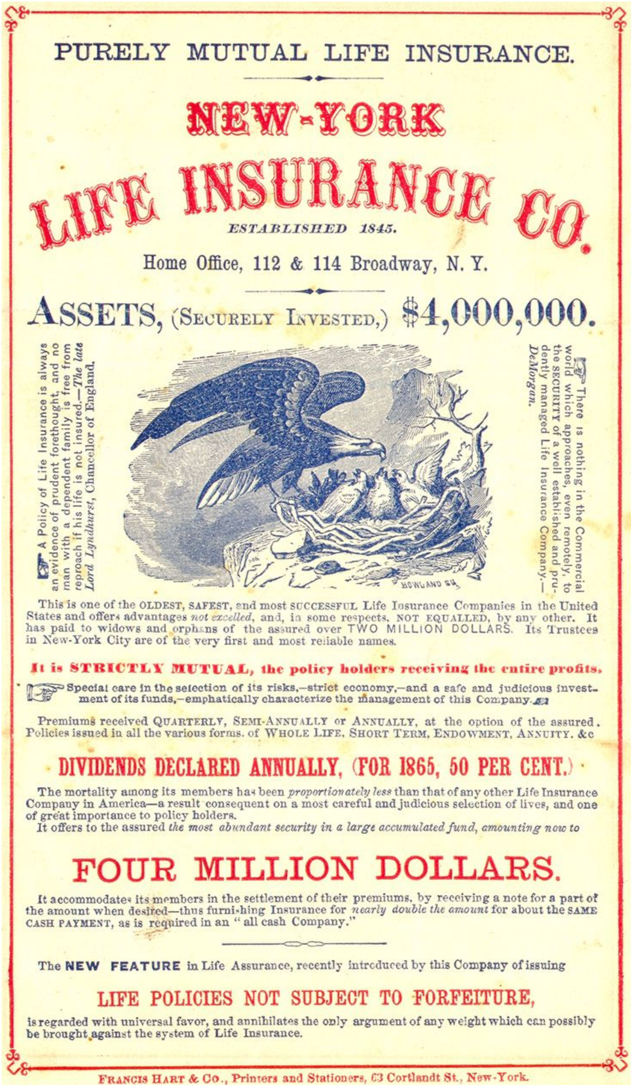
Competitive insurance markets in a context of big data and machine learning

A. Charpentier (Université de Rennes 1)

ESSEC, Paris

November 2017



PURELY MUTUAL LIFE INSURANCE.

NEW-YORK LIFE INSURANCE CO.

ESTABLISHED 1845.

Home Office, 112 & 114 Broadway, N. Y.

ASSETS, (Securely Invested,) $4,000,000.

☞ A Policy of Life Insurance is always an evidence of prudent forethought, and no man with a dependent family is free from reproach if his life is not insured.—*The late Lord Lyndhurst*, Chancellor of England.

☞ There is nothing in the Commercial world which approaches, even remotely, to the SECURITY of a well established and prudently managed Life Insurance Company.—*DeMorgan.*

This is one of the OLDEST, SAFEST, and most SUCCESSFUL Life Insurance Companies in the United States and offers advantages *not excelled*, and, in some respects, NOT EQUALLED, by any other. It has paid to widows and orphans of the assured over TWO MILLION DOLLARS. Its Trustees in New-York City are of the very first and most reliable names.

It is STRICTLY MUTUAL, the policy holders receiving the entire profits.

☞ Special care in the selection of its risks,—strict economy,—and a safe and judicious investment of its funds,—emphatically characterize the management of this Company. ☜

Premiums received QUARTERLY, SEMI-ANNUALLY or ANNUALLY, at the option of the assured. Policies issued in all the various forms. of WHOLE LIFE, SHORT TERM, ENDOWMENT, ANNUITY, &c

DIVIDENDS DECLARED ANNUALLY, (FOR 1865, 50 PER CENT.)

The mortality among its members has been *proportionately less* than that of any other Life Insurance Company in America—a result consequent on a most careful and judicious selection of lives, and one of great importance to policy holders.

It offers to the assured *the most abundant security in a large accumulated fund, amounting now to*

FOUR MILLION DOLLARS.

It accommodates its members in the settlement of their premiums, by receiving a note for a part of the amount when desired—thus furnishing Insurance for *nearly double the amount* for about the SAME CASH PAYMENT, as is required in an "all cash Company."

The **NEW FEATURE** in Life Assurance, recently introduced by this Company of issuing

LIFE POLICIES NOT SUBJECT TO FORFEITURE,

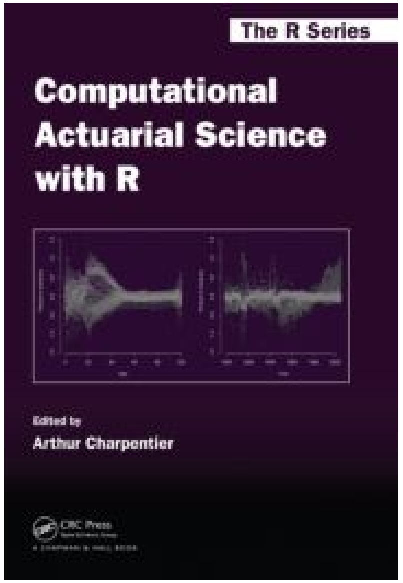
is regarded with universal favor, and annihilates the only argument of any weight which can possibly be brought against the system of Life Insurance.

Francis Hart & Co., Printers and Stationers, 63 Cortlandt St., New-York.

Brief Introduction

A. Charpentier (Université de Rennes 1)

Professor Economics Department, Université de Rennes 1
Director Data Science for Actuaries Program, Institute of Actuaries
(previously Actuarial Sciences, UQàM & ENSAE Paristech
actuary in Hong Kong, IT & Stats FFA)
PhD in Statistics (KU Leuven), Fellow of the Institute of Actuaries
MSc in Financial Mathematics (Paris Dauphine) & ENSAE
Research Chair :
ACTINFO (valorisation et nouveaux usages actuariels de l'information)
Editor of the freakonometrics.hypotheses.org's blog
Editor of Computational Actuarial Science, CRC
Author of Mathématiques de l'Assurance Non-Vie (2 vol.), Economica



Insurance vs. Credit

click to visualize the construction

Insurance Pricing in a Nutshell

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

Finance: risk neutral valuation $\pi = \mathbb{E}_{\mathbb{Q}}\big[S_1\big|\mathcal{F}_0\big] = \mathbb{E}_{\mathbb{Q}_0}\big[S_1\big]$, where $S_1 = \displaystyle\sum_{i=1}^{N_1} Y_i$

Insurance: risk sharing (pooling) $\pi = \mathbb{E}_{\mathbb{P}}\big[S_1\big]$

or, with segmentation / price differentiation $\pi(\omega) = \mathbb{E}_{\mathbb{P}}\big[S_1\big|\Omega = \omega\big]$ for some (unobservable?) risk factor Ω

imperfect information given some (observable) risk variables $\boldsymbol{X} = (X_1, \cdots, X_k)$
$\pi(\boldsymbol{x}) = \mathbb{E}_{\mathbb{P}}\big[S_1\big|\boldsymbol{X} = \boldsymbol{x}\big] = \mathbb{E}_{\mathbb{P}_{\boldsymbol{X}}}\big[S_1|\boldsymbol{x}\big]$

Insurance pricing is not only data driven, it is also essentially model driven (see Pricing Game)

Insurance Pricing in a Nutshell

Premium is $\pi = \mathbb{E}_{\mathbb{P}_{\boldsymbol{X}}}\big[S_1\big]$

It is datadriven (or portfolio driven) since $\mathbb{P}_{\boldsymbol{X}}$ is based on the portfolio.

click to visualize the construction

Insurance Pricing in a Nutshell

Premium is $\pi \approx \mathbb{E}\big[S_1|\boldsymbol{X}=\boldsymbol{x}\big] = \mathbb{E}\left[\sum_{i=1}^{N} Y_i \middle| \boldsymbol{X}=\boldsymbol{x}\right] = \mathbb{E}\left[N|\boldsymbol{X}=\boldsymbol{x}\right] \cdot \mathbb{E}\left[Y_i|\boldsymbol{X}=\boldsymbol{x}\right]$

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

- depends on the model used to approximate $\mathbb{E}\left[N|\boldsymbol{X}=\boldsymbol{x}\right]$ and $\mathbb{E}\left[Y_i|\boldsymbol{X}=\boldsymbol{x}\right]$
- depends on the choice of meta-parameters
- depends on variable selection / feature engineering

Try to avoid overfit

Risk Sharing in Insurance

Important formula $\mathbb{E}[S] = \mathbb{E}[\mathbb{E}[S|\boldsymbol{X}]]$ and its empirical version

$$\frac{1}{n}\sum_{i=1}^{n} S_i \sim \frac{1}{n}\sum_{i=1}^{n} \pi(\boldsymbol{X}_i) \quad \text{(as } n \to \infty\text{, from the law of large number)}$$

interpreted as on average what we pay (losses) is the sum of what we earn (premiums).

This is an ex-post statement, where premiums were calculated ex-ante.

Risk Transfert without Segmentation

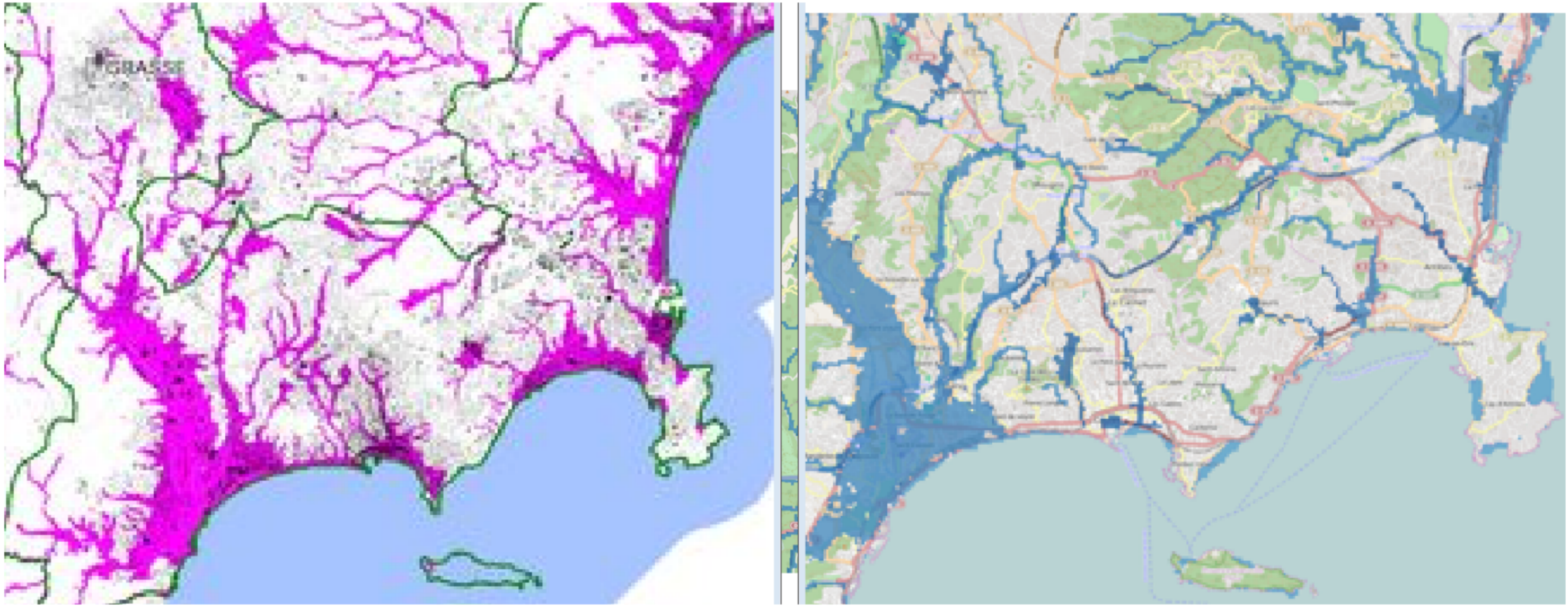
	Insured	Insurer
Loss	$\mathbb{E}[S]$	$S - \mathbb{E}[S]$
Average Loss	$\mathbb{E}[S]$	0
Variance	0	$\text{Var}[S]$

All the risk - $\text{Var}[S]$ - is kept by the insurance company.

Remark: all those interpretation are discussed in Denuit & Charpentier (2004).

Insurance, Risk Pooling and Solidarity

“La Nation proclame la solidarité et l’égalité de tous les Français devant les charges qui résultent des calamités nationales” (alinéa 12, préambule de la Constitution du 27 octobre 1946)



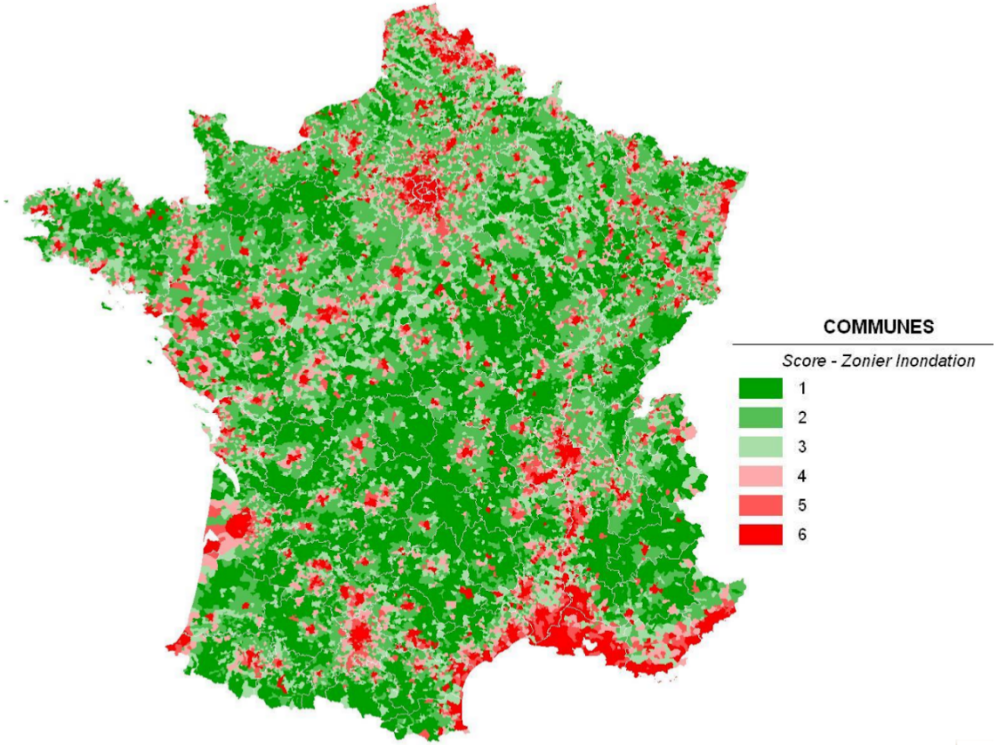
31 zones TRI (Territoires à Risques d’Inondation) on the left, and flooded areas.

Insurance, Risk Pooling and Solidarity

Here is a map with a risk score - $\{1, 2, \cdots, 6\}$ scale

One can look at "Lorenz curve"

	South	Other	Total
% portfolio	11%	89%	100%
% claims	51%	49%	100%
Premium	463	55	100



Risk Transfert with Segmentation and Perfect Information

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

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

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

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

Risk Transfert with 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

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

Risk Transfert with Segmentation and Imperfect Information

With imperfect information, we have the popular risk decomposition

$$\begin{aligned}
\text{Var}[S] &= \mathbb{E}\Big[\text{Var}[S|\boldsymbol{X}]\Big] + \text{Var}\Big[\mathbb{E}[S|\boldsymbol{X}]\Big] \\
&= \underbrace{\underbrace{\mathbb{E}\Big[\text{Var}[S|\boldsymbol{\Omega}]\Big]}_{\text{pooling}} + \underbrace{\mathbb{E}\Big[\text{Var}\Big[\mathbb{E}[S|\boldsymbol{\Omega}]\Big|\boldsymbol{X}\Big]\Big]}_{\text{solidarity}}}_{\to\text{insurer}} \\
&\quad + \underbrace{\text{Var}\Big[\mathbb{E}[S|\boldsymbol{X}]\Big]}_{\to\text{ insured}}.
\end{aligned}$$

More and more price differentiation ?

Consider $\pi_1 = \mathbb{E}\big[S_1\big]$ and $\pi_2(\boldsymbol{x}) = \mathbb{E}\big[S_1|\boldsymbol{X} = \boldsymbol{x}\big]$

Observe that $\mathbb{E}\big[\pi(\boldsymbol{X})\big] = \sum_{\boldsymbol{x}\in\mathcal{X}} \pi(\boldsymbol{x}) \cdot \mathbb{P}[\boldsymbol{x}]$

$$= \sum_{\boldsymbol{x}\in\mathcal{X}_1} \pi(\boldsymbol{x}) \cdot \mathbb{P}[\boldsymbol{x}] + \sum_{\boldsymbol{x}\in\mathcal{X}_2} \pi(\boldsymbol{x}) \cdot \mathbb{P}[\boldsymbol{x}]$$

- Insured with $\boldsymbol{x} \in \mathcal{X}_1$: choose Ins1
- Insured with $\boldsymbol{x} \in \mathcal{X}_2$: choose Ins2

Ins1: $\sum_{\boldsymbol{x}\in\mathcal{X}_1} \pi_1(\boldsymbol{x}) \cdot \mathbb{P}[\boldsymbol{x}] \neq \mathbb{E}[S|\boldsymbol{X} \in \mathcal{X}_1]$

Ins2: $\sum_{\boldsymbol{x}\in\mathcal{X}_2} \pi_2(\boldsymbol{x}) \cdot \mathbb{P}[\boldsymbol{x}] = \mathbb{E}[S|\boldsymbol{X} \in \mathcal{X}_2]$

Price Differentiation, a Toy Example

Claims frequency Y (average cost = 1,000)

		X_1			
		Young	Experienced	Senior	Total
X_2	Town	12% (500)	9% (2,000)	9% (500)	9.5% (3,000)
	Outside	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)

from C., Denuit & Élie (2015)

Price Differentiation, a Toy Example

	Y-T	Y-O	E-T	E-O	S-T	S-O
	(500)	(500)	(2,000)	(1,000)	(500)	(500)
none	82.3	82.3	82.3	82.3	82.3	82.3
$X_1 \times X_2$	120	80	90	66.7	90	40
market	82.3	80	82.3	66.7	82.3	40

none	82.3	82.3	82.3	82.3	82.3	82.3
X_1	100	100	82.2	82.2	65	65
X_2	95	63.3	95	63.3	95	63.3
$X_1 \times X_2$	120	80	90	66.7	90	40
market	82.3	63.3	82.2	63.3	65	40

Price Differentiation, a Toy Example

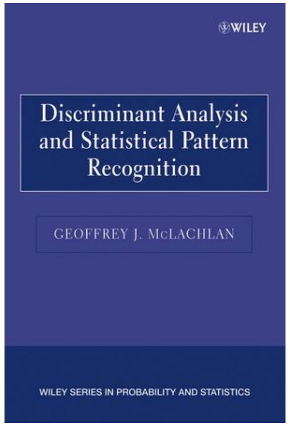
	premium	losses	loss		99.5%	Market
			ratio		quantile	Share
none	247	285	115.4%	(±8.9%)		66.1%
$X_1 \times X_2$	126.67	126.67	100.0%	(±10.4%)		33.9%
market	373.67	411.67	110.2%	(±5.1%)		
none	41.17	60	145.7%	(±34.6%)	189%	11.6%
X_1	196.94	225	114.2%	(±11.8%)	140%	55.8%
X_2	95	106.67	112.3%	(±15.1%)	134%	26.9%
$X_1 \times X_2$	20	20	100.0%	(±41.9%)	160%	5.7%
market	353.10	411.67	116.6%	(±5.3%)	130%	

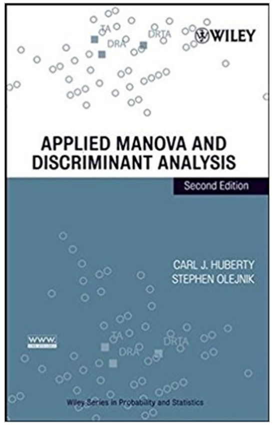
Model Comparison (and Inequalities)

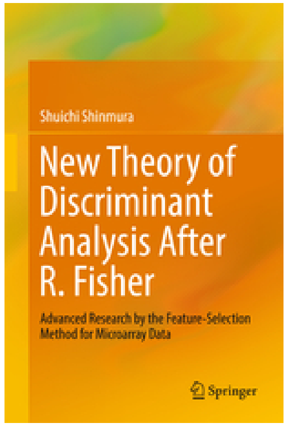
Use of statistical techniques to get price differentiation see discriminant analysis, Fisher (1936)

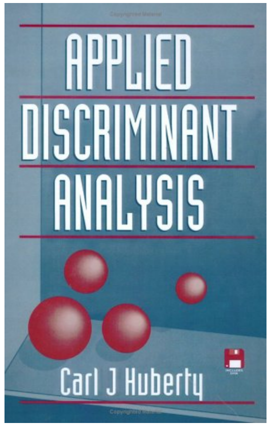
“In human social affairs, discrimination is treatment or consideration of, or making a distinction in favor of or against, a person based on the group, class, or category to which the person is perceived to belong rather than on individual attributes” (wikipedia)

For legal perspective, see Canadian Human Rights Act

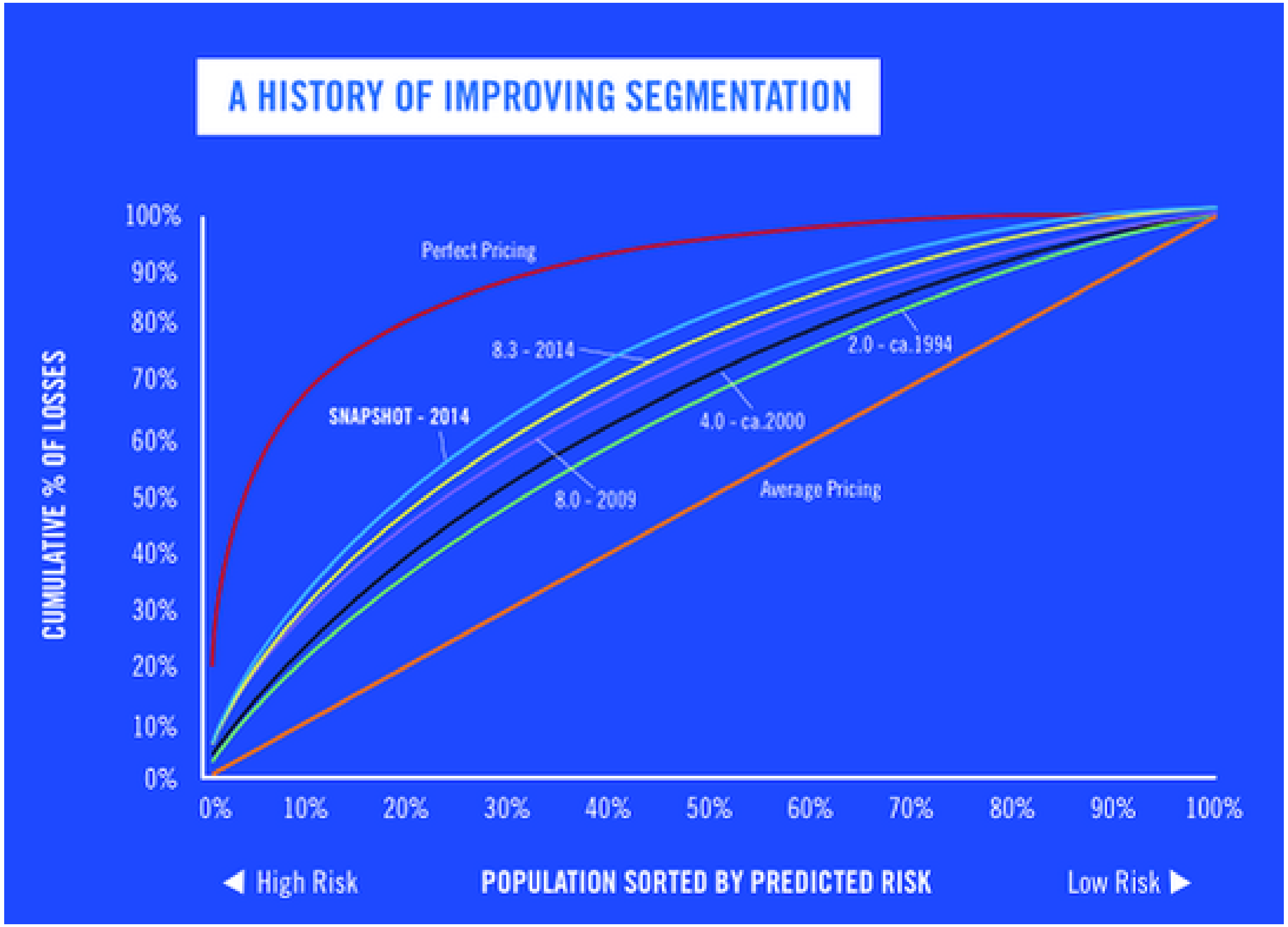

WILEY
Discriminant Analysis and Statistical Pattern Recognition
GEOFFREY J. McLACHLAN
WILEY SERIES IN PROBABILITY AND STATISTICS


WILEY
APPLIED MANOVA AND DISCRIMINANT ANALYSIS
Second Edition
CARL J. HUBERTY
STEPHEN OLEJNIK
Wiley Series in Probability and Statistics


Shuichi Shinmura
New Theory of Discriminant Analysis After R. Fisher
Advanced Research by the Feature-Selection Method for Microarray Data
Springer



Model Comparison and Lorenz curves

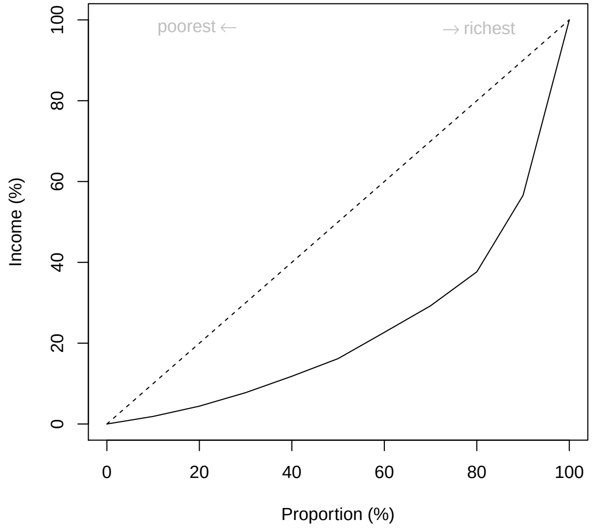


Source: Progressive Insurance

Model Comparison and Lorenz curves

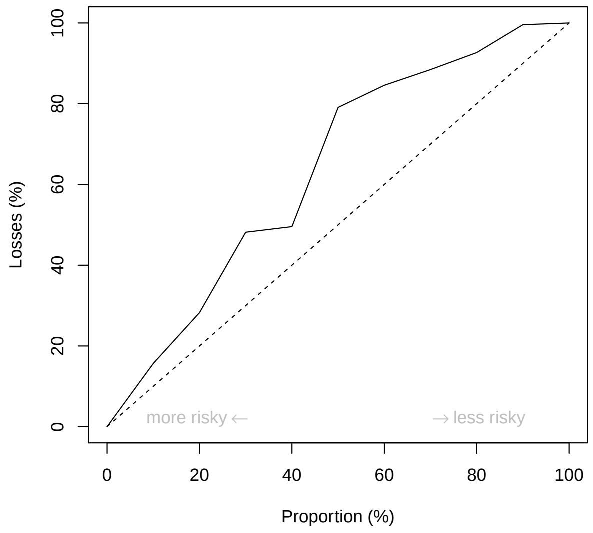
Consider an ordered sample $\{y_1, \cdots, y_n\}$ of incomes, with $y_1 \leq y_2 \leq \cdots \leq y_n$, then Lorenz curve is

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



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

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



Model Comparison for Life Insurance Models

Consider the case of a death insurance contract, that pays 1 if the insured deceased within the year.

$\pi(x) = \mathbb{E}\big[T_x \le t + 1 | T_x > t\big]$

— No price discrimination $\pi = \mathbb{E}[\pi(X)]$

— Perfect discrimination $\pi(x)$

— Imperfect discrimination

$\pi_- = \mathbb{E}[\pi(X)|X < s]$ and $\pi_+ = \mathbb{E}[\pi(X)|X > s]$

click to visualize the construction

From Econometric to 'Machine Learning' Techniques

In a competitive market, insurers can use different sets of variables and different models, e.g. GLMs, $N_t|\boldsymbol{X} \sim \mathcal{P}(\lambda_{\boldsymbol{X}} \cdot t)$ and $Y|\boldsymbol{X} \sim \mathcal{G}(\mu_{\boldsymbol{X}}, \varphi)$

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

that can be extended to GAMs,

$$\widehat{\pi}_j(\boldsymbol{x}) = \underbrace{\exp\left(\sum_{k=1}^{d} \widehat{s}_k(x_k)\right)}_{\text{Poisson } \mathcal{P}(\lambda_{\boldsymbol{x}})} \cdot \underbrace{\exp\left(\sum_{k=1}^{d} \widehat{t}_k(x_k)\right)}_{\text{Gamma } \mathcal{G}(\mu_{\boldsymbol{X}},\varphi)}$$

or some Tweedie model on S_t (compound Poisson, see Tweedie (1984)) conditional on $\boldsymbol{X}$ (see C. & Denuit (2005) or Kaas *et al.* (2008)) or any other statistical model

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

From Econometric to 'Machine Learning' Techniques

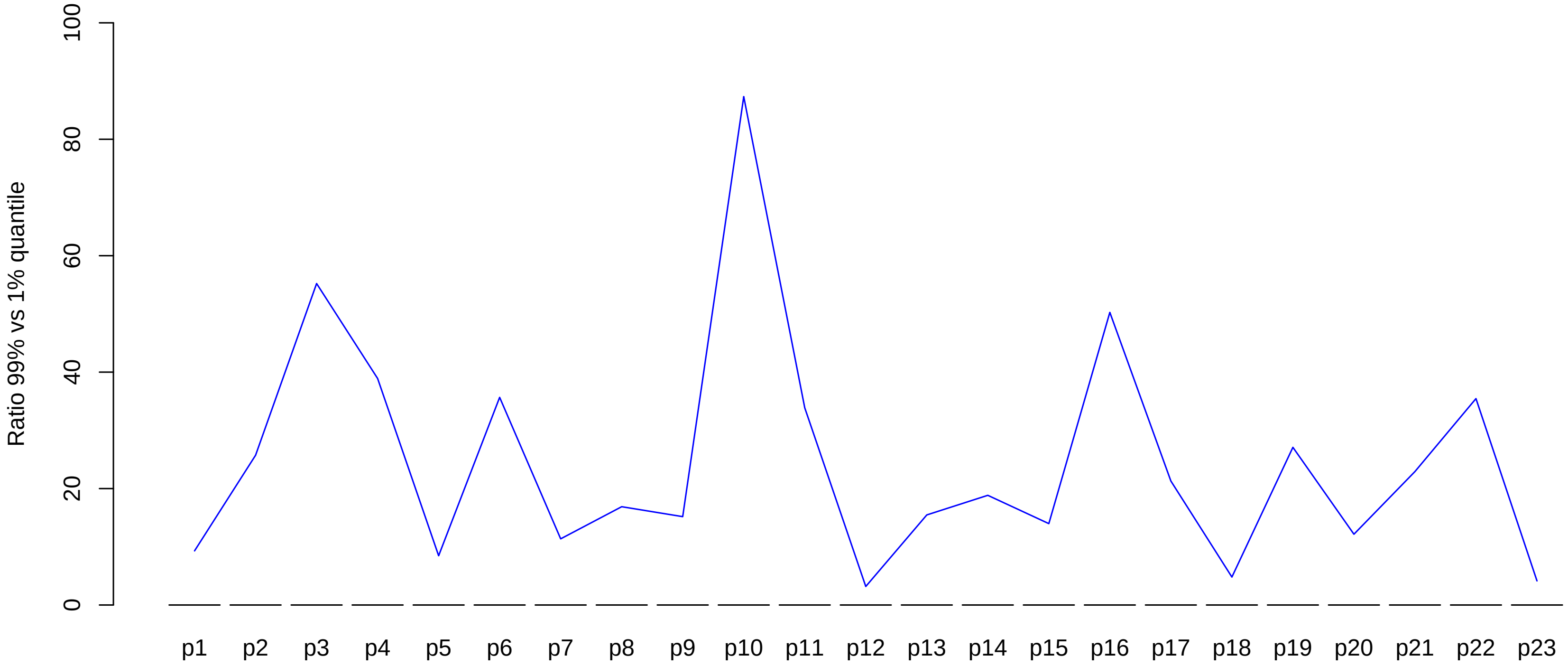
For some loss function $\ell : \mathbb{R}^2 \to \mathbb{R}_+$ (usually an L_2 based loss, $\ell(s,y) = (s-y)^2$ since $\text{argmin}\{\mathbb{E}[\ell(S,m)],\ m \in \mathbb{R}\}$ is $\mathbb{E}(S)$, interpreted as the pure premium).

For instance, consider regression trees, forests, neural networks, or boosting based techniques to approximate $\pi(\boldsymbol{x})$, and various techniques for variable selection, such as LASSO (see Hastie *et al.* (2009) or C., Flachaire & Ly (2017) for a description and a discussion).

With d competitors, each insured i has to choose among d premiums, $\boldsymbol{\pi}_i = \big(\widehat{\pi}_1(\boldsymbol{x}_i), \cdots, \widehat{\pi}_d(\boldsymbol{x}_i)\big) \in \mathbb{R}_+^d$.

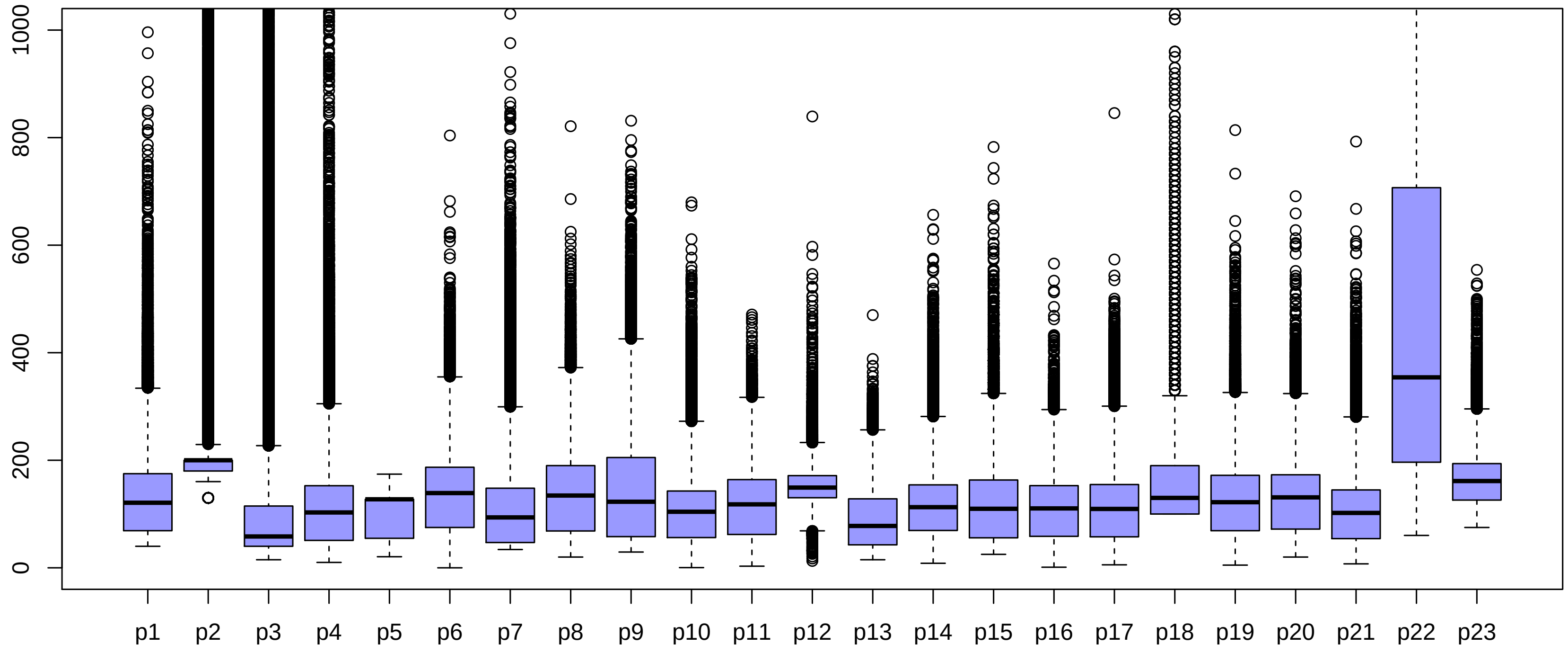
Insurance and Risk Segmentation: Pricing Game

Insurance and Risk Segmentation: Pricing Game

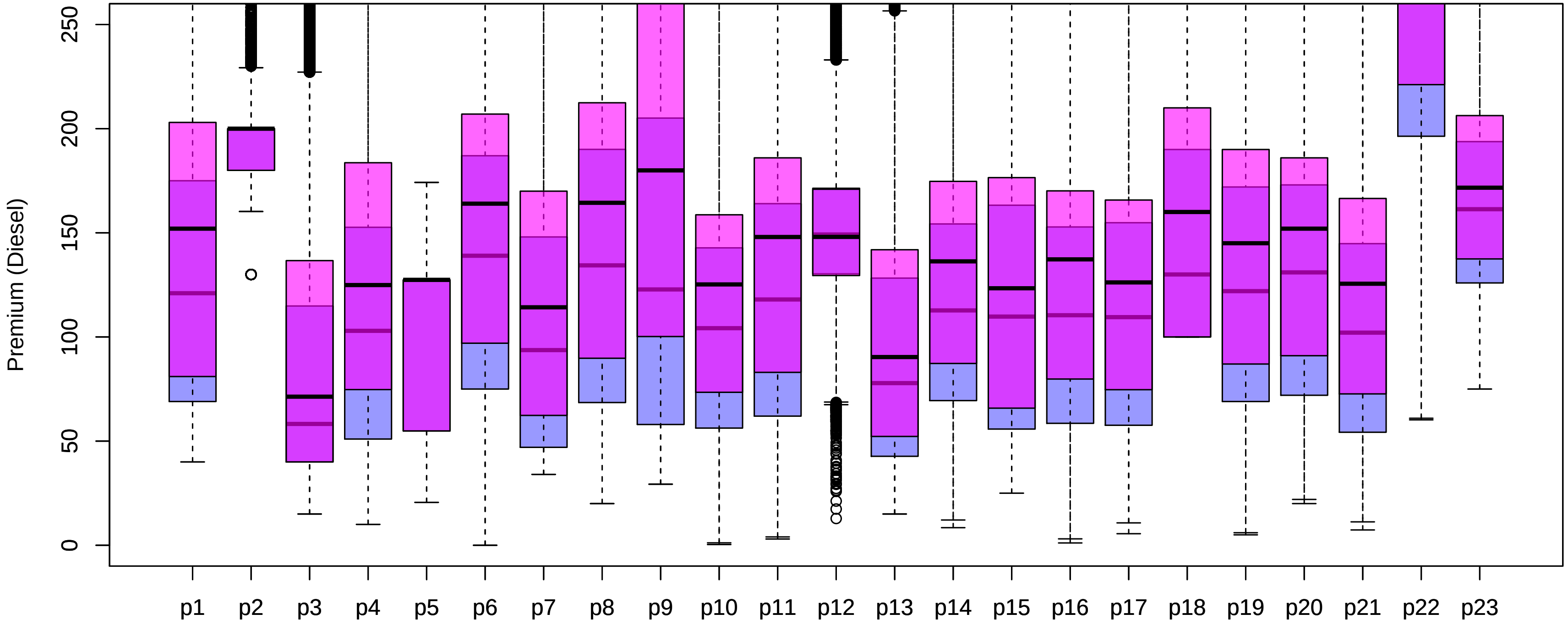


Insurance Ratemaking Before Competition

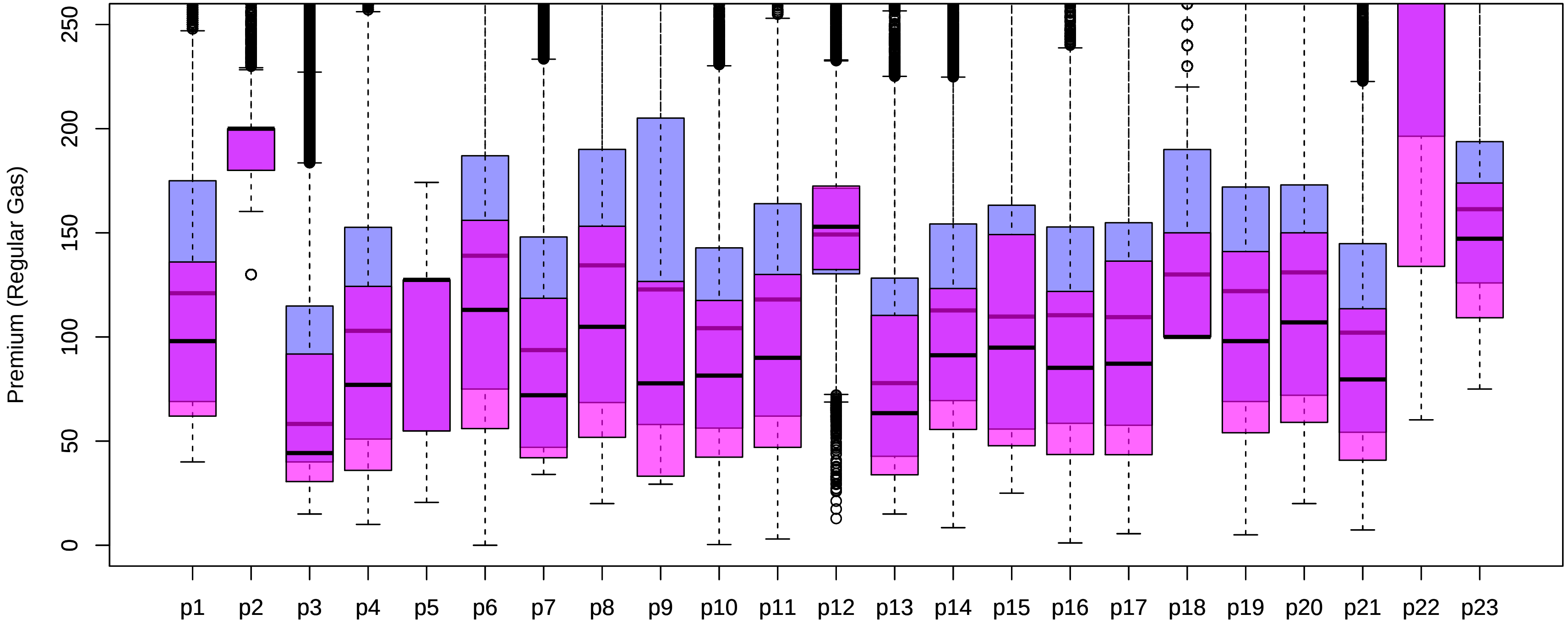
Insurance Ratemaking Before Competition



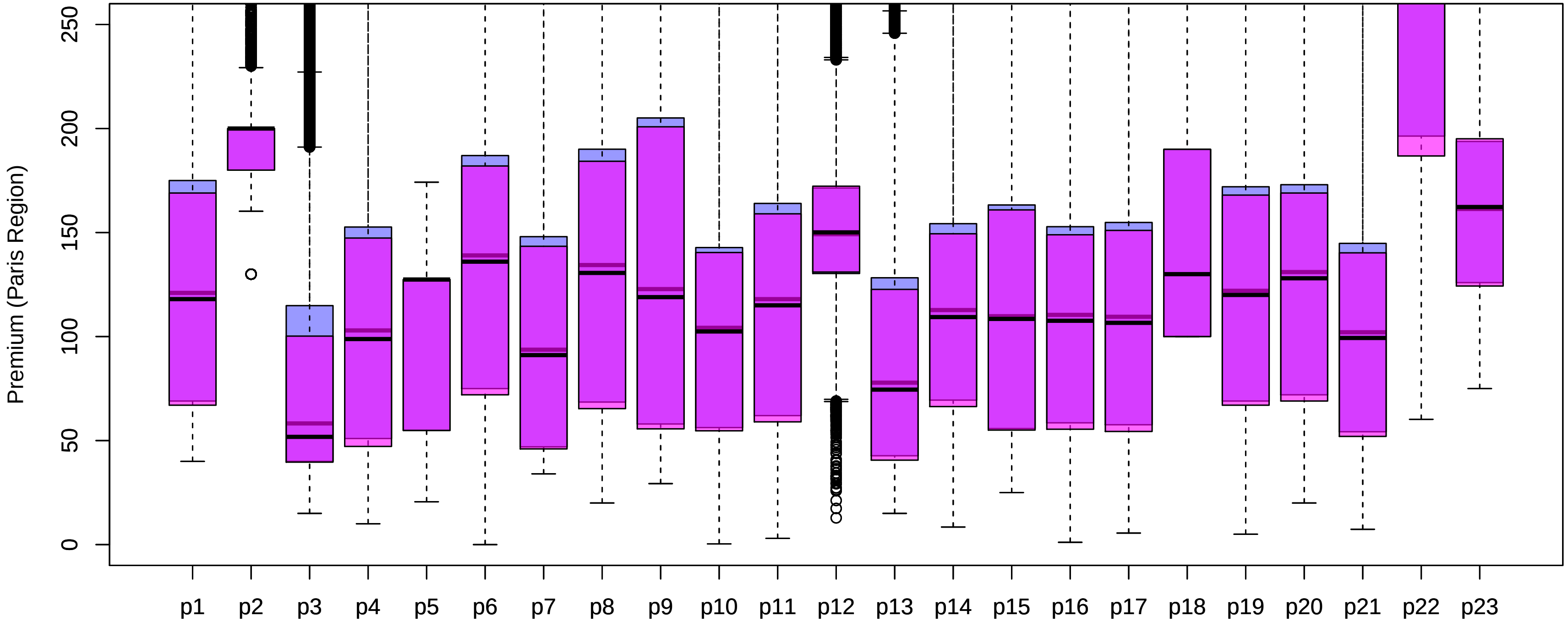
Insurance Ratemaking Before Competition Gas Type Diesel



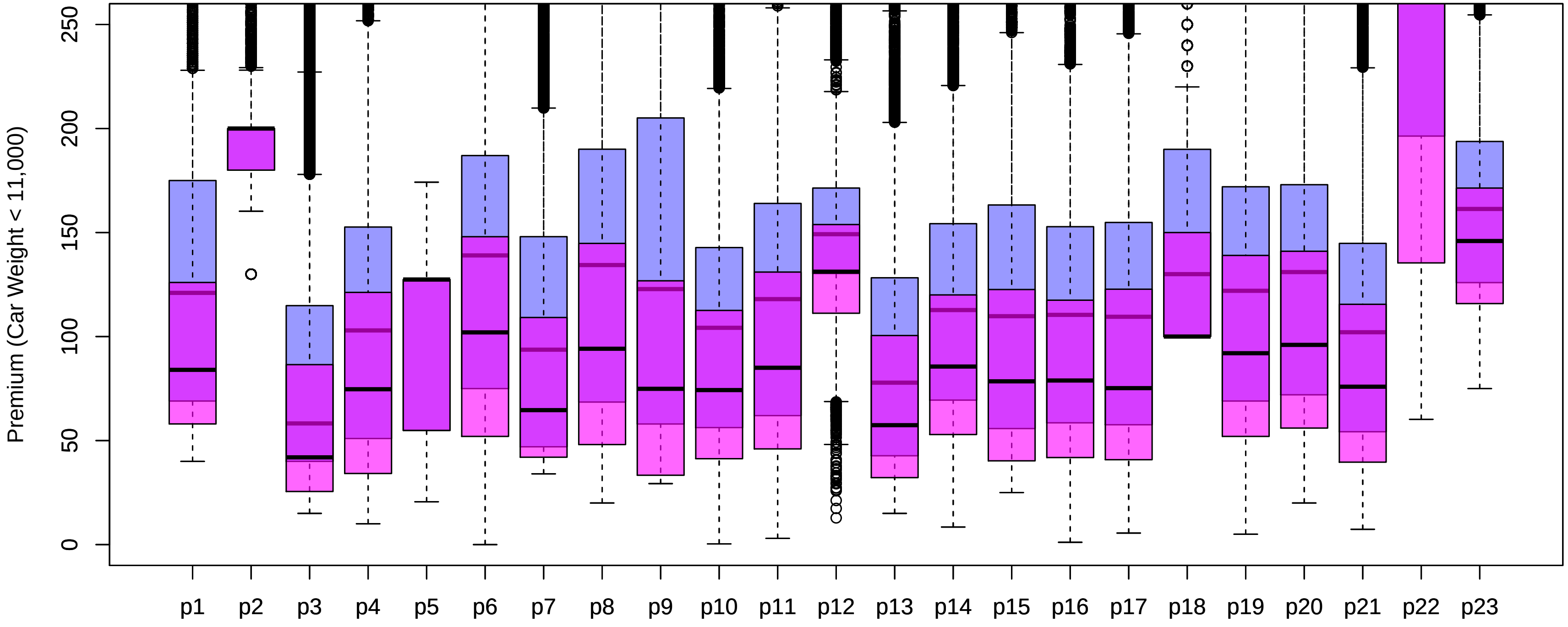
Insurance Ratemaking Before Competition Gas Type Regular



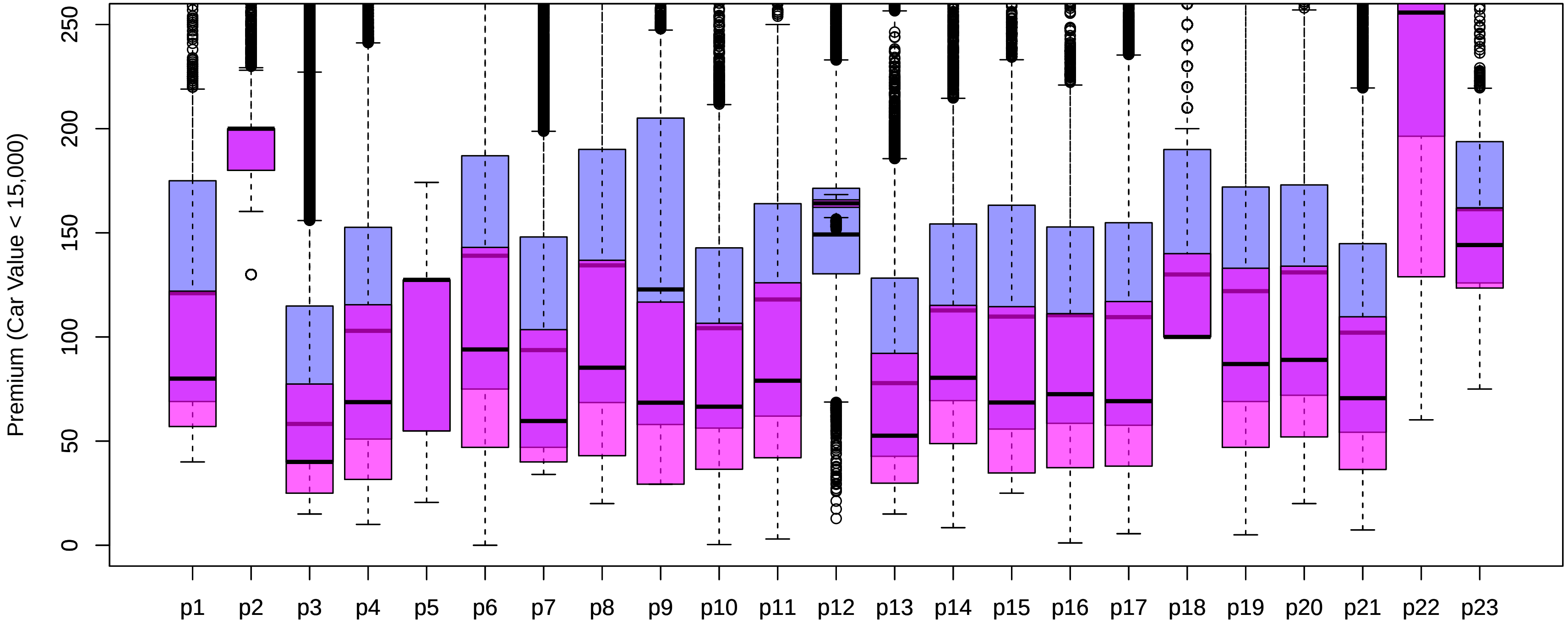
Insurance Ratemaking Before Competition Paris Region



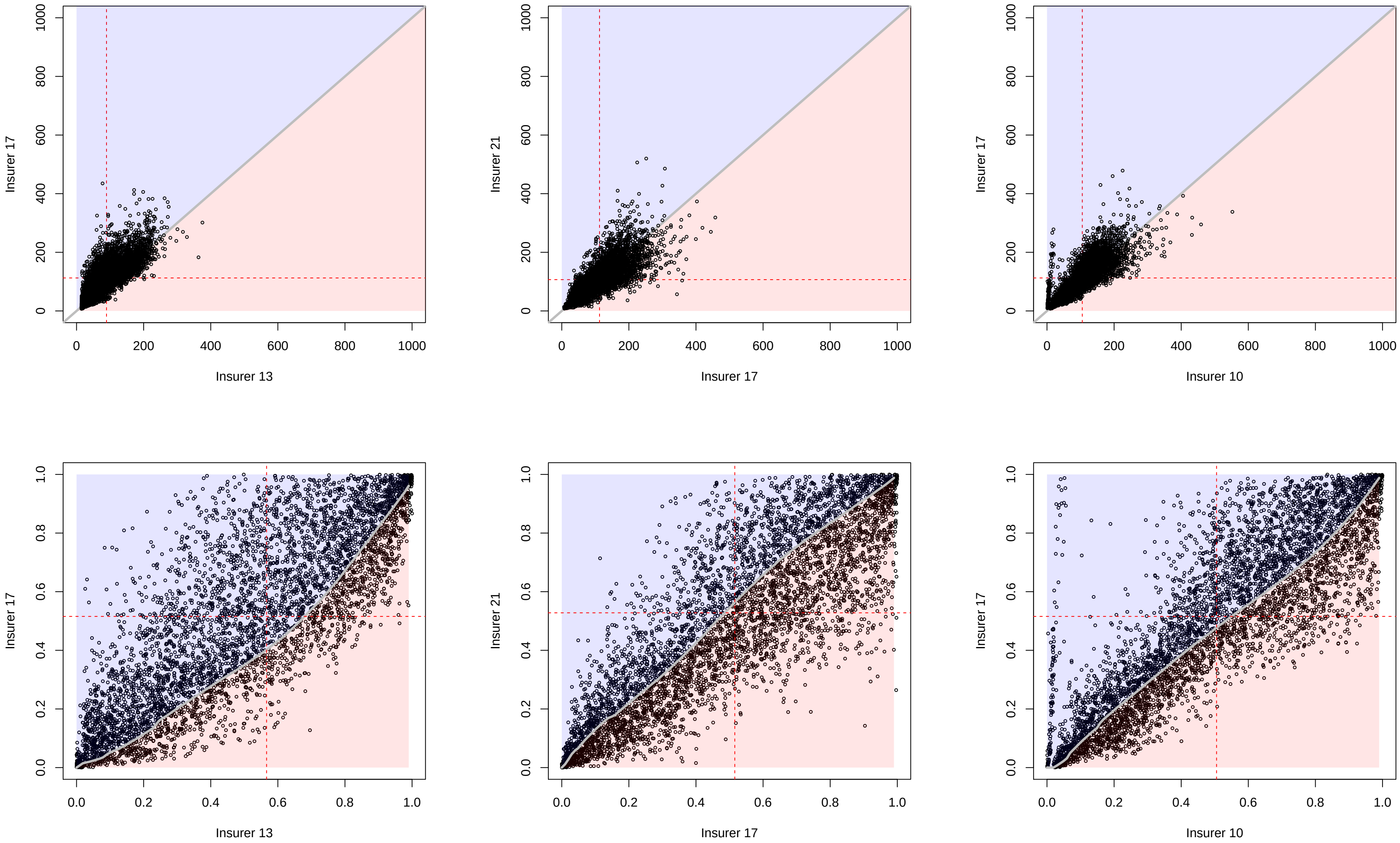
Insurance Ratemaking Before Competition Car Weight



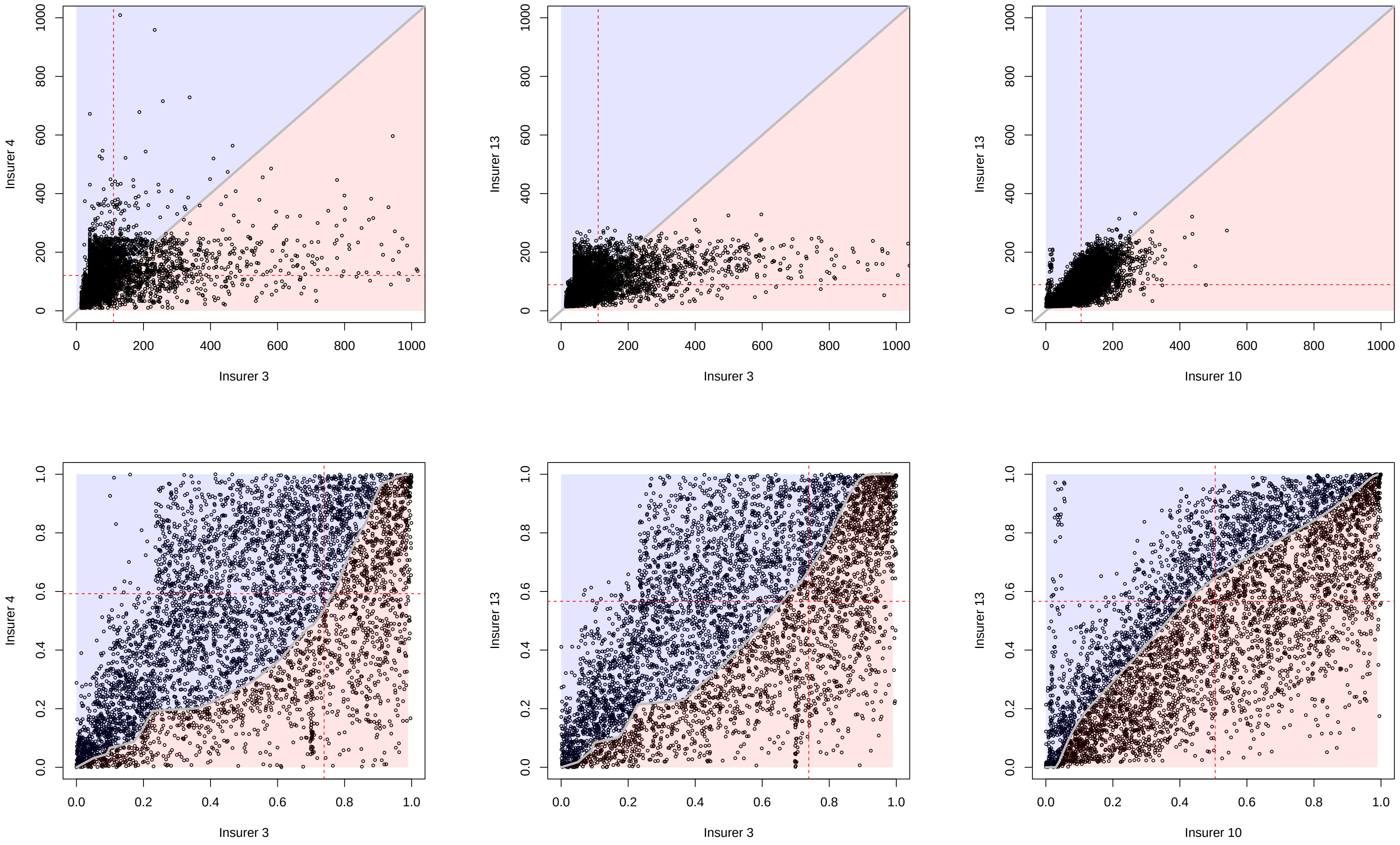
Insurance Ratemaking Before Competition Car Value



Insurance Ratemaking Competition : Comonotonicity?



Insurance Ratemaking Competition : Comonotonicity?



Insurance Ratemaking Competition

We need a **Decision Rule** to select premium chosen by insured i

	Ins1	Ins2	Ins3	Ins4	Ins5	Ins6
	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

Insurance Ratemaking Competition

Basic ‘**rational rule**’ $\pi_i = \min\{\widehat{\pi}_1(\boldsymbol{x}_i),\cdots,\widehat{\pi}_d(\boldsymbol{x}_i)\} = \widehat{\pi}_{1:d}(\boldsymbol{x}_i)$

	Ins1	Ins2	Ins3	Ins4	Ins5	Ins6
	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

Insurance Ratemaking Competition

A more **realistic rule** $\pi_i \in \{\widehat{\pi}_{1:d}(\boldsymbol{x}_i), \widehat{\pi}_{2:d}(\boldsymbol{x}_i), \widehat{\pi}_{3:d}(\boldsymbol{x}_i)\}$

	Ins1	Ins2	Ins3	Ins4	Ins5	Ins6
	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

A Game with Rules... but no Goal

Two datasets : a training one, and a pricing one (without the losses in the later)

Step 1 : provide premiums to all contracts in the pricing dataset

Step 2 : allocate insured among players

Season 1 13 players

Season 2 14 players

Step 3 [season 2] : provide additional information (premiums of competitors)

Season 3 23 players (3 markets, 8+8+7)

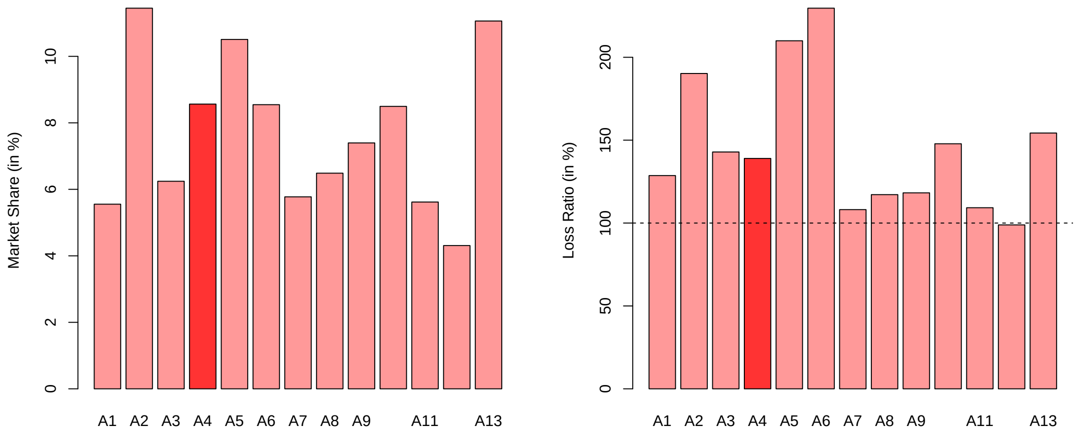
Step 3-6 [season 3] : dynamics, 4 years

Pricing Game in 2015

Insurer 4

GLM for frequency and standard cost (large claimes were removed, above 15k), Interaction Age and Gender

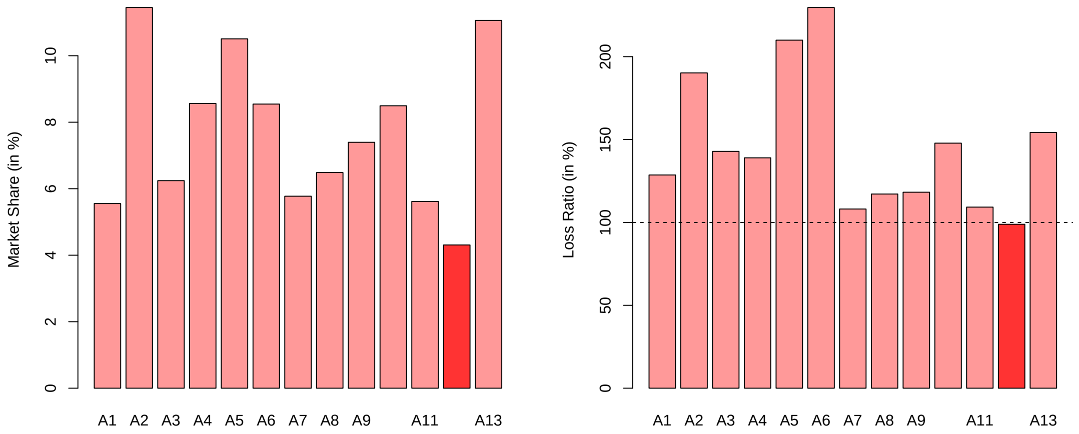
Actuary working for a *mutuelle* company



Insurer 11

Use of two XGBoost models (bodily injury and material), with correction for negative premiums

Actuary working for a private insurance company

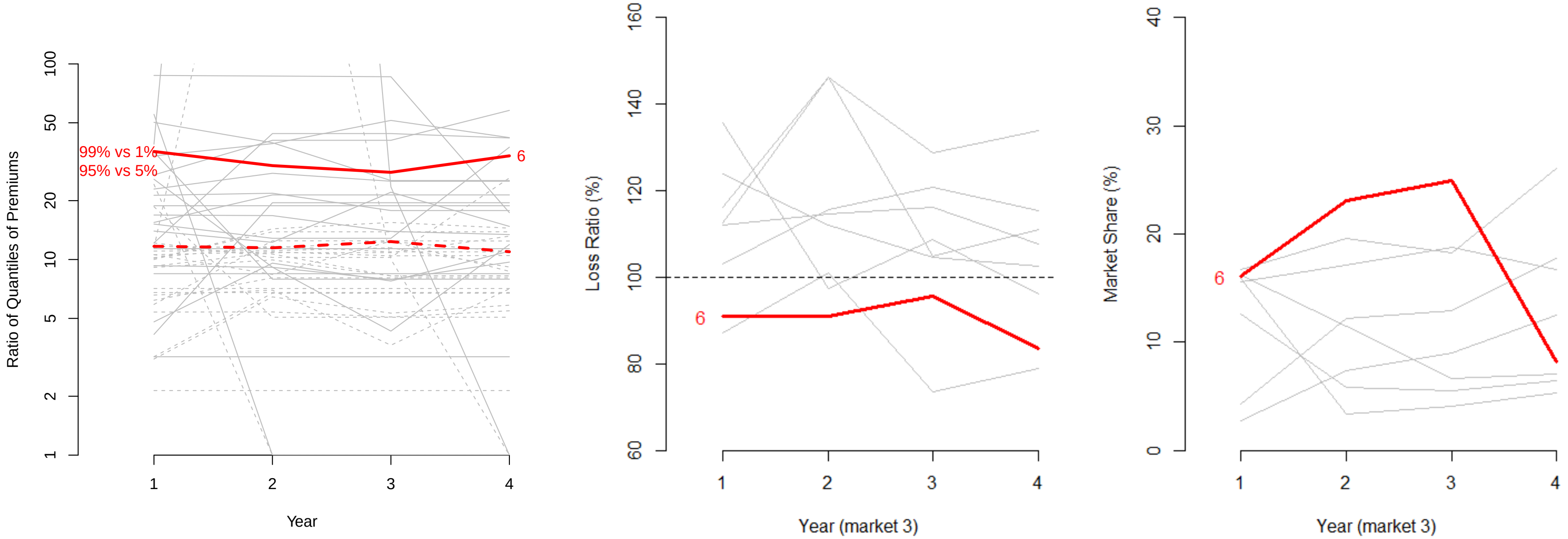


Pricing Game in 2017

Insurer 6 (market 3)

Team of two actuaries (degrees in Engineering and Physics), in Vancouver, Canada. Used GLMs (Tweedie), no territorial classification, no use of information about other competitors

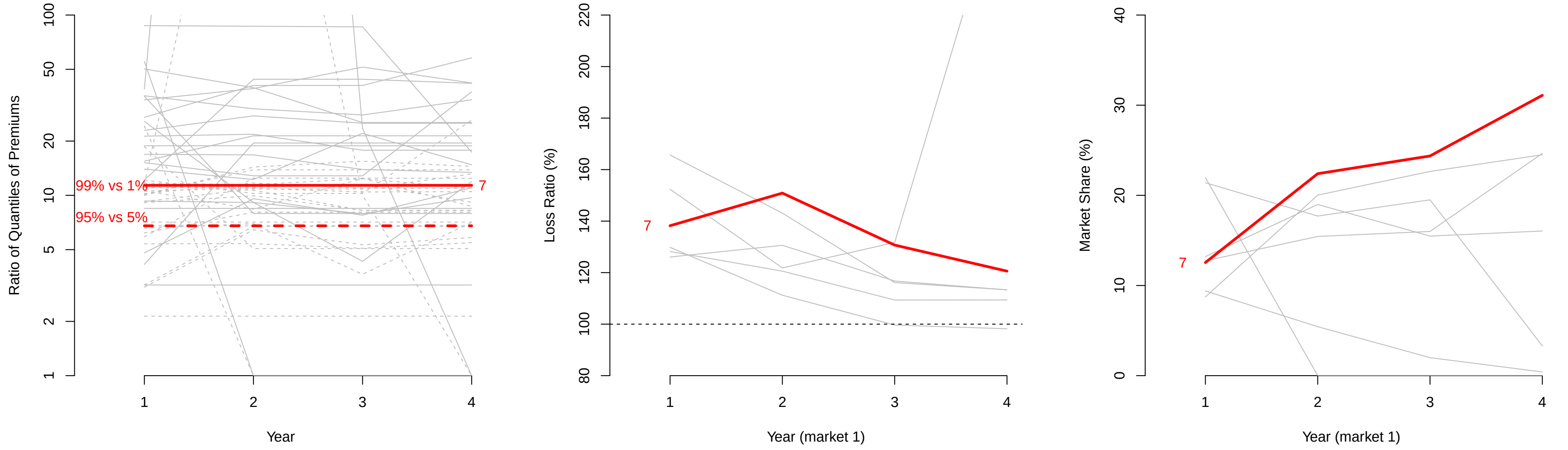
"*Segments with high market share and low loss ratios were also given some premium increase*"



Pricing Game in 2017

Insurer 7 (market 1)

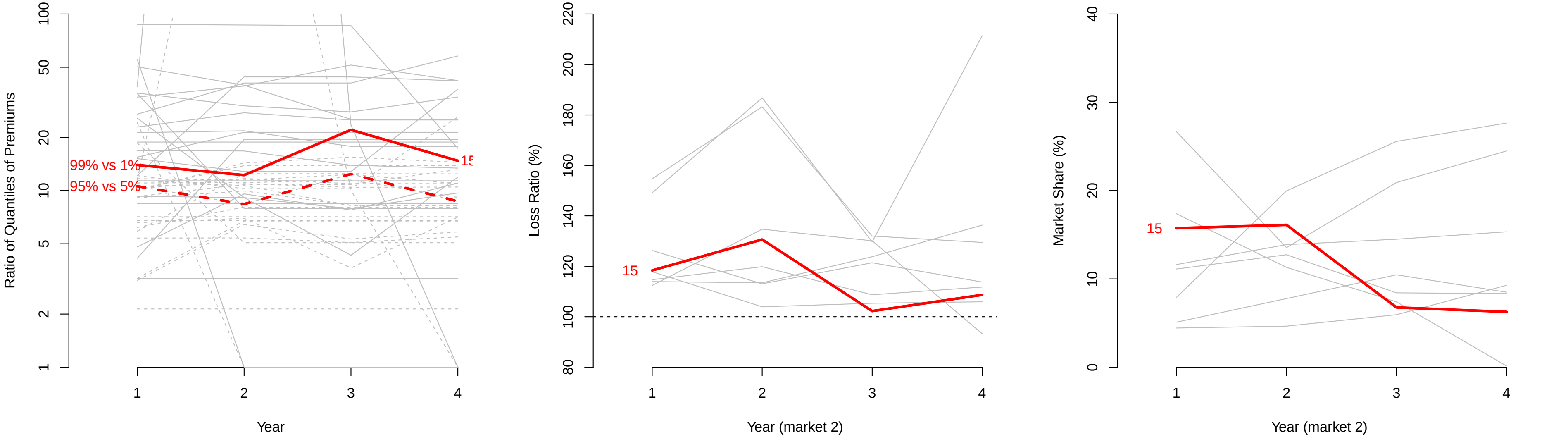
Actuary in France, used random forest for variable selection, and GLMs



Pricing Game in 2017

Insurer 15 (market 2)

Actuary,working as a consultant, Margin Method with iterations, MS Access & MS Excel

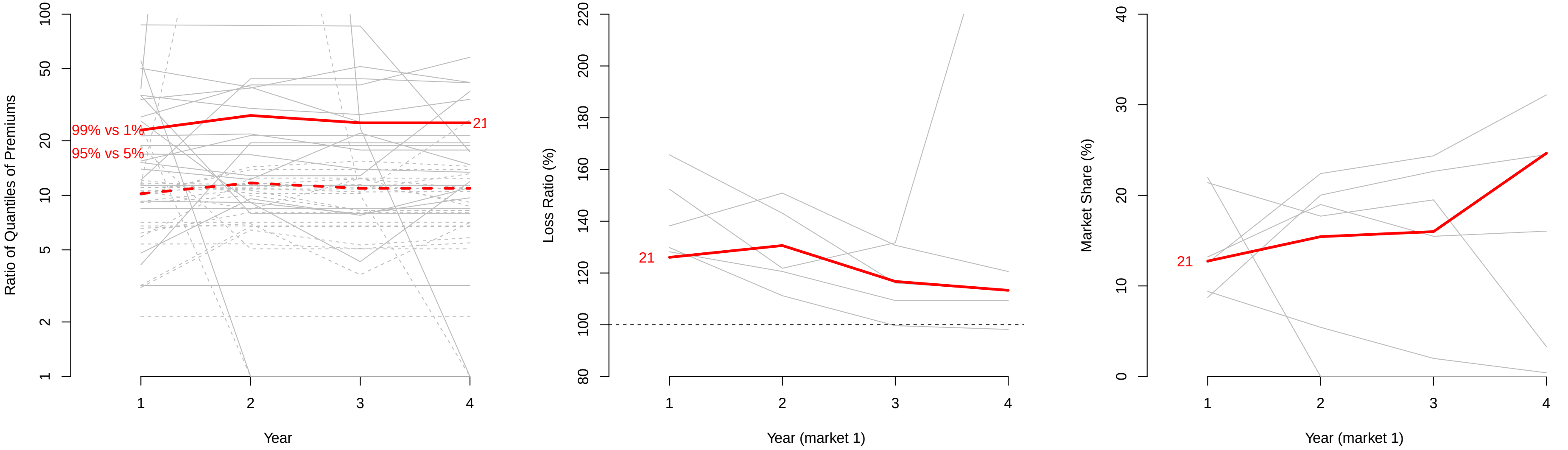


Pricing Game in 2017

Insurer 21 (market 1)

Actuary, working as a consultant, used GLMs, with variable selection using LARS and LASSO

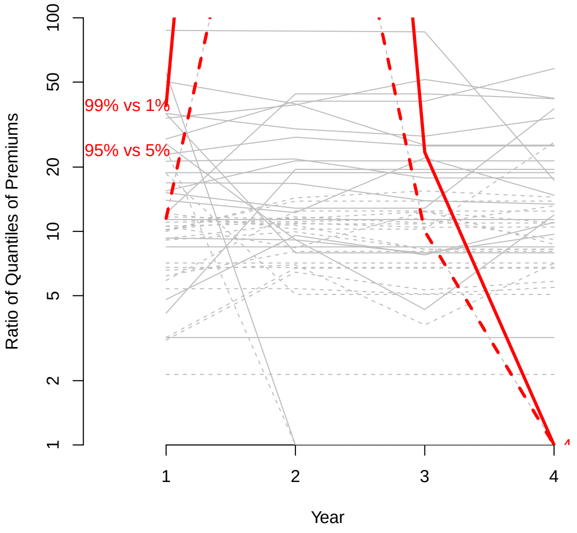
Iterative learning algorithm (codes available on github)

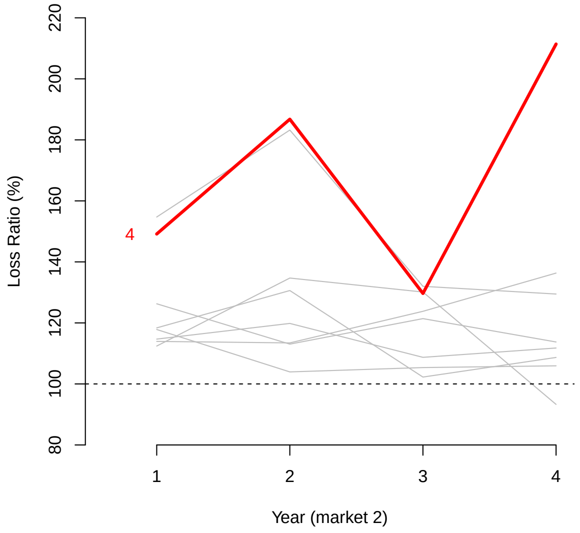


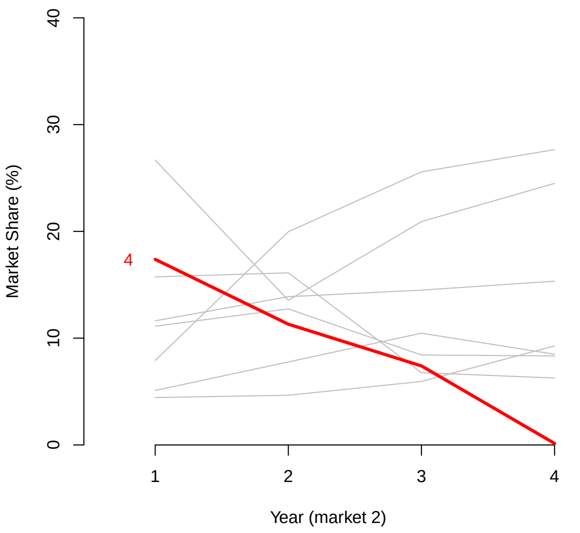
Pricing Game in 2017

Insurer 4 (market 2)

Actuary, working as a consultat,used XGBOOST, used GLMs for year 3.





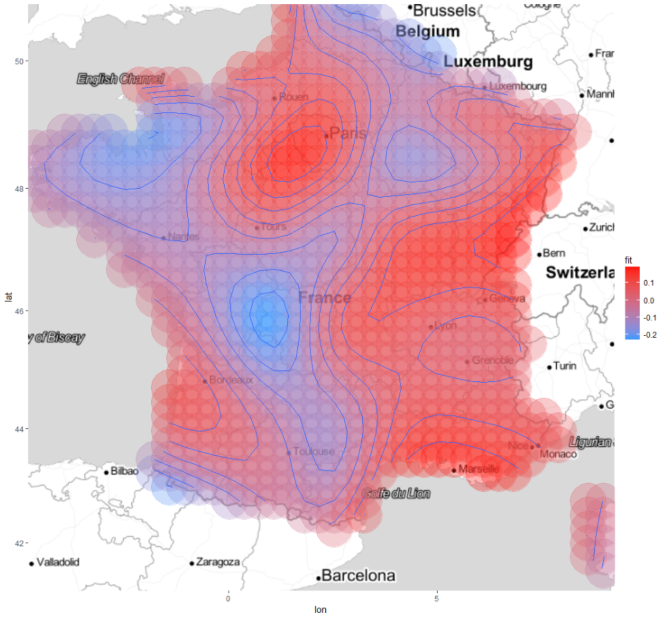


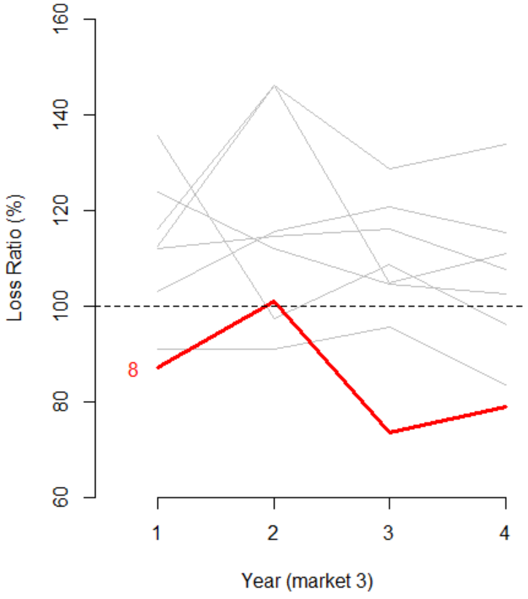
Pricing Game in 2017

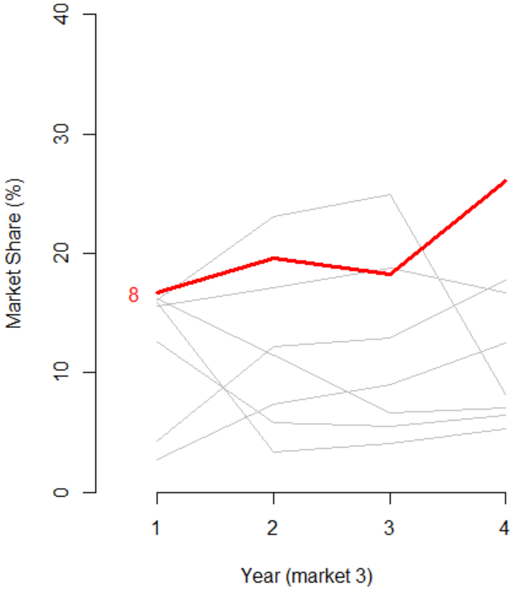
Insurer 8 (market 3)

Mathematician, working on Solvency II sofware in Austria

Generalized Additive Models with spatial variable







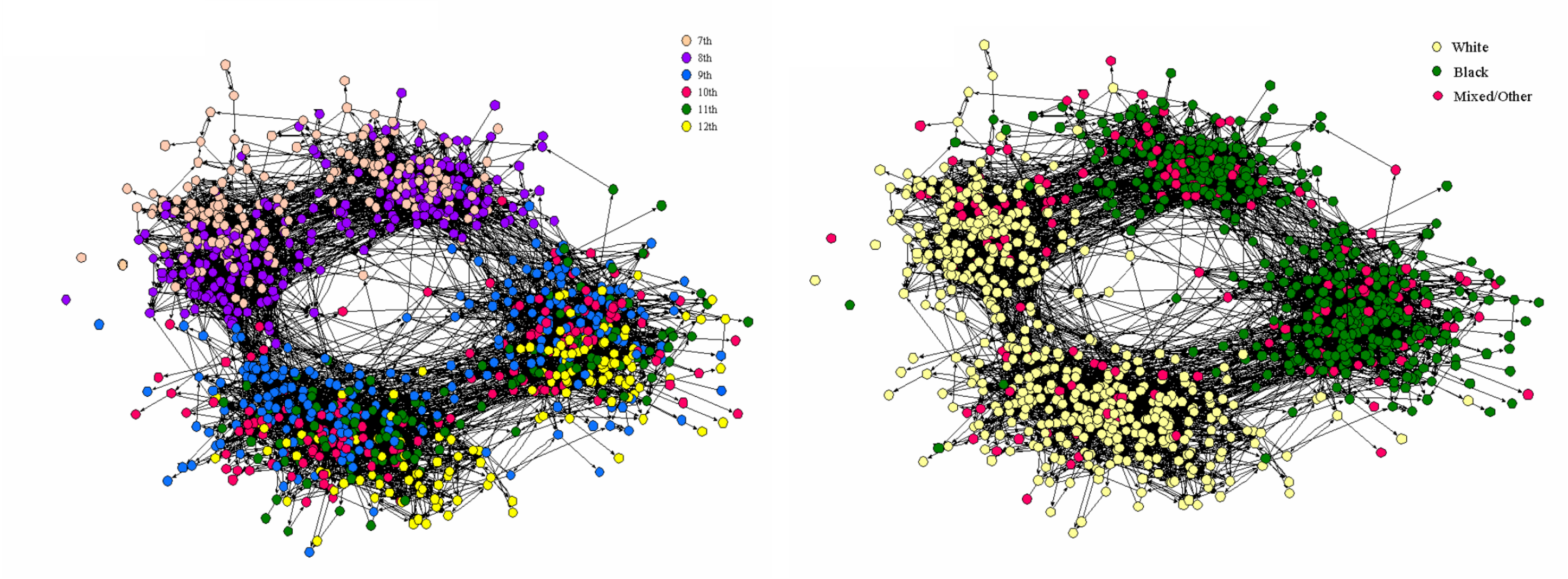
Cluster, Segmentation and (Social) Networks

Social networks could be used to get additional information about insured people...

Why not using social networks to create (more) solidarity ?

Cluster, Segmentation and (Social) Networks

Homophily is the tendency of individuals to associate and bond with similar others, “birds of a feather flock together”



from Moody (2001) Race, School Integration and Friendship Segregation in America

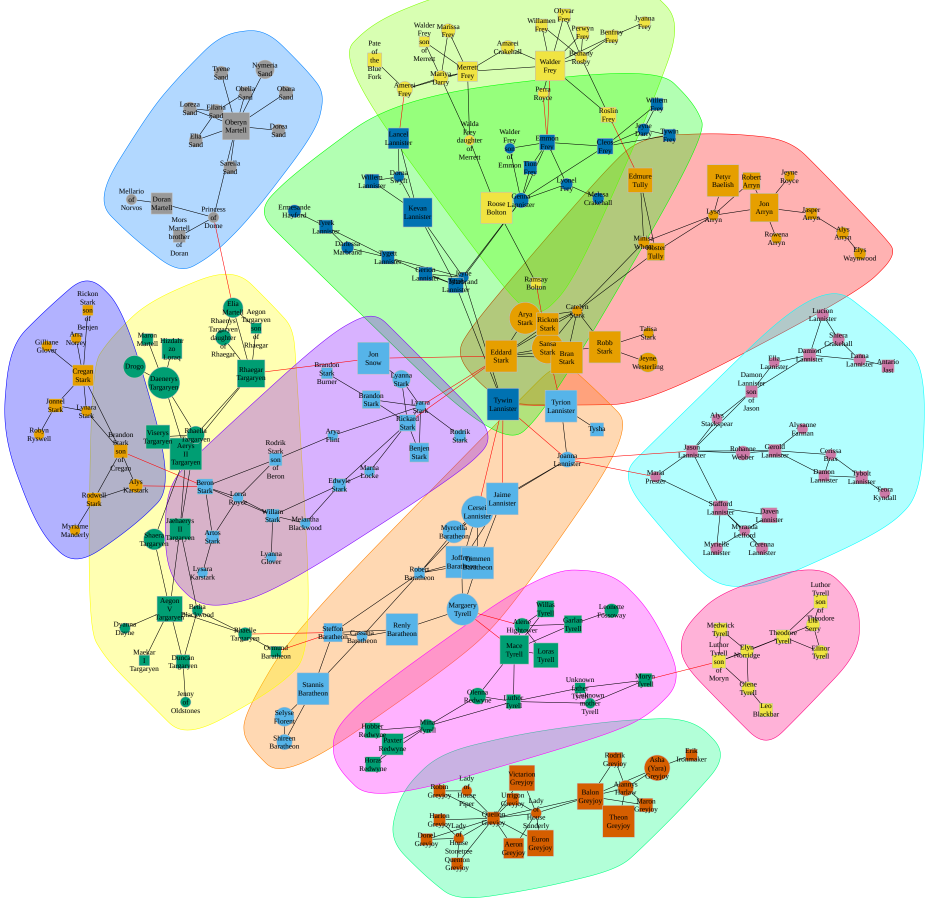
Cluster, Segmentation and (Social) Networks

So far, risk classes are based on covariates $\boldsymbol{X}$, correlated (causal effect?) with claims occurence (or severity).

Why not consider clusters in (social) networks, too?

A lot of cofounding variables (age, profession, location, etc.)

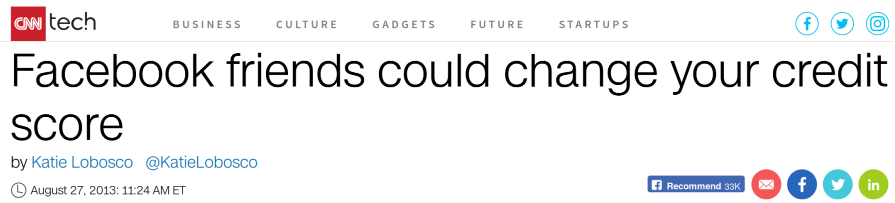
See InsPeer experience.



via shiring.github.io

(Social) Networks and Credit

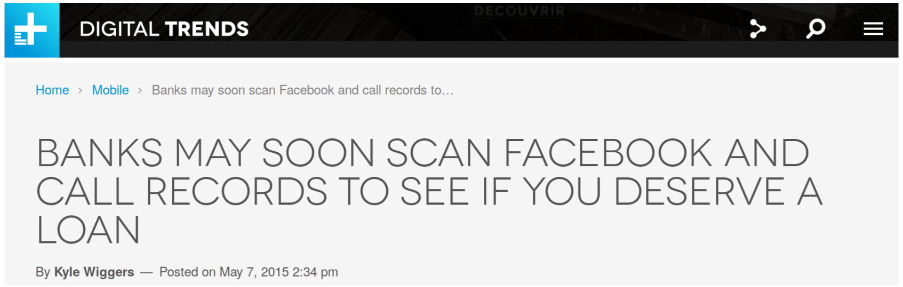
Used already on credit
(see cnn or or digitaltrends)
E.g Lenddo or Lendup
It does mean that homophily can be seen as a substitute to standard credit ‘explanatory’ variales...



Facebook friends could change your credit score

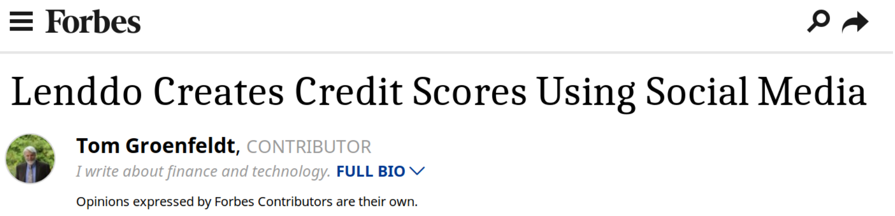
by Katie Lobosco @KatieLobosco

August 27, 2013: 11:24 AM ET


DIGITAL TRENDS
Home › Mobile › Banks may soon scan Facebook and call records to...

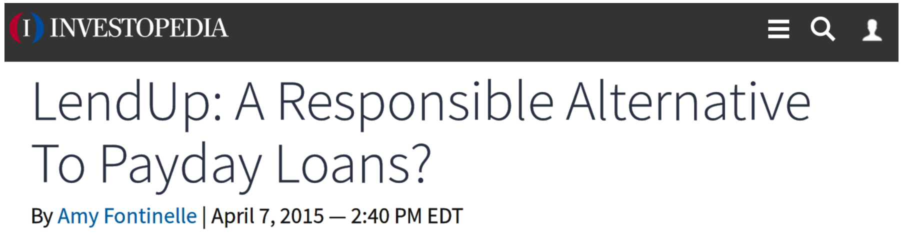
BANKS MAY SOON SCAN FACEBOOK AND CALL RECORDS TO SEE IF YOU DESERVE A LOAN

By Kyle Wiggers — Posted on May 7, 2015 2:34 pm



Lenddo Creates Credit Scores Using Social Media

Tom Groenfeldt, CONTRIBUTOR
I write about finance and technology. FULL BIO
Opinions expressed by Forbes Contributors are their own.

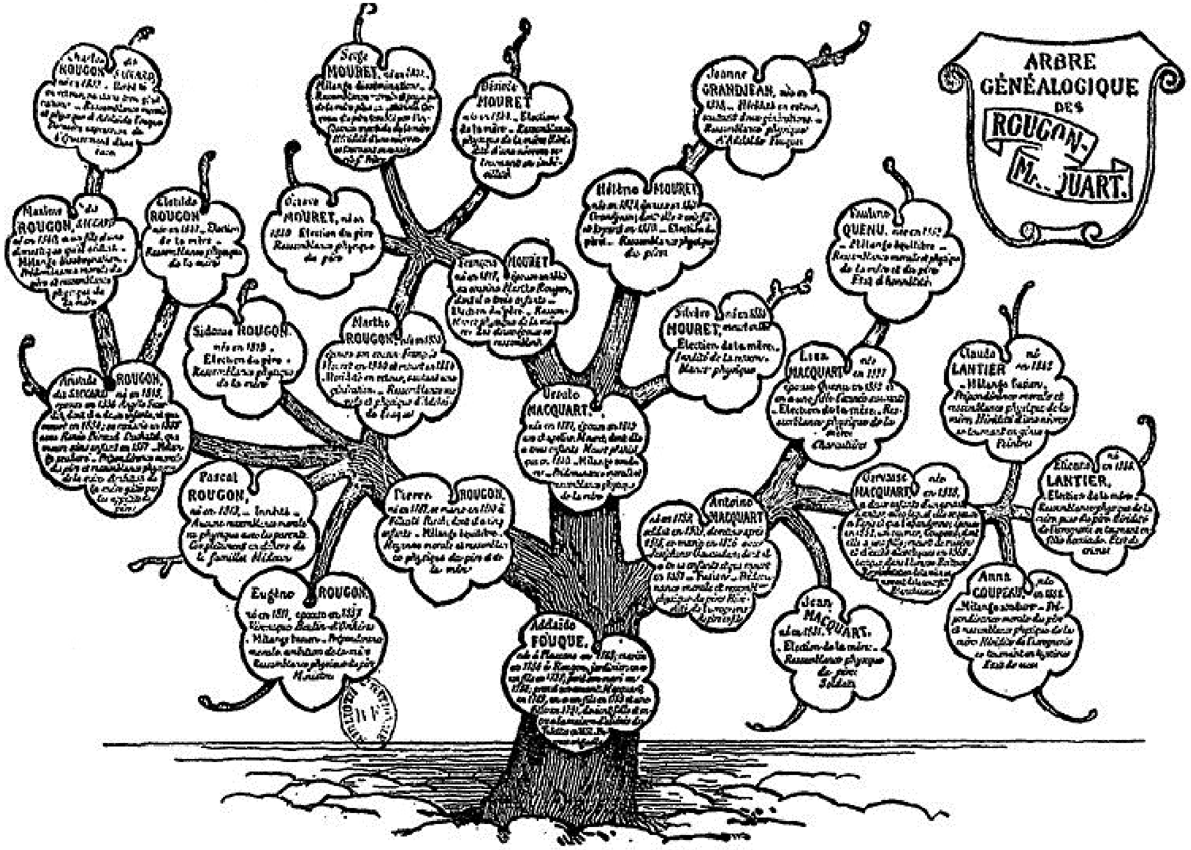


LendUp: A Responsible Alternative To Payday Loans?

By Amy Fontinelle | April 7, 2015 — 2:40 PM EDT

Information and Networks

But other kinds of networks can be used, e.g. (genealogical) trees



See Ewen Gallic's ongoing work (actinfo chair).

Privacy Issues

See General Data Protection Regulation (EU 2016/679) : what about aggregation ?

Consider a population $\{1, \cdots, n\}$ and a partition $\{\mathcal{I}_1, \cdots, \mathcal{I}_k\}$ (e.g. geographical areas Z), with respective sizes $\{n_1, \cdots, n_k\}$. Set $\overline{Y}_j = \dfrac{1}{n_j}\displaystyle\sum_{i\in I_j} Y_i$.

For continous covariates, set $\overline{X}_{k,j} = \dfrac{1}{n_k}\displaystyle\sum_{i\in I_j} X_{k,i}$,

For categorical variables, consider the associate composition variable $\overline{\boldsymbol{X}}_{k,j} = (\overline{X}_{k,1,j}, \cdots, \overline{X}_{k,d_k,j})$ where $\overline{X}_{k,\ell,j} = \dfrac{1}{n_k}\displaystyle\sum_{i\in I_j} \mathbf{1}(X_{k,i} = \ell)$.

See e.g. C. & Pigeon (2016) on micro-macro models and Enora Belz's ongoing work.

Privacy Issues

See Verbelen, Antonio & Claeskens (2016) and Antonio & C. (2017) on GPS data

	Predictor	Classic		Time-hybrid		Meter-hybrid		Telematics	
Policy	Time	×	offset	×	offset				
	Age								
	Experience	×	×	×	×	×	×		
	Sex	×	×						
	Material	×	×	×	×	×	×		
	Postal code	×	×	×	×	×	×		
	Bonus-malus	×	×	×	×	×	×		
	Age vehicle	×	×	×	×	×	×		
	Kwatt			×	×	×	×		
	Fuel	×	×	×		×			
Telematics	Distance					×	offset	×	offset
	Yearly distance			×	×				
	Average distance			×	×	×	×		
	Road type 1111			×	×	×	×	×	×
	Road type 1110			×	×	×	×	×	×
	Time slot			×	×	×	×	×	×
	Week/weekend			×	×	×	×	×	×