## Insurance: Risk Pooling and Price Segmentation

- Using Information in a 'Big Data' Context -
A. Charpentier (Université de Rennes 1)
on-going work with A. Farzanehfar \& Y.A. de Montjoye (Imperial College)

Telecom Paristech
January 2018

## 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 Pricing in a Nutshell

Insurance is the contribution of the many to the misfortune of the few
Finance: risk neutral valuation $\pi=\mathbb{E}_{\mathbb{Q}}\left[S_{1} \mid \mathcal{F}_{0}\right]=\mathbb{E}_{\mathbb{Q}_{0}}\left[S_{1}\right]$, where $S_{1}=\sum_{i=1}^{N_{1}} Y_{i}$
Insurance: risk sharing (pooling) $\pi=\mathbb{E}_{\mathbb{P}}\left[S_{1}\right]$
or, with segmentation / price differentiation $\pi(\omega)=\mathbb{E}_{\mathbb{P}}\left[S_{1} \mid \Omega=\omega\right]$ for some (unobservable?) risk factor $\Omega$
imperfect information given some (observable) risk variables $\boldsymbol{X}=\left(X_{1}, \cdots, X_{k}\right)$ $\pi(\boldsymbol{x})=\mathbb{E}_{\mathbb{P}}\left[S_{1} \mid \boldsymbol{X}=\boldsymbol{x}\right]=\mathbb{E}_{\mathbb{P}_{\boldsymbol{X}}}\left[S_{1} \mid \boldsymbol{x}\right]$
Insurance pricing is not only data driven, it is also essentially model driven (see Pricing Game)

Insurance Pricing in a Nutshell
Premium is $\pi=\mathbb{E}_{\mathbb{P}_{\boldsymbol{X}}}\left[S_{1}\right]$
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 $\pi \approx \mathbb{E}\left[S_{1} \mid \boldsymbol{X}=\boldsymbol{x}\right]=\mathbb{E}\left[\sum_{i=1}^{N} Y_{i} \mid \boldsymbol{X}=\boldsymbol{x}\right]=\mathbb{E}[N \mid \boldsymbol{X}=\boldsymbol{x}] \cdot \mathbb{E}\left[Y_{i} \mid \boldsymbol{X}=\boldsymbol{x}\right]$
Statistical and modeling issues to approximate based on some training datasets, with claims frequency $\left\{n_{i}, \boldsymbol{x}_{i}\right\}$ and individual losses $\left\{y_{i} \boldsymbol{x}_{i}\right\}$

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

Try to avoid overfit

## Risk Sharing in Insurance

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

$$
\frac{1}{n} \sum_{i=1}^{n} S_{i} \sim \frac{1}{n} \sum_{i=1}^{n} \pi\left(\boldsymbol{X}_{i}\right) \quad(\text { as } n \rightarrow \infty, \text { from the law of large number })
$$

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 | $\operatorname{Var}[S]$ |

All the risk - $\operatorname{Var}[S]$ - 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 | $\mathbb{E}[S \mid \boldsymbol{\Omega}]$ | $S-\mathbb{E}[S \mid \boldsymbol{\Omega}]$ |
| Average Loss | $\mathbb{E}[S]$ | 0 |
| Variance | $\operatorname{Var}[\mathbb{E}[S \mid \boldsymbol{\Omega}]]$ | $\operatorname{Var}[S-\mathbb{E}[S \mid \boldsymbol{\Omega}]]$ |

Observe that $\operatorname{Var}[S-\mathbb{E}[S \mid \boldsymbol{\Omega}]]=\mathbb{E}[\operatorname{Var}[S \mid \boldsymbol{\Omega}]]$, so that

$$
\operatorname{Var}[S]=\underbrace{\mathbb{E}[\operatorname{Var}[S \mid \boldsymbol{\Omega}]]}_{\rightarrow \text { insurer }}+\underbrace{\operatorname{Var}[\mathbb{E}[S \mid \boldsymbol{\Omega}]]}_{\rightarrow \text { insured }} .
$$

Risk Transfert with Segmentation and Imperfect Information
Assume that $\boldsymbol{X} \subset \boldsymbol{\Omega}$ is observable

|  | Insured | Insurer |
| :--- | :---: | :---: |
| Loss | $\mathbb{E}[S \mid \boldsymbol{X}]$ | $S-\mathbb{E}[S \mid \boldsymbol{X}]$ |
| Average Loss | $\mathbb{E}[S]$ | 0 |
| Variance | $\operatorname{Var}[\mathbb{E}[S \mid \boldsymbol{X}]]$ | $\mathbb{E}[\operatorname{Var}[S \mid \boldsymbol{X}]]$ |

Now

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

Risk Transfert with Segmentation and Imperfect Information
With imperfect information, we have the popular risk decomposition

$$
\begin{aligned}
\operatorname{Var}[S]= & \mathbb{E}[\operatorname{Var}[S \mid \boldsymbol{X}]]+\operatorname{Var}[\mathbb{E}[S \mid \boldsymbol{X}]] \\
= & \underbrace{\mathbb{E}[\operatorname{Var}[S \mid \boldsymbol{\Omega}]]}_{\text {pooling }}+\underbrace{\mathbb{E}[\operatorname{Var}[\mathbb{E}[S \mid \boldsymbol{\Omega}] \mid \boldsymbol{X}]]}_{\rightarrow \text { insurer }} \\
& +\underbrace{\operatorname{Var}[\mathbb{E}[S \mid \boldsymbol{X}]]}_{\rightarrow \text { insured }} .
\end{aligned}
$$

## More and more price differentiation ?

Consider $\pi_{1}=\mathbb{E}\left[S_{1}\right]$ and $\pi_{2}(\boldsymbol{x})=\mathbb{E}\left[S_{1} \mid \boldsymbol{X}=\boldsymbol{x}\right]$
Observe that $\mathbb{E}[\pi(\boldsymbol{X})]=\sum_{\boldsymbol{x} \in \mathcal{X}} \pi(\boldsymbol{x}) \cdot \mathbb{P}[\boldsymbol{x}]$
$=\sum_{\boldsymbol{x} \in \mathcal{X}_{1}} \pi(\boldsymbol{x}) \cdot \mathbb{P}[\boldsymbol{x}]+\sum_{\boldsymbol{x} \in \mathcal{X}_{2}} \pi(\boldsymbol{x}) \cdot \mathbb{P}[\boldsymbol{x}]$

- Insured with $\boldsymbol{x} \in \mathcal{X}_{1}$ : choose Ins1
- Insured with $\boldsymbol{x} \in \mathcal{X}_{2}$ : choose Ins2

Ins1: $\sum_{\boldsymbol{x} \in \mathcal{X}_{1}} \pi_{1}(\boldsymbol{x}) \cdot \mathbb{P}[\boldsymbol{x}] \neq \mathbb{E}\left[S \mid \boldsymbol{X} \in \mathcal{X}_{1}\right]$
Ins2: $\sum_{\boldsymbol{x} \in \mathcal{X}_{2}} \pi_{2}(\boldsymbol{x}) \cdot \mathbb{P}[\boldsymbol{x}]=\mathbb{E}\left[S \mid \boldsymbol{X} \in \mathcal{X}_{2}\right]$


## Price Differentiation, a Toy Example

Claims frequency $Y$ (average cost $=1,000$ )

|  |  |  | $X_{1}$ |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Young | Experienced | Senior | Total |
| $X_{2}$ | Town | 12\% | $9 \%$ | $9 \%$ | 9.5\% |
|  |  | (500) | $(2,000)$ | (500) | $(3,000)$ |
|  | Outside | 8\% | 6.67\% | $4 \%$ | 6.33\% |
|  |  | (500) | $(1,000)$ | (500) | $(2,000)$ |
| Total |  | 10\% | 8.22\% | 6.5\% | 8.23\% |
|  |  | $(1,000)$ | $(3,000)$ | $(1,000)$ | $(5,000)$ |

from C., Denuit \& Élie (2015)

Price Differentiation, a Toy Example

|  | $\begin{gathered} \text { Y-T } \\ (500) \end{gathered}$ | $\begin{gathered} \text { Y-O } \\ (500) \end{gathered}$ | $\begin{gathered} \text { E-T } \\ (2,000) \end{gathered}$ | $\begin{gathered} \text { E-O } \\ (1,000) \end{gathered}$ | $\begin{gathered} \text { S-T } \\ (500) \end{gathered}$ | $\begin{gathered} \text { S-O } \\ (500) \end{gathered}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 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 <br> ratio |  | $99.5 \%$ <br> quantile | Market <br> Share |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| none | 247 | 285 | $115.4 \%$ | $( \pm 8.9 \%)$ |  | $66.1 \%$ |
| $X_{1} \times X_{2}$ | 126.67 | 126.67 | $100.0 \%$ | $( \pm 10.4 \%)$ |  | $33.9 \%$ |
| market | 373.67 | 411.67 | $110.2 \%$ | $( \pm 5.1 \%)$ |  |  |
| none | 41.17 | 60 | $145.7 \%$ | $( \pm 34.6 \%)$ | $189 \%$ | $11.6 \%$ |
| $X_{1}$ | 196.94 | 225 | $114.2 \%$ | $( \pm 11.8 \%)$ | $140 \%$ | $55.8 \%$ |
| $X_{2}$ | 95 | 106.67 | $112.3 \%$ | $( \pm 15.1 \%)$ | $134 \%$ | $26.9 \%$ |
| $X_{1} \times X_{2}$ | 20 | 20 | $100.0 \%$ | $( \pm 41.9 \%)$ | $160 \%$ | $5.7 \%$ |
| market | 353.10 | 411.67 | $116.6 \%$ | $( \pm 5.3 \%)$ | $130 \%$ |  |

## 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


## Model Comparison and Lorenz curves



Source: Progressive Insurance

## Model Comparison and Lorenz curves

Consider an ordered sample $\left\{y_{1}, \cdots, y_{n}\right\}$ of incomes, with $y_{1} \leq y_{2} \leq \cdots \leq y_{n}$, then Lorenz curve is

$$
\left\{F_{i}, L_{i}\right\} \text { with } F_{i}=\frac{i}{n} \text { and } L_{i}=\frac{\sum_{j=1}^{i} y_{j}}{\sum_{j=1}^{n} y_{j}}
$$



We have observed losses $y_{i}$ and premiums $\widehat{\pi}\left(\boldsymbol{x}_{i}\right)$. Consider an ordered sample by the model, see Frees, Meyers \& Cummins (2014), $\widehat{\pi}\left(\boldsymbol{x}_{1}\right) \geq \widehat{\pi}\left(\boldsymbol{x}_{2}\right) \geq \cdots \geq \widehat{\pi}\left(\boldsymbol{x}_{n}\right)$, then plot

$$
\left\{F_{i}, L_{i}\right\} \text { with } F_{i}=\frac{i}{n} \text { and } L_{i}=\frac{\sum_{j=1}^{i} y_{j}}{\sum_{j=1}^{n} y_{j}}
$$



## Model Comparison for Life Insurance Models

Consider the case of a death insurance contract, that pays 1 if the insured deceased within the year.
$\pi(x)=\mathbb{E}\left[T_{x} \leq t+1 \mid T_{x}>t\right]$

- No price discrimination $\pi=\mathbb{E}[\pi(X)]$
- Perfect discrimination $\pi(x)$
- Imperfect discrimination
$\pi_{-}=\mathbb{E}[\pi(X) \mid X<s]$ and $\pi_{+}=\mathbb{E}[\pi(X) \mid X>s]$



## From Econometric to 'Machine Learning’ Techniques

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

$$
\widehat{\pi}_{j}(\boldsymbol{x})=\widehat{\mathbb{E}}\left[N_{1} \mid \boldsymbol{X}=\boldsymbol{x}\right] \cdot \widehat{\mathbb{E}}[Y \mid \boldsymbol{X}=\boldsymbol{x}]=\underbrace{\exp \left(\widehat{\boldsymbol{\alpha}}^{\top} \boldsymbol{x}\right)}_{\text {Poisson } \mathcal{P}\left(\lambda_{\boldsymbol{x}}\right)} \cdot \underbrace{\exp \left(\widehat{\boldsymbol{\beta}}^{\top} \boldsymbol{x}\right)}_{\text {Gamma } \mathcal{G}\left(\mu_{\boldsymbol{X}}, \varphi\right)}
$$

that can be extended to GAMs,

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

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

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

## From Econometric to 'Machine Learning’ Techniques

For some loss function $\ell: \mathbb{R}^{2} \rightarrow \mathbb{R}_{+}$(usually an $L_{2}$ based loss, $\ell(s, y)=(s-y)^{2}$ since $\operatorname{argmin}\{\mathbb{E}[\ell(S, m)], m \in \mathbb{R}\}$ 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, $\boldsymbol{\pi}_{i}=\left(\widehat{\pi}_{1}\left(\boldsymbol{x}_{i}\right), \cdots, \widehat{\pi}_{d}\left(\boldsymbol{x}_{i}\right)\right) \in \mathbb{R}_{+}^{d}$.

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?






@freakonometrics
(6) freakonometrics

I freakonometrics.hypotheses.org

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' $\pi_{i}=\min \left\{\widehat{\pi}_{1}\left(\boldsymbol{x}_{i}\right), \cdots, \widehat{\pi}_{d}\left(\boldsymbol{x}_{i}\right)\right\}=\widehat{\pi}_{1: d}\left(\boldsymbol{x}_{i}\right)$

Ins1 Ins2 Ins3 Ins4 Ins5 Ins6
$\begin{array}{llllll}787.93 & 706.97 & 1032.62 & 907.64 & 822.58 & 603.83\end{array}$
$\begin{array}{llllll}170.04 & 197.81 & 285.99 & 212.71 & 177.87 & 265.13\end{array}$
$473.15 \quad 447.58 \quad 343.64 \quad 410.76 \quad 414.23 \quad 425.23$
$\begin{array}{llllll}337.98 & 336.20 & 468.45 & 339.33 & 383.55 & 672.91\end{array}$

Insurance Ratemaking Competition
A more realistic rule $\pi_{i} \in\left\{\widehat{\pi}_{1: d}\left(\boldsymbol{x}_{i}\right), \widehat{\pi}_{2: d}\left(\boldsymbol{x}_{i}\right), \widehat{\pi}_{3: d}\left(\boldsymbol{x}_{i}\right)\right\}$

|  | Ins1 | Ins2 | Ins3 | Ins4 | Ins5 | Ins6 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| \%en | 787.93 | 706.97 | 1032.62 | 907.64 | 822.58 | 603.83 |
|  | 170.04 | 197.81 | 285.99 | 212.71 | 177.87 | 265.13 |
| $0^{\circ}$ | 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 113 players
Season 214 players
Step 3 [season 2] : provide additional informa-
tion (premiums of competitors)
Season 323 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

Facebook friends could change your credit score
by Katie Lobosco @KatieLobosco
(L) August 27, 2013: 11:24 AM ET
(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...

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

三 Forbes
Lenddo Creates Credit Scores Using Social Media
Tom Groenfeldt, CONTRIBUTOR
I write about finance and technology. FULL BIO $\checkmark$
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 $\left\{\mathcal{I}_{1}, \cdots, \mathcal{I}_{k}\right\}$ (e.g. geographical areas $Z$ ), with respective sizes $\left\{n_{1}, \cdots, n_{k}\right\}$. Set $\bar{Y}_{j}=\frac{1}{n_{j}} \sum_{i \in I_{j}} Y_{i}$.
For continous covariates, set $\bar{X}_{k, j}=\frac{1}{n_{k}} \sum_{i \in I_{j}} X_{k, i}$,
For categorical variables, consider the associate composition variable
$\overline{\boldsymbol{X}}_{k, j}=\left(\bar{X}_{k, 1, j}, \cdots, \bar{X}_{k, d_{k}, j}\right)$ where $\bar{X}_{k, \ell, j}=\frac{1}{n_{k}} \sum_{i \in I_{j}} \mathbf{1}\left(X_{k, i}=\ell\right)$.
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 |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Time | $\times$ | offset | $\times$ | offset |  |  |  |  |
|  | Age |  |  |  |  |  |  |  |  |
|  | Experience | $\times$ | $\times$ | $\times$ | $\times$ | $\times$ | $\times$ |  |  |
|  | Sex | $\times$ | $\times$ |  |  |  |  |  |  |
| . | Material | $\times$ | $\times$ | $\times$ | $\times$ | $\times$ | $\times$ |  |  |
| 0 | Postal code | $\times$ | $\times$ | $\times$ | $\times$ | $\times$ | $\times$ |  |  |
|  | Bonus-malus | $\times$ | $\times$ | $\times$ | $\times$ | $\times$ | $\times$ |  |  |
|  | Age vehicle | $\times$ | $\times$ | $\times$ | $\times$ | $\times$ | $\times$ |  |  |
|  | Kwatt |  |  | $\times$ | $\times$ | $\times$ | $\times$ |  |  |
|  | Fuel | $\times$ | $\times$ | $\times$ |  | $\times$ |  |  |  |
|  | Distance |  |  |  |  | $\times$ | offset | $\times$ | offset |
| \% | Yearly distance |  |  | $\times$ | $\times$ |  |  |  |  |
| * | Average distance |  |  | $\times$ | $\times$ | $\times$ | $\times$ |  |  |
| 即 | Road type 1111 |  |  | $\times$ | $\times$ | $\times$ | $\times$ | $\times$ | $\times$ |
| O | Road type 1110 |  |  | $\times$ | $\times$ | $\times$ | $\times$ | $\times$ | $\times$ |
| $\square$ | Time slot |  |  | $\times$ | $\times$ | $\times$ | $\times$ | $\times$ | $\times$ |
|  | Week/weekend |  |  | $\times$ | $\times$ | $\times$ | $\times$ | $\times$ | $\times$ |

