Autocalibration & Insurance Pricing

Arthur Charpentier (UQAM)

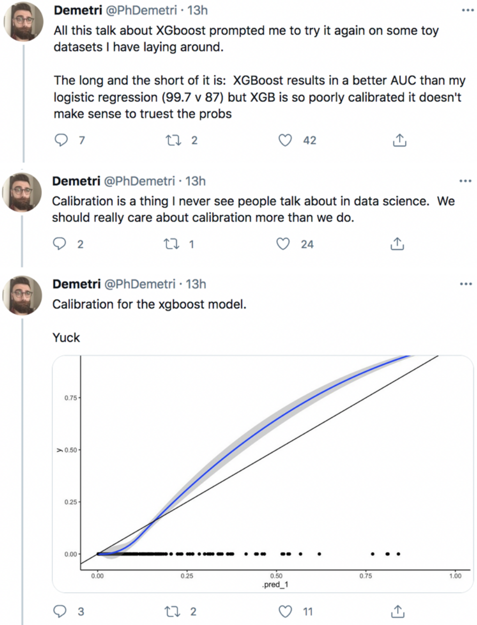
with Michel Denuit (UCL) & Julien Trufin (UCB)

IME (Insurance: Mathematics & Economics) 2021

Machine Learning in Insurance I - Parallel Session C4-I

Agenda

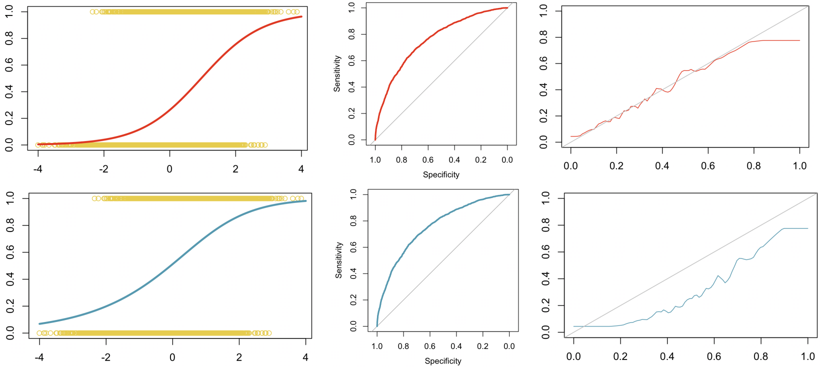
- On insurance pricing
- What is a model ?
- Bias of a model
- Correcting from bias
- Application
- Theoretical properties



Machine Learning & Accuracy Measures (AUC = 0.7565)

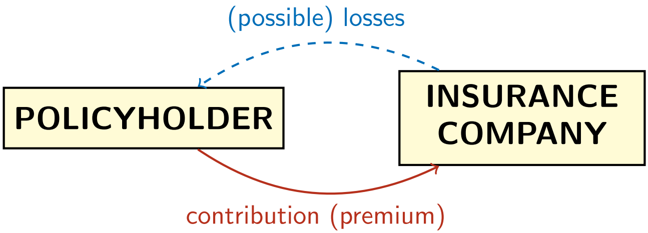
"*XGBoost results in a better AUC than my logistic regression but XGBoost is so poorly calibrated it doesn't make sense to trust the probs*"

- (x_i, y_i) (where $y_i \in \{0, 1\}$) with $\widehat{\pi}_1(x)$ or $\widehat{\pi}_2(x)$ $(= \sqrt{\widehat{\pi}_1(x)})$
- ROC curve of $\widehat{\pi}_1$ or $\widehat{\pi}_2$ are identical (and same AUC)
- $\mathbb{E}[Y|\widehat{\pi}_1(X) = s]$ or $\mathbb{E}[Y|\widehat{\pi}_2(X) = s]$,
  ie. regression of y_i on $\pi(x_i)$ (expected? $\mathbb{E}[Y|\widehat{\pi}(X) = s] \sim s$)



Insurance

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



The "*contribution*" is obtained using predictive models (interpretability / black box / etc issues)

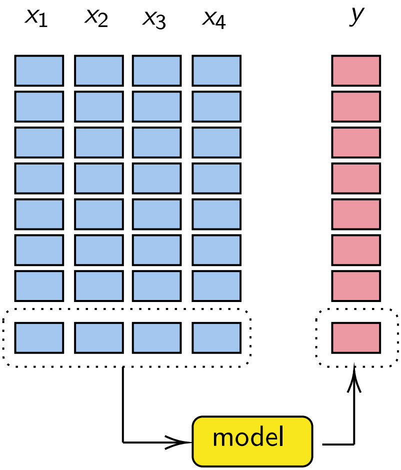
A model, $m : \boldsymbol{x} \mapsto y$

To train / estimate a model m, we need a dataset, i.e. a collection of observations $(\boldsymbol{x}_i, y_i)$

Usually y_i denotes the annual loss

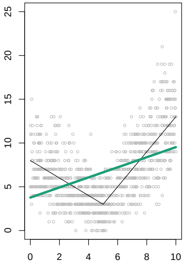
$$y_i = \sum_{j=1}^{n_i} z_{i,j}$$

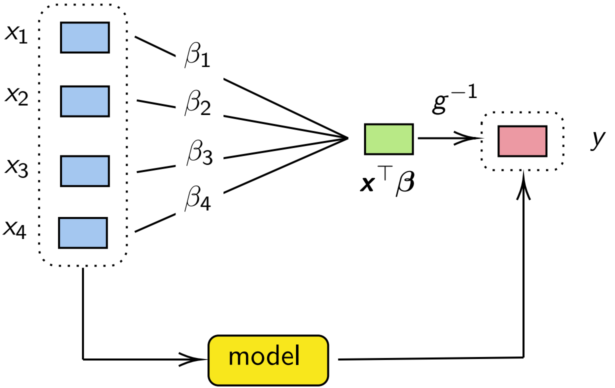
n is the annual frequency
z is the cost of a single claim
(see Tweedie models)
To illustrate, y is a counting variable



A model, $m : \boldsymbol{x} \mapsto y$ (GLM)

Poisson regression

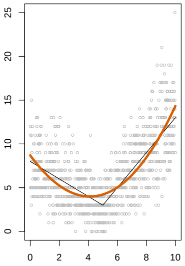


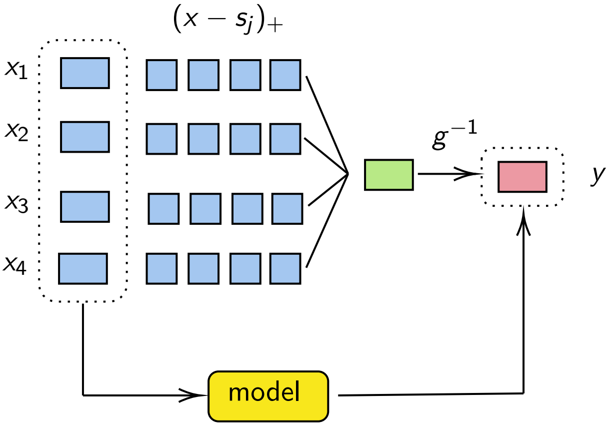


$$m(\boldsymbol{x}) = g^{-1}(\boldsymbol{x}^\top\boldsymbol{\beta}) = g^{-1}\left(\sum_{j=1}^{p} \beta_j x_j\right)$$

A model, $m : \boldsymbol{x} \mapsto y$ (GAM)

Poisson regression

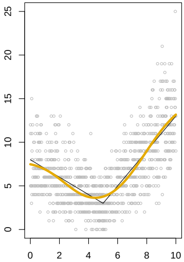


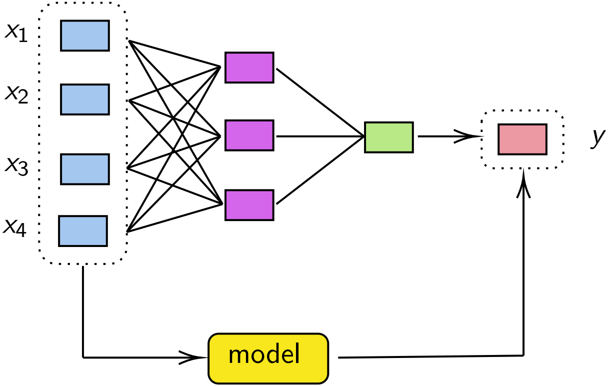


$$m(\boldsymbol{x}) = g^{-1}\left(\sum_{j=1}^{p} \beta_j \psi_j(x_j)\right), \text{ where } \psi_j(x) = \sum_{k=1}^{5} \alpha_{j,k}(x - s_{j,k})_+$$

A model, $m : \boldsymbol{x} \mapsto y$ (neural nets)

Poisson loss

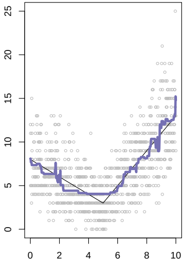


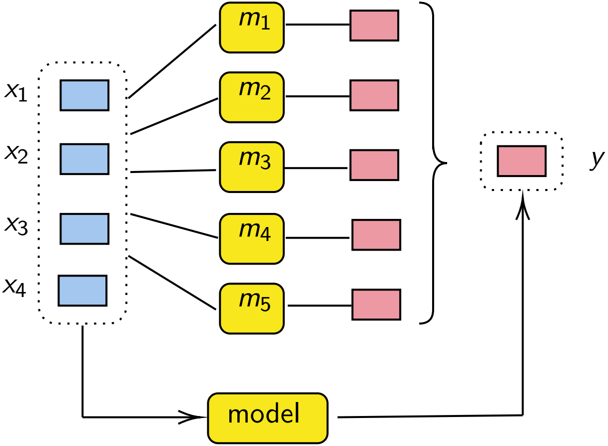


$$\text{E.g. } m(\boldsymbol{x}) = \sum_{j=1}^{3} \omega_{1:j}\; h(\boldsymbol{x}^\top \boldsymbol{\omega}_{2:j})$$

A model, $m : \boldsymbol{x} \mapsto y$ (ensemble parallel, bagging)

Random forest

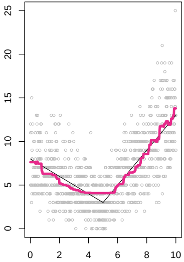


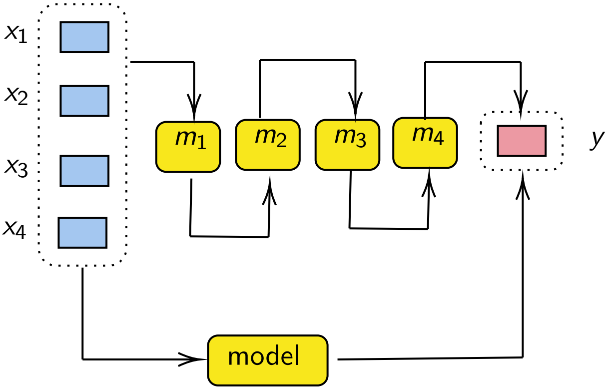


E.g. $m(\boldsymbol{x}) = \dfrac{1}{B}\displaystyle\sum_{b=1}^{B} m_b(\boldsymbol{x})$ where $m_b(\boldsymbol{x}) = \displaystyle\sum_{k=1}^{K} \alpha_{b,k} \mathbf{1}(\boldsymbol{x} \in \mathcal{X}_{b,k})$

A model, $m : \boldsymbol{x} \mapsto y$ (ensemble sequential, boosting)

Boosting





$m(\boldsymbol{x}) = \sum_{t=1}^{T} m_t(\boldsymbol{x})$ where m_t is a (weak) model on $y_i - m_{t-1}(\boldsymbol{x}_i)$
(residuals from the previous step) such as (not too deep) trees

Not a model, $\widehat{m} : \boldsymbol{x} \mapsto y$ (local regression)

To approximate $\mathbb{E}[Y]$ use

$$\widehat{m} = \underset{\mu \in \mathbb{R}}{\text{argmin}} \left\{ \sum_{i=1}^{n} \frac{1}{n} \, [y_i - \mu]^2 \right\}$$

To approximate $\mathbb{E}[Y|\boldsymbol{X} = \boldsymbol{x}]$, use

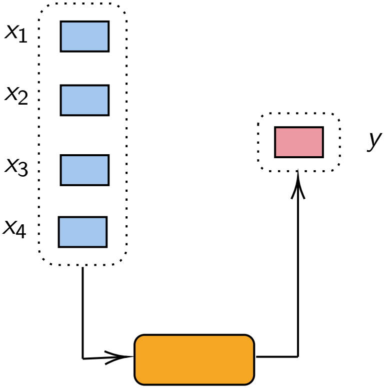
$$\widehat{m}(\boldsymbol{x}) = \underset{\mu \in \mathbb{R}}{\text{argmin}} \left\{ \sum_{i=1}^{n} \omega_i(\boldsymbol{x})[y_i - \mu]^2 \right\}$$

where, see Loader (1999),

$$\omega_i(\boldsymbol{x}) \propto k_\alpha \left(\|\boldsymbol{x} - \boldsymbol{x}_i\| \right)$$

or k nearest neighbors indicator...

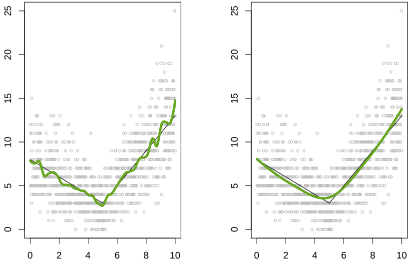
Behaves poorly in high dimension (but efficient in dimension 1)

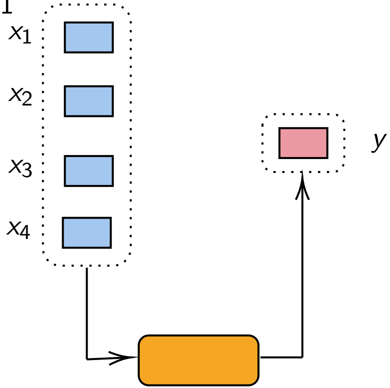


Not a model, $\widehat{m}: \boldsymbol{x} \mapsto y$ (local regression)

$$\widehat{m}(\boldsymbol{x}) = \sum_{i=1}^{n} \omega_i(\boldsymbol{x})\, y_i, \text{ with } \sum_{i=1}^{n} \omega_i(\boldsymbol{x}) = 1$$

see `locfit` in R





(depends on a bandwidth parameter α)
Provides a local estimate of $\mathbb{E}[Y|X=x]$ on a neighborhood of x

GLM, Bias, & Economic Interpretation

For GLMs, $f(y_i) = \exp\left(\dfrac{y_i\theta_i - b(\theta_i)}{\varphi} + c(y_i,\varphi)\right)$

The natural parameter for y_i : θ_i

Prediction for y_i : $\widehat{y}_i = \mu_i = \mathbb{E}(Y_i) = b'(\theta_i)$

Score associated with y_i : $\eta_i = \boldsymbol{x}_i^\top\boldsymbol{\beta}$

Link function : g such that $\eta_i = g(\mu_i) = g(b'(\theta_i))$

$$\log\mathcal{L}(\boldsymbol{\theta},\varphi,\boldsymbol{y}) = \sum_{i=1}^n \log\mathcal{L}_i = \sum_{i=1}^n\left[\frac{y_i\theta_i - b(\theta_i)}{\varphi} + c(y_i,\varphi)\right]$$

First order conditions: $\dfrac{\partial\log\mathcal{L}_i}{\partial\beta_j} = \dfrac{\partial\log\mathcal{L}_i}{\partial\theta_i}\cdot\dfrac{\partial\theta_i}{\partial\mu_i}\cdot\dfrac{\partial\mu_i}{\partial\eta_i}\cdot\dfrac{\partial\eta_i}{\partial\beta_j}$

$$\frac{\partial\log\mathcal{L}_i}{\partial\theta_i} = \frac{y_i-\mu_i}{\varphi},\ \ \frac{\partial\theta_i}{\partial\mu_i} = \left(\frac{\partial\mu_i}{\partial\theta_i}\right)^{-1} = \frac{1}{b''(\theta_i)} = \frac{1}{V(\mu_i)}$$

$$\frac{\partial\eta_i}{\partial\beta_j} = \frac{\partial\boldsymbol{x}_i^\top\boldsymbol{\beta}}{\partial\beta_j} = x_{i,j},\ \ \frac{\partial\mu_i}{\partial\eta_i} = (\frac{\partial\eta_i}{\partial\mu_i})^{-1} = (g'(\mu_i))^{-1}$$

GLM, Bias, & Economic Interpretation

With canonical link $g_\star = b'^{-1}$, i.e. $\eta_i = \theta_i$,

$$\boldsymbol{X}^\top(\boldsymbol{y}-\boldsymbol{\mu}) = 0, \quad \text{where } \widehat{\boldsymbol{y}} = \boldsymbol{\mu}$$

so, if there is an intercept, $\mathbf{1}^\top(\boldsymbol{y}-\boldsymbol{\mu}) = 0$, i.e. $\displaystyle\sum_{i=1}^n y_i = \sum_{i=1}^n \widehat{y}_i$

which is the empirical version (training dataset) of $\mathbb{E}[Y] = \mathbb{E}[\widehat{Y}]$

(see logistic regression or Poisson with log-link)

But more generally, the first order condition is

$$\boldsymbol{X}^\top \boldsymbol{W}\Delta(\boldsymbol{y}-\boldsymbol{\mu}) = 0,$$

where $\boldsymbol{W} = \text{diag}\big((V(\mu_i)g'(\mu_i)^2)^{-1}\big)$ and $\Delta = \text{diag}\big(g'(\mu_i)\big)$.

But usually not an important issue in ML classification problems

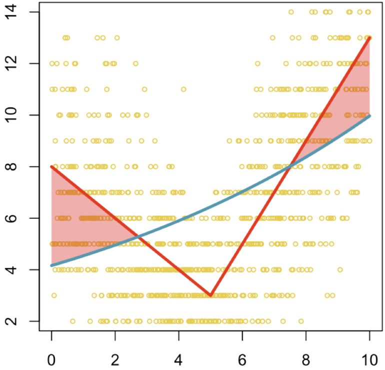
GLM, Bias, & Economic Interpretation

Model $\widehat{\pi}$ is globally unbiased if $\mathbb{E}[\widehat{\pi}(\boldsymbol{X})] = \mathbb{E}[Y]$, $\displaystyle\sum_{i=1}^{n} \widehat{y}_i = \sum_{i=1}^{n} y_i$

Model $\widehat{\pi}$ is locally unbiased if $\mathbb{E}[Y|\widehat{\pi}(\boldsymbol{X}) = s] = s$

GLM $\widehat{\pi}$ globally unbiased,
but possibly locally biased
Major economic impact

- $\widehat{\pi}(\boldsymbol{x}) < \mu(\boldsymbol{x})$
  attractive, but underpriced
- $\widehat{\pi}(\boldsymbol{x}) > \mu(\boldsymbol{x})$
  not attractive

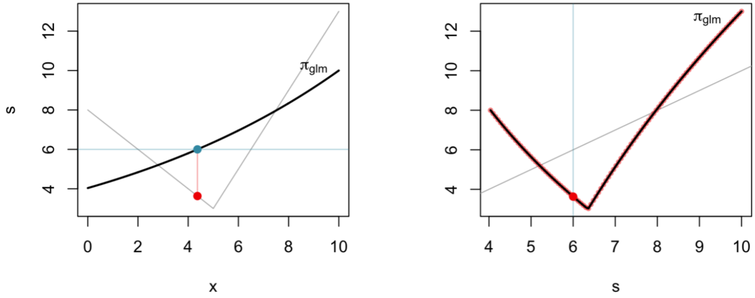


Natural idea: plot $s \mapsto \mathbb{E}[Y|\widehat{\pi}(\boldsymbol{X}) = s]$ " $= \mu(\widehat{\pi}^{-1}(s))$ "

Computing $\mathbb{E}[Y|\widehat{\pi}(\boldsymbol{X}) = s]$

True model $\mu(x)$ and GLM model $\widehat{\pi}(x)$

- select s, e.g. $s = 6$
- compute $x = \widehat{\pi}^{-1}(s)$ ($\widehat{\pi}$ is strictly increasing), here $x = 4.2$
- compute $\mu(\widehat{\pi}^{-1}(s))$, here 3.8
- plot $(s, \mu(\widehat{\pi}^{-1}(s)))$

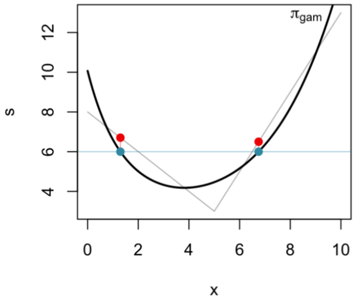


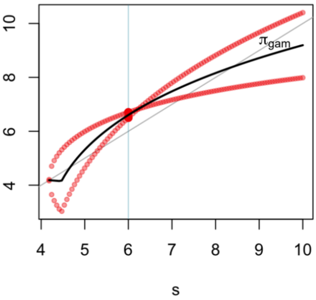
$$s \mapsto \mathbb{E}[Y|\widehat{\pi}(X) = s] = \mu(\widehat{\pi}^{-1}(s)) \text{ since } \mu(x) = \mathbb{E}[Y|X = x].$$

Computing $\mathbb{E}[Y|\widehat{\pi}(\boldsymbol{X})=s]$

True model $\mu(x)$ and GAM model $\widehat{\pi}(x)$

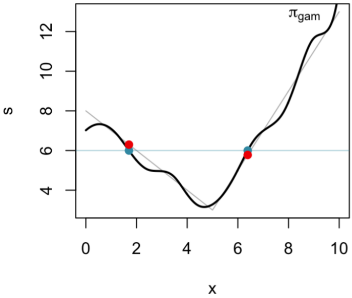
- select s, e.g. $s=6$
- compute $\mathcal{X}_s^{\widehat{\pi}}=\{\boldsymbol{x}\in\mathcal{X},\widehat{\pi}(\boldsymbol{x})=s\}$, here $\{1.6;6.7\}$
- compute $\mu(\widehat{\pi}^{-1}(s))$, $\{6.4,6.2\}$ and its mean $\overline{\mu}(\widehat{\pi}^{-1}(s))$, 6.3
- plot $(s,\overline{\mu}(\widehat{\pi}^{-1}(s)))$

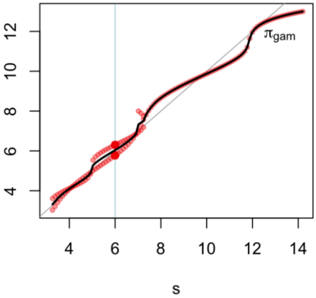




Computing $\mathbb{E}[Y|\widehat{\pi}(\boldsymbol{X})=s]$

True model $\mu(x)$ and GAM model $\widehat{\pi}(x)$ (more degrees of freedom)
Plot $s\mapsto\mathbb{E}[Y|\widehat{\pi}(X)=s]$, should be close to the first diagonal





Seems locally unbiased...

- impossible to get the figure on the left in higher dimension
- we used here μ but in practice, μ is unknown !

$\mathbb{E}[Y|\widehat{\pi}(\boldsymbol{X}) = s]$: empirical version

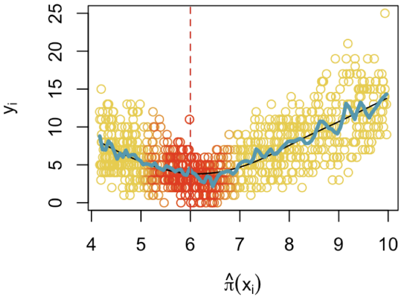
How to plot $s \mapsto \mathbb{E}[Y|\widehat{\pi}(X) = s]$?
but in real life, μ is unknown
Consider $\{(\widehat{\pi}(\boldsymbol{x}_i), y_i)\}$
and fit a local regression

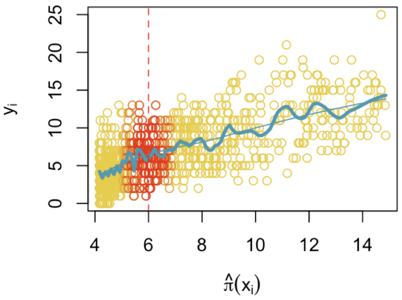
- fit a model $\widehat{\pi}$
- estimate $\mathbb{E}[Y|\widehat{\pi}(X) = s]$

  local regression on $\{(\widehat{\pi}(\boldsymbol{x}_i), y_i)\}$
- local (multiplicative) correction

  $$\lambda_\alpha(s) = \frac{\mathbb{E}[Y|\widehat{\pi}(X) = s]}{s}$$
- correct $\widehat{\pi}$

  $$\widehat{\pi}_{BC}(\boldsymbol{x}) = \lambda_\alpha(\widehat{\pi}(\boldsymbol{x})) \cdot \widehat{\pi}(\boldsymbol{x})$$



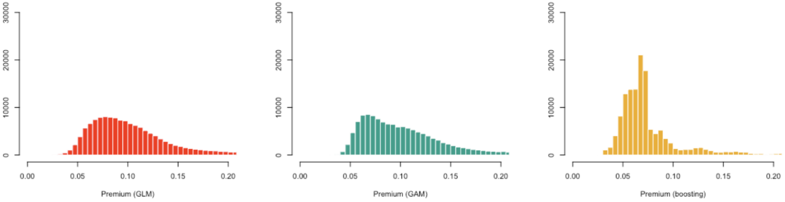


Application of a motor-insurance dataset

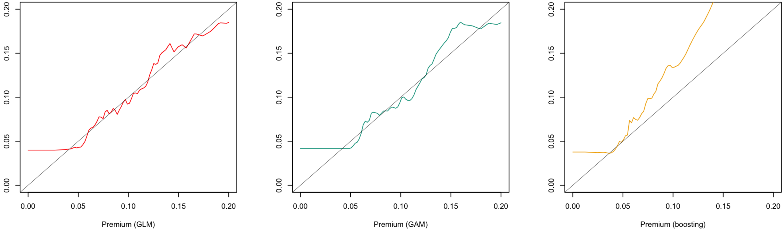
Here we focus only on claims (annual) frequency, corrected from the exposure, `freMTPL2freq` from `CASDataset` package

	π^{glm}	π^{gam}	π^{bst}
average $\overline{\pi}$	0.1087	0.1092	0.0820
10% quantile	0.0605	0.0598	0.0498
90% quantile	0.1682	0.1713	0.1244

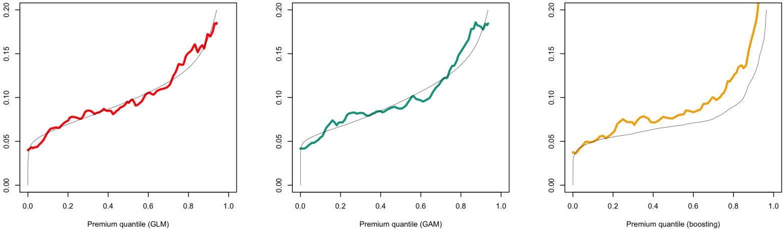
Table 1: Summary statistics on $\{\pi(\boldsymbol{x}_1),\cdots,\pi(\boldsymbol{x}_n)\}$, on the validation dataset (assuming an exposure of 1 to provide annualized predictions).



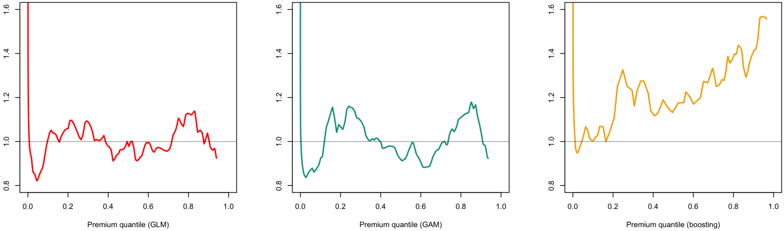
Application of a motor-insurance dataset



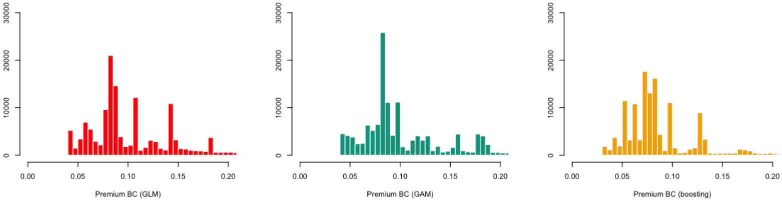
Evolution of $s \mapsto \mathbb{E}[Y|\widehat{\pi}(\boldsymbol{X}) = s]$ and $u \mapsto \mathbb{E}[Y|\widehat{\pi}(\boldsymbol{X}) = F_{\pi}^{-1}(u)]$



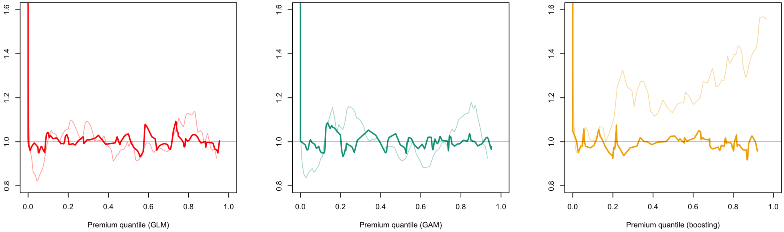
Application of a motor-insurance dataset



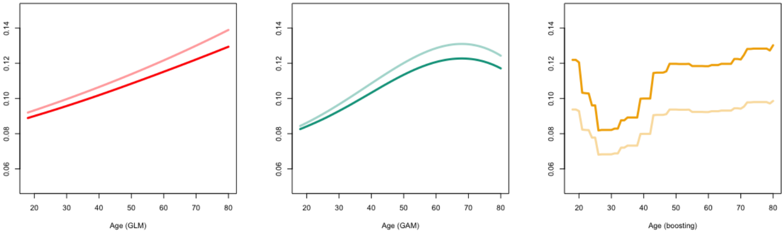
Multificative correction $\lambda_\alpha(u) = \mathbb{E}[Y|\widehat{\pi}(\boldsymbol{X}) = F_{\widehat{\pi}}^{-1}(u)]/F_{\widehat{\pi}}^{-1}(u)$



Application of a motor-insurance dataset



λ_α on the corrected model $\widehat{\pi}_{BC}$, and some partial dependence plot (on the age of the driver)



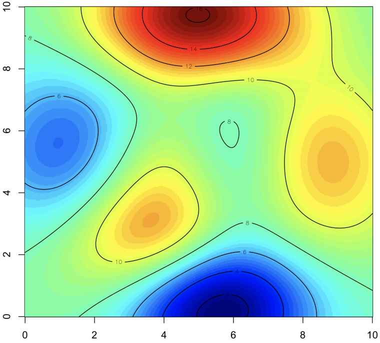
Wrap-Up

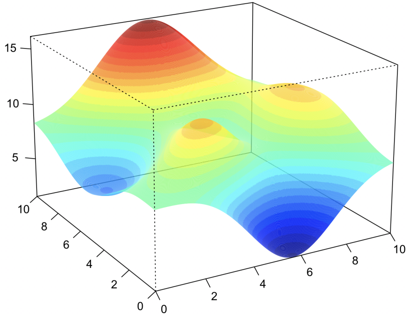
- fit a model $\widehat{\pi}$ using an ML algorithm
- estimate $\mathbb{E}[Y|\widehat{\pi}(X)]$ with local regression on $\{(\widehat{\pi}(\boldsymbol{x}_i), y_i)\}$
- local (multiplicative) correction $\lambda_\alpha(s) = \dfrac{\mathbb{E}[Y|\widehat{\pi}(X) = s]}{s}$
- correct $\widehat{\pi}$ by setting $\widehat{\pi}_{BC}(\boldsymbol{x}) = \lambda_\alpha(\widehat{\pi}(\boldsymbol{x})) \cdot \widehat{\pi}(\boldsymbol{x})$

Denuit, M., Charpentier, A. & Trufin, J. (2021). Autocalibration and Tweedie-dominance for Insurance Pricing with Machine Learning, ArXiv:2103.03635

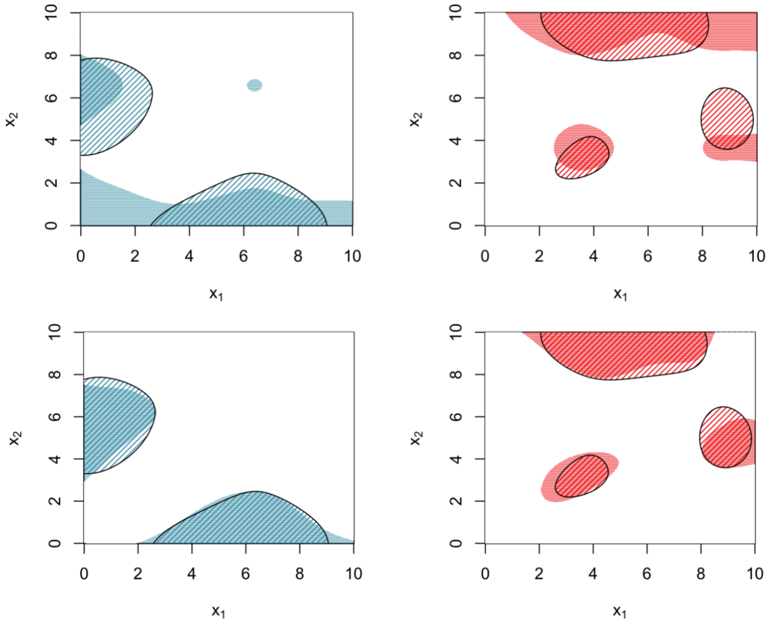
Appendix: Bivariate simulated data

Generate some data $(x_{1,i}, x_{2,i}, y_i)$ where $Y|\boldsymbol{X}$ has some distribution with mean $\mathbb{E}[Y|\boldsymbol{x}] = \mu(\boldsymbol{x})$





Model inversion, $\mathcal{X}_s^{\widehat{\pi}} = \{\boldsymbol{x} \in \mathcal{X}, \widehat{\pi}(\boldsymbol{x}) \lesseqgtr s\}$



$\mathcal{X}_s^{\widehat{\pi}}$, $Y|\boldsymbol{X} \in \mathcal{X}_s^{\widehat{\pi}}$ and $\mathbb{E}[Y|\widehat{\pi}(\boldsymbol{X}) = s]$: $s = 10$