How to avoid discrimination in insurance pricing: a probabilistic and learning toolbox

October 2021

Agenda

SPECIAL FOCUS

This thesis documents a probabilistic and hands on learning toolbox to guide insurers in avoiding discrimination in insurance pricing.

Legal framework

“

Any discrimination based on any ground such as sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation shall be prohibited

Article 21 of the EU Charter of Fundamental Rights (2017)

Belgium: Law of May 10 2007

Direct discrimination: direct distinction based on a protected criterion.

Indirect discrimination: The situation that arises when an apparently neutral provision, criterion or course of action could place persons characterized by a particular protected criterion at a particular disadvantage compared to other persons.

Important (discriminatory) characteristics for insurance pricing: **Age**, **Disability** and **Gender**

Exceptions are granted for discrimination when it is justified by a legitimate purpose. However **Gender** does not pass the test for an exception.

Motivating example

$n_{i,j}$	Small	Large	**Row total**
Female	83	14	97
Male	47	60	107
Column total	130	74	204

Table 3.1: Fictional claim counts.

$e_{i,j}$	Small	Large	**Row total**
Female	567	83	650
Male	269	253	522
Column total	836	336	1172

Table 3.2: Fictional exposures.

$\hat{\lambda}_{i,j}$	Small	Large
Female	0.162	0.169
Male	0.175	0.237

Table 3.3: Gender discrimination: estimated claim frequencies per unit of exposure

$$\hat{\lambda}_{\bullet,0} = \frac{130}{836} = 0.156$$

$$\hat{\lambda}_{\bullet,1} = \frac{74}{336} = 0.220.$$

$\Longleftrightarrow$

$$\begin{aligned}\hat{\lambda}_{\bullet,1} = \frac{n_{0,1}+n_{1,1}}{e_{0,1}+e_{1,1}} &= \hat{\lambda}_{0,1}\frac{e_{0,1}}{e_{0,1}+e_{1,1}} + \hat{\lambda}_{1,1}\frac{e_{1,1}}{e_{0,1}+e_{1,1}} \\ &= \hat{\lambda}_{0,1}\hat{\mathbb{P}}(\text{Female} \mid \text{Large}) + \hat{\lambda}_{1,1}\hat{\mathbb{P}}(\text{Male} \mid \text{Large}) \\ &= 0.169 \cdot 0.25 + 0.237 \cdot 0.75 \\ &= 0.22.\end{aligned}$$

Solution:

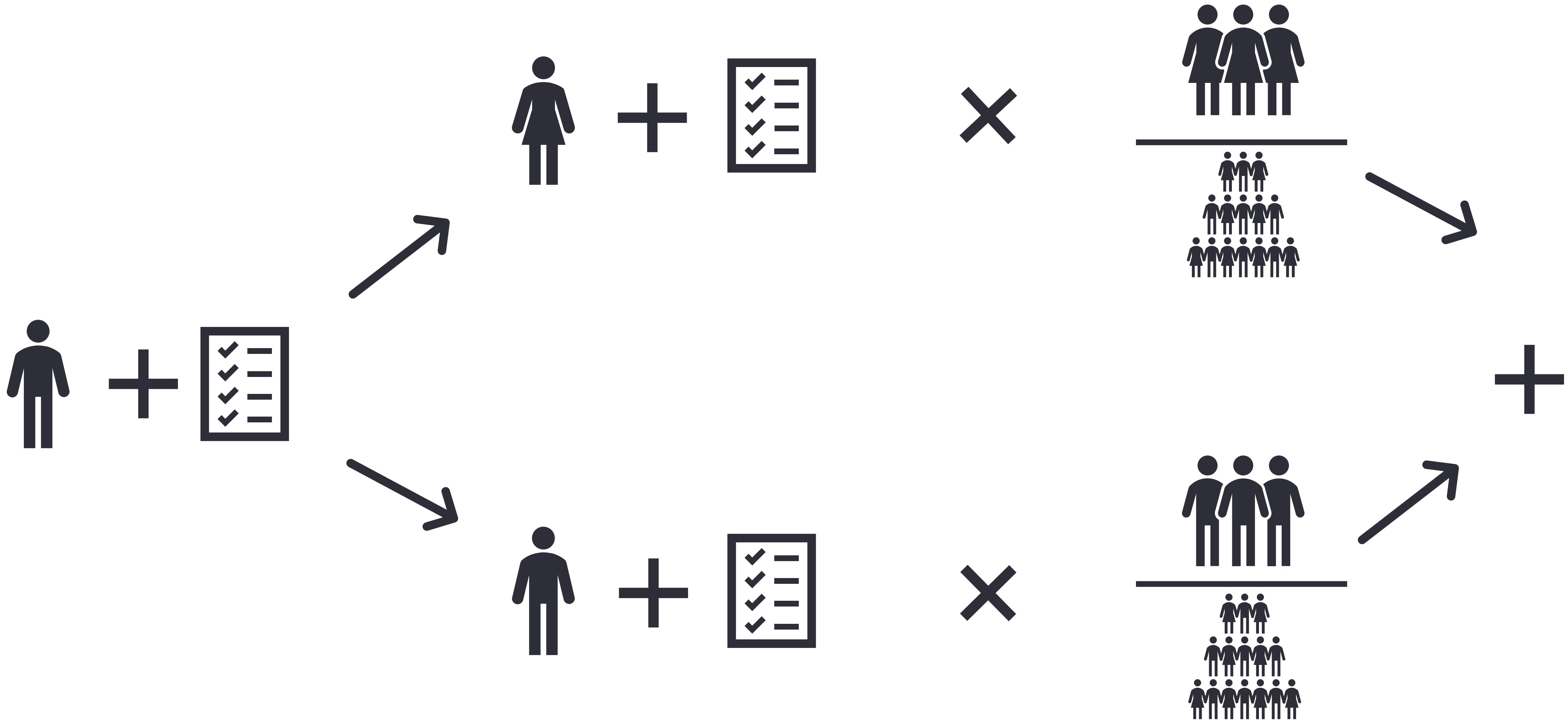
$\Longrightarrow$

$$\hat{\lambda}^{\text{DF}}_{\bullet,j} = \hat{\lambda}_{0,j}\hat{\mathbb{P}}(\text{Female}) + \hat{\lambda}_{1,j}\hat{\mathbb{P}}(\text{Male}).$$

$$\hat{\lambda}^{\text{DF}}_{\bullet,0} = 0.162 \cdot \frac{650}{1172} + 0.175 \cdot \frac{522}{1172} = 0.168$$

$$\hat{\lambda}^{\text{DF}}_{\bullet,1} = 0.169 \cdot \frac{650}{1172} + 0.237 \cdot \frac{522}{1172} = 0.199.$$

Motivating example - intuition



Probabilistic framework

- Best-estimate approach

$$\mu(\boldsymbol{X}, \boldsymbol{D}) := \mathbb{E}[Y \mid \boldsymbol{X}, \boldsymbol{D}].$$

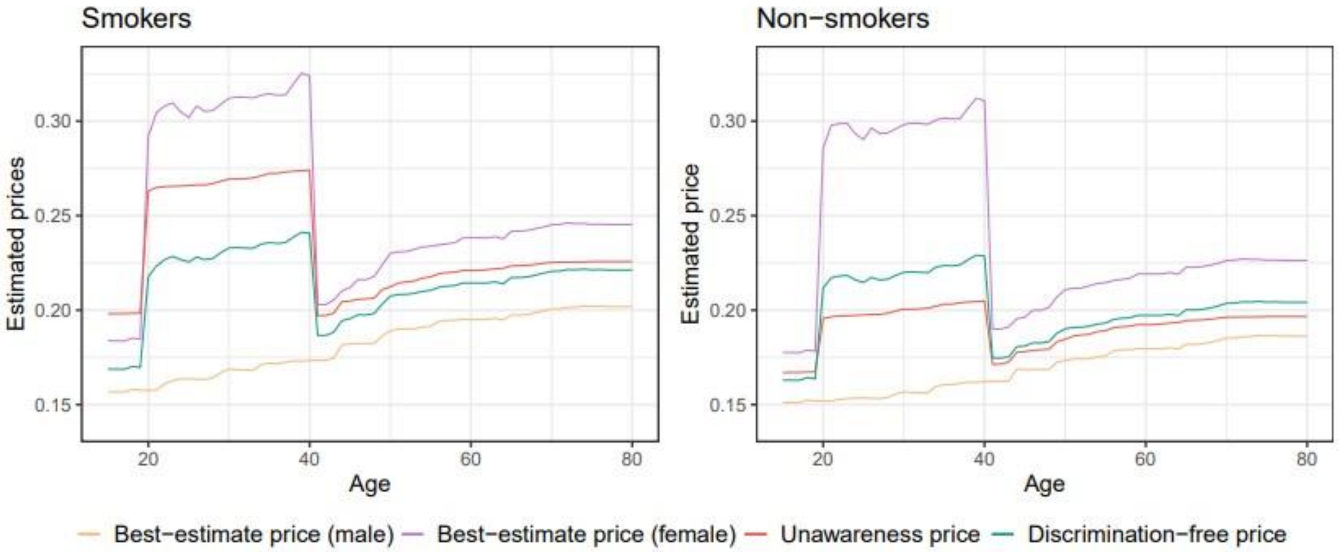
- Unawareness approach

$$\mu(\boldsymbol{X}) := \mathbb{E}[Y \mid \boldsymbol{X}].$$

$$\mu(\boldsymbol{X}) = \int_{\boldsymbol{d}} \mu(\boldsymbol{X}, \boldsymbol{d})\, d\mathbb{P}(\boldsymbol{D} = \boldsymbol{d} \mid \boldsymbol{X}).$$

- Discrimination-free approach

$$h^*(\boldsymbol{X}) := \int_{\boldsymbol{d}} \mu(\boldsymbol{X}, \boldsymbol{d})\, d\mathbb{P}^*(\boldsymbol{D} = \boldsymbol{d}).$$



Application: Private Motor Insurance

PG15training dataset: focus on **third party liability claims**

The males in dataset have a **higher average frequency and severity** compared to the females in the dataset

Two possible underlying models used: **GBM** and **GLM**

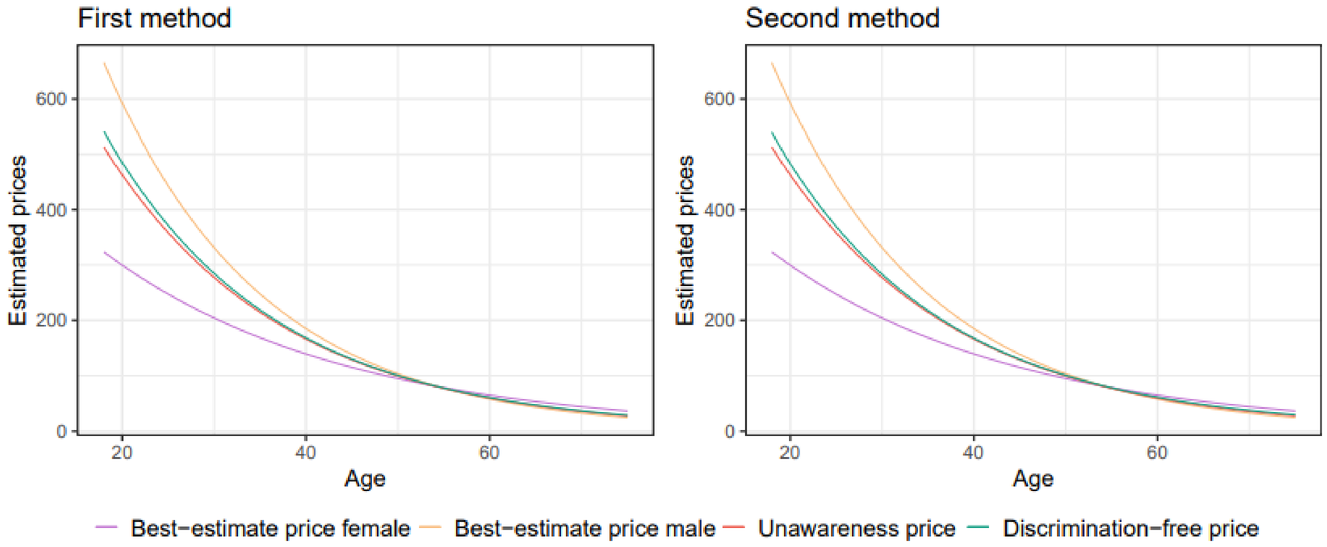
Possible considerations of discriminatory variable(s):

- **Gender**
- **Gender + home region of the driver**
- **Age**

Two types of posibilities to implement method

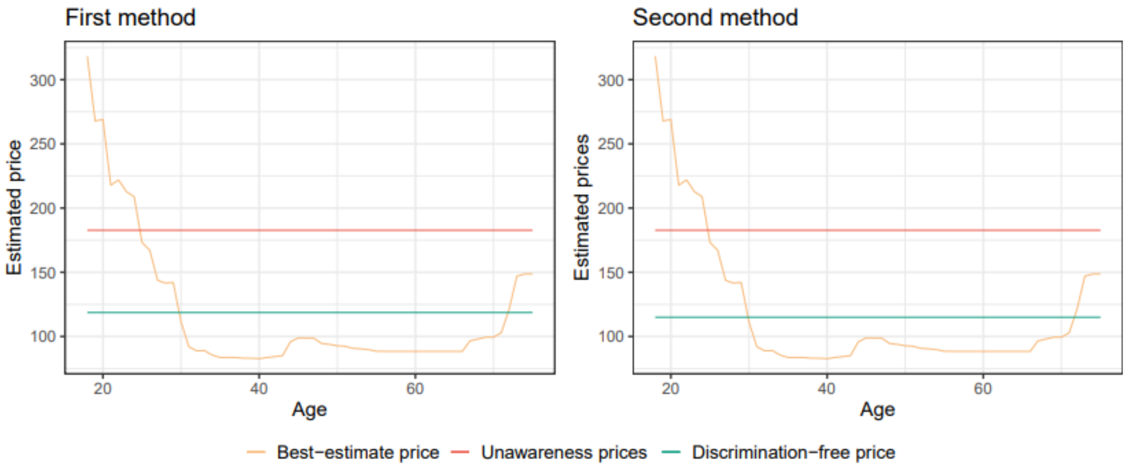
Application: Private Motor Insurance

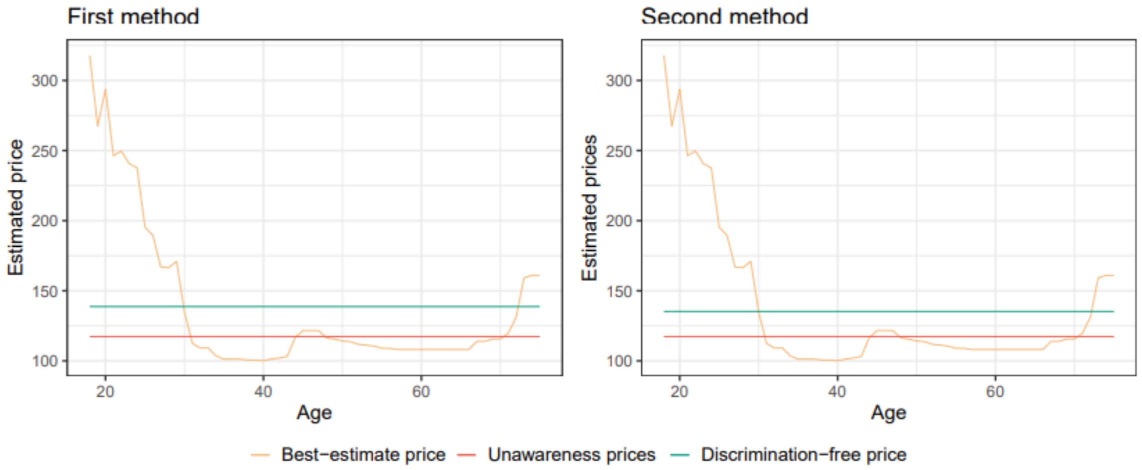
Gender:



Variable	Value
Exposure	1
Group1	19
Bonus	-10
Poldur	6
Group2	L
Density	150
Category	Small
Occupation	Employed
Type	B

Age:





Food for thought and key takeaways