Appendix 1 – Technical details of discrimination-free prices

In this chapter, a rigorous mathematical formulation of the discrimination-free price by Lindhom et. al. (2022) is given.

Let $\boldsymbol{X}_i$ denote the vector of non-discriminatory variables for policyholder i, and $\boldsymbol{D}_i$ the vector of respective discriminatory variables. A claim for policyholder i is denoted with random variable $\boldsymbol{Y}_i$.

The emphasis is on three different prices, which are presented in Table 1.

Table 1

Name	Explanation	Mathematical formulation
Best estimate price	Estimate of the claims cost using all the available information (including discriminatory variables).	$\mathbb{E}(\boldsymbol{Y}_i \mid \boldsymbol{X}_i, \boldsymbol{D}_i)$
Unawareness price	Estimate of the claims cost avoiding direct discrimination.	$\mathbb{E}(\boldsymbol{Y}_i \mid \boldsymbol{X}_i)$
Discrimination-free price	Estimate of the claims cost avoiding both direct and indirect discrimination.	$\int_{\boldsymbol{d}} \mathbb{E}(\boldsymbol{Y}_i \mid \boldsymbol{X}_i, \boldsymbol{d}) d\mathbb{P}^*(\boldsymbol{d})$

The unawareness price corresponds to the ideology the insurers are currently applying to prevent discrimination in their models, that is, developing a model (typically a GLM) which estimates the expected claims costs with given non-discriminatory variables. However, a different equivalent formulation of the unawareness prices indicates that it is possible to perform inference of discriminatory features from non-discriminatory variables[1].

As presented in Table 1, the discrimination-free price is obtained by "averaging" the best estimate prices over the discriminatory variable using a (potentially arbitrary) marginal distribution independent of **X**. A natural choice for such a distribution is the empirical probability distribution obtained from the data. Since the distribution is not dependent on **X**, this approach avoids both direct and indirect discrimination.

One of the advantages of the proposed methodology is that it is not dependent on the underlying models. Whether the best estimate prices are obtained with traditional GLMs or with more advanced methods like neural networks, the methodology works similarly, as it only uses the outputs of the models.

To illustrate this methodology, a simple example is presented in Appendix 2.

[1] The unawareness price can be mathematically equivalently written as $\mathbb{E}(\boldsymbol{Y} \mid \boldsymbol{X}) = \int_{\boldsymbol{d}} \mathbb{E}(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{d}) d\mathbb{P}(\boldsymbol{d} \mid \boldsymbol{X})$. The potential for discrimination arises from the conditional probabilities $\mathbb{P}(\boldsymbol{d} \mid \boldsymbol{X})$, as they enable inference of the discriminatory features from non-discriminatory variables.

ÖSTERMALMSTORG 1, 5TR
SE-114 42, STOCKHOLM, SVERIGE
FCG.GLOBAL