Sensitivity-based measures of discrimination in insurance pricing

Mathias Lindholm
Stockholm University

IDSC, June 19, 2025

Outline

- One slide on non-life pricing
- Discrimination and proxy-discrimination
- Measuring proxy-discrimination

Outline

This presentation is based on joint work with

- Ron Richman
  (insureAI & University of the Witwatersrand, South Africa)
- Andreas Tsanakas
  (Bayes Business School, City St George's, University of London)
- Mario V. Wüthrich
  (ETH Zürich)

Outline

This presentation is based on joint work with

- Ron Richman
  (insureAI & University of the Witwatersrand, South Africa)
- Andreas Tsanakas
  (Bayes Business School, City St George's, University of London)
- Mario V. Wüthrich
  (ETH Zürich)

Particular focus will be on [11]

"*Sensitivity-Based Measures of Discrimination in Insurance Pricing.*"

available at SSRN, Manuscript ID 4897265.

Non-life pricing

Let

- $Y \in \mathbb{R}$ be the response of interest, e.g. claim cost
- $X \in \mathbb{X}$ be a covariate vector (characteristics/rating factors/features/...)
- $\mu(X) := \mathbb{E}[Y \mid X]$ be the actuarial price

Remark.
Model agnostic: use your favourite model class to describe $\mu(X)$

Discrimination

(EU-style)

Discrimination

Definition 1

Direct discrimination: where one person is treated less favourably, on grounds of sex, than another is, has been or would be treated in a comparable situation;

Discrimination

Definition 1

Direct discrimination: where one person is treated less favourably, on grounds of sex, than another is, has been or would be treated in a comparable situation;

Definition 2

Indirect discrimination: where an apparently neutral provision, criterion or practice would put persons of one sex at a particular disadvantage compared with persons of the other sex, unless that provision, criterion or practice is objectively justified by a legitimate aim and the means of achieving that aim are appropriate and necessary;

Discrimination

In other words:

- ☞ "apparently neutral" – **proxy-discrimination**
- ☞ "disadvantage" – materiality of the procedure

  $\Longrightarrow$ "**measures**"

Discrimination

As before, let

- $Y \in \mathbb{R}$ be the response of interest, e.g. claim cost
- $X \in \mathbb{X}$ be non-protected characteristics

Discrimination

As before, let

- $Y \in \mathbb{R}$ be the response of interest, e.g. claim cost
- $X \in \mathbb{X}$ be non-protected characteristics

In addition, let

- $D \in \mathbb{D}$ be protected characteristics
- $\mu(X, D) := \mathbb{E}[Y \mid X, D]$ be the best-estimate (BE) price
- $\mu(X) := \mathbb{E}[Y \mid X]$ be the unawareness price

Discrimination

As before, let

- $Y \in \mathbb{R}$ be the response of interest, e.g. claim cost
- $X \in \mathbb{X}$ be non-protected characteristics

In addition, let

- $D \in \mathbb{D}$ be protected characteristics
- $\mu(X, D) := \mathbb{E}[Y \mid X, D]$ be the best-estimate (BE) price
- $\mu(X) := \mathbb{E}[Y \mid X]$ be the unawareness price

Henceforth, focus is on conditional expectations ("fair prices")

Discrimination

Given the above:

- the BE price $\mu(X, D)$ is **directly discriminatory**, since it depends on D
- the unawareness price $\mu(X)$ is potentially **indirectly discriminatory**

Discrimination

Given the above:

- the BE price $\mu(X, D)$ is **directly discriminatory**, since it depends on D
- the unawareness price $\mu(X)$ is potentially **indirectly discriminatory**

Thus, the tricky part is the situation with $\mu(X)$

Proxy-discrimination and discrimination-free pricing

- Note that $\mu(X)$ can be re-written according to

$$\mu(X) = \sum_d \mu(X, d)\mathbb{P}(D = d \mid X), \qquad (1)$$

where

  - $\mu(X, D)$ describes the impact of X and D on Y
  - $\mathbb{P}(D = d \mid X)$ describes the dependence between X and D

Proxy-discrimination and discrimination-free pricing

- Note that $\mu(X)$ can be re-written according to

$$\mu(X) = \sum_d \mu(X, d)\mathbb{P}(D = d \mid X), \tag{1}$$

  where
  - $\mu(X, D)$ describes the impact of X and D on Y
  - $\mathbb{P}(D = d \mid X)$ describes the dependence between X and D

- In order for $\mu(X)$ to be proxy-discriminatory it is **necessary that both** of the following two conditions hold:
  - ☞ $\mu(X, D) \neq \mu(X)$
  - ☞ $\mathbb{P}(D = d \mid X) \neq \mathbb{P}(D = d)$, for some d

Proxy-discrimination and discrimination-free pricing

- Consider the following adjusted price:

$$\mu^*(X) = \sum_d \mu(X, d)\mathbb{P}^*(D = d), \tag{2}$$

where $\mathbb{P}^*$ is any marginal distribution of D

Proxy-discrimination and discrimination-free pricing

- Consider the following adjusted price:

$$\mu^*(X) = \sum_d \mu(X, d)\mathbb{P}^*(D = d), \tag{2}$$

  where $\mathbb{P}^*$ is any marginal distribution of D

- By using $\mathbb{P}^*$ instead of $\mathbb{P}$ in (2) any potential statistical dependence between X and D is removed – $\mu^*(X)$ **is (proxy) discrimination-free**

Proxy-discrimination and discrimination-free pricing

- Consider the following adjusted price:

$$\mu^*(X) = \sum_d \mu(X, d)\mathbb{P}^*(D = d), \tag{2}$$

  where $\mathbb{P}^*$ is any marginal distribution of D
- By using $\mathbb{P}^*$ instead of $\mathbb{P}$ in (2) any potential statistical dependence between X and D is removed
  – $\mu^*(X)$ **is (proxy) discrimination-free**
- The discrimination-free insurance price (DFIP) $\mu^*(X)$ from (2) was introduced in [7], where more details are discussed

Proxy-discrimination and discrimination-free pricing

- Consider the following adjusted price:

$$\mu^*(X) = \sum_d \mu(X, d)\mathbb{P}^*(D = d), \qquad (2)$$

where $\mathbb{P}^*$ is any marginal distribution of D

- By using $\mathbb{P}^*$ instead of $\mathbb{P}$ in (2) any potential statistical dependence between X and D is removed
  – $\mu^*(X)$ **is (proxy) discrimination-free**
- The discrimination-free insurance price (DFIP) $\mu^*(X)$ from (2) was introduced in [7], where more details are discussed

A lot can be said about DFIP, see e.g. [7, 8, 9, 10] discussing various properties, estimation, relation to notions of algorithmic fairness, causality etc.

Example

Proxy-discrimination and discrimination-free pricing

Example 3.2 in [10]

Assume

- we have two-dimensional covariates (X, D) according to

$$(X, D) \sim f(x, d) = \frac{1}{2} \frac{1}{\sqrt{2\pi\tau^2}} \exp\left\{ -\frac{1}{2\tau^2} (x - x_d)^2 \right\},$$

  with $d \in \mathbb{D} = \{0, 1\}$, $x \in \mathbb{R}$, $\tau^2 > 0$, $x_0 > 0$, $\rho > 0$, and where we set

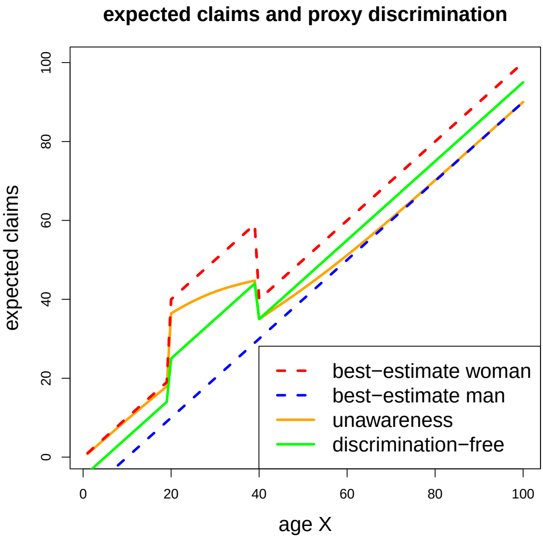
$$x_d = x_0 + \rho d,$$

  where $D = 0$ corresponds to woman, $D \sim \text{Bernoulli}(1/2)$
- that the conditional distribution of Y given (X, D) is given by

$$Y \mid_{(X,D)} \sim \mathcal{N}\left(X + 20(1 - D)\mathbb{1}_{X \in [20,40]} - 10D, 100\right)$$

Proxy-discrimination and discrimination-free pricing

Example 3.2 in [10]



Proxy-discrimination and discrimination-free pricing

Note the following:

- Eq. (2) illustrates that in order to be able to adjust for discrimination, **you need information about D!**
- Collecting and storing data about D can be problematic in itself (see e.g. [8])
- None of the above is a specific problem related to DFIP!!!

Proxy-discrimination and discrimination-free pricing

Definition 3 ([10])

A pricing functional π on $\mathcal{X} \times \mathcal{P}$ avoids proxy-discrimination if for any two portfolios $\mathbb{P}, \mathbb{Q}$ that satisfy $\mathbb{P}(Y \mid X, D) = \mathbb{Q}(Y \mid X, D)$, $\mathbb{P}(D) = \mathbb{Q}(D)$ and $\mathbb{P}(X) = \mathbb{Q}(X)$, we have

$$\pi(X; \mathbb{P}) = \pi(X; \mathbb{Q})$$

Proxy-discrimination and discrimination-free pricing

Definition 3 ([10])

A pricing functional π on $\mathcal{X} \times \mathcal{P}$ avoids proxy-discrimination if for any two portfolios $\mathbb{P}, \mathbb{Q}$ that satisfy $\mathbb{P}(Y \mid X, D) = \mathbb{Q}(Y \mid X, D)$, $\mathbb{P}(D) = \mathbb{Q}(D)$ and $\mathbb{P}(X) = \mathbb{Q}(X)$, we have

$$\pi(X; \mathbb{P}) = \pi(X; \mathbb{Q})$$

N.B. By construction DFIP satisfies Definition 3

Measuring proxy-discrimination

Measuring proxy-discrimination

Materiality of discrimination

- Given a price predictor $\pi(X)$, how can we measure proxy-discrimination?

Measuring proxy-discrimination

Materiality of discrimination

- Given a price predictor $\pi(X)$, how can we measure proxy-discrimination?
- **Idea:** use reference prices $\mu(X, D)$

Measuring proxy-discrimination

Definition 4 ([11])

The pricing functional $X \mapsto \pi(X)$ *avoids proxy discrimination* with respect to $\mu(X, D)$, if for $\mathbb{P}$-almost every X we can write

$$\pi(X) = c + \sum_{d \in \mathfrak{D}} \mu(X, d) v_d, \qquad (3)$$

for some $c \in \mathbb{R}$ and $\boldsymbol{v} \in \mathcal{V}, \mathcal{V} := \{\boldsymbol{v} \in [0,1]^{|\mathfrak{D}|} \; : \; \sum_{d \in \mathfrak{D}} v_d \leq 1\}$, that do not depend on X. If π does not have that structure, we say that it is *proxy-discriminatory*.

Measuring proxy-discrimination

Definition 5 ([11])

The *proxy discrimination metric* PD is defined as

$$\mathrm{PD}(\pi) = \frac{\min_{c \in \mathbb{R},\ \boldsymbol{v} \in \mathcal{V}} \mathbb{E}\left[\left(\pi(\boldsymbol{X}) - c - \sum_{d \in \mathfrak{D}} \mu(\boldsymbol{X}, d) v_d\right)^2\right]}{\mathrm{Var}(\pi(\boldsymbol{X}))}, \quad (4)$$

with the convention that if $\mathrm{Var}(\pi(X)) = 0$, then $\mathrm{PD}(\pi) = 0$.

Measuring proxy-discrimination

Definition 5 ([11])

The *proxy discrimination metric* PD is defined as

$$\mathrm{PD}(\pi) = \frac{\min_{c\in\mathbb{R},\ \boldsymbol{v}\in\mathcal{V}} \mathbb{E}\left[\left(\pi(\boldsymbol{X}) - c - \sum_{d\in\mathfrak{D}} \mu(\boldsymbol{X}, d) v_d\right)^2\right]}{\mathsf{Var}(\pi(\boldsymbol{X}))}, \quad (4)$$

with the convention that if $\mathsf{Var}(\pi(X)) = 0$, then $\mathrm{PD}(\pi) = 0$.

Remarks.

- This is related to the residual variance for the constrained regression of $\pi(\boldsymbol{X})$ on $\mu(\boldsymbol{X}, d),\ d \in \mathfrak{D}$
- This is a type of global sensitivity measure

Measuring proxy-discrimination

Proposition 1 ([11])

The proxy discrimination metric PD *satisfies the following properties.*

i) $0 \leq \mathrm{PD}(\pi) \leq 1$. *Furthermore, for all* $a \in \mathbb{R}, b \in \mathbb{R}_+$ *it holds that* $\mathrm{PD}(a + b\pi) = \mathrm{PD}(\pi)$.

ii) $\mathrm{PD}(\pi) = 0$ *if and only if* π *avoids proxy discrimination with respect to* $\mu(X, D)$.

iii) *If* $\pi(X)$ *is uncorrelated with* $\mu(X, d)$ *for all* $d \in \mathfrak{D}$, *then* $\mathrm{PD}(\mu) = 1$.

Measuring proxy-discrimination

Example, real data in [11]

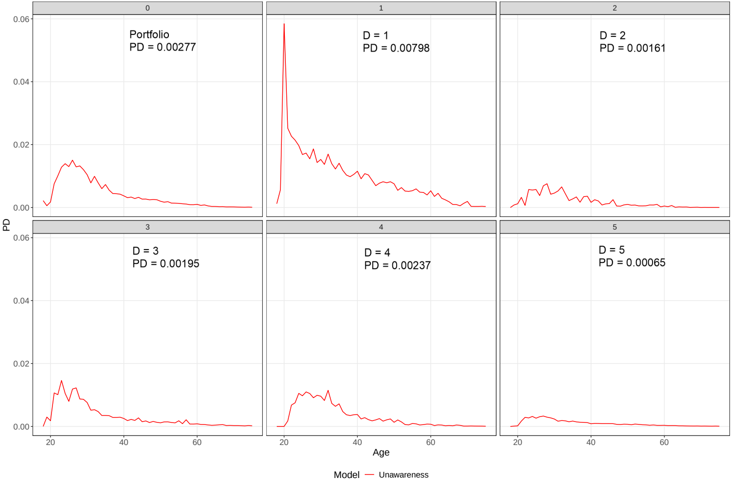


Figure: Real data, $D \in \{1, 2, 3, 4, 5\}$

Summary

- We have discussed definitions of proxy-discrimination
- We have introduced a sensitivity based measure of proxy-discrimination
- This measure relies on a reference model / prices

Summary

- We have discussed definitions of proxy-discrimination
- We have introduced a sensitivity based measure of proxy-discrimination
- This measure relies on a reference model / prices

More things in the paper:

- ☞ How to attribute proxy-discrimination to features
- ☞ More on measuring algorithmic unfairness

Summary

- We have discussed definitions of proxy-discrimination
- We have introduced a sensitivity based measure of proxy-discrimination
- This measure relies on a reference model / prices

More things in the paper:

- ☞ How to attribute proxy-discrimination to features
- ☞ More on measuring algorithmic unfairness

Related research:

- Sensitivity measures, see e.g. [4, 3]
- Algorithmic fairness, see e.g. [2, 6]
- Causality, see e.g. [7, 1, 5]
- Welfare implications, regulation etc, see e.g. [12]

Thank you for your attention!

References I

Araiza Iturria, C.A., Hardy, M., Marriott, P. (2024). A discrimination-free premium under a causal framework. *North American Actuarial Journal*, **28(4)**, 801-821.

Barocas, S., Hardt, M., Narayanan, A. (2019). *Fairness and Machine Learning: Limitations and Opportunities.* https://fairmlbook.org/

Bénesse, C., Gamboa, F., Loubes, J.-M., Boissin, T. (2024). Fairness seen as global sensitivity analysis. *Machine Learning*, **113(5)**, 3205 - 3232.

Borgonovo, E. and Plischke, E. (2016). Sensitivity analysis: A review of recent advances. *European Journal of Operational Research*, **248(3)**, 869 - 887.

References II

Côté, O., Côté, M.-P., and Charpentier, A. (2025). A fair price to pay: Exploiting causal graphs for fairness in insurance. *Journal of Risk and Insurance*, **92(1)**, 33-75.

Dwork, C., Hardt, M., Pitassi, T., Reingold, O., Zemel, R. (2012). Fairness through awareness. *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, 214-226.

Lindholm, M., Richman, R., Tsanakas, A., Wüthrich, M.V. (2022). Discrimination-free insurance pricing. *ASTIN Bulletin* **52(2)**, 55-89.

Lindholm, M., Richman, R., Tsanakas, A., Wüthrich, M.V. (2023). Insurance pricing: Discrimination, Causality, and Fairness. *The European Actuary* **No. 33, March**, 26 - 29.

References III

- Lindholm, M., Richman, R., Tsanakas, A., Wüthrich, M.V. (2024). A multi-output network approach for calculating discrimination-free insurance prices. *European Actuarial Journal*, **14**, 329 - 369.
- Lindholm, M., Richman, R., Tsanakas, A., Wüthrich, M.V. (2024). What is Fair? Proxy Discrimination vs. Demographic Disparities in Insurance Pricing. *Scandinavian Actuarial Journal*, **2024(9)**, 935 - 970.
- Lindholm, M., Richman, R., Tsanakas, A., Wüthrich, M.V. (2024). Sensitivity-Based Measures of Discrimination in Insurance Pricing. *SSRN Manuscript* ID 4897265.
- Xin, X., Huang, F. (2024). Antidiscrimination insurance pricing: Regulations, fairness criteria, and models. *North American Actuarial Journal*, **28(2)**, 285-319.