From Risk Prediction to Fair Insurance
Optimal transport, calibration, and hidden trade-offs

Arthur Charpentier (UQAM)
Agathe Fernandes Machado, Charlotte Jamotton, Sang Huyngh, Lucas Offroy,
Marouane Il Idrissi, François Hu, Philipp Ratz, Oliver Côté, Marie-Pier Côté

Mathematical Advances on Emerging Risks, Merida (Mexico) — November 2026

1. One problem, three different questions

Prediction

What is the expected loss?

$$\mu(\boldsymbol{x}, \boldsymbol{s}) := \mathbb{E}[Y \mid \boldsymbol{X} = \boldsymbol{x}, \boldsymbol{S} = \boldsymbol{s}]$$

See "actuarial fairness".

Insurance as risk transfert, Arrow (1963).

Fairness (DP)

Which distinctions should we allow?

$$\mathcal{L}(\widehat{\mu}(\boldsymbol{X}, \boldsymbol{S}) \mid \boldsymbol{S} = \boldsymbol{s}) \stackrel{?}{=} \mathcal{L}(\widehat{\mu}(\boldsymbol{X}, \boldsymbol{S}) \mid \boldsymbol{S} = \boldsymbol{s}')$$

Pricing

What should each insured pay?

$$\pi(\boldsymbol{x}, \boldsymbol{s}) \stackrel{?}{=} \widehat{\mu}(\boldsymbol{x}, \boldsymbol{s})$$

See "cake-cutting"

Insurance as risk sharing Charpentier (2026a).

A more accurate predictor does not automatically produce a fairer price.

See Charpentier (2024, 2026b).

An actuarial fairness problem: notation

possibly multiple coverage $\boldsymbol{Y}$, Huynh and Charpentier (2026)

$$(\boldsymbol{X}, S, Y) \sim \mathbb{P}, \qquad S \in \{0,1\}, \quad Y \geq 0, \qquad Z = f(\boldsymbol{X}, S) \in \mathbb{R}.$$

possibly multiple sensitive attributes $\boldsymbol{S}$, Fernandes Machado et al. (2026a)

Two statistical objectives

Predictive accuracy
$\mathcal{R}(f) = \mathbb{E}[\ell(Y, f(\boldsymbol{X}, S))]$

Distributional demographic parity
$Z \perp\!\!\!\perp S$

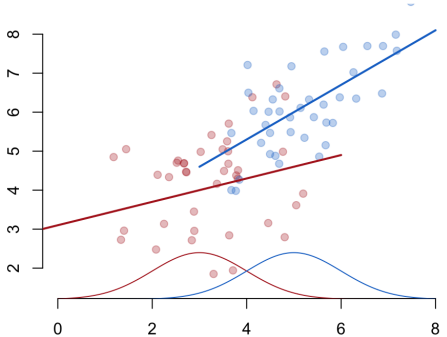
Two different interpretations

Proxy discrimination: $\boldsymbol{X}$ may reveal S.

Causal discrimination: direct / indirect pathways require a structural model.

Disparity of score distributions is a *statistical diagnostic*, not by itself a measure of harm or legal discrimination.

Fairness by awarness ("actuarial fairness")



- best estimate (on full information)

$$\begin{cases} \mathbb{E}[Y \mid X, s=1] = \alpha_1 + \beta_1 x \\ \mathbb{E}[Y \mid X, s=0] = \alpha_0 + \beta_0 x \end{cases}$$

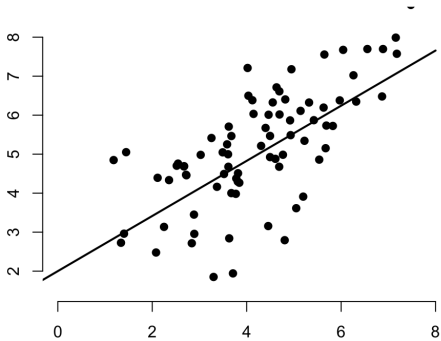
obtained solving

$$(\alpha_1, \beta_1) = \arg\min_{a,b} \sum_{i:s_i=1} (y_i - (a+bx_i))^2$$

$$(\alpha_0, \beta_0) = \arg\min_{a,b} \sum_{i:s_i=0} (y_i - (a+bx_i))^2$$

See e.g., Denuit et al. (2019); Charpentier (2024)

Fairness by **un**awarness



See e.g., Charpentier (2024)

- "blind" with respect to s
- best estimate (on x only)

$$\mathbb{E}[Y \mid X] = \alpha + \beta x$$

obtained solving

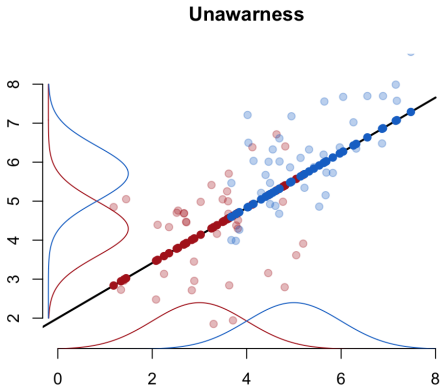
$$(\alpha, \beta) = \arg\min_{a,b} \sum_i \big(y_i - (a + bx_i)\big)^2$$

- if $X \mid S = s \sim \mathcal{N}(\mu_s, \sigma_s^2)$,

$$b = \frac{\pi_0\beta_0\sigma_0^2 + \pi_1\beta_1\sigma_1^2 + \pi_0\pi_1(\mu_1 - \mu_0)(m_1 - m_0)}{\pi_0\sigma_0^2 + \pi_1\sigma_1^2 + \pi_0\pi_1(\mu_1 - \mu_0)^2},$$

$$a = \bar{m} - b\bar{\mu}, \text{ where } m_s = \alpha_s + \beta_s\mu_s,$$
$$\bar{\mu} = \pi_0\mu_0 + \pi_1\mu_1,\ \bar{m} = \pi_0 m_0 + \pi_1 m_1.$$

Fairness by **un**awarness



- Hence, in general

$$\underbrace{(\widehat{Y} \mid S=0)}_{\sim\mathcal{N}\left(a+b\mu_0,\, b^2\sigma_0^2\right)} \overset{\mathcal{L}}{\neq} \underbrace{(\widehat{Y} \mid S=1)}_{\sim\mathcal{N}\left(a+b\mu_1,\, b^2\sigma_1^2\right)}$$

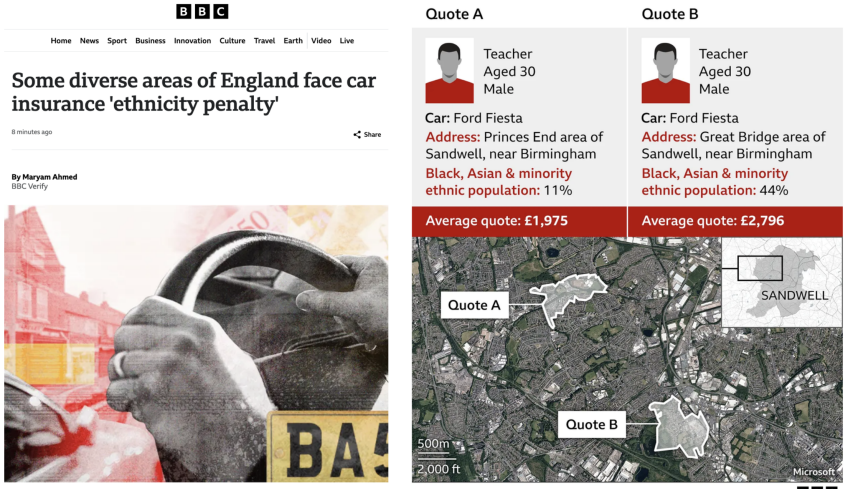
(no strong Demographic Parity)

$$\mathbb{E}[\widehat{Y} \mid S=0] \neq \mathbb{E}[\widehat{Y} \mid S=1]$$

(not even weak Demographic Parity)

Idea $\mathbb{P}[S \mid X] \neq \mathbb{P}[S]$, Lindholm et al. (2022) (propensity score).

Some policyholders face insurance “ethnicity penalty” ?



Source: https://www.bbc.com/news/business-68349396.

An old question: where does a disparity come from?

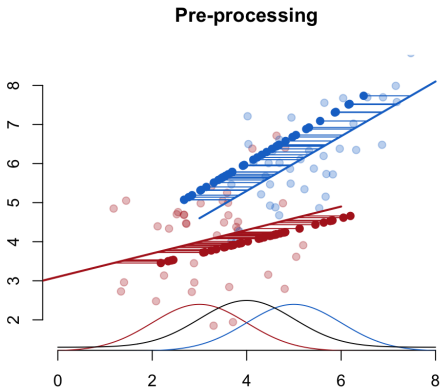
For two groups, (Kitagawa)–Oaxaca–Blinder separates an average outcome gap into differences in *characteristics* and differences in *coefficients*:

$$\bar{Y}_0 - \bar{Y}_1 = \underbrace{(\bar{X}_0 - \bar{X}_1)^\top \beta_1}_{\text{composition / characteristics}} + \underbrace{(\alpha_0 - \alpha_1) + \bar{X}_0^\top (\beta_0 - \beta_1)}_{\text{returns / coefficients}}.$$

Modern fairness work keeps this intuition, but changes the object: from observed outcome gaps to disparities generated by a fitted model. See Kitagawa (1955); Oaxaca (1973); Blinder (1973).

With GLMs, the nonlinear inverse link adds *curvature coupling* and *curvature amplification*, see Tierny et al. (2026a); GLM extension and Oaxaca–Blinder connection: Tierny et al. (2026b).

Route A — change the information used by the model



See e.g., Charpentier (2024)

- Shift the covariates in $x \longrightarrow x'$ so that

$$(X' \mid S = 0) \overset{\mathcal{L}}{=} (X' \mid S = 1)$$

- If $m_s = \alpha_s + \beta_s \mu_s$ and $\tau_s = \beta_s \sigma_s$,

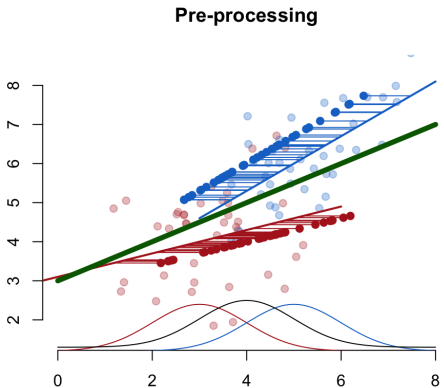
$$x' \longleftarrow \mu^\star + \frac{\sigma^\star}{\sigma_s}(x - \mu_s).$$

- Then

$$(X' \mid S = s) \sim \mathcal{N}(m^\star, \tau^\star), \ \forall s,$$

$$m^\star = \sum_s \pi_s m_s \text{ and } \tau^\star = \sum_s \pi_s \tau_s.$$

Route A — change the information used by the model



- Fit a model on $(x', \hat{y})$

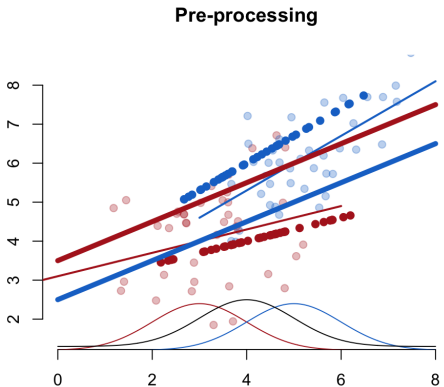
$$(\alpha_\star, \beta_\star) = \arg\min_{a,b} \sum_i \left(y_i - (a + bx_i')\right)^2$$

i.e.,

$$\underbrace{m^\star - \frac{\tau^\star}{\sigma^\star}\mu^\star}_{=\alpha_\star} + \underbrace{\frac{\tau^\star}{\sigma^\star}}_{=b_\star} x'.$$

See e.g., Charpentier (2024)

Route A — change the information used by the model



- Then shift back $x' \longrightarrow x$

$$\widehat{y}_{s,\star}(x) = m^\star + \frac{\tau^\star}{\sigma_s}(x - \mu_s)$$

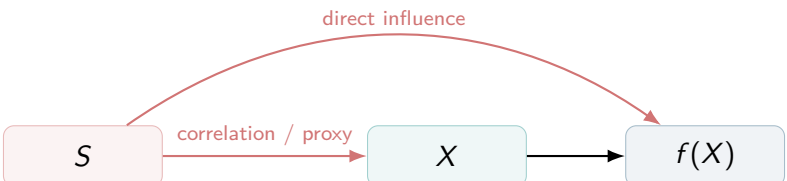
i.e,

$$\begin{cases} \widehat{y}_{s=1,\star}(x) = \alpha_1^\star + \beta_1^\star x \\ \widehat{y}_{s=0,\star}(x) = \alpha_0^\star + \beta_0^\star x \end{cases}$$

See e.g., Charpentier (2024) or Côté et al. (2025).

Route A — change the information used by the model

But if X carries information about S, the unaware approach does not remove proxy discrimination, see Causal Mediation Analysis, Fernandes Machado et al. (2026b).



- Lindholm et al. (2022): construct prices free of direct and proxy discrimination.
- Lindholm et al. (2024): proxy discrimination and demographic disparities are distinct objects.
- Input-side adversarial / representation approaches provide another way to reduce recoverability of S from X, Grari et al. (2022).

Route B — change the optimization problem

- Minimize quadratic risk

$$(\alpha_\star, \beta_\star) = \underset{(a_0,a_1),(b_0,b_1)}{\arg\min} \sum_i \big(y_i - (a_{s_i} + b_{s_i} x_i)\big)^2$$

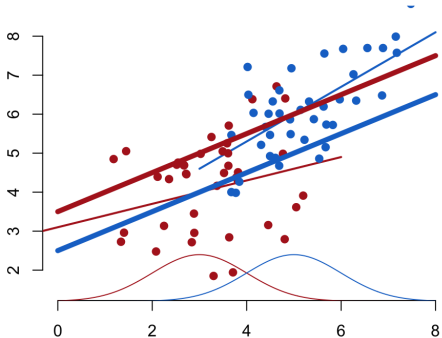
under the constraint that

$$\underbrace{(\widehat{Y} \mid S=0)}_{\sim\mathcal{N}(a_0+b_0\mu_0, b_0\sigma_0)} \overset{\mathcal{L}}{=} \underbrace{(\widehat{Y} \mid S=1)}_{\sim\mathcal{N}(a_1+b_1\mu_1, b_1\sigma_1)}$$

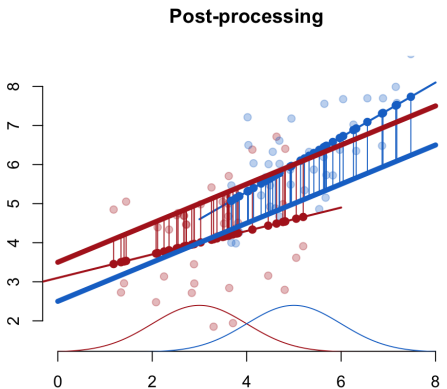
i.e., $\gamma_s + \delta_s\mu_s = m$ and $\delta_s\sigma_s = \tau$.

$$\min_{m,\tau} \sum_s \pi_s \sum_{i:s_i=s} \left(y_i - m - \frac{\tau}{\sigma_s}(x_i - \mu_s) \right)^2$$

$$\widehat{y}_{s,\star}(x) = m^\star + \frac{\tau^\star}{\sigma_s}(x - \mu_s)$$



Route C — change the distribution of prices



- Use best estimate (on full information)

$$\begin{cases} \widehat{y}_{s=1}(x) = \alpha_1 + \beta_1 x \\ \widehat{y}_{s=0}(x) = \alpha_0 + \beta_0 x \end{cases}$$

- Shift the predictions $\widehat{y} \xrightarrow{T_s} \widehat{y}'$

$$\underbrace{(\widehat{Y}' \mid S=0)}_{\sim\mathcal{N}(a_0+b_0\mu_0, b_0\sigma_0)} \overset{\mathcal{L}}{=} \underbrace{(\widehat{Y}' \mid S=1)}_{\sim\mathcal{N}(a_1+b_1\mu_1, b_1\sigma_1)}$$

and again, since $\widehat{y}_{s,\star}(x) = T_s(\widehat{y}_s(x))$,

$$\widehat{y}_{s,\star}(x) = m^\star + \frac{\tau^\star}{\sigma_s}(x - \mu_s)$$

Connections with Wasserstein geometry

For probability laws $\mathbb{P}, \mathbb{Q}$ with finite second moments and quantile functions $F_{\mathbb{P}}^{-1}, F_{\mathbb{Q}}^{-1}$:

$$W_2^2(\mathbb{P}, \mathbb{Q}) = \int_0^1 \left| F_{\mathbb{P}}^{-1}(u) - F_{\mathbb{Q}}^{-1}(u) \right|^2 \mathrm{d}u.$$

Let $\mathbb{P}_s = \mathcal{L}(Z \mid S = s)$, $\pi_s = \mathbb{P}(S = s) > 0$, $\mathbb{Q}^* \in \arg\min_{\mathbb{Q}} \sum_s \pi_s W_2^2(\mathbb{P}_s, \mathbb{Q})$,

$$F_{\mathbb{Q}^*}^{-1}(u) := \sum_s \pi_s F_{\mathbb{P}_s}^{-1}(u) := \overline{Q}(u)$$

Define the distributional disparity of a score function f

$$U(Z) = \min_{\mathbb{Q}} \sum_s \pi_s W_2^2(\mathbb{P}_s, \mathbb{Q}), \text{ where } Z = f(\boldsymbol{X}, S).$$

A remarkably simple contraction law

$$U(Z) = \min_{\mathbb{Q}} \sum_s \pi_s W_2^2(\mathbb{P}_s, \mathbb{Q}), \quad U(Z) = 0 \iff Z = f(\boldsymbol{X}, S) \perp\!\!\!\perp S$$

Interpolate between the original (unfair) score and its barycentric repair:

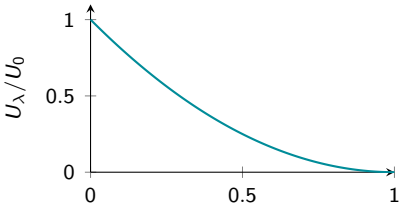
$$Z_\lambda = (1-\lambda)Z + \lambda T_{Z|S}(Z), \qquad \lambda \in [0,1] \text{ and } T_s(z) = \overline{Q}(F_s(z)), \qquad Z^* = T_{Z|S}(Z).$$

$$U(Z_\lambda) = (1-\lambda)^2\, U(Z).$$

(sketch) The transported conditional quantile is

$$Q_{s,\lambda} = (1-\lambda)Q_s + \lambda \overline{Q}.$$

Its barycenter remains $\overline{Q}$.



Three routes to "fair" scores — different guarantees

Pre-processing

Transform inputs:

$$\boldsymbol{X} \overset{T_{x|s}}{\longmapsto} \boldsymbol{X}^*$$

Changes the information given to the model.

In-processing

Constrain learning:

$$\min_f \ \mathcal{R}(f) \text{ s.t. } U(f) \leq \varepsilon$$

Balances prediction and disparity.

Post-processing

Transport predictions:

$$Z \overset{T_{z|s}}{\longmapsto} Z^*$$

May change predictive calibration.

A distributional guarantee is not a causal guarantee or a calibration guarantee.

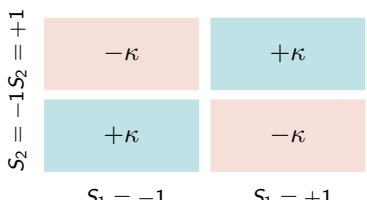
See Hu et al. (2024), Fernandes Machado et al. (2025).

2. Marginal fairness is not joint fairness

Take independent Rademacher attributes $S_1, S_2 \in \{-1, +1\}$ and

$$Z = \kappa S_1 S_2 + \varepsilon, \qquad \varepsilon \perp\!\!\!\perp (S_1, S_2), \quad \mathbb{E}[\varepsilon] = 0.$$

$$Z \perp\!\!\!\perp S_1, \quad Z \perp\!\!\!\perp S_2,$$
$$Z \not\perp\!\!\!\perp (S_1, S_2) \quad (\kappa \neq 0).$$



Conditional laws within either marginal group are the same mixture; intersectional cells are not, see Fernandes Machado et al. (2026a).

How do we audit multiple protected attributes?

Let $\boldsymbol{S} = (S_1, \ldots, S_k)$ be finite-valued, $Z \in L^2$, and denote $\mathbb{P}_{\boldsymbol{s}} = \mathcal{L}(Z \mid \boldsymbol{S} = \boldsymbol{s})$.

$$U_{\text{joint}}(Z) = \inf_{\mathbb{Q}} \sum_{\boldsymbol{s}} \pi_{\boldsymbol{s}} W_2^2(\mathbb{P}_{\boldsymbol{s}}, \mathbb{Q}), \qquad \pi_{\boldsymbol{s}} = \mathbb{P}(\boldsymbol{S} = \boldsymbol{s}).$$

Marginal audits:

$$U_j(Z) = \inf_{\mathbb{Q}} \sum_{s_j} \pi_{j,s_j} W_2^2(\mathcal{L}(Z \mid S_j = s_j), \mathbb{Q})$$

Two useful facts

$U_{\text{joint}} \geq \max_j U_j$.

But no general identity $U_{\text{joint}} = \sum_j U_j$.

Marginal distributions are mixtures of joint-cell distributions; *quantiles of mixtures* are not generally *mixtures of quantiles*.

See Jamotton et al. (2026)

An exact Hilbert-space representation

Let $H = L^2(0,1)$ and define the full-cell conditional quantile

$$Q_{\boldsymbol{S}} : u \mapsto F^{-1}_{Z|\boldsymbol{S}}(u) \in H, \qquad \overline{Q} = \mathbb{E}[Q_{\boldsymbol{S}}] \in H.$$

The joint Wasserstein disparity is precisely its Hilbert-space variance:

$$U_{\text{joint}}(Z) = \mathbb{E}\,\|Q_{\boldsymbol{S}} - \overline{Q}\|^2_H.$$

We can now use the geometry of $L^2(0,1)$, conditional expectations, and orthogonal martingale increments to decompose one *fixed* joint disparity.

See Fernandes Machado et al. (2026a)
For $A \subseteq \{1,\dots,k\}$, write $\boldsymbol{S}_A = (S_j)_{j\in A}$ and define

$$v(A) = \mathbb{E}\Big\|\mathbb{E}[Q_{\boldsymbol{S}} \mid \boldsymbol{S}_A] - \mathbb{E}[Q_{\boldsymbol{S}}]\Big\|^2_H.$$

Following Hu et al. (2024), we wanted to consider a sequential approach $A \to A \cup \{j\}$.

From information order to Shapley attribution

We have

$$\underbrace{U_{\text{joint}}}_{\text{total}} = \underbrace{v(A)}_{\text{explained by } A} + \underbrace{\mathbb{E}\left[\|Q_{\boldsymbol{S}} - \mathbb{E}[Q_{\boldsymbol{S}} \mid \boldsymbol{S}_A]\|_H^2\right]}_{\text{residual given } A}$$

Information game

$0 = v(\varnothing) \leq v(A) \leq v(B)$ for $A \subseteq B$;

$$v(\{1, \ldots, k\}) = U_{\text{joint}}.$$

Every order yields *orthogonal increments* summing to U_{joint}.

Order-free allocation

$$\phi_j = \sum_{A \not\ni j} \frac{|A|!(k - |A| - 1)!}{k!} \big[v(A \cup \{j\}) - v(A)\big].$$

$$\sum_j \phi_j = U_{\text{joint}}.$$

Important distinction: $v(\{j\})$ is *not generally* the marginal Wasserstein audit U_j (conditional averaging and mixture quantiles differ).

Attribution is not intervention

Suppose a correction targets S_1 and produces scores Z_λ.

$$\mathcal{E}_{1\to 2}(\lambda) = U_2(Z_\lambda) - U_2(Z), \qquad \mathcal{E}_{1\to \text{joint}}(\lambda) = U_{\text{joint}}(Z_\lambda) - U_{\text{joint}}(Z).$$

Attribution: a fixed disparity

How much of the joint disparity should be assigned to each dimension?

$$\sum_j \phi_j = U_{\text{joint}}(Z).$$

Intervention: a change in disparity

What happens when we repair a particular dimension?

$$U_1(Z_\lambda) = (1-\lambda)^2 U_1(Z).$$

Shapley values *allocate* disparity; externalities measure changes *induced by a specified repair*.

Gaussian attributes, mitigating S_1

Let (S_1, S_2) be jointly Gaussian, with unit variances and $\text{Cov}(S_1, S_2) = \rho \in (-1, 1)$. Consider

$$Z = \beta_1 S_1 + \beta_2 S_2 + R_S, \qquad R_S \perp\!\!\!\perp (S_1, S_2), \quad \mathbb{E}[R_S] = 0, \quad \mathbb{E}[R_S^2] < \infty.$$

Set $c_1 = \beta_1 + \rho\beta_2$, $c_2 = \beta_2 + \rho\beta_1$.

Conditional laws are translations $\Rightarrow$ exact barycentric repair:

$$Z_\lambda = Z - \lambda c_1 S_1.$$

$$\begin{aligned} U_1(Z_\lambda) &= (1-\lambda)^2 c_1^2, \\ U_2(Z_\lambda) &= (c_2 - \lambda\rho c_1)^2, \\ U_{\text{joint}}(Z_\lambda) &= U_{\text{joint}}(Z) - \lambda(2-\lambda)c_1^2. \end{aligned}$$

β_j may be projection coefficients: S_2 need not be an explicit input to the predictor.

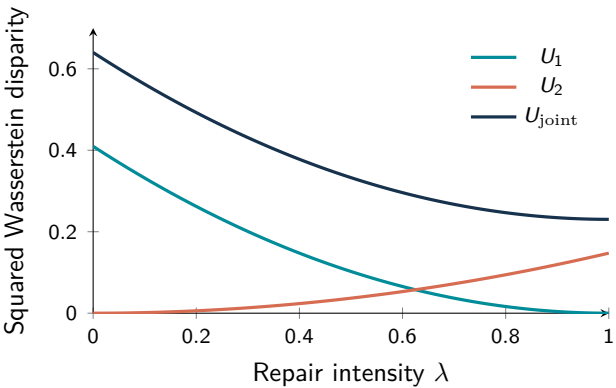
Robbing Peter to pay Paul, mitigating S_1

Parameters: $\beta_1 = 1$, $\beta_2 = -0.6$, $\rho = 0.6$

Audit	$\lambda = 0$	$\lambda = 1$
U_1	0.4096	0
U_2	0	0.1475
U_{joint}	0.6400	0.2304

Masking and unmasking

The targeted and joint disparities decrease, **but an initially absent S_2 disparity emerges**.

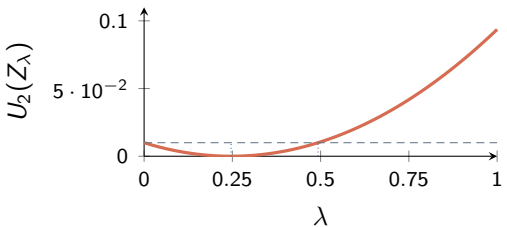


The trajectories do not depend on the distribution of the independent residual R_S., Jamotton et al. (2026).

An even subtler effect, mitigating S_1, non-monotone externalities

For $\beta_1 = 1$, $\beta_2 = -0.6$, $\rho = 0.7$, we have

$$c_1 = 0.58, \qquad c_2 = 0.10, \qquad U_2(Z_\lambda) = (0.10 - 0.406\lambda)^2.$$



U_2 first **decreases**, then **increases**.

$$\lambda_* = \frac{c_2}{\rho c_1} \simeq 0.246.$$

It returns to its initial level at

$$\lambda_0 = \frac{2c_2}{\rho c_1} \simeq 0.493,$$

$\mathcal{E}_{1\to 2}(\lambda) > 0$ beyond λ_0.

More repair of S_1 can first help, then harm the S_2 audit.

A hidden attribute can become less fair

S_2 is unavailable and never enters the predictor

Assume standardized jointly Gaussian $(S_1, S_2, X^\star)$, with

$$\widetilde{Z} = \alpha_1 S_1 + \alpha_2 X^\star, \qquad X^\star \perp\!\!\!\perp S_1, \qquad \mathrm{Cov}(S_1, S_2) = \rho, \quad \mathrm{Cov}(X^\star, S_2) = \tau.$$

Repair targeting only S_1:

$$\widetilde{Z}_\lambda = \widetilde{Z} - \lambda\alpha_1 S_1, \qquad U_2(\widetilde{Z}_\lambda) = \big[(1-\lambda)\alpha_1\rho + \alpha_2\tau\big]^2.$$

An exact masking example

$\alpha_1 = \alpha_2 = 1$, $\rho = 0.6$, $\tau = -0.6$.

The two contributions to the hidden S_2 audit initially cancel:

$$U_2(\widetilde{Z}_\lambda) = 0.36\lambda^2.$$

Audit	$\lambda = 0$	$\lambda = 1$
U_1	1	0
U_2	0	0.36
U_{joint}	1.5625	0.5625

Mitigating S_1 disparity can create S_2 disparity, although never used to compute the score.

Hidden attribute: what can we identify?

Known: α_1, α_2 and ρ (external data); unknown: τ

Positive semidefiniteness gives $|\tau| \leq \sqrt{1-\rho^2}$.

$$\mathcal{E}_{1\to 2}(\lambda) = m_\lambda - 2\lambda\alpha_1\alpha_2\rho\tau, \qquad m_\lambda = -\lambda(2-\lambda)\alpha_1^2\rho^2.$$

$$r_\lambda = 2\lambda|\alpha_1\alpha_2\rho|\sqrt{1-\rho^2}.$$

$$\mathcal{E}_{1\to 2}(\lambda) \in [m_\lambda - r_\lambda,\ m_\lambda + r_\lambda] \qquad \text{(sharp bounds)}$$

A central scenario

$\tau = 0$ (conditional independence) gives

$$\mathcal{E}_{1\to 2}(\lambda) = m_\lambda \leq 0.$$

Certificate for full repair

If $\alpha_1\rho \neq 0$, no deterioration is possible iff

$$2\alpha_2\sqrt{1-\rho^2} \leq |\alpha_1\rho|.$$

Assume $\alpha_2 \geq 0$, $|\rho| < 1$; if $\alpha_1\rho = 0$, the externality is zero. Bounds concern *identification*, not sampling.

3. What does an individual probability mean?

For a binary claim indicator $Y \in \{0, 1\}$, the predictive score is

$$\widehat{p} = f(X) \in [0, 1], \qquad \text{e.g. } \widehat{p}_i = 0.02.$$

One policyholder

A claim occurs or it does not.
A single realization cannot validate the quoted 2%.

A population of forecasts

Calibration is testable collectively:

$$\mathbb{E}[Y \mid \widehat{p}] = \widehat{p}.$$

Individualized forecasts can be empirically accountable without an observable "true individual probability".

Dawid (1982, 2017), Roth and Tolbert (2025), Barry and Charpentier (2026).

Calibration and strong demographic parity can conflict

A short incompatibility result

Assume the *strong* group-calibration condition

$$\mathbb{E}[Y \mid Z, S] = Z,$$

and demographic parity $Z \perp\!\!\!\perp S$. Then, for every group s with positive mass,

$$\begin{aligned}\mathbb{E}[Y \mid S = s] &= \mathbb{E}\big[\mathbb{E}[Y \mid Z, S] \mid S = s\big] \\ &= \mathbb{E}[Z \mid S = s] \\ &= \mathbb{E}[Z].\end{aligned}$$

If group base rates differ, both conditions cannot hold simultaneously.

This argument assumes *group* calibration $\mathbb{E}[Y \mid Z, S] = Z$, not merely pooled calibration $\mathbb{E}[Y \mid Z] = Z$.

Calibration is not enough: the resolution problem

For a binary claim indicator, let $m = \mathbb{E}[Y]$ and assume $\mathbb{E}[Y \mid Z] = Z$.

$$\begin{aligned}\mathbb{E}[(Y - Z)^2] &= \mathbb{E}[Y] - \mathbb{E}[Z^2] \\ &= m(1 - m) - \mathrm{Var}(Z).\end{aligned}$$

Constant predictor

$Z = m$ is calibrated.

$$\mathrm{Var}(Z) = 0.$$

No risk differentiation.

Informative calibrated score

Greater score resolution can lower Brier risk:

$$\mathrm{Var}(Z) > 0.$$

But what are legitimate rating factors?

Dawid (1982), Fernandes Machado et al. (2024).

From fair predictions to fair insurance systems

A closing distinction

A predictive rule, a correction rule, and a pricing rule are not interchangeable:

$$\underbrace{\mu(x,s)=\mathbb{E}[Y\mid x,s]}_{\text{forecast}} \longrightarrow \underbrace{Z^*(x,s)}_{\text{adjusted score}} \longrightarrow \underbrace{\pi(x,s)}_{\text{premium}}.$$

Statistics
Prediction and calibration

Geometry
Distributional corrections and trade-offs

Risk sharing
Who pays? Who remains protected?

Which predictive differences should become price differences — and which should remain collectively shared?

Three mathematical messages to remember

1. **Optimal transport produces interpretable distributional corrections.**

$$U(Z_\lambda) = (1-\lambda)^2 U(Z).$$

2. **Several attributes change the geometry of the problem.**

$$U_{\text{joint}} = \text{Var}_H(Q_{\boldsymbol{S}}), \qquad \phi_1 + \cdots + \phi_k = U_{\text{joint}}.$$

3. **Improving one notion of fairness can worsen another.**

$$\mathcal{E}_{1\to 2}(\lambda) > 0 \quad \text{is possible even if} \quad U_1(Z_\lambda) < U_1(Z).$$

References

Arrow, K. J. (1963). Uncertainty and the welfare economics of medical care. *The American Economic Review*, 53(5):941–973.

Barry, L. and Charpentier, A. (2026). Beyond the issue of reference classes: what does an individual actuarial score mean? *In progress*.

Blinder, A. S. (1973). Wage discrimination: Reduced form and structural estimates. *Journal of Human Resources*, 8(4):436–455.

Charpentier, A. (2024). *Insurance, Biases, Discrimination and Fairness*. Springer Actuarial. Springer.

Charpentier, A. (2026a). Ex ante premiums as expected risk-sharing allocations: Pricing beliefs, cooperative games, and partial identification. *Submitted*.

Charpentier, A. (2026b). *Risk-Priced Society*. https://freakonometrics.github.io/risk-priced-society/.

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*, 92:33–75.

Dawid, A. P. (1982). The well-calibrated bayesian. *Journal of the American statistical Association*, 77(379):605–610.

Dawid, P. (2017). On individual risk. *Synthese*, 194(9):3445–3474.

References

Denuit, M., Hainaut, D., and Trufin, J. (2019). *Effective statistical learning methods for actuaries I: GLMs and extensions*, volume 1. Springer.

Fernandes Machado, A., Charpentier, A., Flachaire, E., Gallic, E., and Hu, F. (2024). Probabilistic scores of classifiers, calibration is not enough. *arXiv*, 2408.03421.

Fernandes Machado, A., Charpentier, A., and Gallic, E. (2025). Sequential conditional transport on probabilistic graphs for interpretable counterfactual fairness. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(18):19358–19366.

Fernandes Machado, A., Jamotton, C., and Charpentier, A. (2026a). Multi-sensitive distributional fairness: Joint wasserstein audits and shapley allocation. Preprint; under review.

Fernandes Machado, A., Voitsitska, I., Charpentier, A., and Gallic, E. (2026b). Graph-aligned sequential transport for distributional disparity decomposition. *ArXiv*, 2603.15182.

Grari, V., Charpentier, A., and Detyniecki, M. (2022). A fair pricing model via adversarial learning.

Hu, F., Ratz, P., and Charpentier, A. (2024). A sequentially fair mechanism for multiple sensitive attributes. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

Huynh, S. and Charpentier, A. (2026). A transport framework for fair insurance premiums across multiple coverages. Working paper.

References

Jamotton, C., Fernandes Machado, A., and Charpentier, A. (2026). Robbing peter to pay paul: Cross-attribute externalities of fairness interventions. *Submitted*.

Kitagawa, E. M. (1955). Components of a difference between two rates. *Journal of the american statistical association*, 50(272):1168–1194.

Lindholm, M., Richman, R., Tsanakas, A., and Wüthrich, M. V. (2022). Discrimination-free insurance pricing. *ASTIN Bulletin*, 52(1):55–89.

Lindholm, M., Richman, R., Tsanakas, A., and Wüthrich, M. V. (2024). What is fair? proxy discrimination vs. demographic disparities in insurance pricing. *Scandinavian Actuarial Journal*, (9):935–970.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review*, 14(3):693–709.

Roth, A. and Tolbert, A. W. (2025). Resolving the reference class problem at scale. *Philosophy of Science*, 92(4):868–882.

Tierny, B., Charpentier, A., and Hu, F. (2026a). Decomposing direct and indirect biases in linear models under demographic parity constraint. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

Tierny, B., Charpentier, A., and Hu, F. (2026b). Direct and indirect discrimination in generalized linear models. *ArXiv*, 2606.23116.