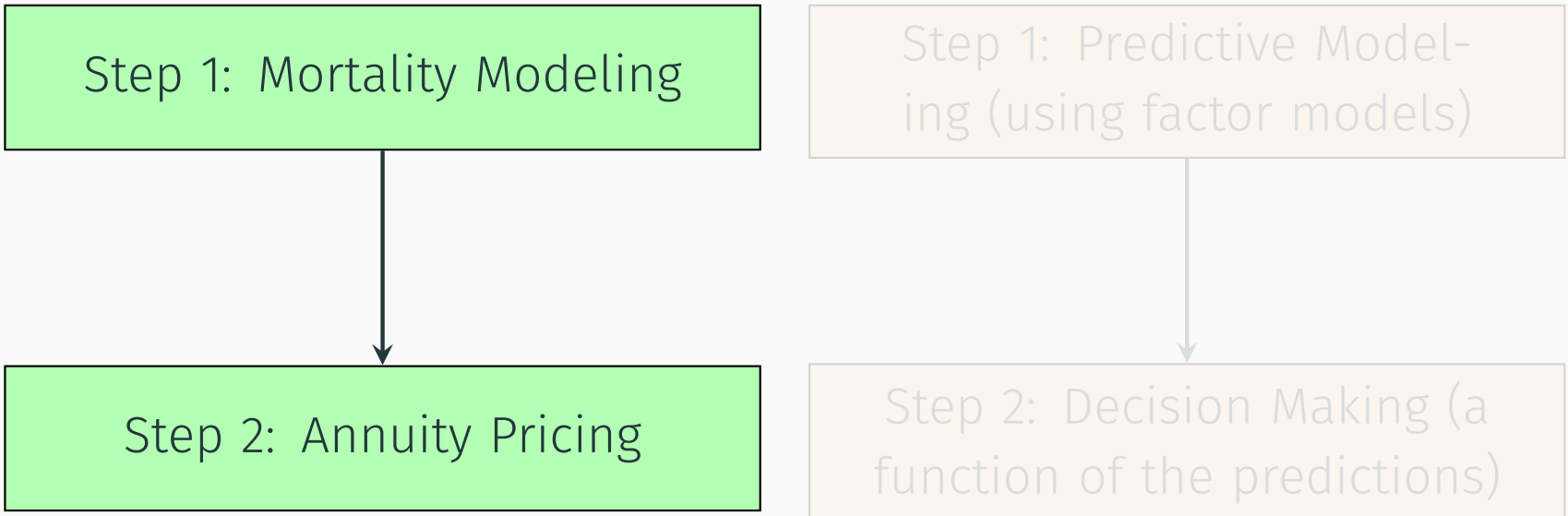
Learning Fair Decisions with factor models: Applications to Annuity Pricing

Ran Zhao (ANU)

Joint work with Fei Huang (UNSW), Junhao Shen, Yanrong Yang (ANU)

Two-step Decision-making Process

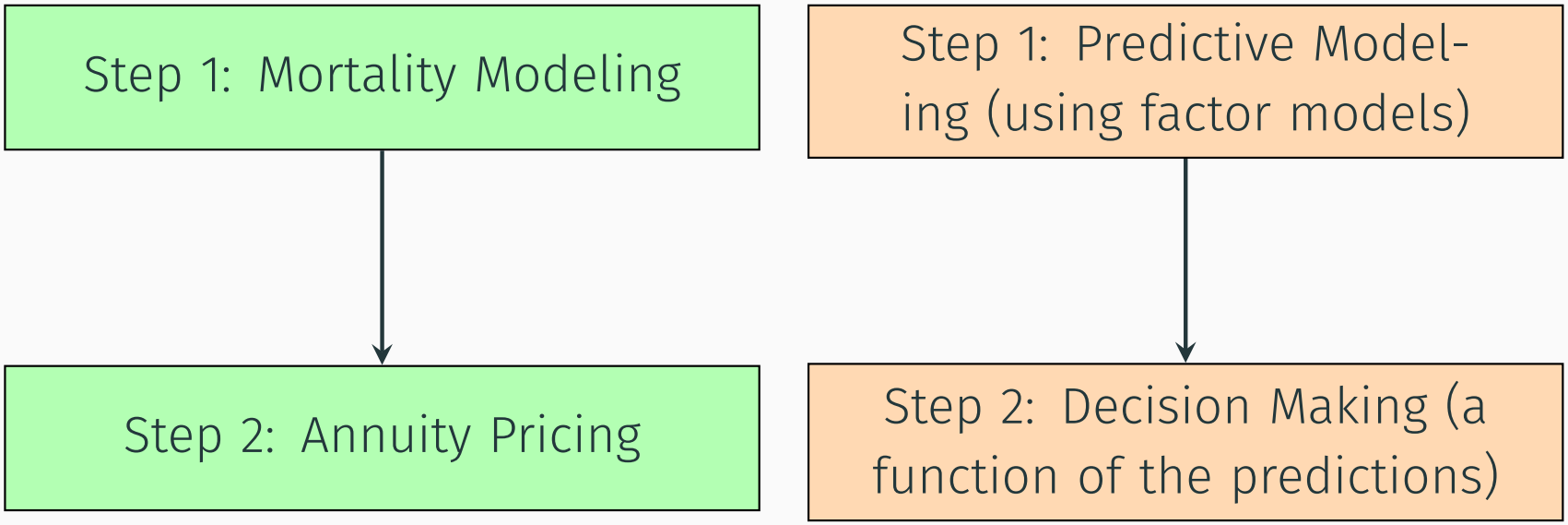


Application

Process

Which step to apply fairness constraints?

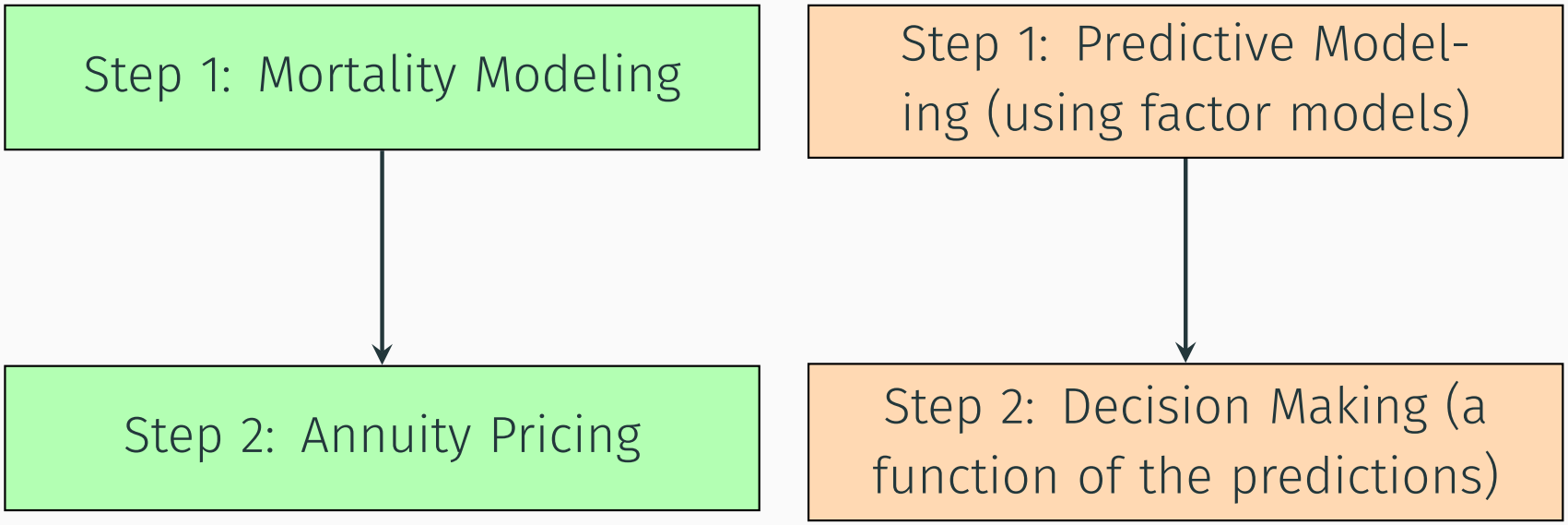
Two-step Decision-making Process



Application **Process**

Which step to apply fairness constraints?

Two-step Decision-making Process



Application **Process**

Which step to apply fairness constraints?

Table of contents

Introduction

Introduction

- Discrimination is rampant in society
  - Think about how you are treated in a store based on your hair color, skin color, the clothes you wear, ethinicity/heritage/religion, and so forth
- Insurance pricing is interesting because the entire industry is based on differentiation.
  - Discrimination means the differentiating treatment or outcome among groups that are characterized by salient features such as hair color, age, ethnicity, heritage, religion, and so forth.
- Although many variables are routinely used, the use of others (e.g., ethinicity/heritage/religion) are frowned upon/forbidden.

Literature Review

- Existing literature predominantly focuses on fairness for step 1: **fair machine learning** (or fair claims cost modeling for insurance) and mitigating **indirect discrimination**, see Lindholm et al. (2022), Frees & Huang (2023), Xin & Huang (2024), Araiza Iturria et al. (2024), Côté et al. (2024).
- A few papers look at fairness in decision-making beyond fairness in machine learning by including the price optimisation step or final pricing outcomes, for example, see Shimao & Huang (2022), Dolman & Semenovich (2018).
- Most research focuses on fair pricing for general insurance.

Existing Fairness Definitions

Most commonly discussed fairness notions can be categorized into two main types: **group fairness notions** and **individual fairness notions** (He & Li 2025).

- **Group fairness**: Demographic parity (independence); Conditional demographic parity (conditional independence); Equalized odds (error rate parity); Sufficiency (well-calibration) and so on;
- **Individual fairness**: Consistency; Counterfactual fairness (causal-based fairness); generalized entropy index.

Fairness-Enhancing Mechanisms

- **Pre-processing**: eliminate discrimination in model output and are applied before the implementation of the target model (modifying the training data distribution, learning fair representations, and decoupling);
- **In-processing**: are applied during the training phase of the model (directly modifying model parameters, architectures, or training procedures);
- **Post-processing**: designed to enforce fairness after the training process is complete (operate on the model's predictions or inferred outcomes).

Main Contributions

1. Extend the existing fair insurance pricing framework from fair machine learning (cost modeling) to fair decision-making using factor models.
2. Propose a new fairness notion, decision error parity, for fair decision making by ensuring that decision errors are equal across demographic groups, when direct discrimination is allowed.
3. Introduce a new fair decision model to achieve decision error parity with factor models.
4. Apply the framework to annuity pricing using empirical data.

Factor Models

Factor Models

Consider $\{y_{it}\}_{1\leq i\leq N, 1\leq t\leq T}$, a N-dimensional time series with T observations. We model $\{y_{it}\}$ using a factor model with r factors:

$$y_{it} = \boldsymbol{\lambda}_i^\top \boldsymbol{f}_t + \epsilon_{it}, 1 \leq i \leq N, 1 \leq t \leq T, \tag{1}$$

where

- $\boldsymbol{f}_t \in \mathbb{R}^r$ is a vector of common factors;
- $\boldsymbol{\lambda}_i \in \mathbb{R}^r$ is a vector of loadings associated with $\boldsymbol{f}_t$ capturing the impact of each common factor on the i-th dimension;
- ϵ_{it} is the idiosyncratic error of y_{it}.

Application to Mortality Modeling

- Let $m_{x,t}$ denote the central death rate for age x in year t, where $x = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$.
- We use log transformation to map the central death rates from $\mathbb{R}^+$ space to $\mathbb{R}$ space for modeling purposes.
- Assume $\boldsymbol{a}_x$ is the age-specific constants, which is the averages over time of the $\ln(m_{x,t})$.

Then $\ln(m_{x,t}) - \boldsymbol{a}_x$ can be modeled using the classical factor model (Lee-Carter-type Model), as follows:

$$\ln(m_{x,t}) - \boldsymbol{a}_x = \boldsymbol{b}_x^\top \boldsymbol{k}_t + \epsilon_{x,t}, \tag{2}$$

Matrix Form of the Factor Model

- The vector form of the factor model is

$$\boldsymbol{y}_t = \boldsymbol{\Lambda} \boldsymbol{f}_t + \boldsymbol{\epsilon}_t,$$

  where $\boldsymbol{y}_t^\top = (y_{1t}, \cdots, y_{Nt})$, $\boldsymbol{\Lambda} = (\boldsymbol{\lambda}_1, \cdots, \boldsymbol{\lambda}_N)^\top \in \mathbb{R}^{N \times r}$, and $\boldsymbol{\epsilon}_t^\top = (\epsilon_{1t}, \cdots, \epsilon_{Nt})$.

- The matrix form of the factor model is

$$\mathbf{Y} = \mathbf{F} \boldsymbol{\Lambda}^\top + \mathbf{E},$$

  where $\mathbf{Y} = (\boldsymbol{y}_1, \cdots, \boldsymbol{y}_T)^\top \in \mathbb{R}^{T \times N}$, $\mathbf{F} = (\boldsymbol{f}_1, \cdots, \boldsymbol{f}_T)^\top \in \mathbb{R}^{T \times r}$, and $\mathbf{E} = (\boldsymbol{\epsilon}_1, \cdots, \boldsymbol{\epsilon}_T)^\top \in \mathbb{R}^{T \times N}$.

- To ensure identifiability, we impose the following conditions:

$$\frac{1}{N} \boldsymbol{\Lambda}^\top \boldsymbol{\Lambda} = \mathbf{I}_r \text{ and } \frac{1}{T} \mathbf{F}^\top \mathbf{F} \text{ is diagonal.}$$

Estimation of the Factor Model

It is natural to estimate $\mathbf{\Lambda}$ and $\mathbf{F} = (\boldsymbol{f}_1, \cdots, \boldsymbol{f}_T^\top)$ by minimizing

$$\min_{\mathbf{\Lambda}, \mathbf{F}} \frac{1}{T} \sum_{t=1}^{T} \|\boldsymbol{y}_t - \mathbf{\Lambda} \boldsymbol{f}_t\|_2^2 = \frac{1}{T} \|\mathbf{Y} - \mathbf{F}\mathbf{\Lambda}^\top\|_F^2$$

$$\text{subject to } \frac{1}{N}\mathbf{\Lambda}^\top\mathbf{\Lambda} = \mathbf{I}_r \text{ and } \frac{1}{T}\mathbf{F}^\top\mathbf{F} \text{ is diagonal.}$$

We have $\widehat{\mathbf{\Lambda}}/\sqrt{N}$, the eigenvectors corresponding to the top r eigenvalues of the matrix $\mathbf{Y}^\top\mathbf{Y}$; $\widehat{\mathbf{F}} = \mathbf{Y}\widehat{\mathbf{\Lambda}}/N$; the common component matrix is estimated by

$$\widehat{\mathbf{F}}\widehat{\mathbf{\Lambda}}^\top = N^{-1}\mathbf{Y}\widehat{\mathbf{\Lambda}}\widehat{\mathbf{\Lambda}}^\top,$$

and finally $\widehat{\mathbf{E}} = \mathbf{Y} - \widehat{\mathbf{F}}\widehat{\mathbf{\Lambda}}^\top = \mathbf{Y} - N^{-1}\mathbf{Y}\widehat{\mathbf{\Lambda}}\widehat{\mathbf{\Lambda}}^\top$.

Estimation of the Factor Model

It is natural to estimate $\mathbf{\Lambda}$ and $\mathbf{F} = (\boldsymbol{f}_1, \cdots, \boldsymbol{f}_T^\top)$ by minimizing

$$\min_{\mathbf{\Lambda}, \mathbf{F}} \frac{1}{T} \sum_{t=1}^{T} \|\boldsymbol{y}_t - \mathbf{\Lambda} \boldsymbol{f}_t\|_2^2 = \frac{1}{T} \|\mathbf{Y} - \mathbf{F}\mathbf{\Lambda}^\top\|_F^2$$

$$\text{subject to } \frac{1}{N}\mathbf{\Lambda}^\top\mathbf{\Lambda} = \mathbf{I}_r \text{ and } \frac{1}{T}\mathbf{F}^\top\mathbf{F} \text{ is diagonal.}$$

We have $\widehat{\mathbf{\Lambda}}/\sqrt{N}$, the eigenvectors corresponding to the top r eigenvalues of the matrix $\mathbf{Y}^\top\mathbf{Y}$; $\widehat{\mathbf{F}} = \mathbf{Y}\widehat{\mathbf{\Lambda}}/N$; the common component matrix is estimated by

$$\widehat{\mathbf{F}}\widehat{\mathbf{\Lambda}}^\top = N^{-1}\mathbf{Y}\widehat{\mathbf{\Lambda}}\widehat{\mathbf{\Lambda}}^\top,$$

and finally $\widehat{\mathbf{E}} = \mathbf{Y} - \widehat{\mathbf{F}}\widehat{\mathbf{\Lambda}}^\top = \mathbf{Y} - N^{-1}\mathbf{Y}\widehat{\mathbf{\Lambda}}\widehat{\mathbf{\Lambda}}^\top$.

Reconstruction Error

Definition (Reconstruction Error)

For any given dataset $\mathbf{Y}$ and any loading matrix $\mathbf{\Lambda}$, the total reconstruction error of $\mathbf{Y}$ using $\mathbf{\Lambda}$ is defined as:

$$\mathcal{L}(\mathbf{\Lambda}) = \frac{1}{T} \left\| \mathbf{Y} - \frac{1}{N} \mathbf{Y} \mathbf{\Lambda} \mathbf{\Lambda}^\top \right\|_F^2 . \tag{3}$$

Fair Factor Models

Setup Notations

Consider a binary protected attribute $\mathcal{A} = \{1, 2\}$. We denote the data matrix for each group as $\mathbf{Y}_k \in \mathbb{R}^{T_k \times N}$, where T_k is the number of samples in group k.

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{F}_1 \\ \mathbf{F}_2 \end{pmatrix} \mathbf{\Lambda}^\top + \begin{pmatrix} \mathbf{E}_1 \\ \mathbf{E}_2 \end{pmatrix},$$

then, the $\mathbf{Y}$ is estimated by

$$\widehat{\mathbf{Y}} = \begin{pmatrix} \widehat{\mathbf{Y}}_1 \\ \widehat{\mathbf{Y}}_2 \end{pmatrix} = \begin{pmatrix} \widehat{\mathbf{F}}_1 \widehat{\mathbf{\Lambda}}^\top \\ \widehat{\mathbf{F}}_2 \widehat{\mathbf{\Lambda}}^\top \end{pmatrix} = \begin{pmatrix} \mathbf{Y}_1 \widehat{\mathbf{\Lambda}} \widehat{\mathbf{\Lambda}}^\top / N \\ \mathbf{Y}_2 \widehat{\mathbf{\Lambda}} \widehat{\mathbf{\Lambda}}^\top / N \end{pmatrix}.$$

The reconstruction error for each group using $\mathbf{\Lambda}$ is denoted as

$$\mathcal{L}_k(\mathbf{\Lambda}) = \frac{1}{T_k} \left\| \mathbf{Y}_k - \frac{1}{N} \mathbf{Y}_k \mathbf{\Lambda} \mathbf{\Lambda}^\top \right\|_F^2, 1 \leq k \leq 2.$$

Reconstruction Error Parity and Fair Factor Model

The reconstruction errors across different demographic groups are often unequal, reflecting differences in how well the model captures group-specific data patterns.

Definition (Reconstruction Error Parity and Fair factor model)

A fair factor model with loading matrix $\mathbf{\Lambda}$ is called fair, if it satisfies Reconstruction Error Parity, that is the expected reconstruction errors are equal across different groups, i.e.,

$$\mathbb{E}(\mathcal{L}_1(\mathbf{\Lambda})) = \mathbb{E}(\mathcal{L}_2(\mathbf{\Lambda})).$$

Note: this fairness notion is similar to demographic parity for reconstruction errors.

Objective Function – Estimating Fair Factor Models

For any loading matrix $\mathbf{\Lambda}$, we define the **unfairness** as the squared difference in reconstruction error between the two groups:

$$\mathcal{U}(\mathbf{\Lambda}) := \left(\mathcal{L}_1(\mathbf{\Lambda}) - \mathcal{L}_2(\mathbf{\Lambda})\right)^2 .$$

We consider the following fair factor model problem:

$$\min_{\mathbf{\Lambda},\mathbf{F}} \frac{1}{T}\|\mathbf{Y} - \mathbf{F}\mathbf{\Lambda}^\top\|_F^2 + \lambda \left(\frac{1}{T_1}\|\mathbf{Y}_1 - \mathbf{F}_1\mathbf{\Lambda}^\top\|_F^2 - \frac{1}{T_2}\|\mathbf{Y}_2 - \mathbf{F}_2\mathbf{\Lambda}^\top\|_F^2\right)^2, \quad (4)$$

subject to $\frac{1}{N}\mathbf{\Lambda}^\top\mathbf{\Lambda} = \mathbf{I}_r$, where $\lambda \geq 0$ is penalty parameter that enforces fairness.

Solving the Optimization Problem

For a given $\mathbf{\Lambda}$, the solutions for the least-square problem is $\widehat{\mathbf{F}} = \mathbf{Y}\mathbf{\Lambda}/N$, where $\widehat{\mathbf{F}}_1 = \mathbf{Y}_1\mathbf{\Lambda}/N, \widehat{\mathbf{F}}_2 = \mathbf{Y}_2\mathbf{\Lambda}/N$. Substituting them in (4), the objective function now becomes

$$\min_{\mathbf{\Lambda}^\top\mathbf{\Lambda}/N=\mathbf{I}_r} \mathcal{L}(\mathbf{\Lambda}) + \lambda(\mathcal{L}_1(\mathbf{\Lambda}) - \mathcal{L}_2(\mathbf{\Lambda}))^2. \tag{5}$$

We can solve it iteratively using the gradient descent method. The projection can be determined using gradient descent by progressively refining an initial solution through iterative updates:

$$\mathbf{\Lambda}_{j+1} = \sqrt{N}\prod_{\mathcal{P}_r}(\mathbf{\Lambda}_j - \eta\mathcal{G}(\mathbf{\Lambda}_j)),$$

where η is the learning rate.

The specific iteration process is summarized in Algorithm 1.

Algorithm for Fair Factor Model

Algorithm 1 The Gradient Descent Algorithm for Fair Factor Model.

Input: Data matrix $\mathbf{Y} = \mathbf{Y}_1 \cup \mathbf{Y}_2$, low rank r, penalty parameter λ, convergence threshold ϵ, the number of iterations $MaxI$

Output: Fair factor loading matrix $\widehat{\mathbf{\Lambda}}$ with $\text{rank}(\mathbf{\Lambda}) = r$, factor matrix $\widehat{\mathbf{F}}_1, \widehat{\mathbf{F}}_2$

1: $t = 0$, draw the initial estimator $\mathbf{\Lambda}_0 \in \mathbb{R}^{d \times r}$ randomly;
2: **repeat**
3: calculate the gradient $\mathcal{G}(\mathbf{\Lambda}_j)$;
4: update $\mathbf{\Lambda}_{j+1}$ as $\mathbf{\Lambda}_{j+1} = \sqrt{N} \prod_{\mathcal{P}_r}(\mathbf{\Lambda}_j - \eta \mathcal{G}(\mathbf{\Lambda}_j))$;
5: **until** $\dfrac{\|\mathbf{Y}\mathbf{\Lambda}_{j+1}\mathbf{\Lambda}_{j+1}^\top/N - \mathbf{Y}\mathbf{\Lambda}_j\mathbf{\Lambda}_j^\top/N\|_F}{\|\mathbf{Y}\mathbf{\Lambda}_j\mathbf{\Lambda}_j^\top/N\|_F} \leq \epsilon$ or $j = MaxI$;
6: output the estimator $\widehat{\mathbf{\Lambda}}$ from the last step and $\widehat{\mathbf{F}}_1 = \mathbf{Y}_1\widehat{\mathbf{\Lambda}}/N, \widehat{\mathbf{F}}_2 = \mathbf{Y}_2\widehat{\mathbf{\Lambda}}/N$.

Multiclass Fair Factor Model

It is straightforward to extend the methodology to multiclass attributes where $\mathcal{A} = \{1, \cdots, K\}$. In this setting, we aim to find a $\mathbf{\Lambda}$ such that $\mathbb{E}(\mathcal{L}_k(\mathbf{\Lambda})) = \mathbb{E}(\mathcal{L}_{k'}(\mathbf{\Lambda}))$ for all $k < k'$. Let $\mathbf{Y}_k \in \mathbb{R}^{T_k \times N}$ denote the data matrix corresponding to group k, where T_k represents the sample size of group $k, k \in \{1, \cdots, K\}$. Let $\mathbf{F}_k \in \mathbb{R}^{T_k \times r}$ denote the factor matrix for the k-th group. $T = T_1 + \cdots + T_K$. We consider the following minimization problem:

$$\min_{\mathbf{\Lambda}, \mathbf{F}_1, \cdots, \mathbf{F}_K} \frac{1}{T} \sum_{k=1}^{K} \|\mathbf{Y}_k - \mathbf{F}_k \mathbf{\Lambda}^\top\|_F^2 + \lambda \sum_{k<k'} \left(\frac{1}{T_k} \|\mathbf{Y}_k - \mathbf{F}_k \mathbf{\Lambda}^\top\|_F^2 - \frac{1}{T_{k'}} \|\mathbf{Y}_{k'} - \mathbf{F}_{k'} \mathbf{\Lambda}^\top\|_F^2 \right)^2.$$

By substituting $\mathbf{F}_1 = \mathbf{Y}_1 \mathbf{\Lambda}/N, \cdots, \mathbf{F}_K = \mathbf{Y}_K \mathbf{\Lambda}/N$ into the objective, we obtain

$$\min_{\mathbf{\Lambda}^\top \mathbf{\Lambda} = \mathbf{I}_r} \mathcal{L}(\mathbf{\Lambda}) + \lambda \sum_{k<k'} \left(\mathcal{L}_k(\mathbf{\Lambda}) - \mathcal{L}_{k'}(\mathbf{\Lambda}) \right)^2.$$

By computing the gradient with respect to $\mathbf{\Lambda}$, the solution can be obtained iteratively using a gradient descent method similar to Algorithm 1. Therefore, we omit the details here.

Fair Decision Models

Decision Error Parity and Fair Decision Model

Definition (Decision Error Parity and Fair Decision Model)

The decision with loading matrix $\mathbf{\Lambda}$ is fair if Decision Error Parity is satisfied, that is the expected decision errors are equal across different groups,

$$\frac{1}{T_1}\mathbb{E}\left\|g\left(\frac{1}{N}\mathbf{Y}_1\mathbf{\Lambda}\mathbf{\Lambda}^\top\right)-g\left(\mathbf{Y}_1\right)\right\|_F^2=\frac{1}{T_2}\mathbb{E}\left\|g\left(\frac{1}{N}\mathbf{Y}_2\mathbf{\Lambda}\mathbf{\Lambda}^\top\right)-g\left(\mathbf{Y}_2\right)\right\|_F^2,$$

where $g(\cdot)$ is a certain functional transformation, $\mathbf{Y}_1=\mathbf{F}_1\mathbf{\Lambda}^\top+\mathbf{E}_1$ and $\mathbf{Y}_2=\mathbf{F}_2\mathbf{\Lambda}^\top+\mathbf{E}_2$.

Note: this criterion is similar to demographic parity for decision errors.

We advocate for decision error parity to achieve fair outcomes.

Objective Function - Estimating Fair Decision Models

We then propose the following fair decision model problem,

$$\min_{\mathbf{\Lambda},\mathbf{F}} \frac{1}{T}\left\|g(\mathbf{F}\mathbf{\Lambda}^\top)-g(\mathbf{Y})\right\|_F^2+\lambda\left(\frac{1}{T_1}\left\|g(\mathbf{F}_1\mathbf{\Lambda}^\top)-g(\mathbf{Y}_1)\right\|_F^2-\frac{1}{T_2}\left\|g(\mathbf{F}_2\mathbf{\Lambda}^\top)-g(\mathbf{Y}_2)\right\|_F^2\right)^2. \tag{6}$$

This problem can be solved using gradient decent algorithm.

Algorithm for Fair Decision Model

Algorithm 2 The Gradient Descent Algorithm for Fair Decision Model.

Input: Data matrix $\mathbf{Y} = \mathbf{Y}_1 \cup \mathbf{Y}_2$, function g, low rank r, penalty parameter λ, convergence threshold ϵ, the number of iterations $MaxI$

Output: Fair factor loading matrix $\mathbf{\Lambda}$ with $\text{rank}(\mathbf{\Lambda}) = r$, factor matrix $\widehat{\mathbf{F}}_1, \widehat{\mathbf{F}}_2$, estimators $g(\widehat{\mathbf{Y}}_1), g(\widehat{\mathbf{Y}}_2)$

1: $t = 0$, draw the initial estimator $\mathbf{\Lambda}_0 \in \mathbb{R}^{N \times r}$ randomly;

2: **repeat**

3: calculate the gradient $\mathcal{G}(\mathbf{\Lambda}_j)$;

4: update $\mathbf{\Lambda}_{j+1}$ as $\mathbf{\Lambda}_{j+1} = \sqrt{N} \prod_{\mathcal{P}_r}(\mathbf{\Lambda}_j - \eta \mathcal{G}(\mathbf{\Lambda}_j))$;

5: **until** $\dfrac{\| \left(g(\mathbf{Y}\mathbf{\Lambda}_{j+1}\mathbf{\Lambda}_{j+1}^\top/N) - g(\mathbf{Y}\mathbf{\Lambda}_j\mathbf{\Lambda}_j^\top/N) \right) \|_F}{\|g(\mathbf{Y}\mathbf{\Lambda}_j\mathbf{\Lambda}_j^\top/N)\|_F} \leq \epsilon$ or $j = MaxI$;

6: output the estimator $\widehat{\mathbf{\Lambda}}$ from the last step, $\widehat{\mathbf{F}}_1 = \mathbf{Y}_1\widehat{\mathbf{\Lambda}}/N, \widehat{\mathbf{F}}_2 = \mathbf{Y}_2\widehat{\mathbf{\Lambda}}/N$, and $g(\mathbf{Y}_1\widehat{\mathbf{\Lambda}}\widehat{\mathbf{\Lambda}}^\top/N), g(\mathbf{Y}_2\widehat{\mathbf{\Lambda}}\widehat{\mathbf{\Lambda}}^\top/N)$.

Multiclass Fair Decision Model

In the multiclass case, we assume the attribute set $\mathcal{A} = \{1, \cdots, K\}$, and aim to obtain $\mathbf{\Lambda}$ such that

$$\frac{1}{T_k}\mathbb{E}\left(\left\|g\left(\frac{1}{N}\mathbf{Y}_k\mathbf{\Lambda}\mathbf{\Lambda}^\top\right) - g\left(\mathbf{Y}_k\right)\right\|_F^2\right) = \frac{1}{T_{k'}}\mathbb{E}\left(\left\|g\left(\frac{1}{N}\mathbf{Y}_{k'}\mathbf{\Lambda}\mathbf{\Lambda}^\top\right) - g\left(\mathbf{Y}_{k'}\right)\right\|_F^2\right)$$

for all $k < k'$. Then, the multiclass Fair Decision Model can be represented as follows:

$$\min_{\mathbf{\Lambda}, \mathsf{F}_1, \cdots, \mathsf{F}_K} \frac{1}{T}\sum_{k=1}^{K}\left\|g\left(\mathsf{F}_k\mathbf{\Lambda}^\top\right) - g\left(\mathsf{Y}_k\right)\right\|_F^2 + \lambda \sum_{k<k'}\left(\frac{1}{T_k}\left\|g\left(\mathsf{F}_k\mathbf{\Lambda}^\top\right) - g\left(\mathsf{Y}_k\right)\right\|_F^2 - \frac{1}{T_{k'}}\left\|g\left(\mathsf{F}_{k'}\mathbf{\Lambda}^\top\right) - g\left(\mathsf{Y}_{k'}\right)\right\|_F^2\right)^2.$$

Substituting $\mathbf{F}_1 = \mathbf{Y}_1\mathbf{\Lambda}/N, \cdots, \mathbf{F}_K = \mathbf{Y}_K\mathbf{\Lambda}/N$ into the above objective function yields:

$$\begin{aligned}\min_{\mathbf{\Lambda}} \frac{1}{T}\sum_{k=1}^{K}&\left\|g\left(\frac{1}{N}\mathbf{Y}_k\mathbf{\Lambda}\mathbf{\Lambda}^\top\right) - g\left(\mathbf{Y}_k\right)\right\|_F^2 \\ &+ \sum_{k<k'}\lambda\left(\frac{1}{T_k}\left\|g\left(\frac{1}{N}\mathbf{Y}_k\mathbf{\Lambda}\mathbf{\Lambda}^\top\right) - g\left(\mathbf{Y}_k\right)\right\|_F^2 - \frac{1}{T_{k'}}\left\|g\left(\frac{1}{N}\mathbf{Y}_{k'}\mathbf{\Lambda}\mathbf{\Lambda}^\top\right) - g\left(\mathbf{Y}_{k'}\right)\right\|_F^2\right)^2.\end{aligned}$$

Fair Decision Model for Annuity Pricing

- We consider the Expected Present Value (EPV) of an annuity-due,

$$g(y_{ti}) = \sum_{s=0}^{n-1} v^s \prod_{k=0}^{s-1} (1 - \exp(y_{t+k,i+k} + a_{i+k}))$$
$$= \sum_{s=0}^{n-1} v^s \prod_{k=0}^{s-1} (1 - m_{t+k,i+k}) := h(m_{ti}),$$

  where

- v is the discount factor per year.
- n denotes the term of the annuity.
- m_{ti} is the death rate at age i in year t.

Fair Decision Model for Annuity Pricing

The fair pricing decision model can be expressed as

$$\begin{aligned}\min_{\mathbf{\Lambda},\mathbf{F}} \frac{1}{T-2n+4} \Bigg\{ & \sum_{t=1}^{T_1-n+2} \sum_{i=1}^{N-n+2} \left[g(\boldsymbol{\lambda}_i^\top \boldsymbol{f}_t^{(1)}) - g\left(y_{ti}^{(1)}\right) \right]^2 \\ & + \sum_{t=1}^{T_2-n+2} \sum_{i=1}^{N-n+2} \left[g(\boldsymbol{\lambda}_i^\top \boldsymbol{f}_t^{(2)}) - g\left(y_{ti}^{(2)}\right) \right]^2 \Bigg\} \\ & + \lambda \Bigg\{ \frac{1}{T_1-n+2} \sum_{t=1}^{T_1-n+2} \sum_{i=1}^{N-n+2} \left[g(\boldsymbol{\lambda}_i^\top \boldsymbol{f}_t^{(1)}) - g\left(y_{ti}^{(1)}\right) \right]^2 \\ & - \frac{1}{T_2-n+2} \sum_{t=1}^{T_2-n+2} \sum_{i=1}^{N-n+2} \left[g(\boldsymbol{\lambda}_i^\top \boldsymbol{f}_t^{(2)}) - g\left(y_{ti}^{(2)}\right) \right]^2 \Bigg\}^2 . \end{aligned} \tag{7}$$

Here, $\boldsymbol{f}_t^{(1)}$ and $\boldsymbol{f}_t^{(2)}$ denote the latent factors for the male ($A = 1$) and female ($A = 2$) subgroups, respectively. We estimate the model by minimizing the proposed objective via gradient descent.

Empirical Results

Data

- Australian mortality data collected from HMD.
- Age: 0-85
- Year: 1921-2021
- Training: 1921-1989; Testing: 1990-2021
- Forecasts are obtained from a factor model with projected factors, $\widehat{\boldsymbol{f}}_{t+1} = \widehat{\boldsymbol{\Lambda}}^{\top}\boldsymbol{y}_{t+1}/N$
- The number of factors: 1
- Term of annuity-due: 10 years
- Interest rate assumption: 5%

Models

- Model 1: Factor Model (Lee-Carter-type)
- Model 2: Fair Factor Model (Fairness applied to the Mortality Modeling step)
- Model 3: Fair Decision Model (Fairness applied to the Annuity Pricing step)

Mortality Prediction & Reconstruction Error

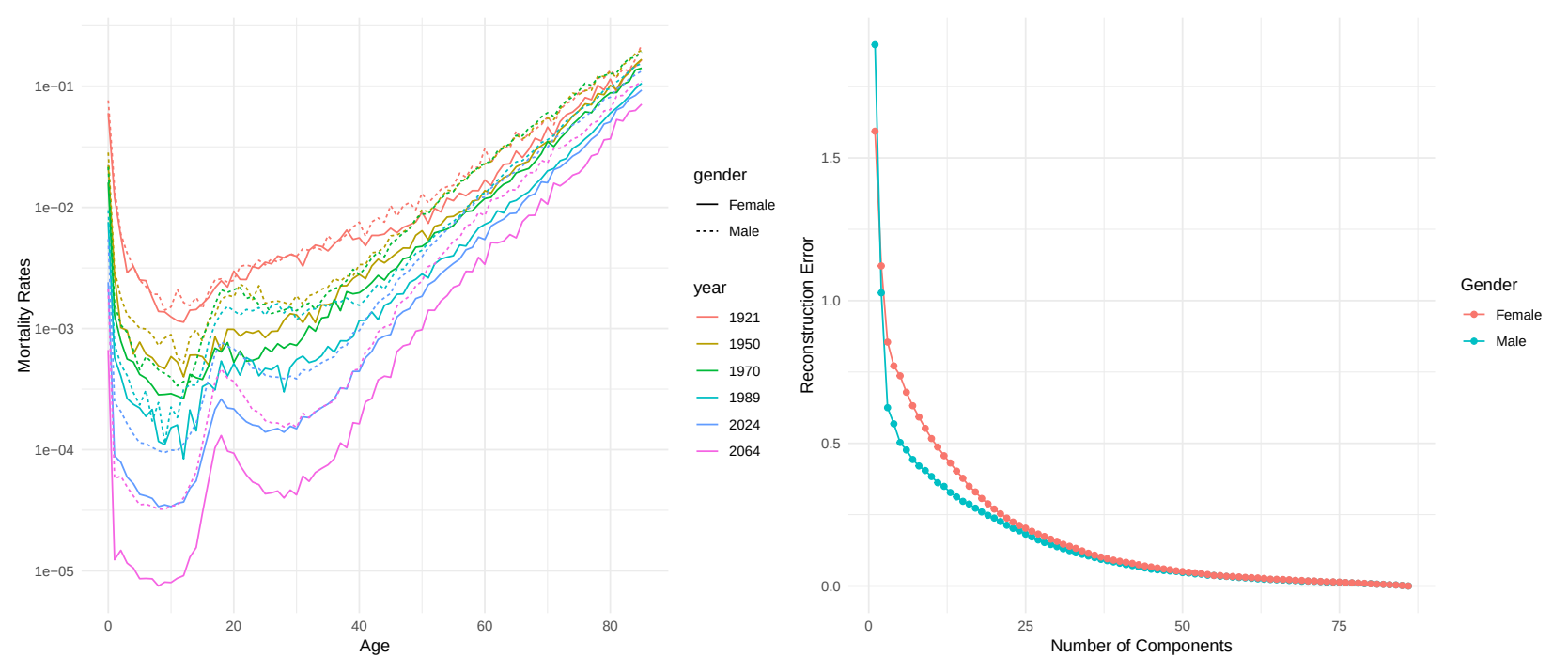


Figure 1: Left: Age–specific mortality rates for males and females at selected historical years and forecasts from the Lee–Carter factor model. Right: Training-sample reconstruction error of the Lee–Carter model under different values of the number of factors *r*.

Log Mortality Prediction –Accuracy v.s. Fairness

- Group–specific RMSEs

$$\mathrm{RMSE}_k = \left\{ \frac{1}{T_k N} \sum_{t=1}^{T_k} \sum_{i=1}^{N} \left(x_{ti}^{(k)} - \widehat{x}_{ti}^{(k)} \right)^2 \right\}^{1/2},$$

- Overall Accuracy Measure:

$$\mathrm{RMSE}_{\text{total}} = \left\{ \frac{1}{TN} \sum_{t=1}^{T} \sum_{i=1}^{N} \left(x_{ti} - \widehat{x}_{ti} \right)^2 \right\}^{1/2},$$

- Fairness Measure:

$$\mathrm{Diff\ (Fairness)} = \left| \mathrm{RMSE}_1 - \mathrm{RMSE}_2 \right|,$$

where $k \in \{1(\text{male}), 2(\text{female})\}$, $T = T_1 + T_2$, and $x_{ti}^{(k)} = \ln(m_{ti}^{(k)}) = y_{ti}^{(k)} + a_i^{(k)}, \widehat{x}_{ti}^{(k)} = \ln(\widehat{m}_{ti}^{(k)}) = \widehat{y}_{ti}^{(k)} + \widehat{a}_i^{(k)}$ denote observed and predicted log–mortality at age i and year t for group k.

Results - Mortality Modeling

Table 1: Factor-model prediction of log mortality: group-specific RMSEs, overall RMSE, and between-sex disparity.

RMSE	Factor Model	Fair Factor Model	Fair Decision Model
Male	0.4327	0.4013	0.9385
Female	0.3729	0.3499	1.0265
Total	0.4039	0.3765	0.9835
Difference	0.0598	0.0514	0.0880

Results - Annuity Pricing

Table 2: Factor-model prediction of annuity values: overall accuracy (RMSE) and between-sex disparity (absolute RMSE difference).

RMSE	Factor Model	Fair Factor Model	Fair Decision Model
Male	0.2086	0.1894	0.0501
Female	0.1581	0.1433	0.0424
Total	0.1851	0.1679	0.0464
Difference	0.0505	0.0461	0.0077

RMSE

We also calculate the RMSE across two dimensions: RMSE by age (RMSE_i), and RMSE by year (RMSE_t). For $k \in \{1, 2\}$, we define

$$\text{RMSE}_i^{(k)} = \left\{ \frac{1}{T_k} \sum_{t=1}^{T_k} (x_{ti}^{(k)} - \widehat{x}_{ti}^{(k)})^2 \right\}^{1/2}, \quad \text{RMSE}_t^{(k)} = \left\{ \frac{1}{N} \sum_{i=1}^{N} (x_{ti}^{(k)} - \widehat{x}_{ti}^{(k)})^2 \right\}^{1/2},$$

where $i = 1, \ldots, N$ and $t = 1, \ldots, T_k$.

Accuracy of Mortality Modeling

Figure 2: Accuracy and fairness of log-mortality predictions across age and year for three models. Top panels: prediction accuracy (RMSE of log-mortality rates) by age and year. Bottom panels: fairness, measured as the absolute RMSE difference between genders across age and year.

Accuracy of Annuity Pricing

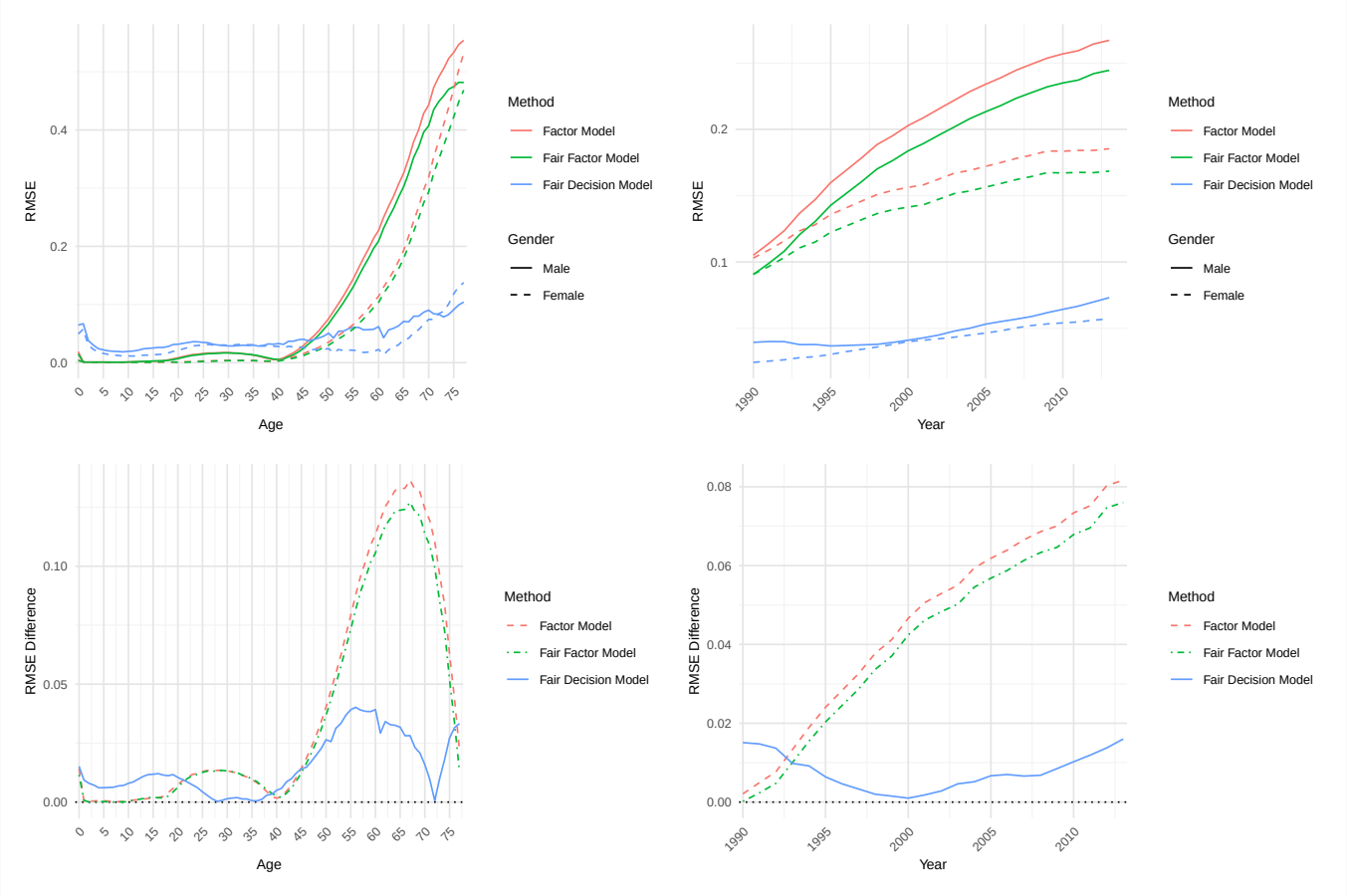


Figure 3: Accuracy and fairness of annuity pricing across age and year for three models. Top panels: prediction accuracy of annuity prices (RMSE of the expected present value of an annuity-due) by age and year. Bottom panels: fairness in annuity pricing, measured as the absolute RMSE difference between genders by age and year.

Conclusion

Conclusion

- Fair Factor Model achieves the smallest RMSE (highest accuracy) and the smallest Difference (highest fairness) in mortality prediction.
- Fair Decision Model has the smallest RMSE (highest accuracy) and the smallest Difference (highest fairness) in annuity pricing.
- Results obtained from time-series forecasts of the factors lead to the same conclusion.
- In this model, both genders are estimated in one factor model (with the same factor loading estimations). We could also model gender separately (with different loading estimations) and apply fairness constraints to achieve decision error parity.

Thank you!

References

Araiza Iturria, C. A., Hardy, M. & Marriott, P. (2024), 'A discrimination-free premium under a causal framework', *North American Actuarial Journal* **0**(0), 1–21.

Côté, O., Côté, M.-P. & Charpentier, A. (2024), 'A fair price to pay: exploiting causal graphs for fairness in insurance', *Available at SSRN 4709243* .

Dolman, C. & Semenovich, D. (2018), Algorithmic fairness: Contemporary ideas in the insurance context, *in* 'IFoA GIRO Conference'.

Frees, E. W. & Huang, F. (2023), 'The discriminating (pricing) actuary', *North American Actuarial Journal* **27**(1), 2–24.

References ii

He, X. & Li, Y. (2025), 'Fairness in machine learning: A review for statisticians', *Journal of the American Statistical Association* **0**(ja), 1–34.

Lindholm, M., Richman, R., Tsanakas, A. & Wüthrich, M. V. (2022), 'Discrimination-free insurance pricing', *ASTIN Bulletin: The Journal of the IAA* **52**(1), 55–89.

Shimao, H. & Huang, F. (2022), 'Welfare cost of fair prediction and pricing in insurance market', *UNSW Business School Research Paper Forthcoming* .

Xin, X. & Huang, F. (2024), 'Antidiscrimination insurance pricing: Regulations, fairness criteria, and models', *North American Actuarial Journal* **28**(2), 285–319.