Institute and Faculty of Actuaries

GIRO Conference 2022

21–23 November, ACC Liverpool

#GiroConf22

Institute and Faculty of Actuaries

A new method for discrimination free insurance pricing and real-world impacts

Ronald Richman FIA

#GiroConf22

Agenda

- **Introduction**
- Synthetic example
- Missing data and multi-task networks
- Real-world example
- Outlook and conclusions

Institute and Faculty of Actuaries

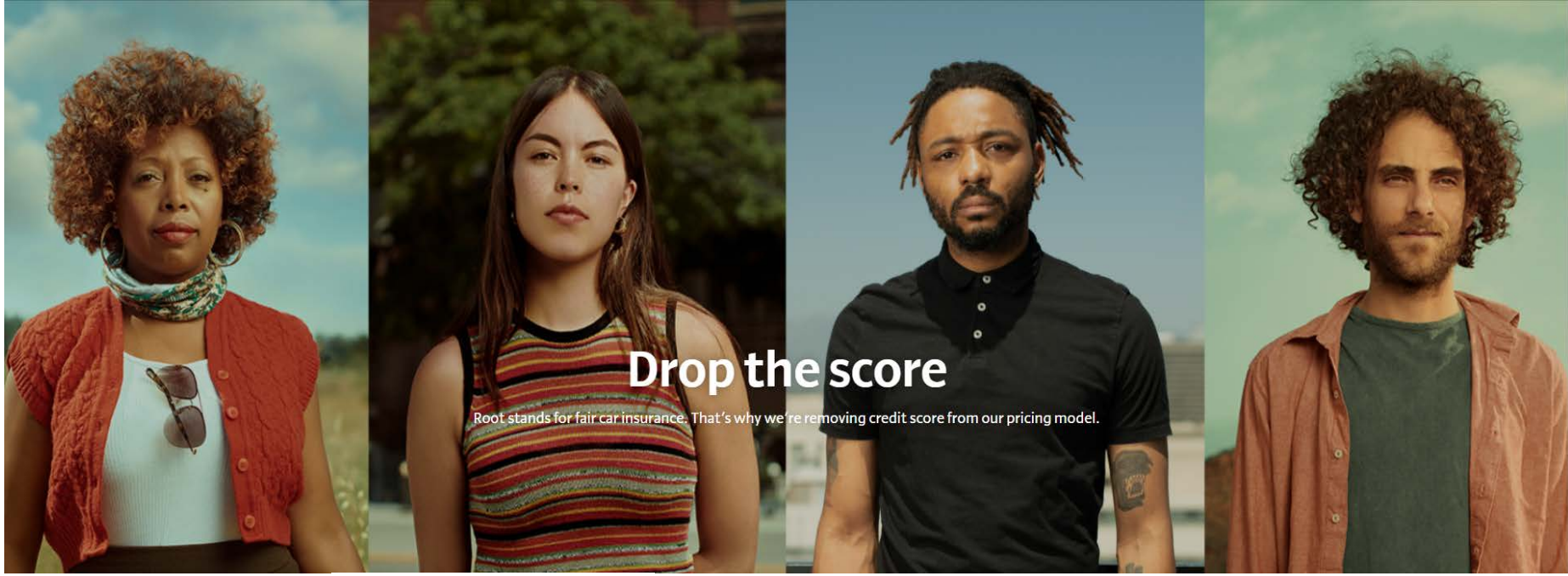
Background

- DFIP - joint work with Mario Wüthrich, Matthias Lindholm, Andreas Tsanakas
- Based on:

  - Lindholm, M., Richman, R., Tsanakas, A., & Wüthrich, M. V. (2022). Discrimination-free insurance pricing. ASTIN Bulletin, 52(1), 55–89. https://doi.org/10.1017/asb.2021.23
  - Lindholm, M., Richman, R., Tsanakas, A., & Wüthrich, M. V. (2022). A multi-task network approach for calculating discrimination-free insurance prices. Retrieved from http://arxiv.org/abs/2207.02799

Institute and Faculty of Actuaries

Rationale-1

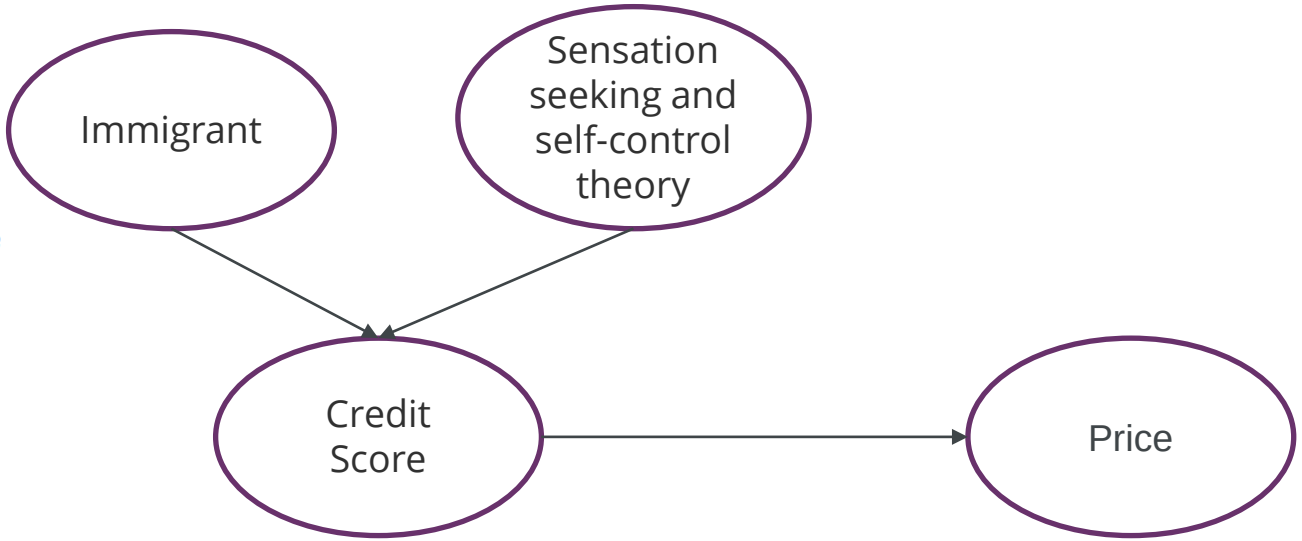


The impact of credit score

Relying on credit scores disproportionately harms certain groups and reinforces the much bigger problem of inherent bias and systemic discrimination facing our country today. This means that certain groups of people are being asked to pay more for car insurance, including **historically under-resourced communities, immigrants, people struggling to pay large medical expenses, people with errors in their credit history information,** and **people who have suffered an economic crisis.**

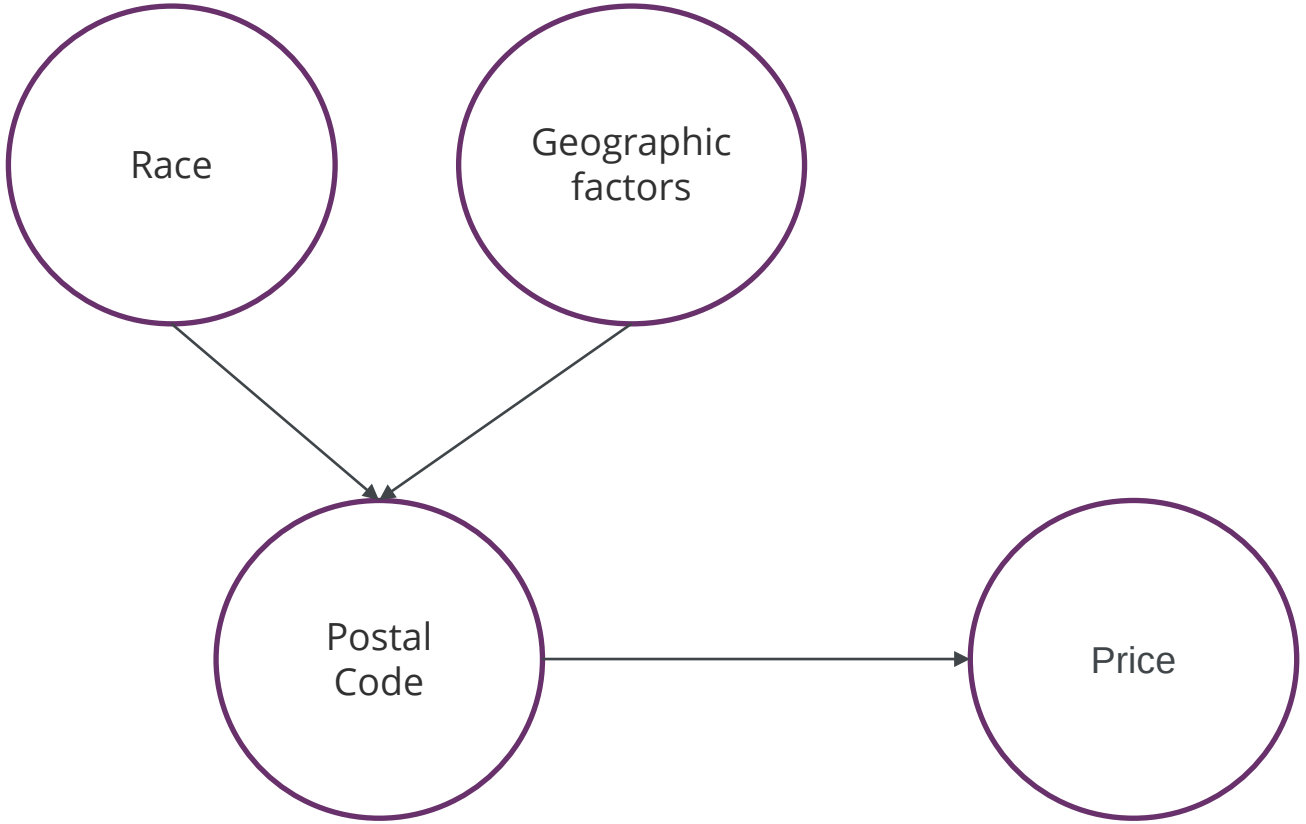
The reality is that the best drivers may sometimes have terrible credit scores, but the risk they present on the road is minimal. People with low credit scores—even if they are the safest drivers—are paying as much as $1,500 more in annual premium payments than people with high credit scores, according to The Zebra.

Root Insurance



Institute and Faculty of Actuaries

Rationale-2


Race
Geographic factors
Postal Code
Price


MAY 13, 2019
TIME
The World's Most Unequal Country
Johannesburg, South Africa, suburbs Primrose, left, and Makause
time.com

Discrimination free pricing

Current environment

- More advanced techniques becoming widely known and used
- Increasing scrutiny internationally on pricing practices (e.g. FCA review and ban on price-walking)

Legal/ethical requirements

- Legal (e.g. EU/UK ban on gender based pricing) and ethical concerns (e.g. postal code ~= race in South Africa)
- How to ensure models are not influenced by discriminatory factors?

Naïve Solution = Unawareness Prices

- Ignore the problem by leaving out discriminatory rating factors
- Could advanced models figure out proxies for these factors?
- Actually, even simple models can do this!

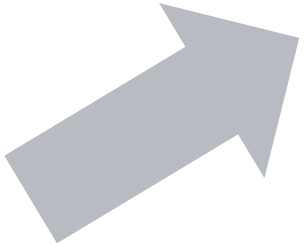
Institute and Faculty of Actuaries

EU Legal Basis

For the purposes of this Directive, the following definitions shall apply:

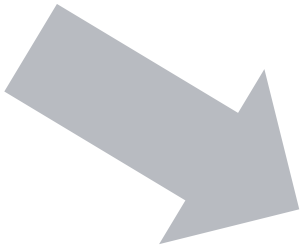
(a) direct discrimination: where one person is treated less favourably, on grounds of sex, than another is, has been or would be treated in a comparable situation;

(b) 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;"



Directly include discriminatory characteristics within pricing models
Usually observe that Gender is a significant risk factor for general/non-life insurance
For life insurance, rates vary clearly with gender.

Note: We rely on society to guide us as to the definition of a discriminatory factor; in this talk we are concerned with methods for correcting pricing once discriminatory factors are defined



Include other factors within pricing model that are highly correlated with the discriminatory factor
Can pick up much of the same effect – e.g. annual driving distance

Definitions

- Insurance pricing models often take the form of best estimates plus loadings.
- Best estimates are usually defined as conditional expectations. Define:
  - $Claims\ costs=Y$
  - $Non-discriminatory\ covariates=X$
  - $Discriminatory\ covariates=D$
- Best estimate prices take account of both X and D:
  - $u(X,D)=E[Y \mid X,D]$
- For complex lines of business, we approximate $E[Y \mid X,D]$ using a regression model
  - $u(X,D)$ discriminates based on D
- A naïve approach – unawareness prices - ignores D and hopes that X and D are uncorrelated:
  - $u(X)=E[Y \mid X]$
- Or relies on proxies for *D* to get closer to the best estimate price…

Institute and Faculty of Actuaries

Example 1 – problem with unawareness?

$n_{i,j}$	woman	man	row total
smoker	32	4	36
non-smoker	28	48	76
column total	60	52	112

$r_{i,j}$	woman	man	row total
smoker	133	24	157
non-smoker	131	301	432
column total	264	325	589

Assume we do not want to price with gender i.e. can only differentiate by smoking status

$$\widehat{\lambda}_{1,\bullet} = \frac{36}{157} = 0.229$$

P(Woman|Smoker) =~85%

$$\begin{aligned}\widehat{\lambda}_{1,\bullet} &= \widehat{\lambda}_{1,1}\frac{r_{1,1}}{r_{1,1}+r_{1,0}} + \widehat{\lambda}_{1,0}\frac{r_{1,0}}{r_{1,1}+r_{1,0}} \\ &= \widehat{\lambda}_{1,1}\widehat{\mathbb{P}}(\text{woman} \mid \text{smoker}) + \widehat{\lambda}_{1,0}\widehat{\mathbb{P}}(\text{man} \mid \text{smoker})\end{aligned}$$

Institute and Faculty of Actuaries

Example 1 – problem with unawareness?

$n_{i,j}$	woman	man	row total
smoker	32	4	36
non-smoker	28	48	76
column total	60	52	112

$r_{i,j}$	woman	man	row total
smoker	133	24	157
non-smoker	131	301	432
column total	264	325	589

Assume we do not want to price with gender i.e. can only differentiate by smoking status

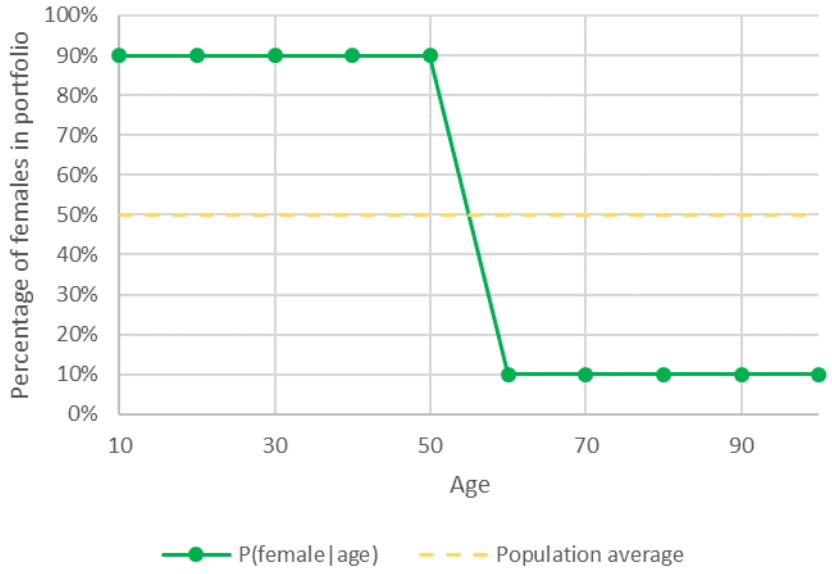
$$\widehat{\lambda}_{1,\bullet} = \frac{36}{157} = 0.229$$

$$\begin{aligned}\widehat{\lambda}_{1,\bullet} &= \widehat{\lambda}_{1,1}\frac{r_{1,1}}{r_{1,1}+r_{1,0}} + \widehat{\lambda}_{1,0}\frac{r_{1,0}}{r_{1,1}+r_{1,0}} \\ &= \widehat{\lambda}_{1,1}\widehat{\mathbb{P}}(\text{woman} \mid \text{smoker}) + \widehat{\lambda}_{1,0}\widehat{\mathbb{P}}(\text{man} \mid \text{smoker})\end{aligned}$$

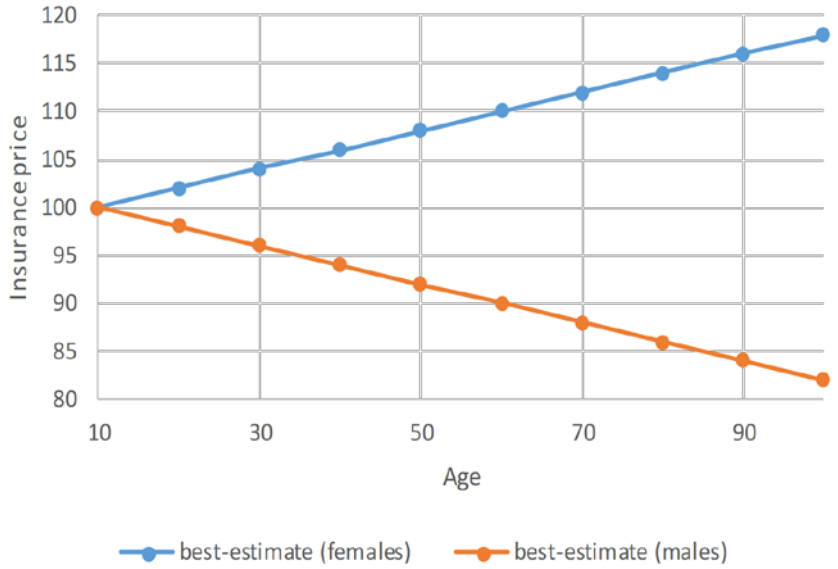
Implicitly inferring gender based on smoking status = indirect discrimination

Institute and Faculty of Actuaries

Example 2 – what is DFIP price?



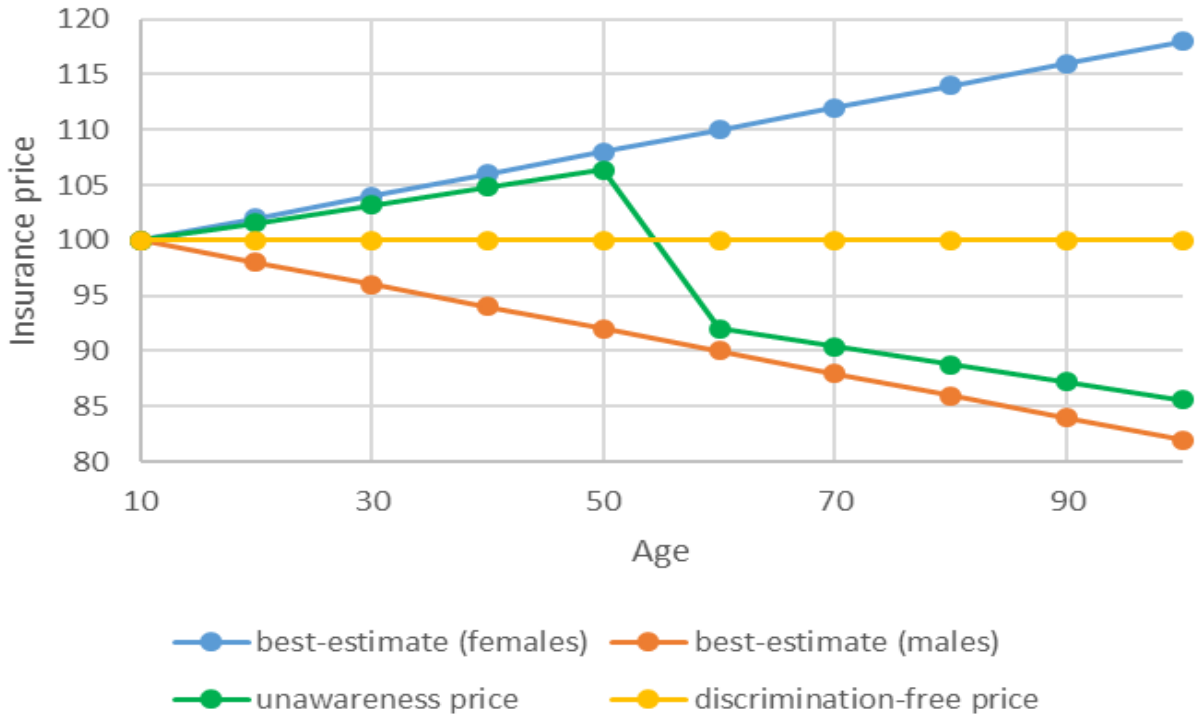
Portfolio proportions: Distribution of gender across age classes and population average



Best estimate costs

Institute and Faculty of Actuaries

Discrimination free prices



Unawareness prices act as proxies for discriminatory factors

Institute and Faculty of Actuaries

Defining DFIP

- Intuition – we need to decouple X and D
- Propose a procedure whereby:
  - Best-estimate prices (including D) are calculated using a model
  - Then take a weighted average of prices where the weights are independent of X
- Formally:
  - $u^*(X) = \sum_d u(X, D = d)P(D = d)$
- It can be shown that:
  - $u(X) = \sum_d u(X, D = d)P(D = d|X)$
- Formal definition of $u^*(X)$ can be given using measure theory; see the paper for details

Institute and Faculty of Actuaries

Back to Example 1

$n_{i,j}$	woman	man	row total
smoker	32	4	36
non-smoker	28	48	76
column total	60	52	112

$r_{i,j}$	woman	man	row total
smoker	133	24	157
non-smoker	131	301	432
column total	264	325	589

$$\begin{aligned}\widetilde{\lambda}_{1,\bullet} &= \widehat{\lambda}_{1,1}\widehat{\mathbb{P}}(\text{woman}) + \widehat{\lambda}_{1,0}\widehat{\mathbb{P}}(\text{man}) \\ &= \frac{32}{133}\cdot\frac{264}{589} + \frac{4}{24}\cdot\frac{325}{589} \\ &= 0.200 \;<\; 0.229 = \widehat{\lambda}_{1,\bullet}.\end{aligned}$$

$$\widetilde{\lambda}_{1,\bullet}(r_{1,1}+r_{1,0}) + \widetilde{\lambda}_{0,\bullet}(r_{0,1}+r_{0,0}) = 110.77 < 112.$$

Institute and Faculty of Actuaries

Back to Example 1

$n_{i,j}$	woman	man	row total
smoker	32	4	36
non-smoker	28	48	76
column total	60	52	112

$r_{i,j}$	woman	man	row total
smoker	133	24	157
non-smoker	131	301	432
column total	264	325	589

$$\begin{aligned}\widetilde{\lambda}_{1,\bullet} &= \widehat{\lambda}_{1,1}\widehat{\mathbb{P}}(\text{woman}) + \widehat{\lambda}_{1,0}\widehat{\mathbb{P}}(\text{man}) \\ &= \frac{32}{133}\cdot\frac{264}{589} + \frac{4}{24}\cdot\frac{325}{589} \\ &= 0.200 \;<\; 0.229 = \widehat{\lambda}_{1,\bullet}.\end{aligned}$$

$$\widetilde{\lambda}_{1,\bullet}(r_{1,1}+r_{1,0}) + \widetilde{\lambda}_{0,\bullet}(r_{0,1}+r_{0,0}) = 110.77 < 112.$$

Institute and Faculty of Actuaries

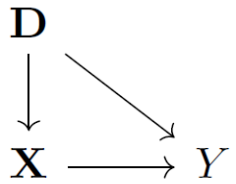
Causal Inference

- Often in actuarial work, we care about prediction, not inference about causality
- Exception: price optimization
- Pricing factors do not need to reflect causal mechanisms i.e. correlation is enough
- Other disciplines (economics, medicine etc) try to make inferences about causation
  - Medicine – randomized controlled trials: does Drug X cause a particular effect?
- Can one infer causality based on observational data?
  - In some circumstances, yes!
  - See the Book of Why by Judea Pearl and Dana Mackenzie
- Need to control for variables that may cause an incorrect inference:
  - Selection bias
  - Confounding

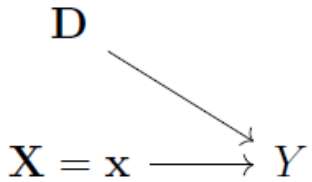
Institute and Faculty of Actuaries

Insights from Causal Inference

- Causal Inference à la Pearl uses directed acyclic graphs (DAGs) to work out how to derive causal quantities from observational data
- Consider the following DAG (where X is confounded by D):



- Non-discriminatory pricing ~= finding the effect of X on Y after removing effect of D on X



- Formula for u *(X) can be recognized as:
- Pearl's Back-door Adjustment to remove confounding effects
- Friedman's Partial Dependence Plot formula

Institute and Faculty of Actuaries

Agenda

- Introduction
- **Synthetic example**
- Missing data and multi-task networks
- Real-world example
- Outlook and conclusions

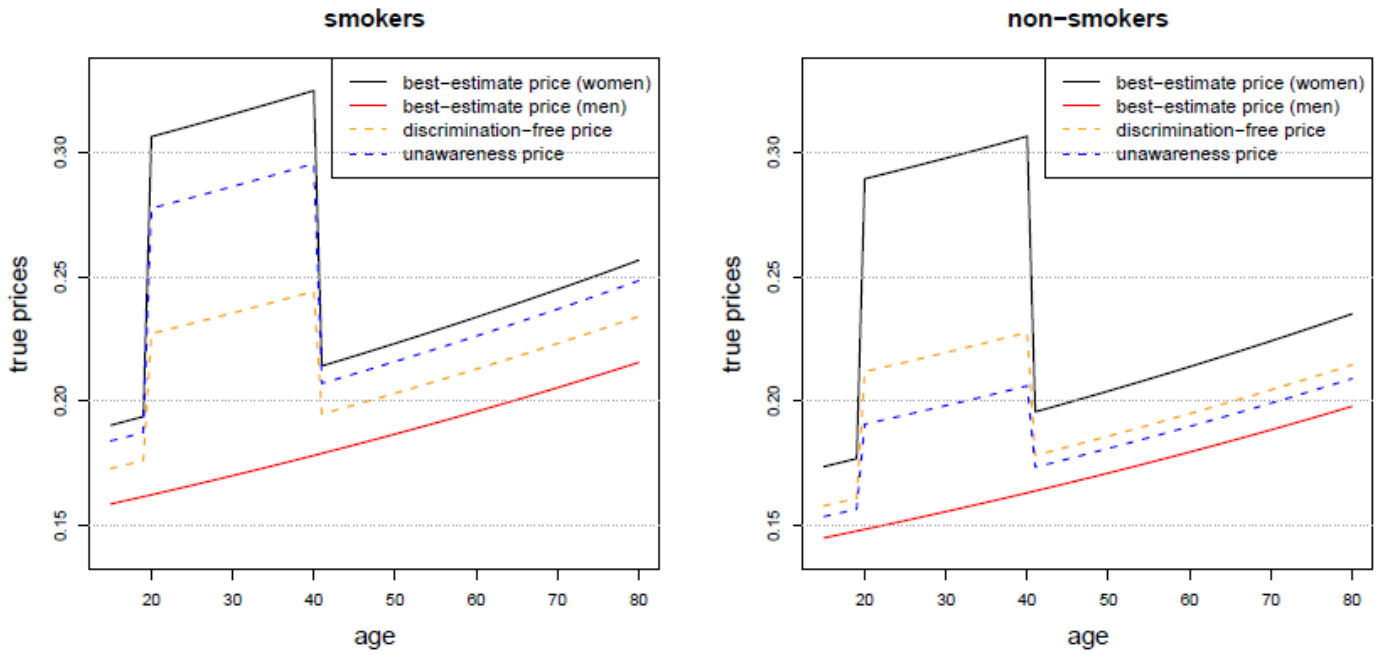
Institute and Faculty of Actuaries

Simulated Example

- Health insurance setup:
  - X = Age, Smoker
  - D = Gender
- Claims costs:
  - Cost 1 – only affect females of ages (20-40)
  - Cost 2 – depend on age and higher frequency for smokers and females
  - Cost 3 – depend on age
  - Total costs = 1+2+3
- Simulated portfolio of 100 000 policyholders
  - Assumed 45% are females
  - P(Smoker) = 0.3
  - P(Smoker|Female) = 0.8

Institute and Faculty of Actuaries

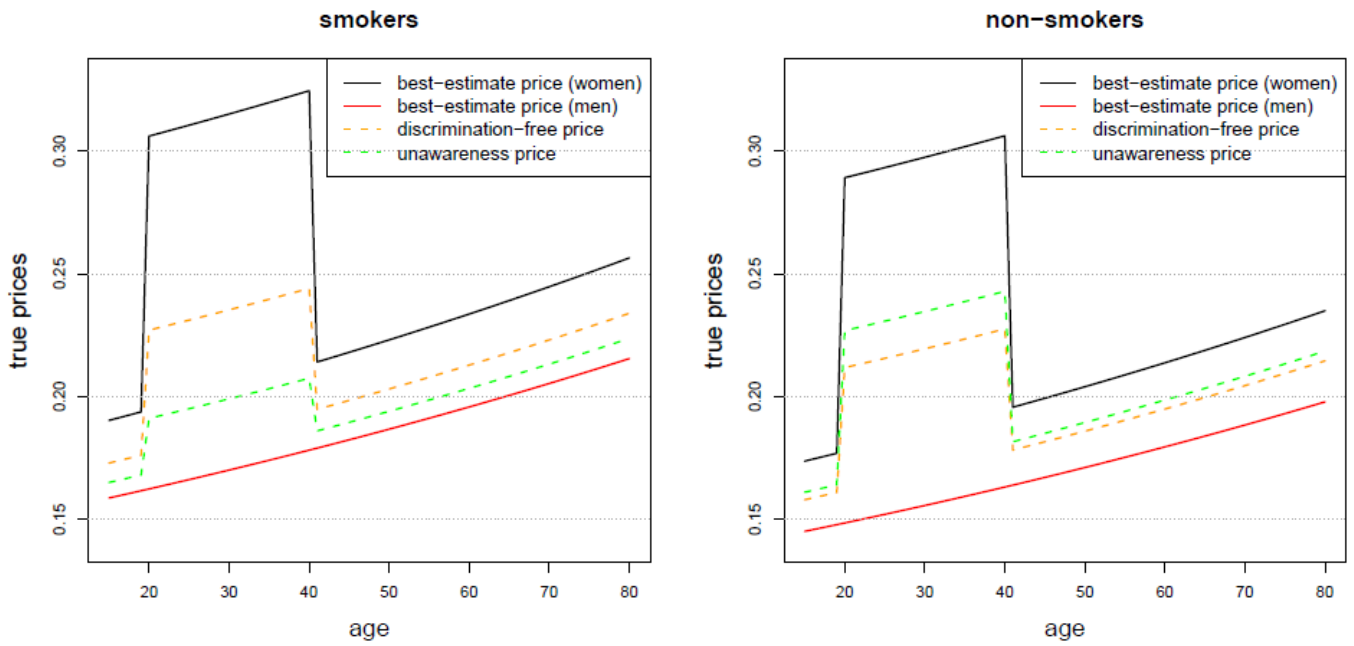
Example: Smoker ~ Female



$$P(D = female|X = smoker) = 0.8$$

Institute and Faculty of Actuaries

Example: Smoker ~ Male



$$P(D = female|X = smoker) = 0.2$$

Institute and Faculty of Actuaries

Bias Correction

- Both best-estimate and unawareness prices are unbiased => prices set for individuals will reproduce the aggregate portfolio costs.
- Discrimination-free prices lose unbiased property as we do not charge the price implied by the experience
- Need to correct for the bias to allocate all costs i.e. base rate adjustment if using DFIP
- Suggest that discrimination free prices be unbiased using several methods:
  - Additive adjustment = add a bias term to each price
  - Proportional adjustment = ratio each price by the proportion of bias
  - By adjusting the probabilities $P(D{=}d)$
- Bias correction allocates costs across portfolio instead of to a particular protected group

$$\widetilde{\lambda}_{1,\bullet}(r_{1,1}+r_{1,0})+\widetilde{\lambda}_{0,\bullet}(r_{0,1}+r_{0,0})=110.77<112$$

Institute and Faculty of Actuaries

Conclusion (1)

- Method presented in the paper can be used as a drop-in method to remove unwanted or illegal discrimination from pricing models
- May have applications wider than insurance pricing – e.g. triage of patients for hospitals using predictive models
- Depends on knowing the discriminatory variables D
  - If D is not measured, then the method cannot be applied directly
  - Hard to ask policyholders questions about ethnicity/race
- Further research needed on:
  - systemic implications of implementing DFIP
  - use of proxies for D

Institute and Faculty of Actuaries

Agenda

- Introduction
- Synthetic example
- **Missing data and multi-task networks**
- Real-world example
- Outlook and conclusions

Institute and Faculty of Actuaries

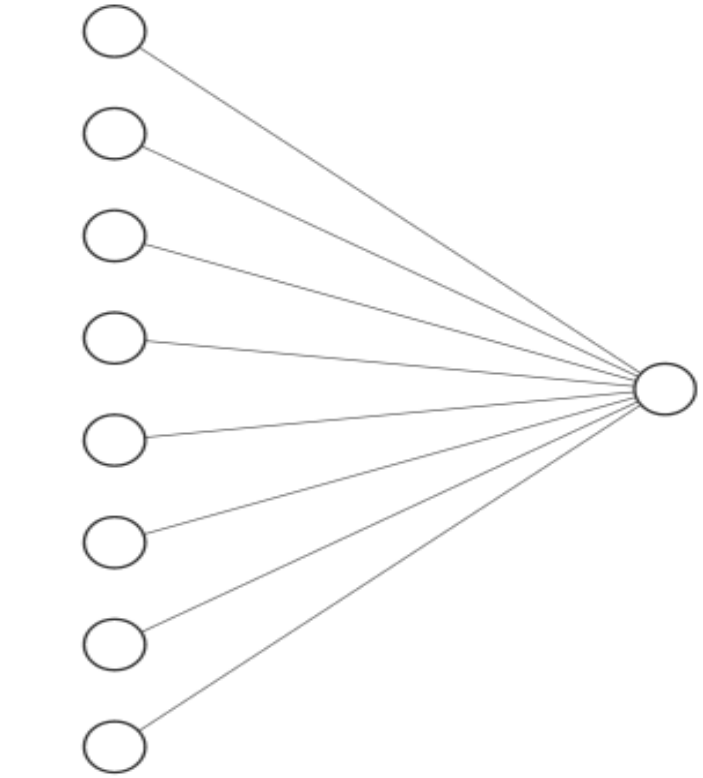
Data needs for DFIP

- DFIP Procedure needs D
  - Best-estimate prices (including D) are calculated using a model
  - Then take a weighted average of prices where the weights are independent of X
- May be the case that D is not available
- E.g. difficult to collect highly sensitive data such as ethnicity…
- … even if the only goal is to reduce potential discrimination!
- How can we then apply DFIP if we do not have access to D for the whole portfolio?
- Proposal: adapt neural networks to work in the case of missing discriminatory information

Institute and Faculty of Actuaries

Single Layer NN = Linear Regression

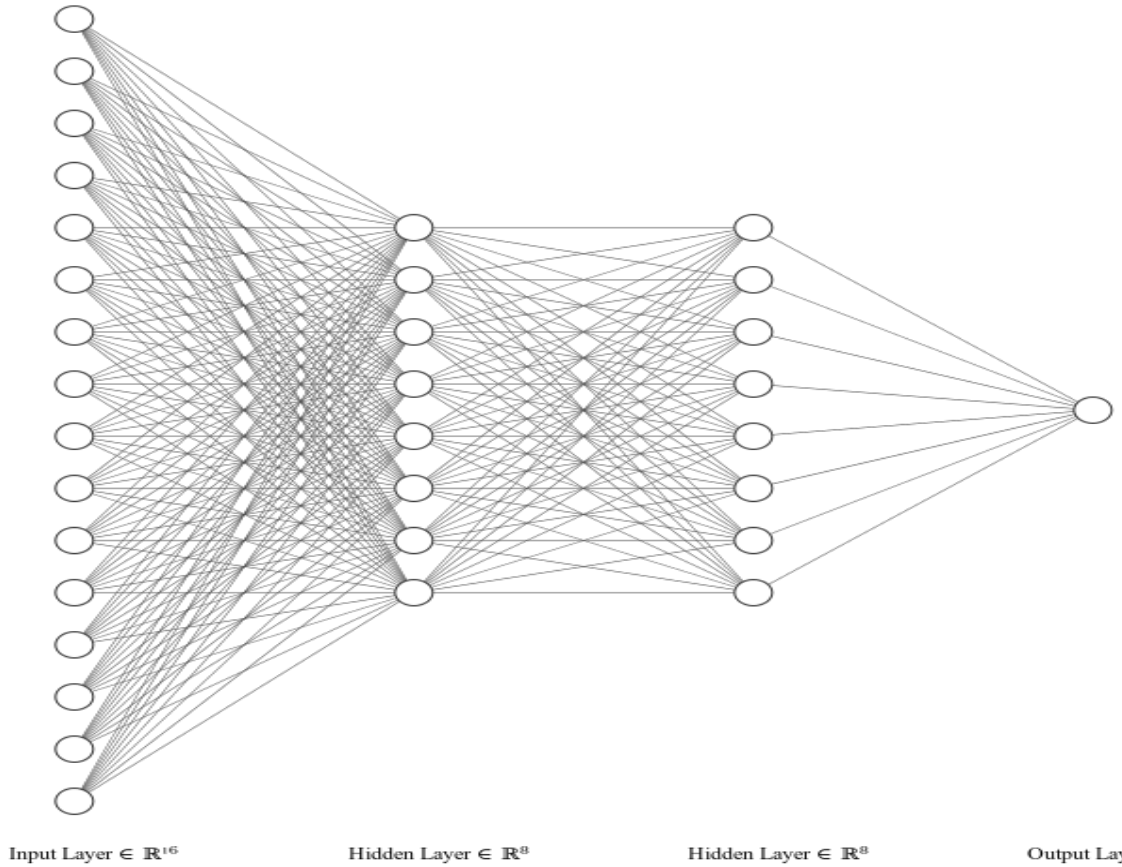
- Single layer neural network
  - Circles = variables
  - Lines = connections between inputs and outputs
- Input layer holds the variables that are input to the network…
- … multiplied by weights (coefficients) to get to result
- Single layer neural network is a GLM!



Input Layer $\in \mathbb{R}^8$

Deep Feedforward Net

- Deep = multiple layers
- Feedforward = data travels from left to right
- Fully connected network (FCN) = all neurons in layer connected to all neurons in previous layer
- More complicated representations of input data learned in hidden layers - subsequent layers represent regressions on the variables in hidden layers

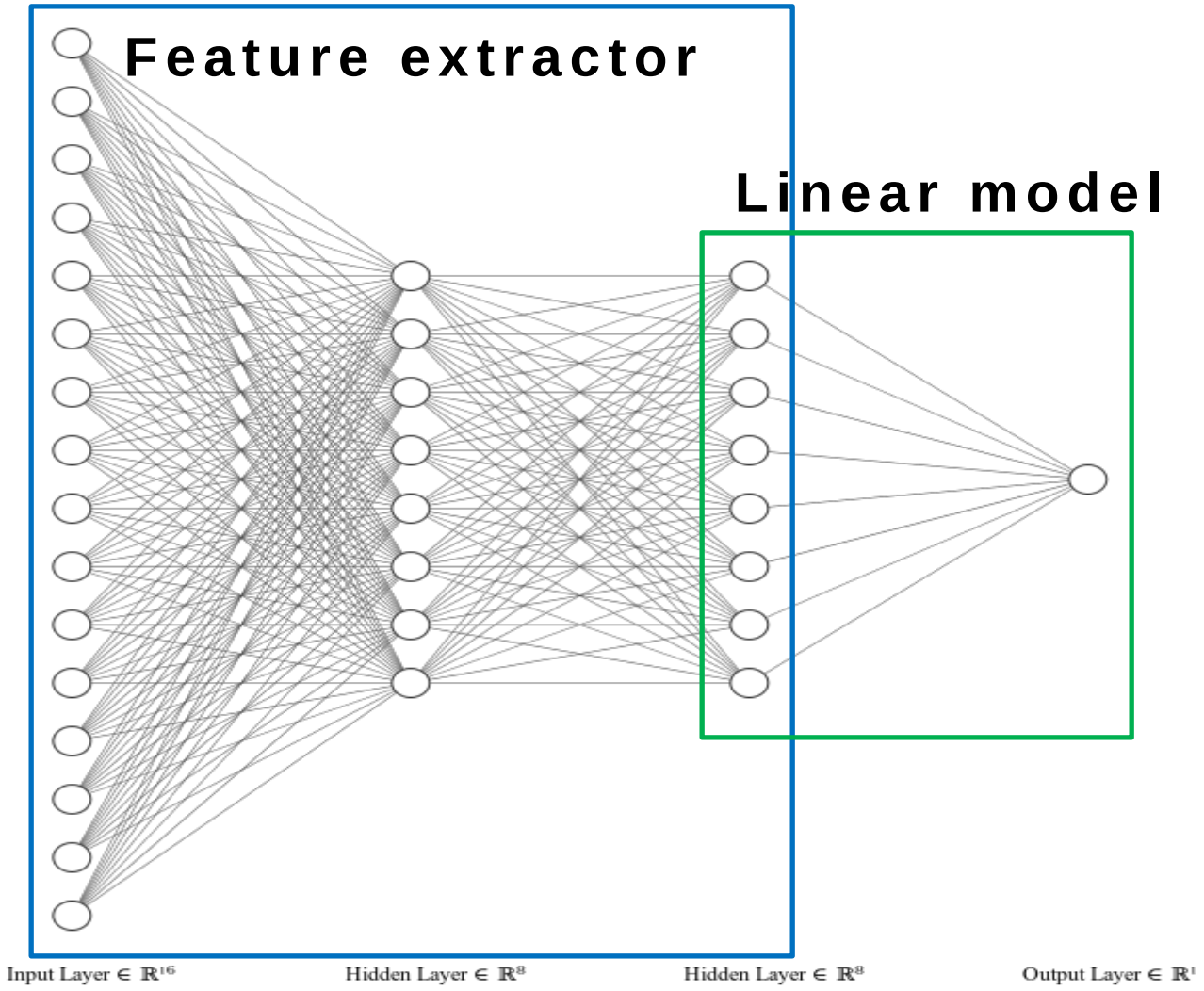


Institute and Faculty of Actuaries

FFN generalizes GLM

- Intermediate layers = representation learning, guided by supervised objective.
- Last layer = (generalized) linear model, where input variables = new representation of data
- No need to use GLM – strip off last layer and use learned features in, for example, XGBoost
- Or mix with traditional method of fitting GLM

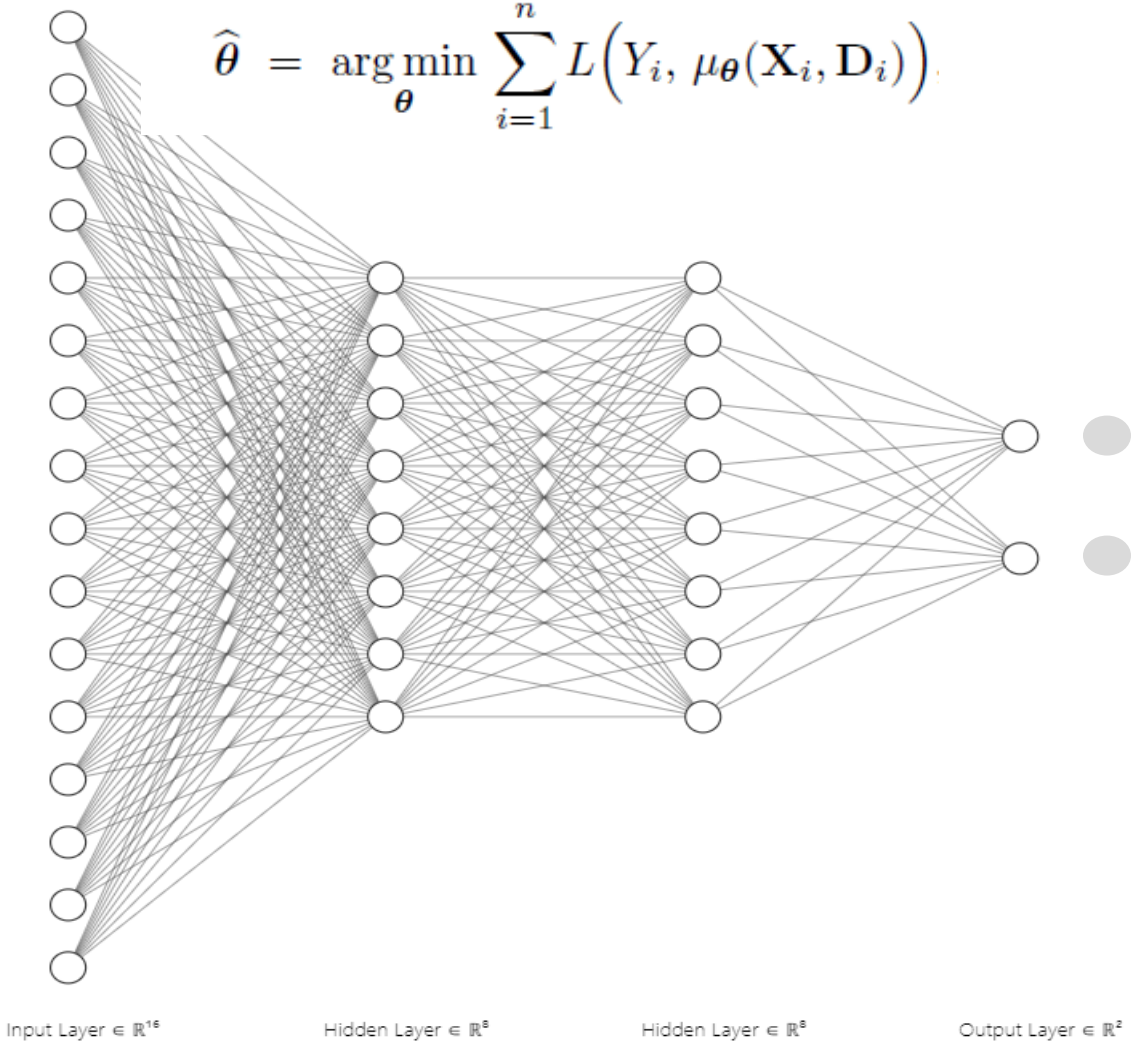
$$\mathbf{x} \quad \mapsto \quad \mathbf{z}^{(m:1)}(\mathbf{x}) = \left(\mathbf{z}^{(m)} \circ \ldots \circ \mathbf{z}^{(1)}\right)(\mathbf{x})$$



Multi-output network

- Most actuarial models output a single variable:
  - Frequency
  - Severity
  - Pure Premium
  - Single LDF
- More general class of models with multivariate outputs
- Benefit from shared representation in last layer
- How to train these models?
  - Usually we supply examples of the same dimension as the output
  - Ensure that network predicts both examples well using a relevant loss function

$$\widehat{\theta} = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} L\Big(Y_i,\ \mu_{\boldsymbol{\theta}}(\mathbf{X}_i, \mathbf{D}_i)\Big)$$

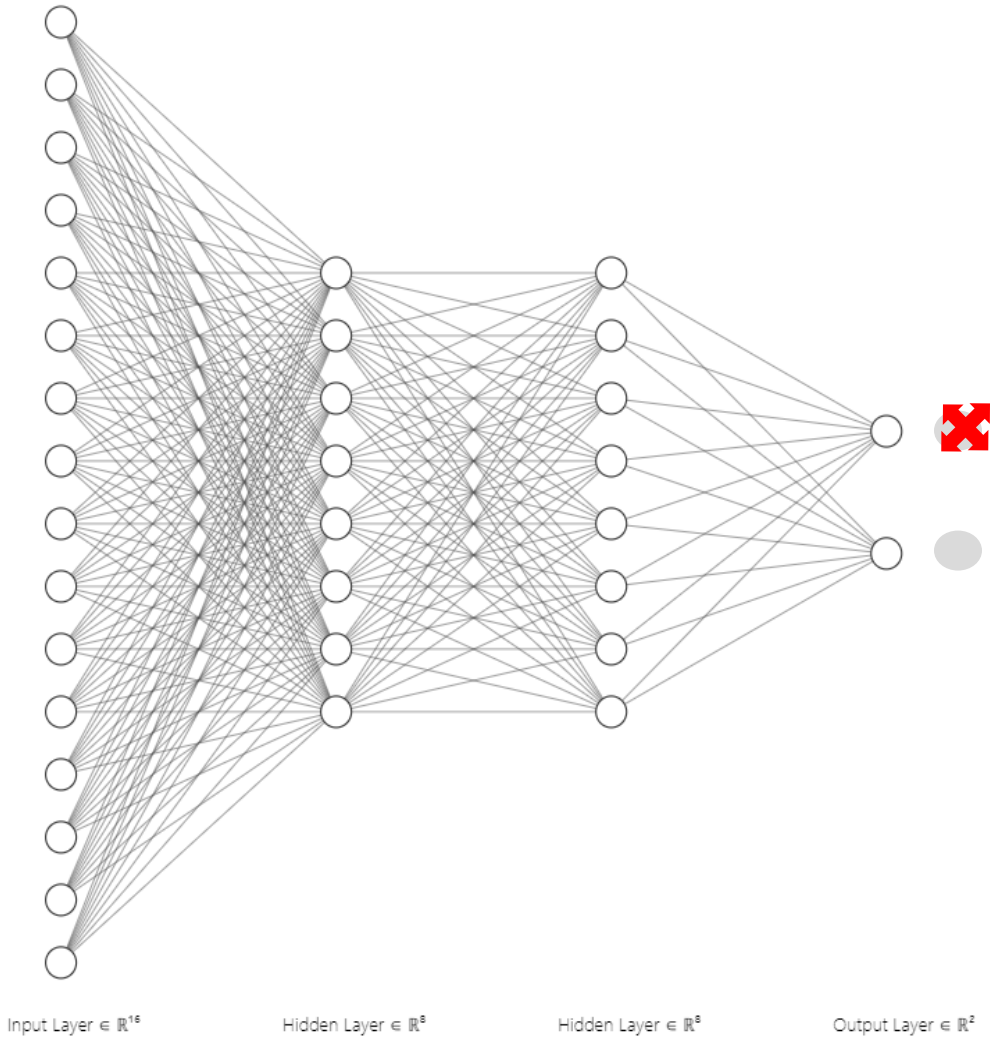


Missing some D

- In the case of DFIP we need to predict a price for each level of d in D
- $u^*(X) = \sum_d u(X, D = d)P(D = d)$
- How can we train a network to provide outputs for all levels of D if we only observe a single level d?
- Adapt the loss function to this special case using an indicator variable:

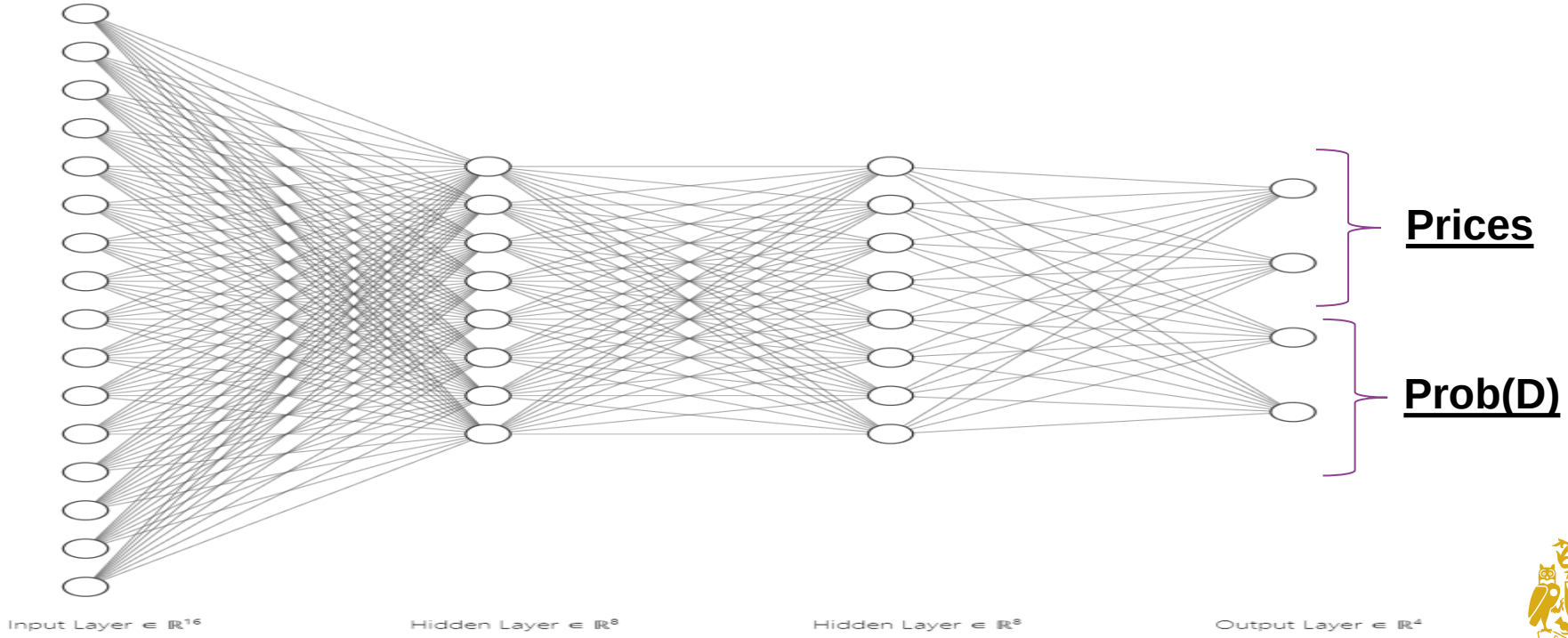
$$\widehat{\theta} = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} \sum_{k=1}^{K} L\left(Y_i, \mu_{\boldsymbol{\theta}}(\mathbf{X}_i, \mathbf{d}_k)\right) \mathbb{1}_{\{\mathbf{D}_i = \mathbf{d}_k\}}$$

- i.e. we can fit a multi-output network to the usual type of data we have in a pricing file!
- DFIP then can be estimated directly from the network outputs



Using all of the available data (1)

- How can we also benefit from using the other records i.e. all the records where D has not been recorded?
- Use network to predict both prices and probabilities for each record!



Institute and Faculty of Actuaries

Using all of the available data (2)

- Case: D **is** available
  - Train the network to predict P(D=d) and match the observed price as closely as possible
- Case: D **is not** available
  - Train the network to predict the unawareness price
- Combining these two cases, we arrive at the loss function
- Conclusion: we can train a multi-output network to provide discrimination free prices using data the both includes and excludes D!

$$p_k(\mathrm{x}) := \mathbb{P}[\mathbf{D} = \mathrm{d}_k \mid \mathbf{X} = \mathrm{x}]$$

$$\mu(\mathrm{x}) = \sum_{k=1}^{\cdot\cdot} \mu(\mathrm{x}, \mathrm{d}_k)\, p_k(\mathrm{x})$$

$$\begin{aligned}\widehat{\theta} \;=\; \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} \Bigg[& \sum_{k=1}^{K} L_\mu\Big(Y_i,\, \mu(\mathbf{X}_i, \mathrm{d}_k)\Big)\, \mathbb{1}_{\{\mathbf{D}_i = \mathrm{d}_k\}} \\ & + L_p\Big(\mathbf{D}_i,\, (p_k(\mathbf{X}_i))_{1\le k\le K}\Big)\, \mathbb{1}_{\{\mathbf{D}_i \neq \mathtt{NA}\}} + L_{\widetilde{\mu}}\Big(\widetilde{\mu}(\mathbf{X}_i),\, \mu(\mathbf{X}_i)\Big)\Bigg]\end{aligned}$$

Institute and Faculty of Actuaries

Agenda

- Introduction
- Synthetic example
- Missing data and multi-task networks
- **Real-world example**
- Outlook and conclusions

Institute and Faculty of Actuaries

Real-world dataset

- Historical dataset of experience in around 2000
  - Contributed by anonymous multinational insurer
  - Usual PL rating factors (policyholder/vehicle) – 19 factors
- Motor coverages (hull/third party property and/or bodily injury)
- ~42 000 claims and ~166 000 years of exposure
- Insurer records ethnicity to track insurance market penetration
  - 5 ethnicity codes (defined in the insurer's jurisdiction)
- Exact coverages and excesses not disclosed
- => not useful for commercial purposes

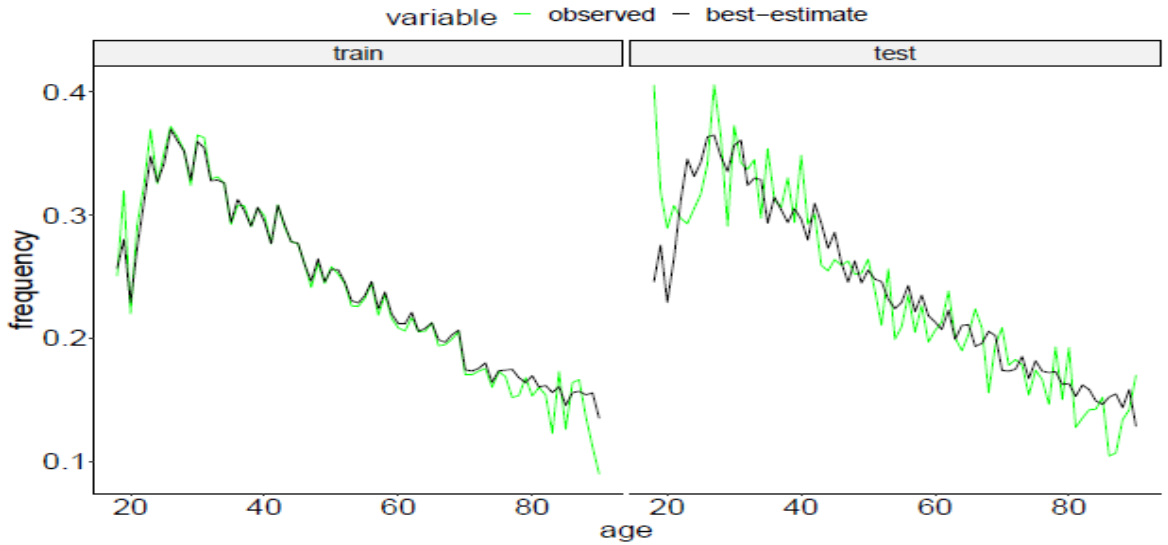
$$\mathbf{D} \in D_5 = \{1, 2, 3, 4, 5\}$$

ethnicity code	number of claims	exposure	frequency
1	5,223	14,317	36.48%
2	965	3,925	24.59%
3	3,354	14,363	23.35%
4	5,249	20,240	25.93%
5	26,817	112,667	23.80%

Institute and Faculty of Actuaries

Modelling Details

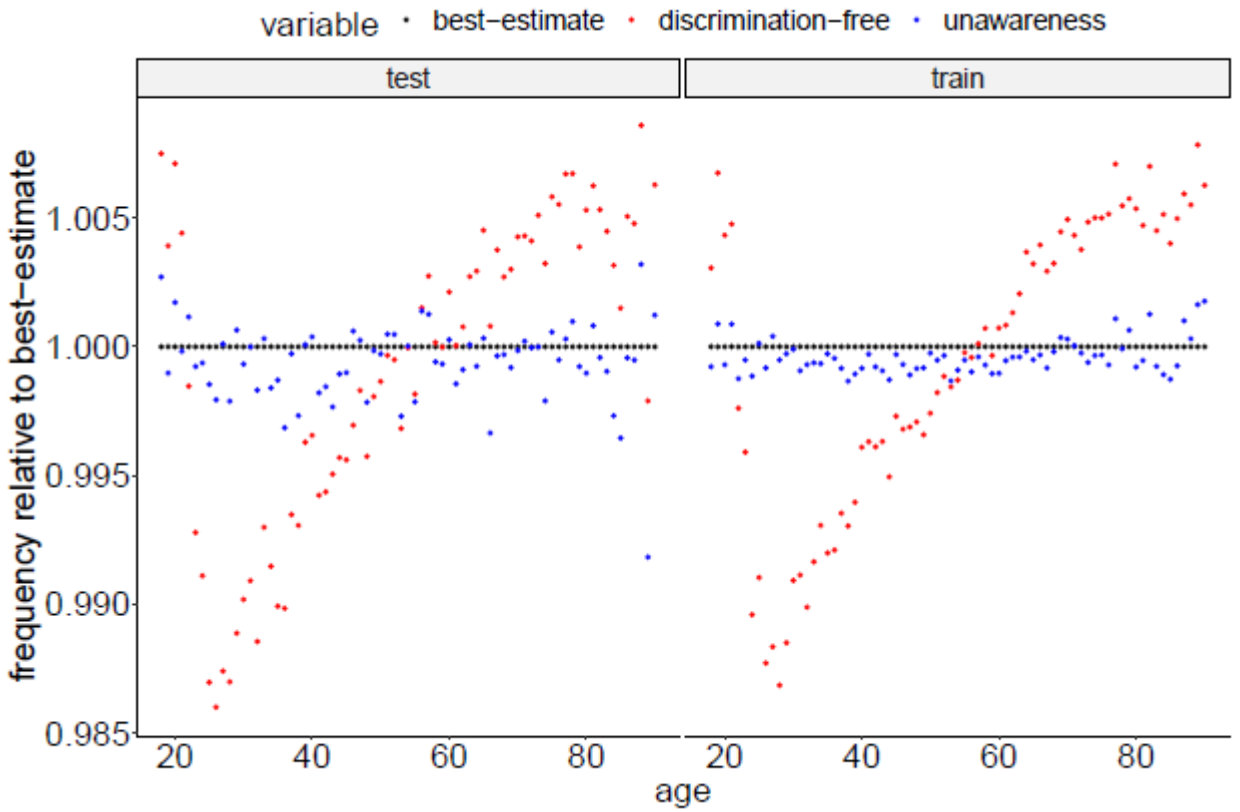
- Used a deep neural network with embedding layers for categorical data
- Averaged over 20 different training runs of the network; see `Nagging Predictors` Richman and Wuthrich (2020)
- Regularized using dropout and batch normalization
- 80%/20% training/test set split; 5% of training set used for validation



Institute and Faculty of Actuaries

Comparing prices

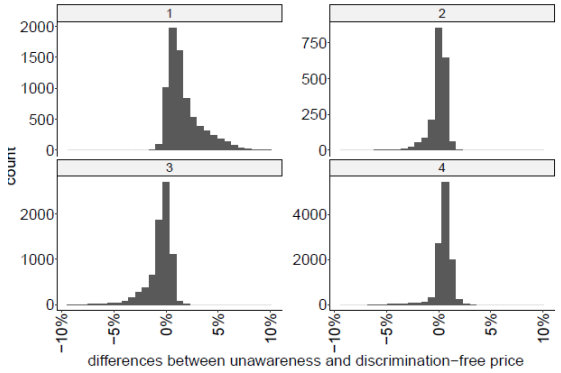
- Comparison of best-estimate prices relative to unawareness and discrimination-free prices
- Narrow range of about 1.5% around best estimate price
- DFIP most different at the youngest ages
- Unawareness price tracks best-estimate closely =>
- Possible to infer ethnicity implicitly from ***X*** i.e. indirect discrimination in this portfolio

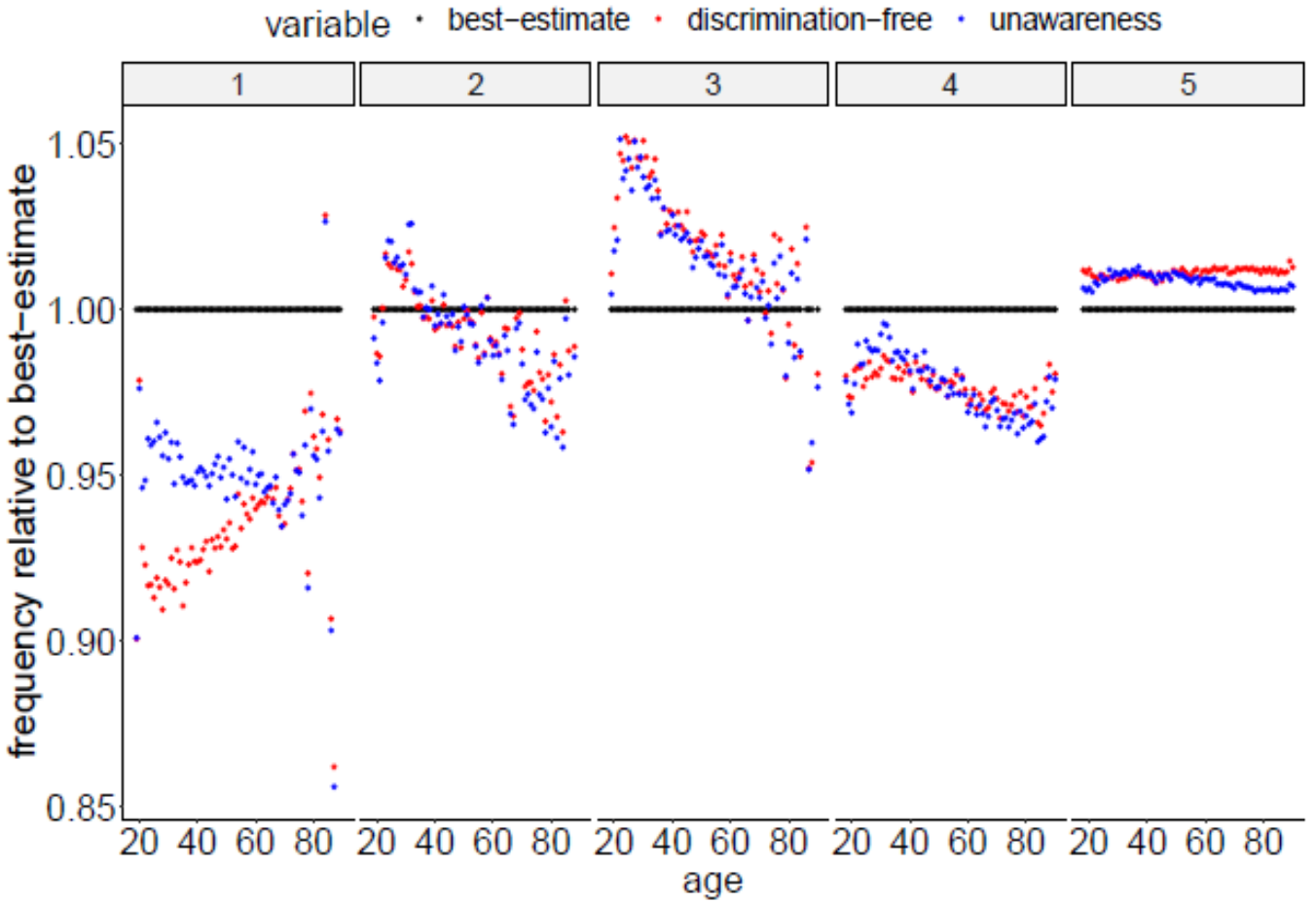


Institute and Faculty of Actuaries

Impact by ethnicity

- D = 5, largest group in the book – not much difference between unawareness and discrimination-free…
- DFIP slightly higher as low frequency for group 5
- Largest divergences occur for D = 1 at younger ages of more than 5%





Institute and Faculty of Actuaries

How predictive is DFIP?

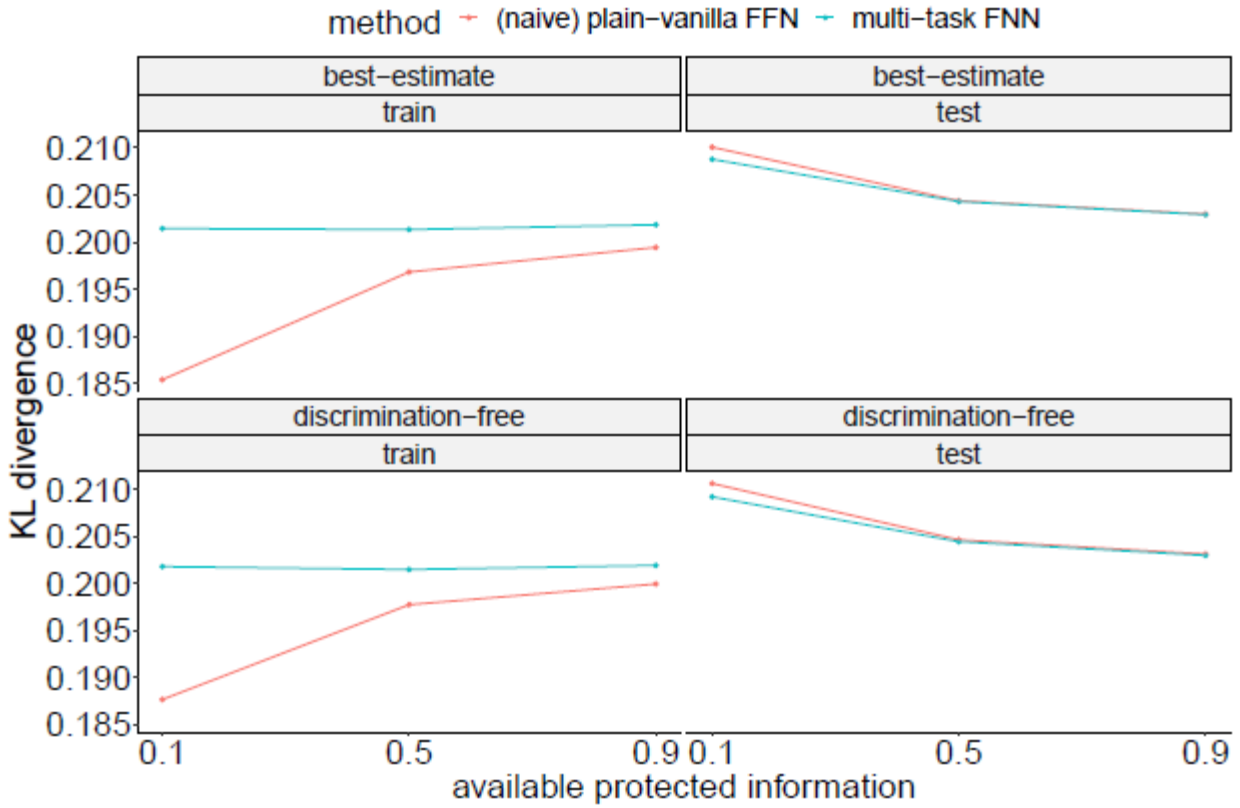
- Measure difference between observed claims and best-estimate/unawareness/discrimination-free prices using Kullback-Leibler divergence
  - KL divergence = 0.5 x Poisson Deviance
- Small differences between 3 sets of prices on training set…
- … and even less difference on test set
- => out-of-sample DFIP overfits a bit less than best-estimate

	training set	test set
plain-vanilla FNN best-estimate price $\mu(\mathbf{x}, \mathbf{d})$	0.20028	0.20295
plain-vanilla FNN unawareness price $\mu(\mathbf{x})$	0.20055	0.20302
plain-vanilla FNN discrimination-free price $\mu^*(\mathbf{x})$	0.20063	0.20304

Institute and Faculty of Actuaries

Missing Discriminatory Information

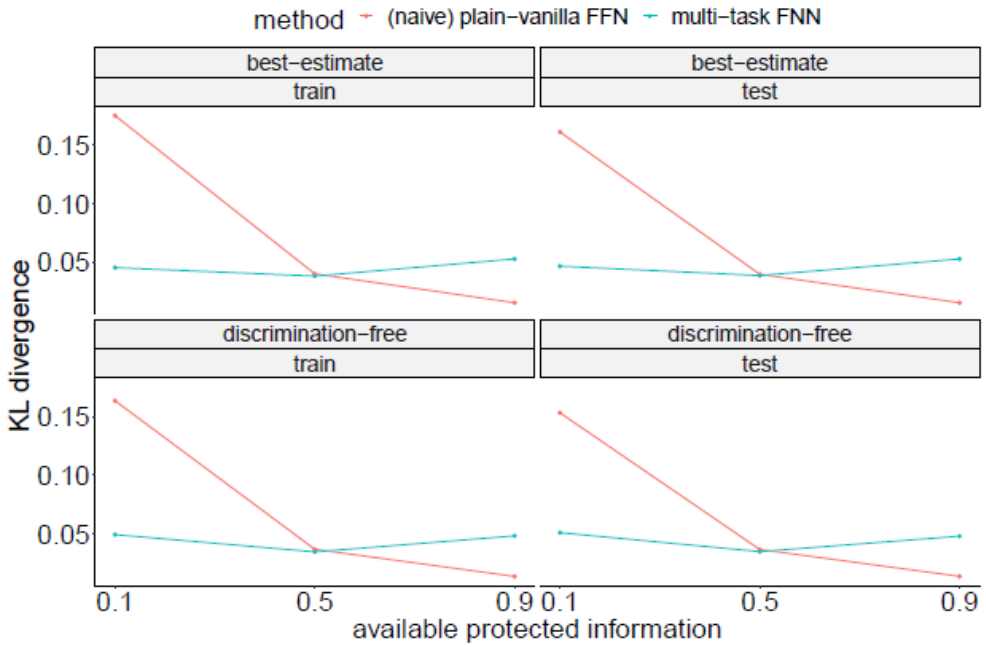
- Drop out D with probabilities of 10%, 50% and 90%
- Apply multi-task network approach
- For comparison, fit FFN **only** to those observations with D (naïve)
- As above, average over 20 network calibrations for each level of drop-out
- Training set – multi-task approach appears worse but…
- … on test set, as good or better than naïve approach (overfitting)



Institute and Faculty of Actuaries

How well does multi-task approach approximate?

- Naïve approach is approximates the estimated rates better than the multi-task approach for drop-out probability = 10%
- Multi-task is as good or better for higher probabilities
- Difference between predictive performance and price charged!



Institute and Faculty of Actuaries

Agenda

- Introduction
- Synthetic example
- Missing data and multi-task networks
- Real-world example
- **Outlook and conclusions**

Institute and Faculty of Actuaries

Outlook and Conclusions (1)

- Only looked at technical price; what about office premiums?
- If loadings additive/multiplicative then DFIP necessary on technical rate
- Personal views: due to competitiveness of personal lines space, relatively unlikely to be implemented without regulatory intervention
- No substantial loss of accuracy but important changes for some groups of policyholders => contradiction?
  - Group for whom this makes significant differences is small
  - Not much of a difference even if some prices different for one group when predicting noisy out of sample claims

Institute and Faculty of Actuaries

Outlook and Conclusions (2)

- For real-world portfolios similar to this example, could use DFIP without losing too much predictive accuracy
- Might be important for some groups but depends on having access to discriminatory information
- In cases of significant missing data, multi-task network by far outperforms normal NN
- How can we get D on some of our portfolio?
  - Commercial scheme – offer discount/bonus to customers willing to disclose D
  - => selection bias?
  - Survey sampling?

Institute and Faculty of Actuaries

Questions

Comments

Expressions of individual views by members of the Institute and Faculty of Actuaries and its staff are encouraged.

The views expressed in this presentation are those of the presenter.

Institute and Faculty of Actuaries

Institute and Faculty of Actuaries

Thank you

#GiroConf22