A New Method for Discrimination Free Pricing and Real-world Impacts

Ronald Richman
FIA FASSA CERA CPCU

Old Mutual Insure

EAA e-Conference on
Data Science & Data Ethics

16 May 2023

AGENDA

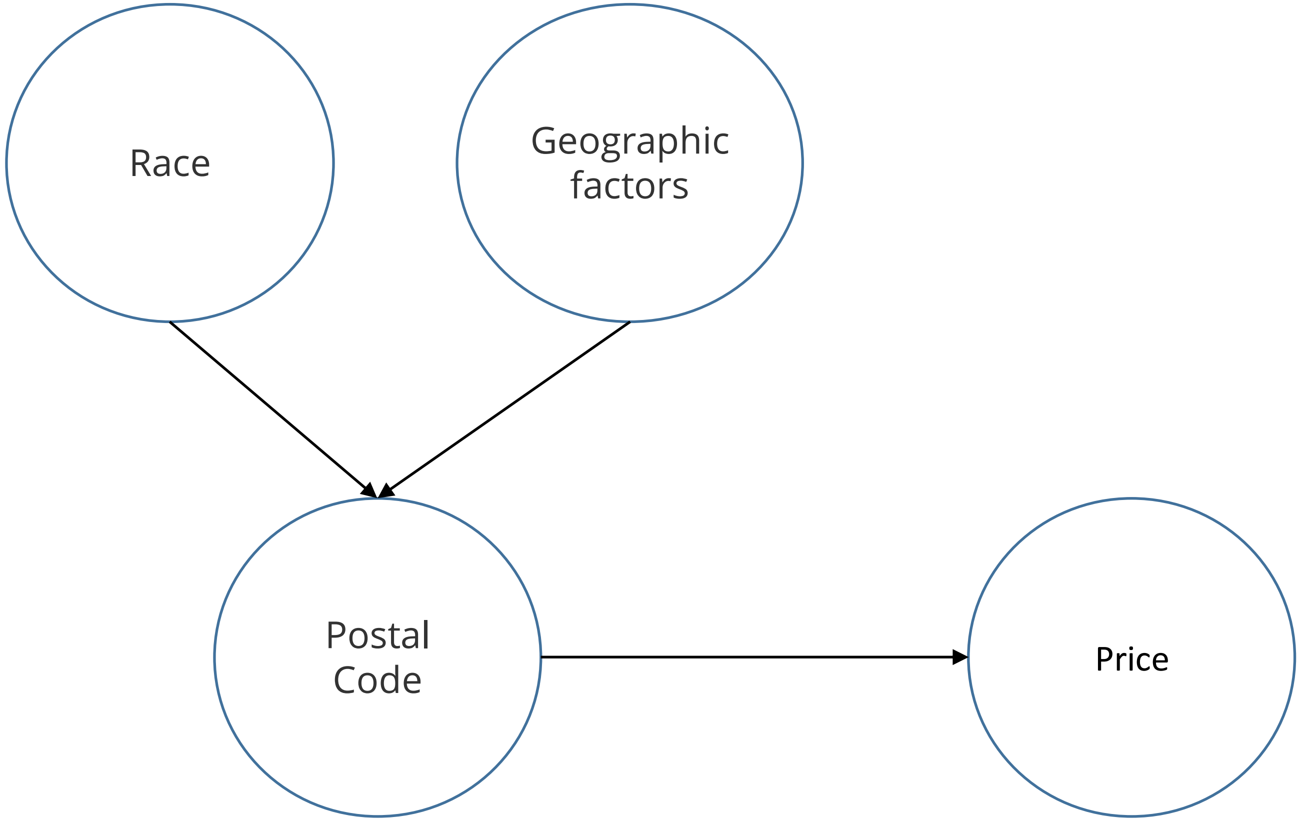
- Introduction
- Missing data and multi-task networks
- Real-world example
- Outlook and conclusions

INTRODUCTION

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

RATIONALE


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

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.

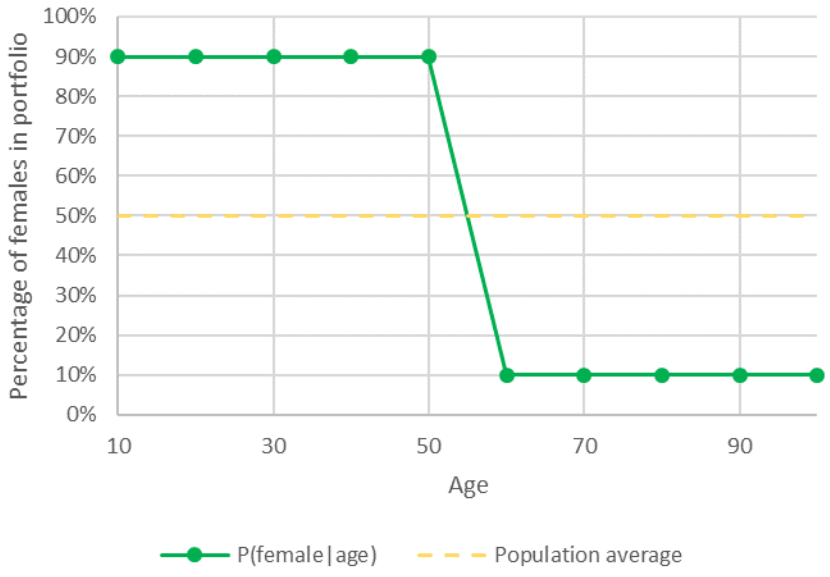
Note: We rely on society to guide us as to the definition of a discriminatory factor; in this talk we are concerned with methods for correcting pricing once discriminatory factors are defined

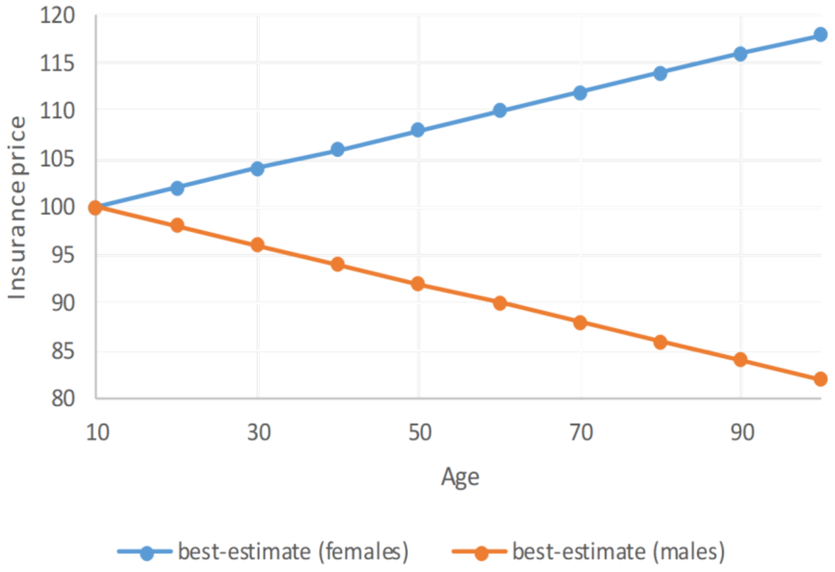
Include other factors within pricing model that are highly correlated with the discriminatory factor
Can pick up much of the same effect – e.g. annual driving distance

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:
  - $u(X,D) = E[Y|X,D]$
- For complex lines of business, we approximate $E[Y|X,D]$ 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:
  - $u(X) = E[Y|X]$
- Or relies on proxies for *D* to get closer to the best estimate price…

EXAMPLE – WHAT IS DFIP PRICE?

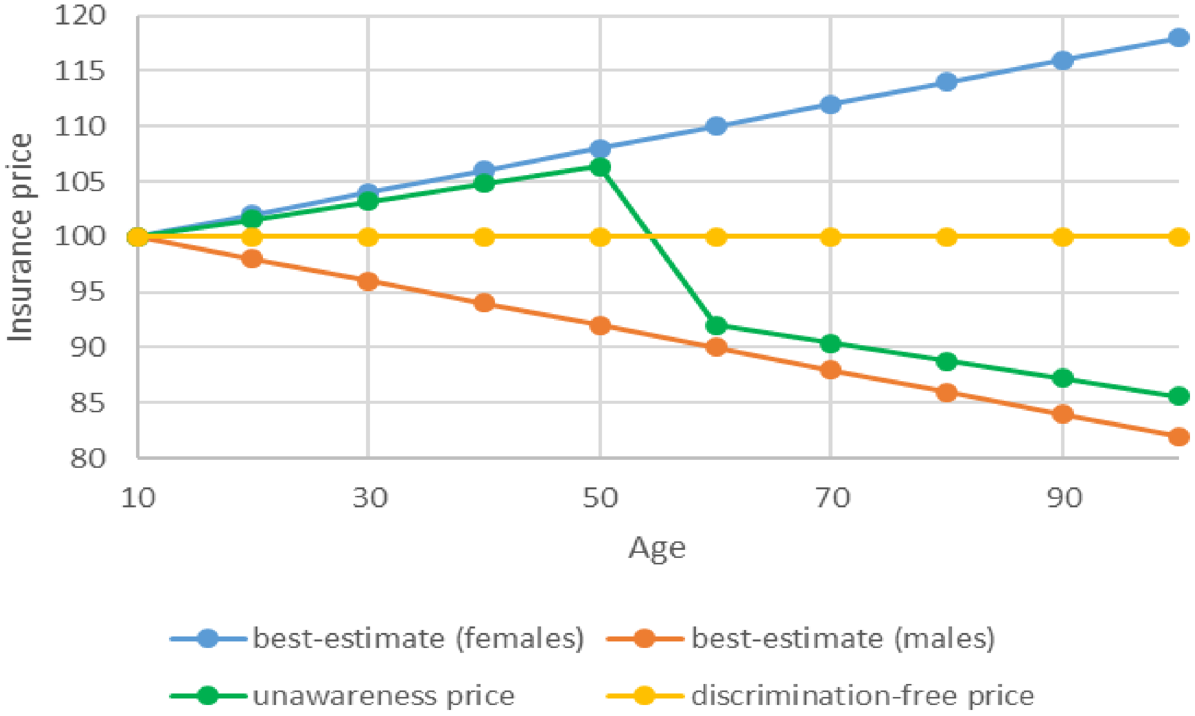




Portfolio proportions: Distribution of gender across age classes and population average

Best estimate costs

DISCRIMINATION FREE PRICES



Unawareness prices act as proxies for discriminatory factors

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

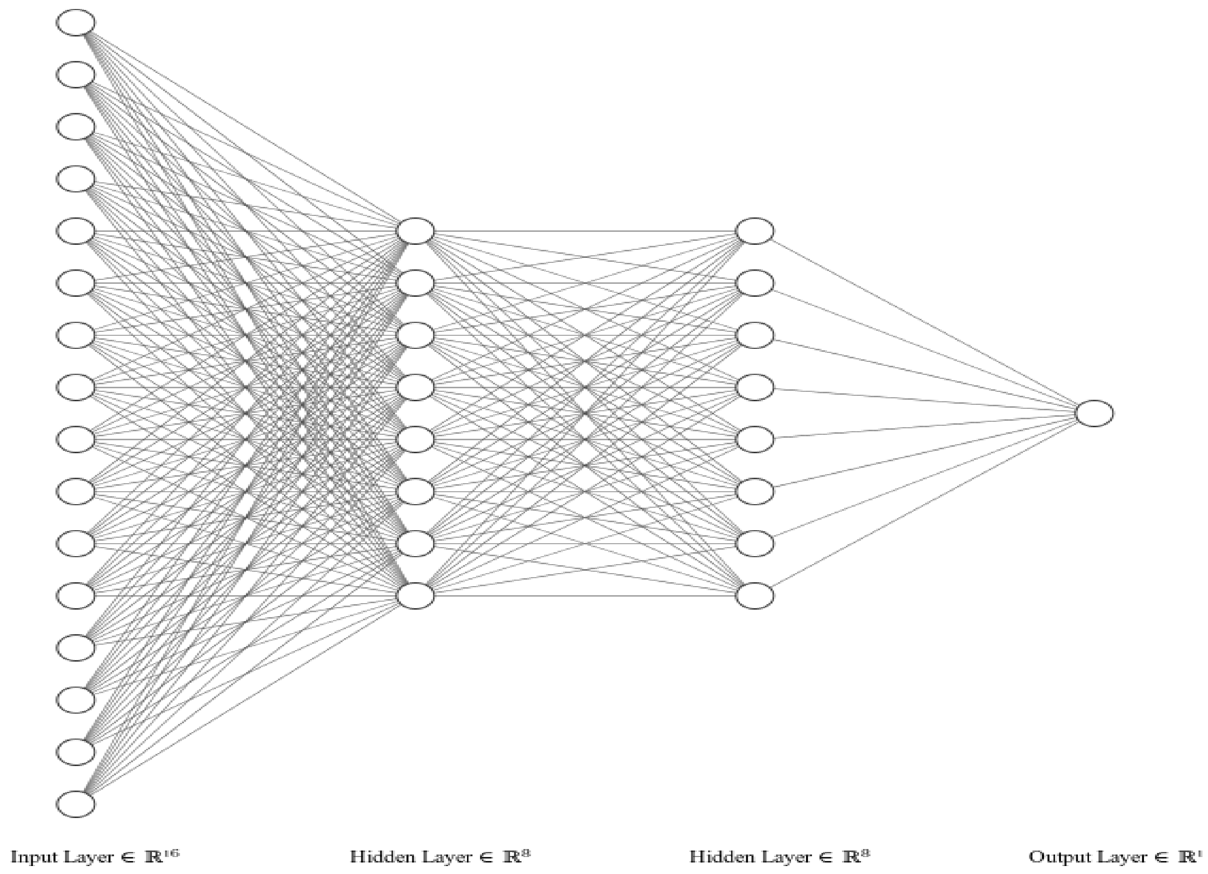
MISSING DATA AND MULTI-TASK NETWORKS

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

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

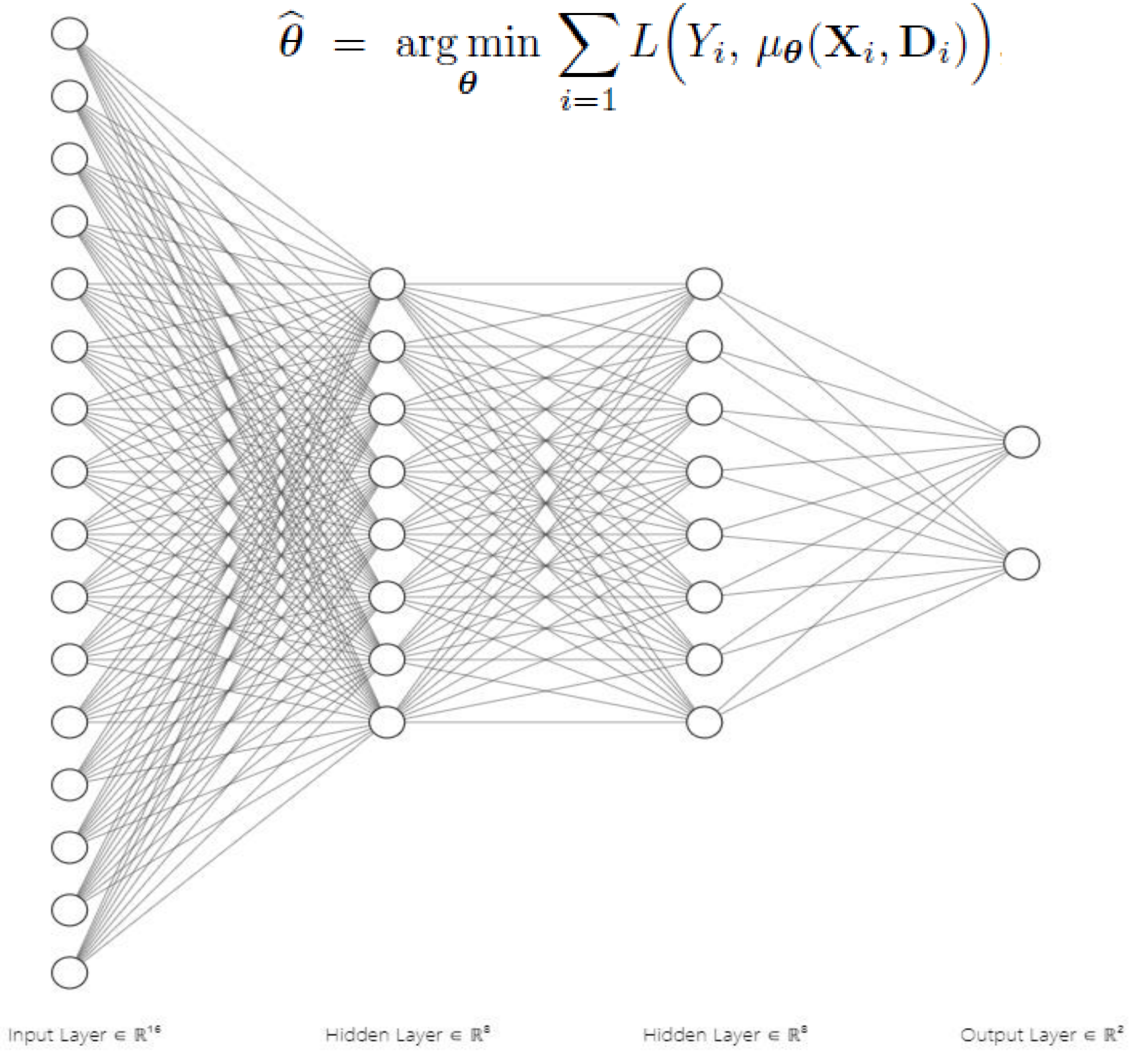


MISSING DATA AND MULTI-TASK NETWORKS

MULTI-OUTPUT NETWORK

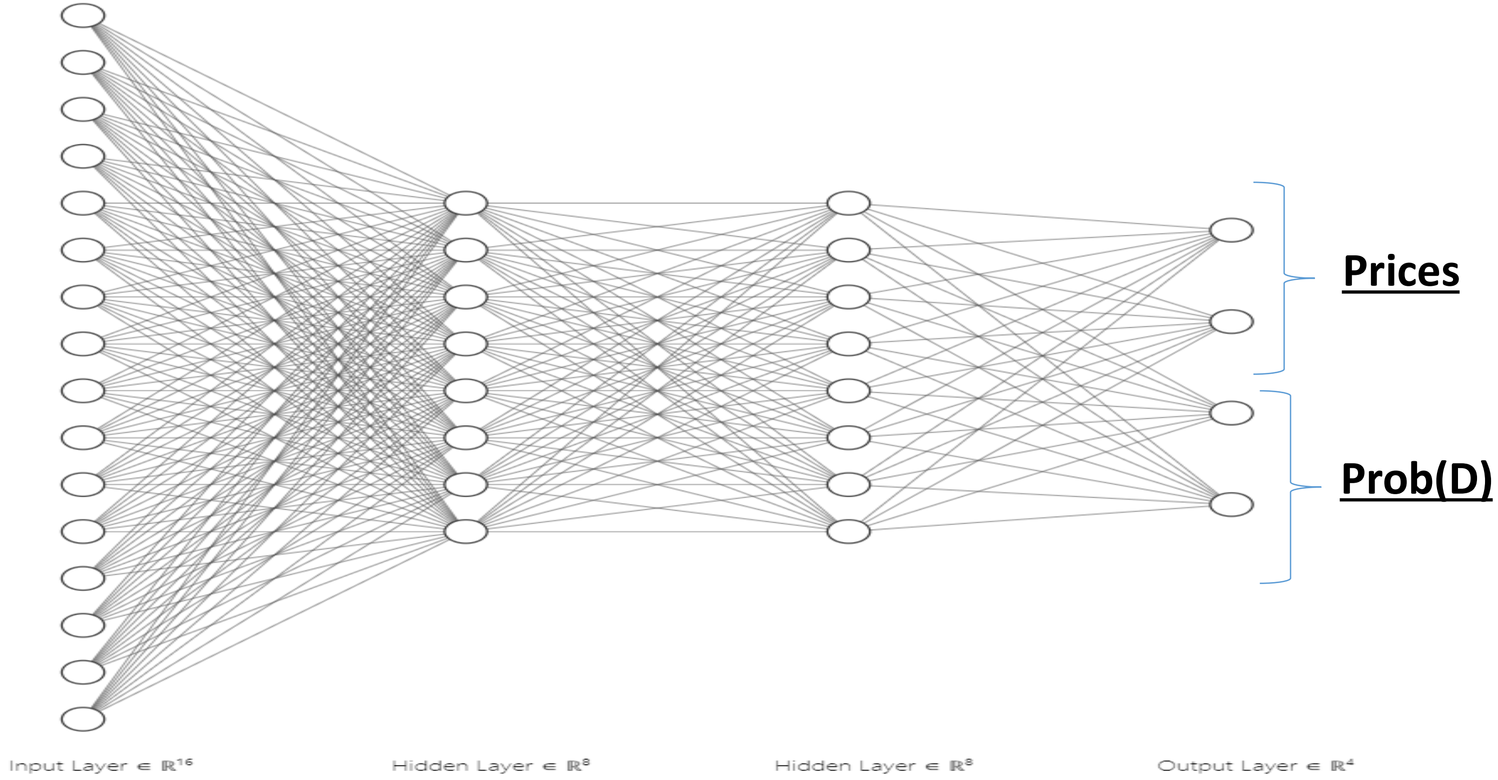
- Most actuarial models output a single variable:
  - Frequency
  - Severity
  - Pure Premium
  - Single LDF
- More general class of models with multivariate outputs
- Benefit from shared representation in last layer
- How to train these models?
  - Usually we supply examples of the same dimension as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

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



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!



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

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

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

- Combining these two cases, we arrive at the loss function

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

- Conclusion: we can train a multi-output network to provide discrimination free prices using data the both includes and excludes D!

REAL-WORLD EXAMPLE

REAL-WORLD DATASET

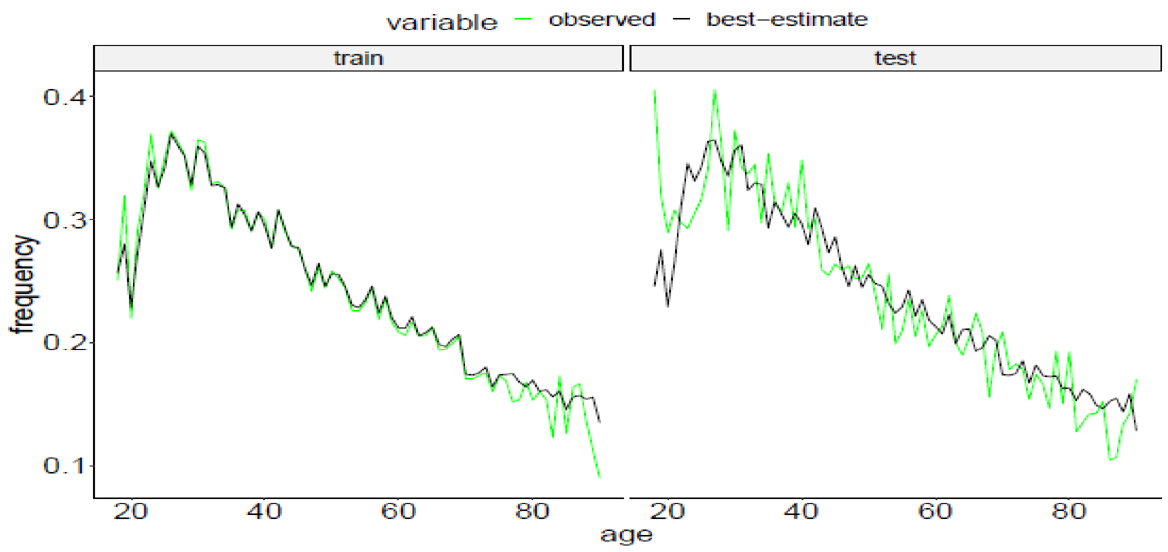
- Historical dataset of experience in around 2000
  - Contributed by anonymous multinational insurer
  - Usual PL rating factors (policyholder/vehicle) – 19 factors
- Motor coverages (hull/third party property and/or bodily injury)
- ~42 000 claims and ~166 000 years of exposure
- Insurer records ethnicity 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%

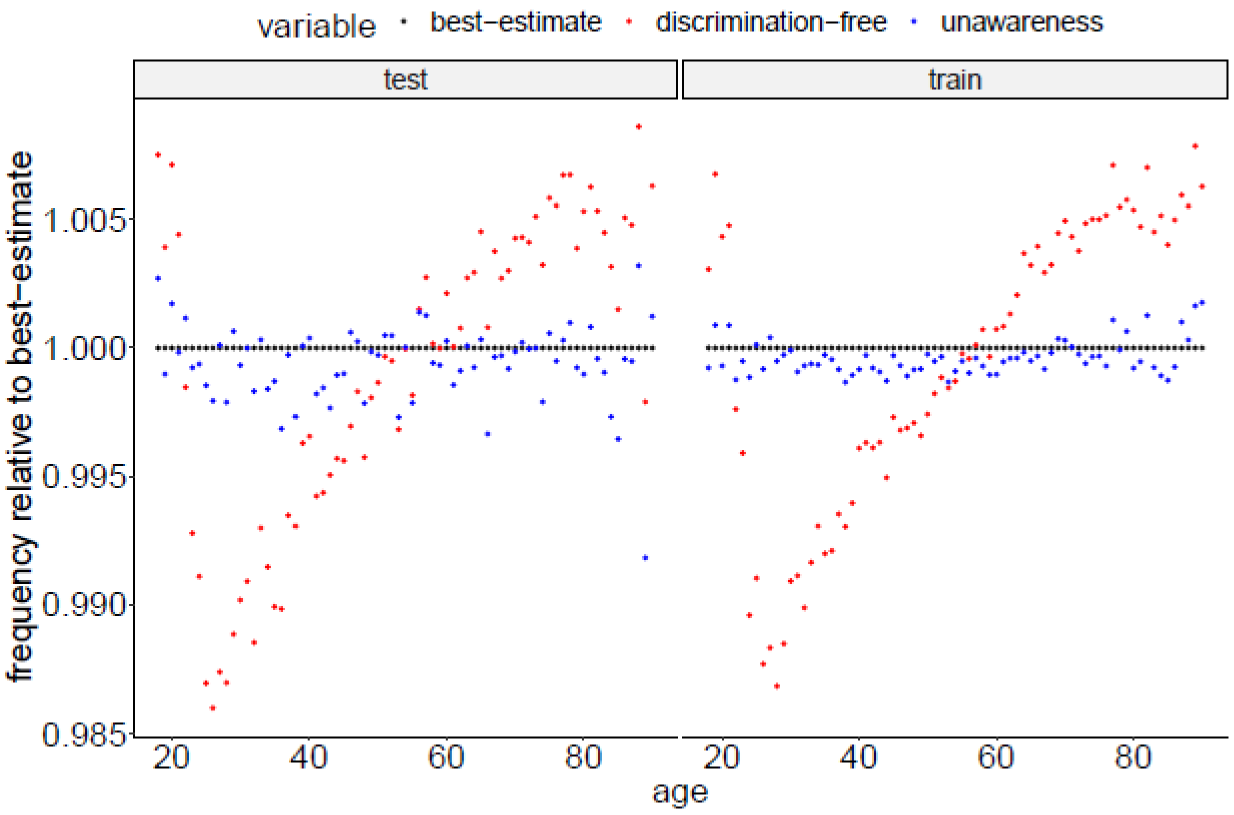
MODELLING DETAILS

- 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



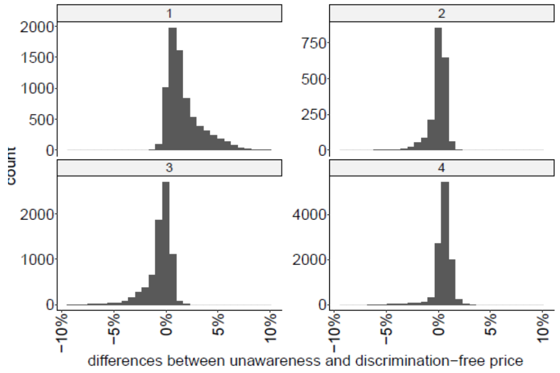
COMPARING PRICES

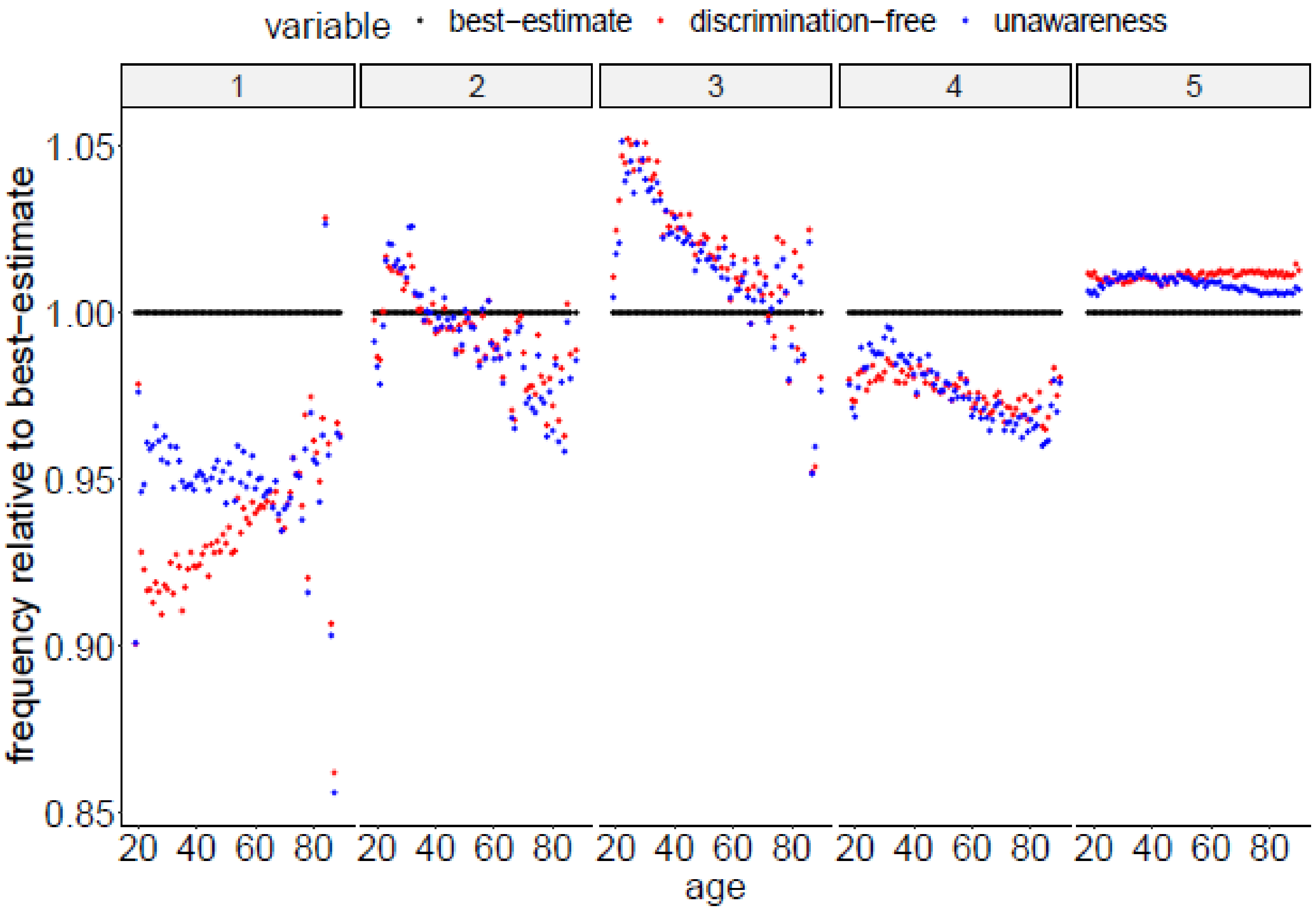
- Comparison of best-estimate prices relative to unawareness and discrimination-free prices
- Narrow range of about 1.5% around best estimate price
- DFIP most different at the youngest ages
- Unawareness price tracks best-estimate closely =>
- Possible to infer ethnicity implicitly from ***X*** i.e.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



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%





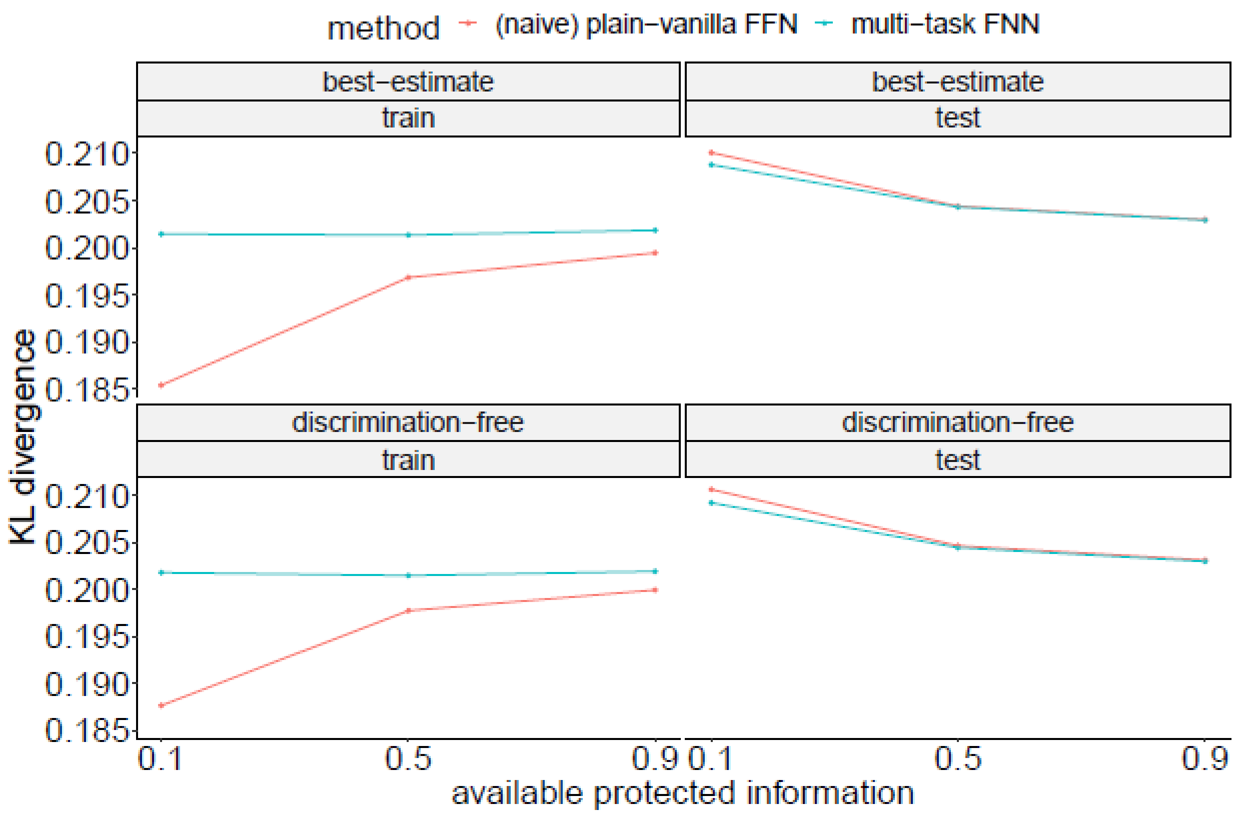
HOW PREDICTIVE IS DFIP?

- 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

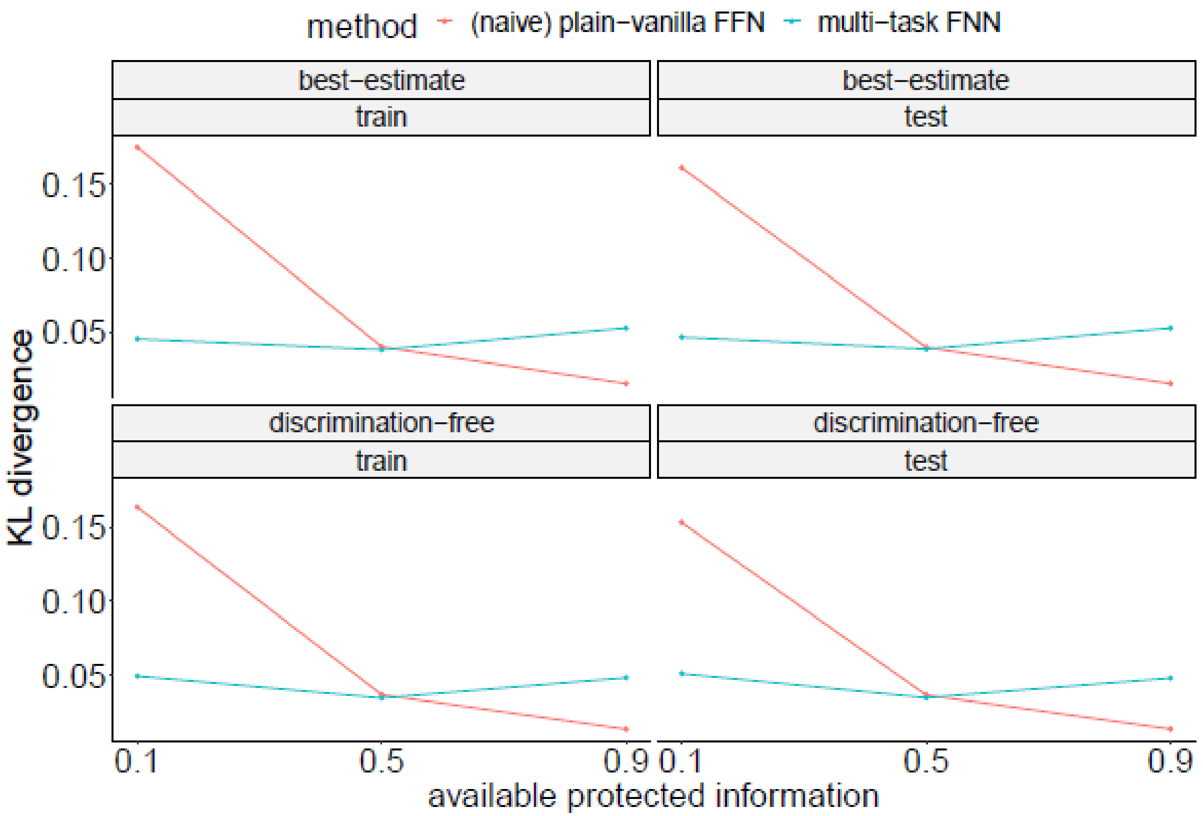
MISSING DISCRIMINATORY INFORMATION

- 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)



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!



OUTLOOK AND CONCLUSIONS

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

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?

QUESTIONS?

Comments?

Ron Richman is Chief Actuary at Old Mutual Insure, where he is responsible for oversight of all actuarial activities in the OMI Group. Ron is a Fellow of the Institute and Faculty of Actuaries (IFoA) and the Actuarial Society of South Africa (ASSA), holds practicing certificates in Short Term Insurance and Life Insurance from ASSA, and a Masters of Philosophy in Actuarial Science, with distinction, from the University of Cape Town. He chairs the Actuarial Society of South Africa's Climate Change committee and is a member of the ASTIN Board.

ABOUT ME

Ronald Richman

FIA FASSA CERA CPCU

Old Mutual Insure

EAA e-Conference on
Data Science & Data Ethics

16 May 2023

Thank you very much for your attention

Contact

Ronald Richman FIA FASSA CERA CPCU

Old Mutual Insure

Ronald.Richman@ominsure.co.za

+27791337248