A Additional Figures

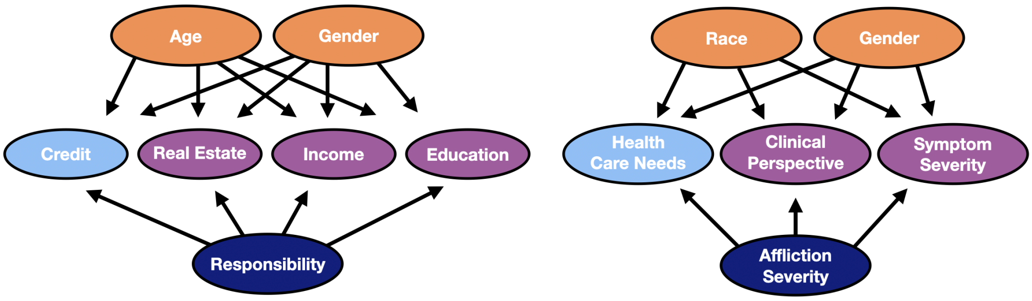

Figure 5 The causal diagrams for the default risk example (on the left) and healthcare example (on the right).

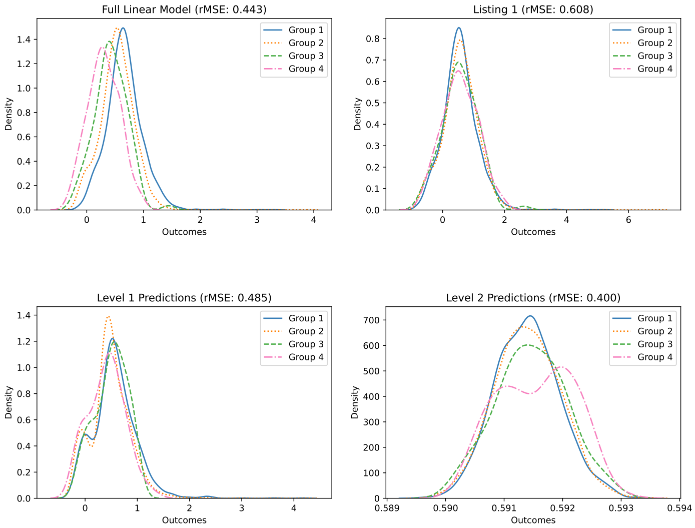

Figure 6 Density of different predictions for the default risk data set. For each group, the distribution of the predictions made by the full linear model are slightly shifted. Surprisingly, the predictions made by the three levels appear to come from different types of distributions. In contrast, our two algorithms make predictions that are essentially the same across groups, satisfying demographic parity.

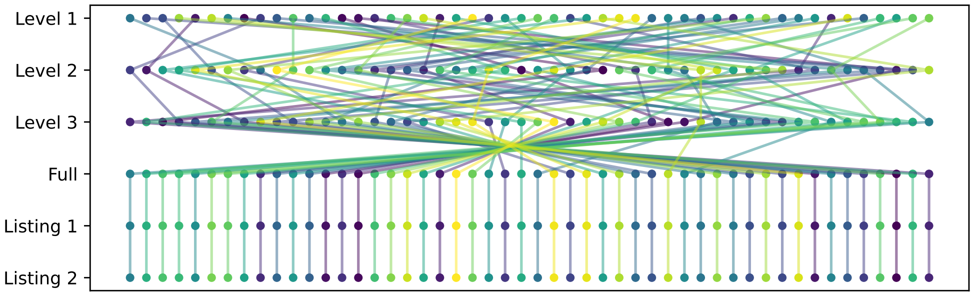

Figure 7 We randomly sample 50 individuals from Group 1 in the default risk example and visualize how the relative order of their predictions changes based on the algorithm making predictions. As we saw with the law school example, the three levels make drastically different predictions at the individual level.

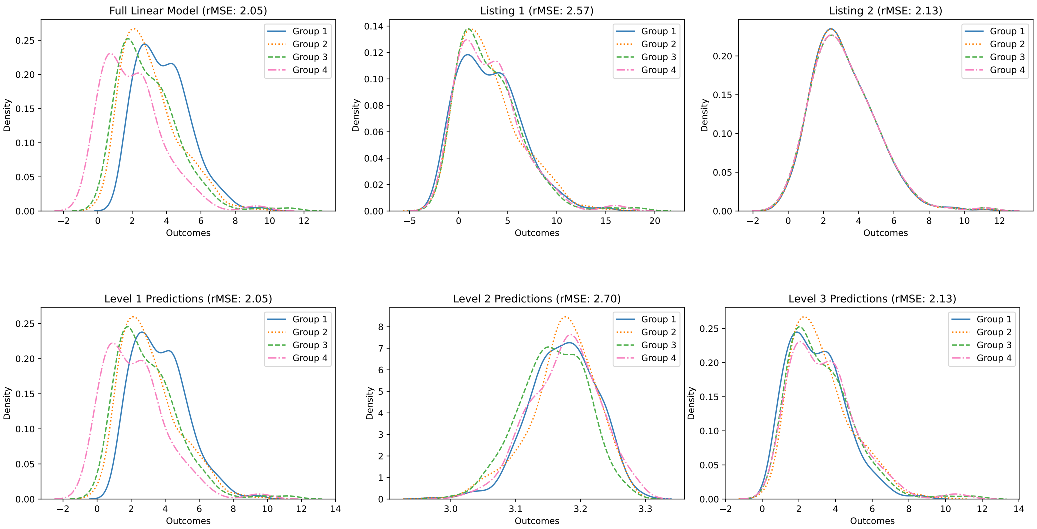

Figure 8 Density of different predictions for the healthcare data set. We intentionally made the distributions between different groups dissimilar, in order to create a challenging context for the algorithms we consider. Since the score distributions are not normal, Listing 2 outperforms Listing 1.

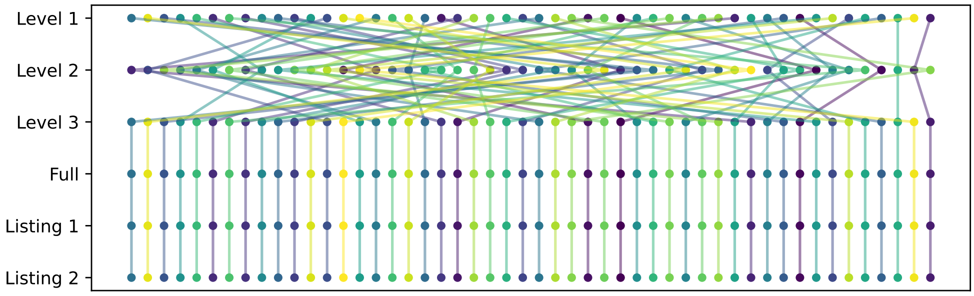

Figure 9 We randomly sample 50 individuals from Group 1 in the healthcare example and visualize how the relative order of their predictions changes based on the algorithm making predictions. Unlike the law school and default risk examples, Level 3 ranks individuals the same as the full linear model. We suspect this is because the data is generated from a causal model with linear functions on each attribute. As a result, the abductions made by Level 3 preserve the order of individual predictions.