Altenburger, Kristen M, and Daniel E Ho. "When Algorithms Import Private Bias into Public Enforcement: The Promise and Limitations of Statistical Debiasing Solutions." *Journal of Institutional and Theoretical Economics*, 2018, pp. 1–25., doi:10.1628/jite-2019-0001.

<u>Possible debiasing solution</u>

Section 4 of the article, which begins on page 12, is the beginning of the part of the article where they talk about a debiasing solution. There is one proposed method by Pope and Sydnor. This method addresses contentious predictors that could proxy for race. There are 3 types of these predictors: socially acceptable predictors (SAPs), socially unacceptable predictors (SUPs), and contentious predictors (CPs). The problem is that CPs may have information that we need but could proxy for SUPs.

The following model is given:

P&S then posit the following data-generating process (DGP), meeting the usual assumptions of ordinary least squares (OLS):

$$y_i = \beta_0 + \beta_1 X_i^{SAP} + \beta_2 X_i^{CP} + \beta_3 X_i^{SUP} + \varepsilon_i,$$

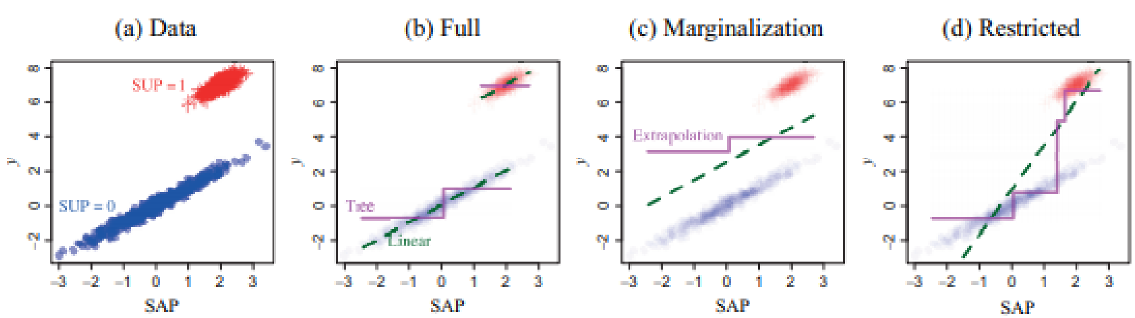
where y_i is the observed outcome, X_i^{SAP} is the vector of SAPs, X_i^{CP} is the vector of CPs, X_i^{SUP} is the vector of SUPs for unit i, and $\varepsilon_i \sim N(0, \sigma)$. SAPs are assumed independent of SUPs and CPs, but SUPs are correlated with CPs:

$$X_i^{SUP} = \delta_0 + \delta_{CP} X_i^{CP} + v_i$$

The article then analyzes different approaches to dealing with contentious predictors. Their "proposed" or "marginalization" model uses the full model given above, but also uses average SUP for predictions, which marginalizes out SUPs.

One approach is <u>"Extrapolation with OLS and Decision Trees".</u>

Illustration of How Decision Tree Can Magnify Extrapolation



The "restricted model" is one that controls only for SAPs. Their conclusion was that with non-identical support (SAP is not independent of SUP) the restricted model can outperform the "proposed" model. This type of analysis helps because it will be hard to have/find data where socially acceptable predictors are independent of socially unacceptable predictors.

This article didn't give exact ways to implement a decision tree or the exact "restricted" model but these are things that I can continue to look into.

Another approach: <u>Extrapolation with Random Forests</u>

$$X_i^{\mathrm{CP}} \sim N(0,1);$$
$$X_i^{*\mathrm{SUP}} = \frac{1}{1+\exp[-(\delta_0+\delta_{\mathrm{CP}}X_i^{\mathrm{CP}}+v_i)]};$$
$$X_i^{\mathrm{SUP}} = \begin{cases} 1 & \text{if } X_i^{*\mathrm{SUP}} > 0.5, \\ 0 & \text{otherwise;} \end{cases}$$
$$X_i^{\mathrm{SAP}} \sim N(\boxed{\eta} \times X_i^{\mathrm{SUP}}, \boxed{\sigma} \times X_i^{\mathrm{SUP}} + (1-X_i^{\mathrm{SUP}}));$$

The first line signifies using the normal distribution for contentious predictors. The last line signifies the shifted mean and variance parameters (in boxes) that are used by the restricted model.

$$y_i = \beta_0+\beta_1X_i^{\mathrm{SAP}}+\beta_2X_i^{\mathrm{CP}}+\beta_3X_i^{\mathrm{SUP}} + \boxed{\beta_4}[X_i^{\mathrm{SUP}}\times((X_i^{\mathrm{SAP}}+2)^2)]+\varepsilon_i$$

pg. 16

This is the model used.

The steps they used for the predictive analysis are as follows.

- Draw 100 simulated data sets (N=10,000)
- Fit the model on a random sample of 80% of the dataset, then test the model on the remaining 20%
    - This is pretty standard statistical practice, I believe.

<u>Conclusions reached</u>

Overall, the "restricted" model used by the authors of the article, outperformed the "proposed" model by Pope and Sydnor. The marginalized model becomes inaccurate when the effect of SUPs (employment, criminal justice, etc..) is strong. Since this is a possible problem we can run into, the given "restricted model" and techniques used can be helpful to us when deciding how to differentiate the significance between SAPs, SUPs, and CPs.