Appendix A

The Oaxaca-Blinder Decomposition.

The Oaxaca-Blinder decomposition (OBD), is a standard econometric technique used in wage gap decomposition analysis plus a range of health outcomes[1 2]. It can be used to reveal the driving factors behind pay gaps[3]. If we took two groups as a starting point, an OBD procedure estimates the earnings structures of group one and group two using separate ordinary least squares regression estimations. This produces estimates of the meaningfulness of differences in personal and job characteristics for group one and group two in terms of earnings, alongside the average values of those characteristics for each group (referred to as "endowments"). Second, the decomposition produces an estimate of rewards, referred to as the 'coefficients' effect. Here the estimate measures differences in financial returns for each group holding equal measures of the same characteristic. For example, in this context it estimates different returns for men and women of differing ethnic groups relating to, for example, being a surgeon, or being 40. This could be considered discrimination. The coefficient effect also includes the constant in the model, which captures unobserved attributes associated with the pay gap.

The original Oaxaca-Blinder decomposition takes the following form[4]:

$$ln(W_{WMi}) = X_{WMi}\beta_{WM} + \varepsilon_{WMi}$$

$$ln(W_{CGi}) = X_{CGi}\beta_{CG} + \varepsilon_{CGi}$$

$$R = \overline{ln(W_{WM})} - \overline{ln(W_{CG})} = (\bar{X}_{WM} - \bar{X}_{CG})\hat{\beta}_M + \bar{X}_F(\hat{\beta}_M - \hat{\beta}_{CG}) \quad (1)$$

Where WM = white male, CG = comparison group and R is the ethnic pay gap in percentage points. The first term on the right-hand side reflects the difference in endowments between white men and the comparison group and the second term reflects the difference in the slope in the regression term of white male and comparison group wages i.e. differences in the structure of rewards to these endowments.

In this paper, we follow an alternative specification of the OBD decomposition commonly used in labour economics. It is assumed that there is a non-discriminatory coefficient vector derived from the data which is used to determine the contribution of the differences in the predictors outcomes[5 6]. Let β^* be such a nondiscriminatory coefficient vector. The outcome difference from (1) can then be rewritten as

$$R = \{E(X_{WM}) - E(X_{CG})\}'\beta^* + \{E(X_{WM})'(\beta_{WM} - \beta^*) + E(X_{CG})'(\beta^* - \beta_{CG})\} \quad (2)$$

[1] Sen B. Using the Oaxaca–Blinder decomposition as an empirical tool to analyze racial disparities in obesity. Obesity. 2014 Jul;22(7):1750-5.
[2] O'Donnell O, Van Doorslaer E, Wagstaff A. Decomposition of inequalities in health and health care. InThe Elgar companion to health economics, Second Edition 2012 Jan 31. Edward Elgar Publishing.
[3] Oaxaca RL, Ransom MR. On discrimination and the decomposition of wage differentials. Journal of econometrics. 1994 Mar 1;61(1):5-21.
[4] Oaxaca R. Male-female wage differentials in urban labor markets. International economic review. 1973 Oct 1:693-709.

[5] Oaxaca RL, Ransom MR. On discrimination and the decomposition of wage differentials. Journal of econometrics. 1994 Mar 1;61(1):5-21.
[6]Jann B. The Blinder–Oaxaca decomposition for linear regression models. The Stata Journal. 2008 Dec;8(4):453-79.

This results in a "twofold" decomposition,

$$R = Q + U$$

where the first component,

$$Q = \{E(X_{WM}) - E(X_{CG})\}'\beta^*$$

is the part of the wage gap that is explained by group differences in the explanatory variables (the "endowment effect"), and the second component;

$$U = E(X_{WM})'(\beta_{WM} - \beta^*) + E(X_{CG})'(\beta^* - \beta_{CG})$$

is the unexplained part, often known as the "coefficient effect". The coefficient effect is sometimes attributed to discrimination or disadvantage stemming from the wage structure, but it is important to highlight that this can also potentially capture unmeasured and unobserved variables. We derive β^* by using a pooled regression for both white males and the comparison group and derive the robust variance covariance matrix of the estimates using the delta method as outlined by Oaxaca and Ransom[7].

7 Oaxaca RL, Ransom M. Calculation of approximate variances for wage decomposition differentials. Journal of Economic and Social Measurement. 1998 Jan 1;24(1):55-61.