Simulation Appendix

Some mathematical details for setting values for parameters a and c in the simulations:
1. Consider the simplified case:

$$y(s) = f(x_1, \cdots, x_4) + \delta(s) = ag(x_1, x_2) + h(x_3, x_4) + \delta(s)$$

where $var(g) = \sigma_g^2$ and $var(h) = \sigma_h^2$ are computed empirically. Find a for a given fixed value of $var(ag)/var(f) = \lambda$ (i.e., the proportion of variance in f explained by the nonlinear component is fixed).

$$\begin{aligned}\lambda = \frac{a^2\sigma_g^2}{a^2\sigma_g^2 + \sigma_h^2} &\implies \lambda a^2\sigma_g^2 + \lambda\sigma_h^2 = a^2\sigma_g^2 \\ &\implies \lambda\sigma_h^2 = a^2\sigma_g^2 - \lambda a^2\sigma_g^2 = a^2(1-\lambda)\sigma_g^2 \\ &\implies a = \sqrt{\frac{\lambda\sigma_h^2}{(1-\lambda)\sigma_g^2}}\end{aligned}$$

2. Now consider the more general case:

$$\begin{aligned}y(s) &= cf(x_1, \cdots, x_4) + \delta(s) \\ &= c[ag(x_1, x_2) + h(x_3, x_4)] + \delta(s)\end{aligned}$$

where $var(f) = \sigma_f^2$ and $var(\delta) = \sigma_d^2$ are computed empirically. Find c for a given fixed value of $R^2 = \rho$ (i.e, the proportion of variance in y explained by the covariates is fixed).

$$\begin{aligned}R^2 = \rho = \frac{var(cf)}{var(y)} &= \frac{c^2\sigma_f^2}{c^2\sigma_f^2 + \sigma_d^2} \\ &\implies \rho(c^2\sigma_f^2 + \sigma_d^2) = c^2\sigma_f^2 \\ &\implies \rho\sigma_d^2 = c^2\sigma_f^2 - \rho c^2\sigma_f^2 = c^2(1-\rho)\sigma_f^2 \\ &\implies c = \sqrt{\frac{\rho\sigma_d^2}{(1-\rho)\sigma_f^2}}\end{aligned}$$

Since $\sigma_f^2 = a^2\sigma_g^2 + \sigma_h^2$ we can also express c as

$$c = \sqrt{\frac{\rho\sigma_d^2}{(1-\rho)(a^2\sigma_g^2 + \sigma_h^2)}}$$

Note the R^2 and λ values vary in R since the above derivation assume independence between model components (which is true in the expectation for the population).