

HW: Adjusted Plus-Minus and RAPM

Obtain NBA play-by-play data from, say, nbastatr.

Adjusted Plus-Minus are the NBA player power scores obtained from the following regression,

i = index of i^{th} play in dataset

y_i = pts scored by offensive team on play i
 $\in \{0, 1, 2, 3, 4\}$

P NBA players $1, \dots, P$

Each player in NBA has power score β_1, \dots, β_P

Model:

$$y_i = \beta_{\text{off. team player 1}} + \beta_{\text{off. team player 2}} + \dots + \beta_{\text{off. team player 5}} \\ - \beta_{\text{def. team player 1}} - \beta_{\text{def. team player 2}} - \dots - \beta_{\text{def. team player 5}} \\ + \beta_0 + \epsilon_i$$

$E\epsilon_i = 0$ (mean zero noise)

$$\text{So, } X_{ij} = \begin{cases} 1 & \text{if intercept } j=0 \\ 1 & \text{if player } j \text{ on offense in play } i \\ -1 & \text{if player } j \text{ on defense in play } i \\ 0 & \text{else} \end{cases}$$

Adjusted Plus Minus (APM): use OLS to estimate the player strength scores β

Regularized Adjusted Plus-Minus (RAPM): use Ridge Regression to estimate β

Which have better predictive performance? Why?

How bad is multicollinearity in basketball?