Diana Taurasi's Scoring: Can She Just Shoot from Anywhere?

Melanie Ackerman
Metis Linear Regression Project
12/15/21

Who is Diana Taurasi?

- WNBA superstar on the Phoenix Mercury
- 3x WNBA champion
- 2x WNBA Finals MVP
- 2009 Regular Season MVP
- 1st all-time in career regular season points (9,174)
- 1st all-time in career playoff points (1,397)
- 5th all-time in career regular season assists (2,032)
- 1st all-time in career regular season 3-pointers (1,205)
- And more…

Does it Matter From Where on the Court She Shoots?

DT *can* shoot from anywhere on the court…

Literally.

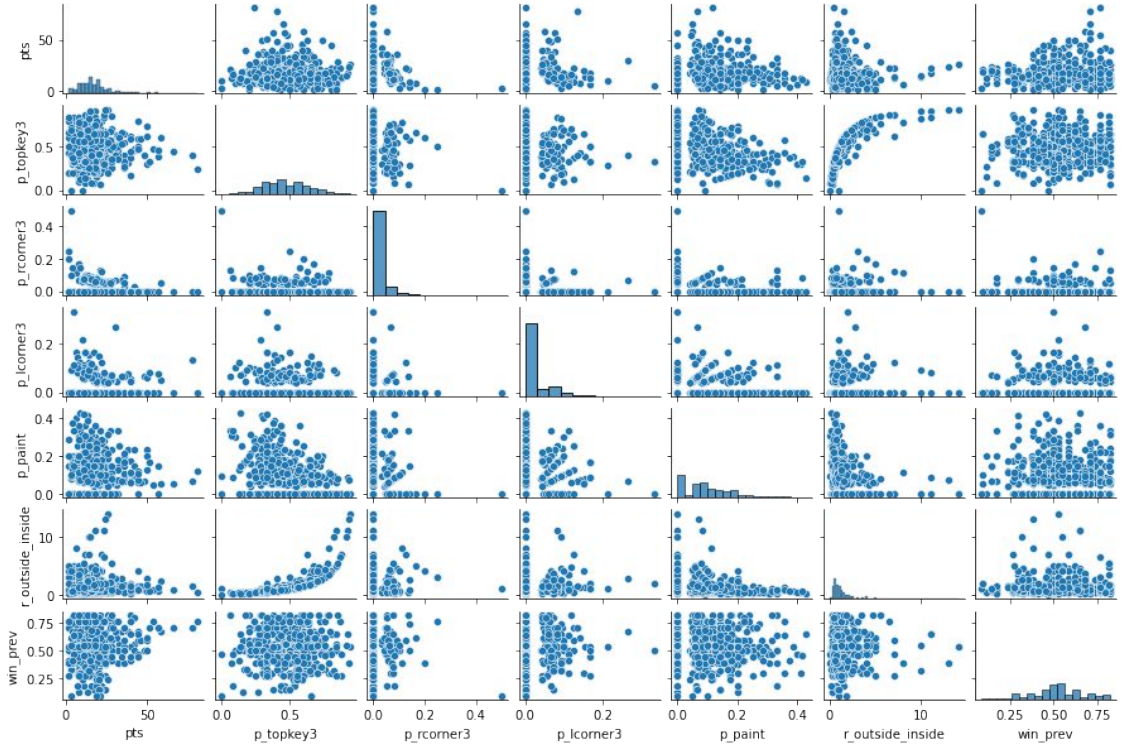
…But can we optimize where she shoots from most often to maximize points scored?

The Data

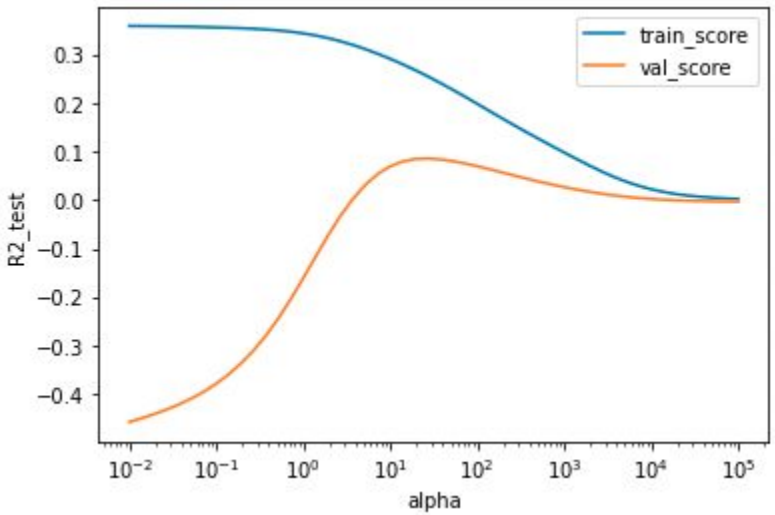
#	Column	Non-Null Count	Dtype
0	GRID_TYPE	8884 non-null	object
1	GAME_ID	8884 non-null	object
2	GAME_EVENT_ID	8884 non-null	int64
3	PLAYER_ID	8884 non-null	int64
4	PLAYER_NAME	8884 non-null	object
5	TEAM_ID	8884 non-null	int64
6	TEAM_NAME	8884 non-null	object
7	PERIOD	8884 non-null	int64
8	MINUTES_REMAINING	8884 non-null	int64
9	SECONDS_REMAINING	8884 non-null	int64
10	EVENT_TYPE	8884 non-null	object
11	ACTION_TYPE	8884 non-null	object
12	SHOT_TYPE	8884 non-null	object
13	SHOT_ZONE_BASIC	8884 non-null	object
14	SHOT_ZONE_AREA	8884 non-null	object
15	SHOT_ZONE_RANGE	8884 non-null	object
16	SHOT_DISTANCE	8884 non-null	int64
17	LOC_X	8884 non-null	int64
18	LOC_Y	8884 non-null	int64
19	SHOT_ATTEMPTED_FLAG	8884 non-null	int64
20	SHOT_MADE_FLAG	8884 non-null	int64
21	GAME_DATE	8884 non-null	object
22	HTM	8884 non-null	object
23	VTM	8884 non-null	object

- Shot-by-shot data scraped from wnba.com during DT's career (17 seasons)
- Use details on shot zones to calculate percentage of field goal attempts taken from various zones on the court, ratio of outside to inside attempts
- Dummy variables for notable seasons for DT or Phoenix Mercury (i.e. MVP season and playoff champion seasons)
- Connect with opponent's winning percentage from previous season

What do the variables look like?

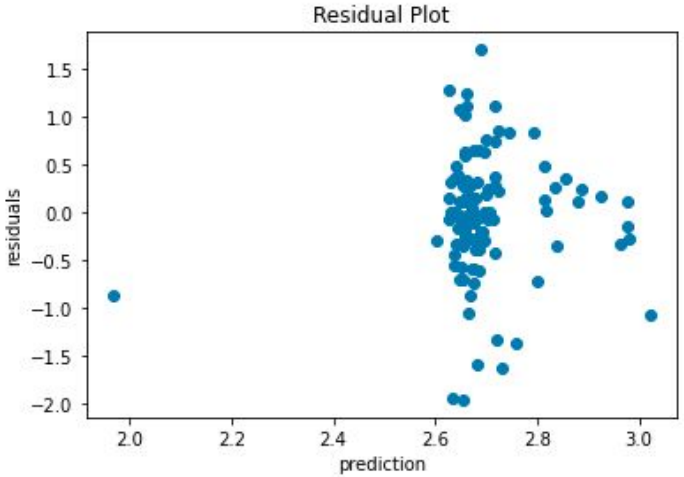


The Model



- Baseline OLS:
  - R-squared = 0.114; Adj. R-squared = 0.076
- Log(pts), only notable season dummies:
  - R-squared = 0.065; Adj. R-squared = 0.047
- Polynomial transformation of features, ridge regression:
  - Train score: 0.058
  - Test score: 0.043

Model Predictiveness



- The residual plot is not as randomly scattered as would be ideal. Points should be centered around y=0.
- Mean absolute error = 0.468
- Notable coefficients transformed:
  - win_prev = 1.0054
  - season06 = 1.0061
  - p_topkey3 x season06 = 1.0058
  - win_prev x season06 = 1.0063

Conclusion: Is DT too good to predict?

- Possibilities for why this regression does not have much predictive power include working with a small dataset and/or neglect of other potential features that would have more impact but were not available in the data.
- Or… maybe Diana Taurasi is really just so good at shooting that it doesn't matter much where she shoots from in determining her scoring in a given game.