Data 8 Connector: Sports Analytics

...

Review - 02/06/18

Course stuff

- OH stuff: after class 2-4 in Evans 458. Don't be afraid!
  Let me know if this time doesn't work
- Homework: coming tonight or tomorrow. 2 weeks
  Goal is enough time to have 2 OH sessions
- Project groups: Issues? Without a group? Talk project ideas or datasets?
- Other stuff
  - Materials on Piazza
  - HW1 solutions on Piazza
  - Pythagorean Expectation notebook online

Recap

Pythagorean Expectation

Last week: Pythagorean Expectation

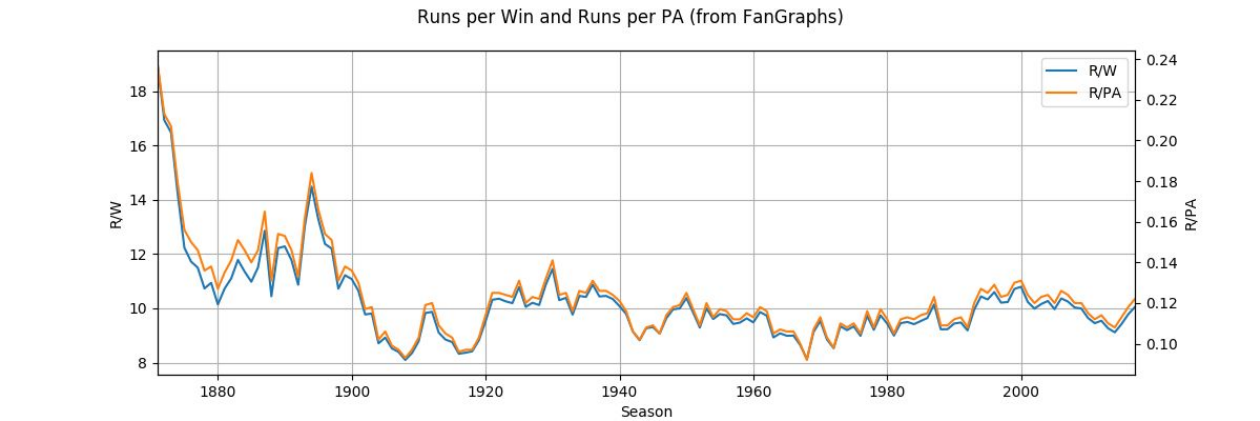
$$\text{Pythagorean Win Pct} = \frac{(\text{Runs Scored} \ / \ \text{Runs Allowed})^2}{1 + (\text{Runs Scored} \ / \ \text{Runs Allowed})^2}$$

- ➔ An empirical relationship between runs and wins
- ➔ Works for basketball too!
- ➔ The exponent defines a feature of the sport (the sensitivity to scoring ratio)

Runs per Win

Runs and Wins

- ➔ Higher run scoring means more runs needed to win
- ➔ Runs per Win over time from FanGraphs:



Classic Stats

Classical Baseball Stats

Batting Average

$$BA = \frac{H}{AB}$$

On-Base Percentage

$$OBP = \frac{H + BB + HBP}{AB + BB + HBP + SF} \approx \frac{H + BB + HBP}{PA}$$

Slugging Percentage

$$SLG = \frac{(1 \times 1B) + (2 \times 2B) + (3 \times 3B) + (4 \times HR)}{AB}$$

Classic Stats

Main idea: measure intuitive ideas offensive performance

Higher BA, OBP, SLG is good: you're getting on and hitting for power

Core features

- ➔ Rate stats: count up a total (like hits) and divide by opportunities (like ABs)
- ➔ You can treat events equally (BA, OBP) or weight by bases (SLG)
- ➔ You can include walks (OBP)
- ➔ But definitely don't consider context (a single is a single, a HR is a HR)

We ignore R and RBI from here on (too context dependent, and not rate stats)

Classic Stats

So how do they actually relate to run scoring?

- ➔ To explore that, we'll return to Lahman dataset
- ➔ We'll look at total run scoring as it directly relates to the metrics, i.e. scatter plots

A Basketball Digression...

The classic stats did some simple things: “total good events” over “opportunities”

Similar ideas exist in basketball, the most noticeable is Field Goal%

$$\text{FG \%} = \frac{\text{Field Goals}}{\text{Field Goals Attempts}}$$

A Basketball Digression...

These are considered “advanced stats” but they’re very close to what we know

Effective Field Goal%

$$\text{Effective FG \%} = \frac{\text{Field Goals} + \frac{1}{2} \cdot \text{3-pt Field Goals}}{\text{Field Goals Attempts}}$$

True Shooting%

$$\text{True Shooting \%} = \frac{\text{Total Points}}{2(\text{Field Goals Attempts} + .44 \cdot \text{Free Throw Attempts})}$$

A Basketball Digression...

“Totals” over “opportunities”

FG% is makes per attempts from the floor

eFG% is points per field goal attempt: your efficiency from the floor (ignores FTs!)

True Shooting% incorporates FTs: Total points per possessions used

➔ Why .44? FTs typically come in pairs so .5 · FTs is approximately the number of possessions used that ended in 2 FT shots. .44 accounts for flagrants, technicals, And-1s, and 3pt fouls.

A Basketball Digression...

Field Goal Pct	
1. Clint Capela • *HOU*	.656
2. DeAndre Jordan • *LAC*	.653
3. Steven Adams • *OKC*	.634
4. Enes Kanter • *NYK*	.607
5. John Collins • *ATL*	.591
6. John Henson • *MIL*	.589
7. Taj Gibson • *MIN*	.574
8. Jonas Valanciunas • *TOR*	.569
9. Hassan Whiteside • *MIA*	.554
10. Andre Drummond • *DET*	.552
11. Julius Randle • *LAL*	.551
12. Derrick Favors • *UTA*	.550
13. Dwight Howard • *CHO*	.547
14. Giannis Antetokounmpo • *MIL*	.544
15. LeBron James • *CLE*	.542
Karl-Anthony Towns • *MIN*	.542
17. Anthony Davis • *NOP*	.542
18. Domantas Sabonis • *IND*	.538
19. Marcin Gortat • *WAS*	.537
20. Ben Simmons • *PHI*	.526

Effective Field Goal Pct	
1. Clint Capela • *HOU*	.656
2. DeAndre Jordan • *LAC*	.653
3. Steven Adams • *OKC*	.634
4. Joe Ingles • *UTA*	.615
5. Stephen Curry • *GSW*	.615
6. Enes Kanter • *NYK*	.607
7. E'Twaun Moore • *NOP*	.604
8. John Collins • *ATL*	.599
9. Karl-Anthony Towns • *MIN*	.595
10. Klay Thompson • *GSW*	.593
11. Joe Harris • *BRK*	.592
12. John Henson • *MIL*	.589
13. Jonas Valanciunas • *TOR*	.589
14. LeBron James • *CLE*	.588
15. Trey Lyles • *DEN*	.587
16. Kevin Durant • *GSW*	.585
17. Taj Gibson • *MIN*	.581
18. Al Horford • *BOS*	.578
19. Gary Harris • *DEN*	.578
20. Kelly Olynyk • *MIA*	.573

True Shooting Pct	
1. Stephen Curry • *GSW*	.673
2. Clint Capela • *HOU*	.661
3. DeAndre Jordan • *LAC*	.656
4. Mike Scott • *WAS*	.645
5. Enes Kanter • *NYK*	.641
6. Steven Adams • *OKC*	.640
7. Darius Miller • *NOP*	.639
Karl-Anthony Towns • *MIN*	.639
9. Kyle Korver • *CLE*	.637
10. Jonas Valanciunas • *TOR*	.635
11. Kevin Durant • *GSW*	.634
12. John Collins • *ATL*	.632
13. Marvin Williams • *CHO*	.624
14. Joe Ingles • *UTA*	.623
15. LeBron James • *CLE*	.621
16. Montrezl Harrell • *LAC*	.621
17. Anthony Davis • *NOP*	.620
18. James Harden • *HOU*	.619
19. Nikola Mirotic • *TOT*	.614
20. Giannis Antetokounmpo • *MIL*	.613
Kevin Love • *CLE*	.613

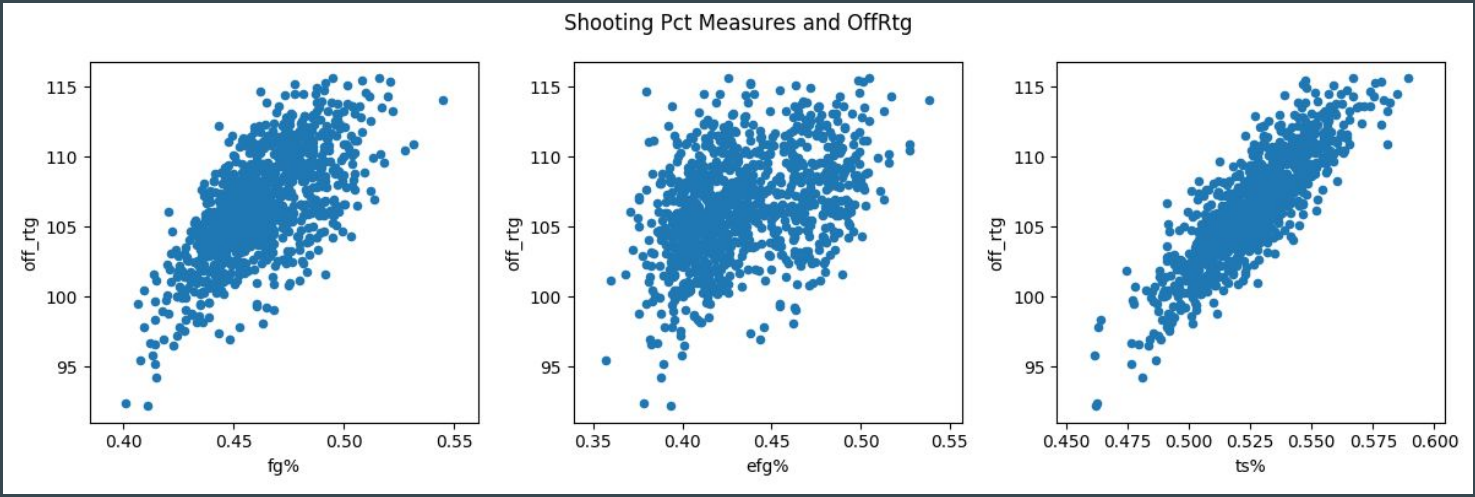
A Basketball Digression...

TS% can change our view of a player's performance a lot

- ➔ FG% loves big men: short shots
  DeAndre Jordan 2016-17: #2 all time in FG% and eFG%, #18 all time in TS%
- ➔ eFG% boosts up 3pt shooters
- ➔ James Harden finally in the top 20 for TS%
- ➔ Makes sense: an attacking player getting a lot of FTs
  - ◆ Inherently shoots tough shots so FG% is going to be low
  - ◆ But if the player earns fouls and scores from the line, that needs to count
    That's a good use of a possession! (this is like counting a walk)

A Basketball Digression...

And how do they relate to scoring? (FG%, eFG%, TS% in order)



Not too bad, of course TS% is obviously the best!