Advanced Basketball Analytics

By: Neil Liberman

Win More Games

OBJECTIVE:

Find metrics that correlate to more wins.

In my presentation I will:

- Prove that offensive efficiency correlates strongly with winning more games.
- Attempt to find features that correlate with a higher offensive efficiency.
- Provide insight as to where on the court quality shots exist and how to structure player development goals.

Defining Terms

- Win %: (Games won) * 100 / (games played)

    Ex: Warriors: 73/82 = .890 = 89.0%

- Field Goal %: (shots made / shots attempted) or (FGM / FGA)

    Max FG% = 100%

- Effective Field Goal %: shots made + (3PM * 0.5) / Shots Attempted

    Max eFG% = 150%

Defining Terms

Offensive Efficiency (OE):

[FGM + Assists] /

[FGA + Assists + TO - ORB]

Adjusted Offensive Efficiency (Adj_OE):

[FGM + (0.5 * 3PM) + (FTM * 0.44) + Assists] /

[FGA + (FTA * 0.44) + Assists + TO - ORB]

From here on out, if I say OE or efficiency, I am referring to my Adjusted Offensive Efficiency metric.

Defining Terms

Net Offensive Efficiency (Net_Adj_OE):

OE - Opponents OE

Efficient Offensive Production (EOP)

OE * (Points + (0.76 * Assists))

Passer gets credit for 38% of assists (approximation of meaningful assists)

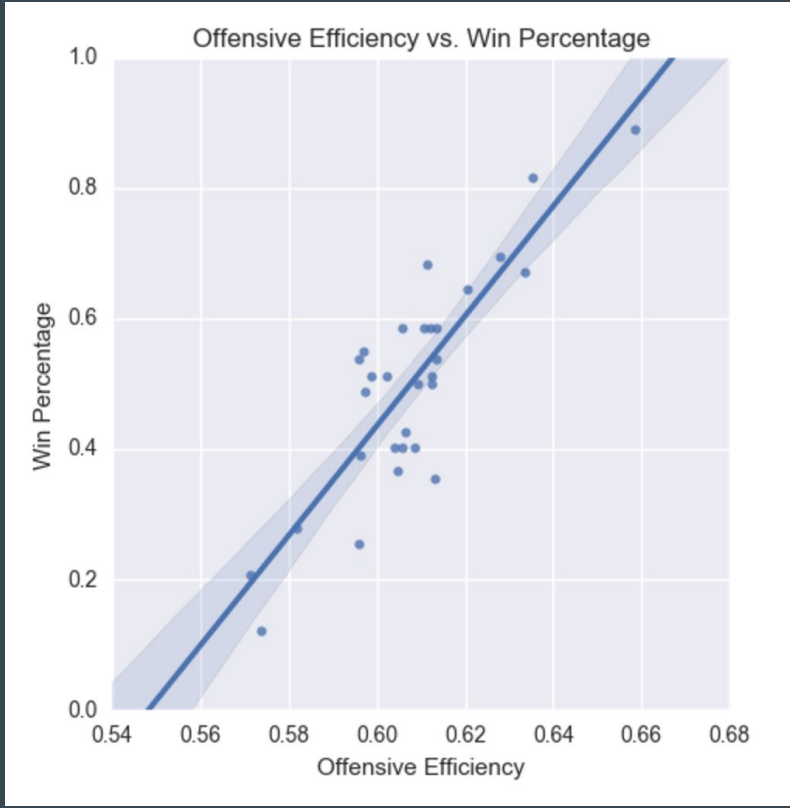
Essentially, EOP is OE times the amount of points the player accounted for.

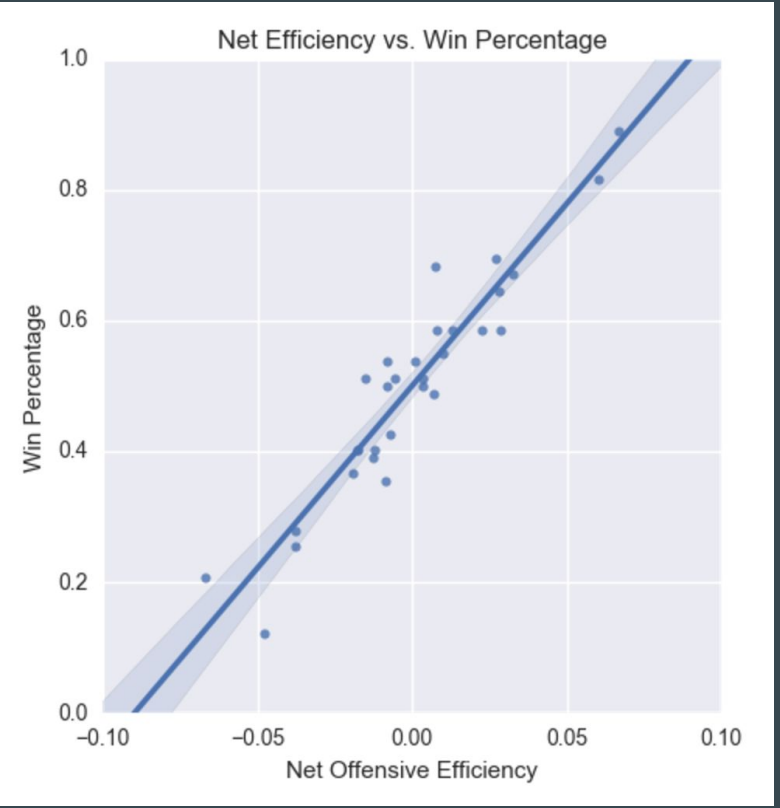
Helps explain how much a player contributes

How Does Efficiency Correlate to Win Percentage?

	Win%	Adj_OE	Opp_Adj_OE	Net_Adj_OE
Win%	1.000000	0.855437	-0.793472	0.939394
Adj_OE	0.855437	1.000000	-0.546196	0.895726
Opp_Adj_OE	-0.793472	-0.546196	1.000000	-0.861670
Net_Adj_OE	0.939394	0.895726	-0.861670	1.000000

Efficiency and Win Percentage





Top Offensive Efficiency Differential

	Win%	Adj_OE	Opp_Adj_OE	Net_Adj_OE
Team				
Golden State Warriors	0.890	0.650946	0.583045	0.067901
San Antonio Spurs	0.817	0.627167	0.567351	0.059816
Oklahoma City Thunder	0.671	0.623912	0.592120	0.031791
Atlanta Hawks	0.585	0.605900	0.576616	0.029284
Los Angeles Clippers	0.646	0.611334	0.582725	0.028608
Cleveland Cavaliers	0.695	0.619649	0.592098	0.027551

Why Analyzing Basketball is Tough

- Many moving parts. The popularity of sports analytics took off with “Moneyball” but analytics suit baseball well. Events are mostly independent of teammates. Basketball poses different problems.
  - Quality of teammates
  - Scheme affects statistics
- Statistics are contextual.
  - I’ve displayed the importance of offensive efficiency, yet it’s easy to point out flaws such as offensive rebounds disproportionate effect.
  - James Jones vs. JJ Redick

Maintaining Offensive Efficiency

Some players may have a high offensive efficiency but how do we know if it's the result of playing with talented players who get them high quality shots?

GOAL:

Find out which statistical categories correlate most to maintaining efficiency when the top EOP teammate is off the court.

To better evaluate players

To tailor player development towards this goal.

OE vs EOP

Player	Adj_OE
Andre Roberson	0.713681
Andre Iguodala	0.696899
Shaun Livingston	0.693182
Stephen Curry	0.686272
Jose Calderon	0.680971
Chris Paul	0.679910

Player	Adj_EOP
Stephen Curry	25.405799
Russell Westbrook	21.365389
James Harden	20.326789
Chris Paul	20.302117
Kevin Durant	20.293603
LeBron James	19.839693

Areas of Exploration

- Ability to make contested shots
- Ability to finish around the hoop
- Ability to make shots off the dribble
- Shot location
- Player movement

Variables Explored

- FG% on Drives
- Points Per Drive
- Assists Per Drive
- Turnovers Per Drive
- FG% inside 5 feet
- Contested eFG%
- Pullup eFG%
- eFG% off the dribble
- Percentage of FG from 2
- Percentage of FG from 3
- Percentage of Points from 2
- Percentage of Points from Mid Range
- Percentage of Points from 3
- Percentage of Points from Fast Break
- Percentage of Points from Free Throw Line
- Percentage of Points off Turnovers
- Percentage of FG were Assisted
- Distance run on offense
- Average speed run on offense

*** distance metrics were converted to per 36 to minutes.

Findings

All metrics analyzed had very little predictive value of its effect on changes in offensive efficiency.

Running principal component analysis, a tactic used to combine variables, was also ineffective at providing any insight.

Aggregating these metrics into a singular model was very ineffective at predicting offensive efficiency.

Why might this be the case:

- Different quality of teammates
- Different roles when playing alongside other players

- There may be a correlation holding all things equal, however it is impossible to do so.

Revisions

Finding factors that correlate to a change in offensive efficiency when the best player is off the court is impractical. The question must be broader.

Revised Problem Statement:

> Are these same features useful in predicting offensive efficiency regardless of who is on the court?

Revised Findings

Some metrics did have mild correlation to offensive efficiency regardless of who was on the court. Features display a large variance but they have general trends.

However, correlations were still very low and many features had no correlation at all.

Maybe because of flaws such as ORB?

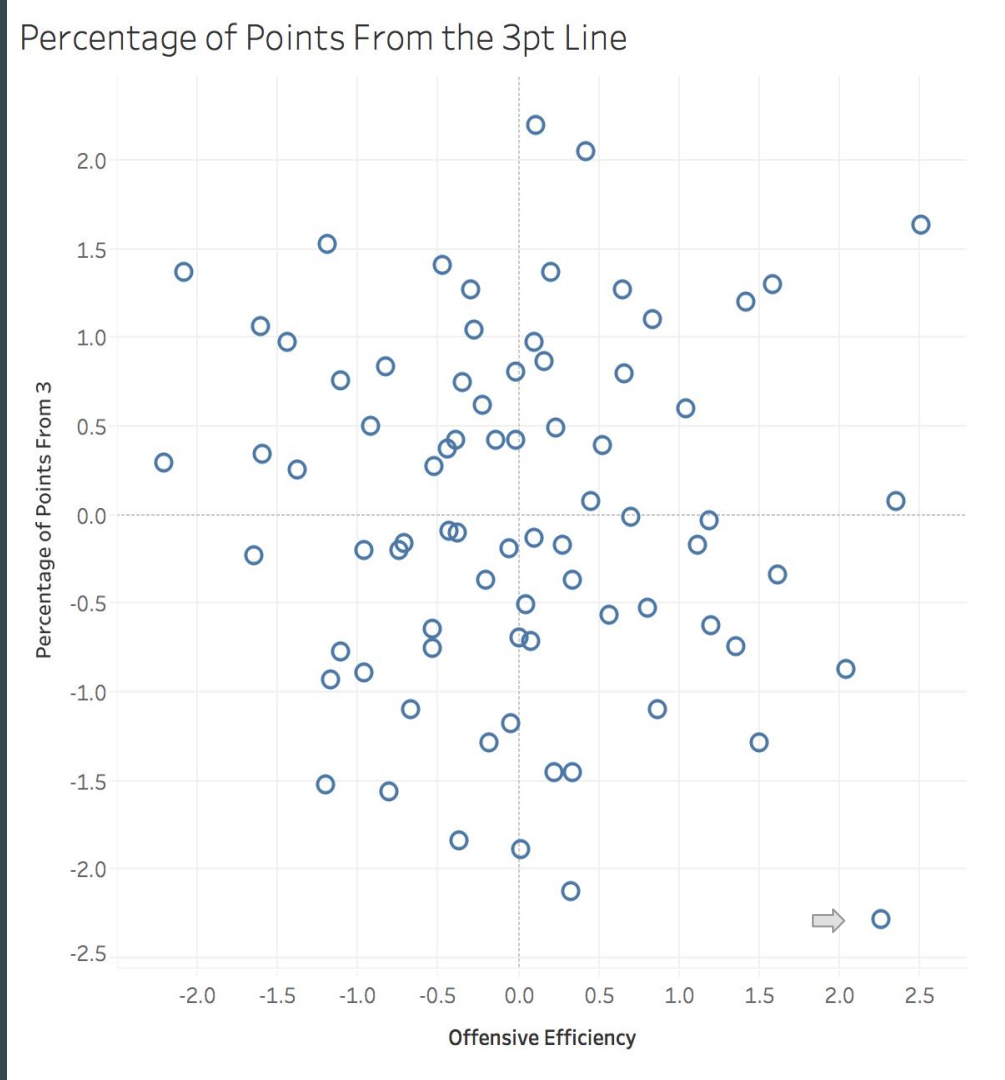
It seems these metrics are too granular to provide predictive insight, but there ***were still takeaways***.

Percentage of Points from 3

This shows zero correlation between percentage of points coming from the three point line and offensive efficiency.

Counter to the massive trend in the NBA.

Defensive adjustment to the NBA trend will possibly reduce the efficiency of 3 point attempts and create the potential to exploit mid range opportunities.



Further Problems with Reliance on Three Point Attempts

- Making 33% from the three point line is the same (in points scored) as going 50% from two point attempts.
- There was a mild correlation between the percentage of points scored off turnovers and a higher offensive efficiency. This may seem obvious but keep in mind a higher percentage of points scored off turnovers could indicate an inability to score in the half court, which would only skew its correlation down. Helps confirm that getting in transition is useful.
- By shooting 33% as opposed to 50%, you have increased transition opportunities and may imply an increase to your opponents OE.

Two point attempts

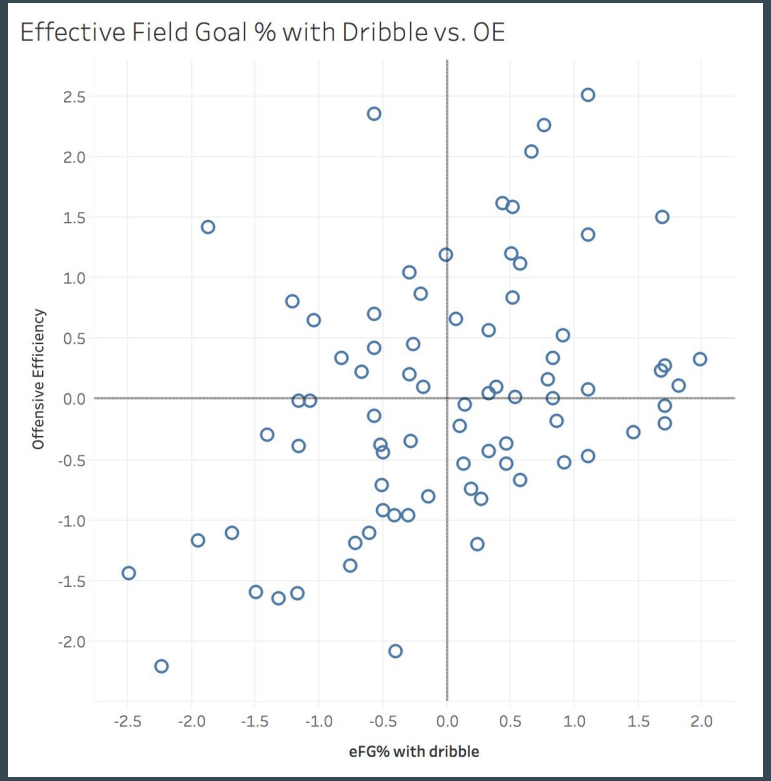
- Points in the paint correlates to higher efficiency.
  - Further demonstrates the overly bullish views on the three point shot.
- Mid Range shots correlate to lower efficiency. However, the trend doesn't seem as severe as movements in the NBA have suggested.

Sub-Categories of Field Goal Percentage

There is a trend correlating eFG% off the dribble to efficiency. There is also a similar trend in FG% driving to the hoop. Seems intuitive.

However, eFG% on contested shots and eFG% on pull-up jumpshots were not correlative at all.

Tells us something about player development goals and player evaluation.



Can These Features Predict Offensive Efficiency?

- Ran regression models using random forests, support vector machines, and linear regression.
  - Utilized Support Vector Regression due to its capacity to effectively handle high dimensions.
  - Ran principal component analysis as a means to reduce dimension for random forest models which do not handle high dimensionality well.

- Unfortunately, the features analyzed do not produce an effective model. While the features display loose trends that can guide our intuition, they are not effective in estimating efficiency.

Next Steps

- Increase the sample size. Rerun all analyses on a larger dataset.
- Explore less granular statistics.
- Take into account defense as net efficiency was most strongly correlated to win percentage.

Questions?