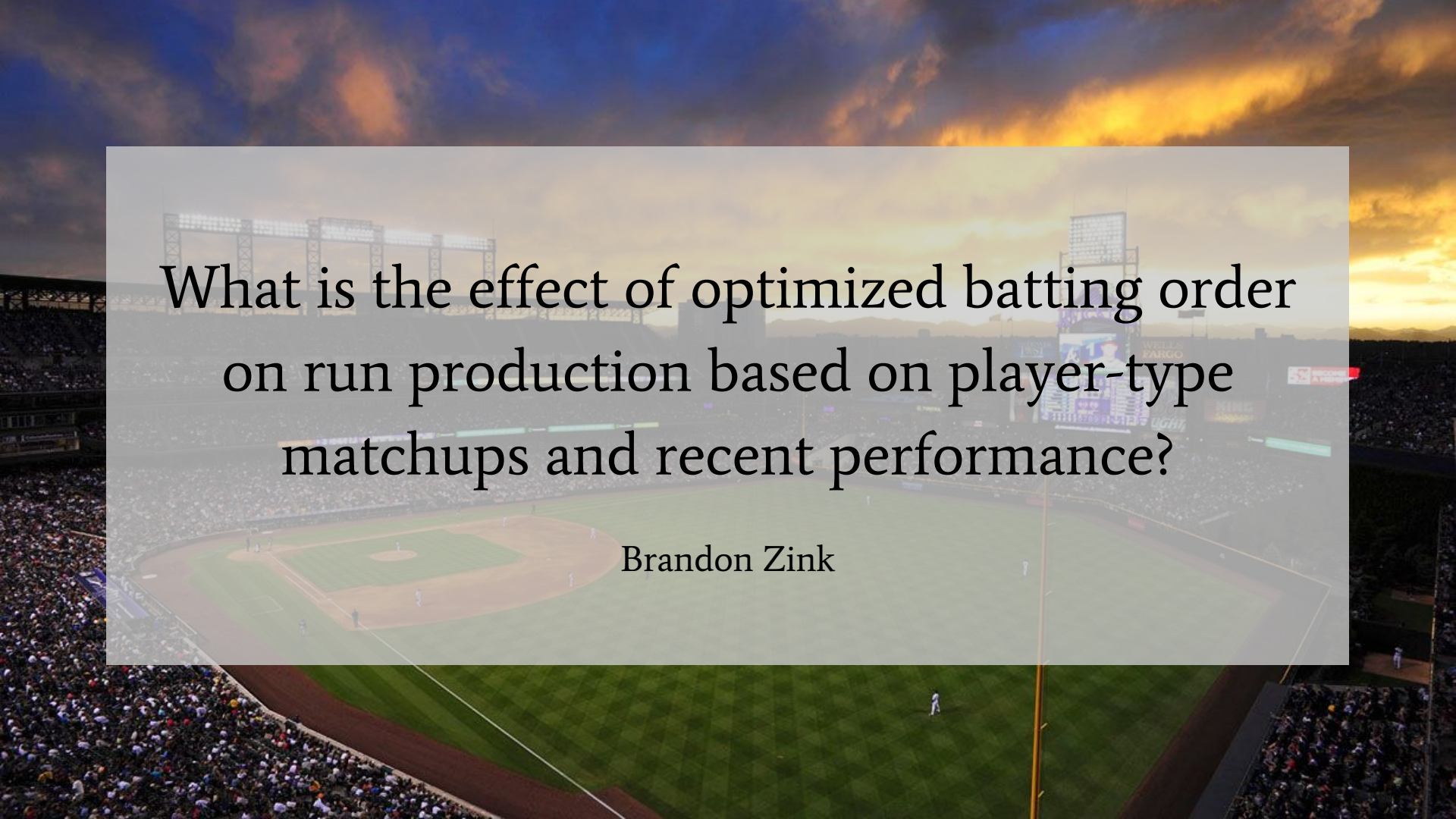


What is the effect of optimized batting order on run production based on player-type matchups and recent performance?

Brandon Zink

Lineup Averages in 2018 (AL)

Order	BA	ISO	BB%	K%	wRC+	Player Type
1	.259	.159	8.9%	19.6%	105	Low strikeout, low power hitter
2	.262	.167	9.0%	20.2%	109	Contact, low power hitter
3	.266	.195	9.7%	20%	118	Best all around hitter
4	.258	.207	8.4%	22.2%	113	High power slugger
5	.244	.183	8.2%	22.4%	100	Generic, league average hitter
6	.248	.157	8.4%	21.9%	97	Fourth worst hitter
7	.235	.151	7.7%	23.6%	87	Third worst hitter
8	.237	.144	7.2%	23.3%	85	Second worst hitter
9	.227	.126	6.4%	24.9%	74	Worst hitter

The Book on Batting Order

Order	Player Type
1	OBP king, speed doesn't matter
2	Best OBP all-around hitter
3	High power, high K%
4	Best power all-around hitter
5	Fourth best hitter
6	Fourth worst hitter
7	Third worst hitter
8	Second worst hitter
9	Worst hitter

The Book on the Pitchers Spot

2018 Pitcher Batting Statistics

Pos	BA	ISO	BB%	K%	wRC+
P	.115	.033	2.9%	42.2%	-25

"The second leadoff hitter theory exists. You can put your pitcher in the eighth slot and gain a couple of extra runs per year." (pg. 151)

What I’m Doing Different

Simulation based approach (no tables of averages as used in *The Book*).

Creating a tool that would allow a team to put in their players and the opposing pitcher and spit out an optimal lineup.

Taking into account recency (hot/cold streaks), player sample size, and player-type matchups in my decision making.

Recency Based Player Modeling

Recency Based Player Modeling

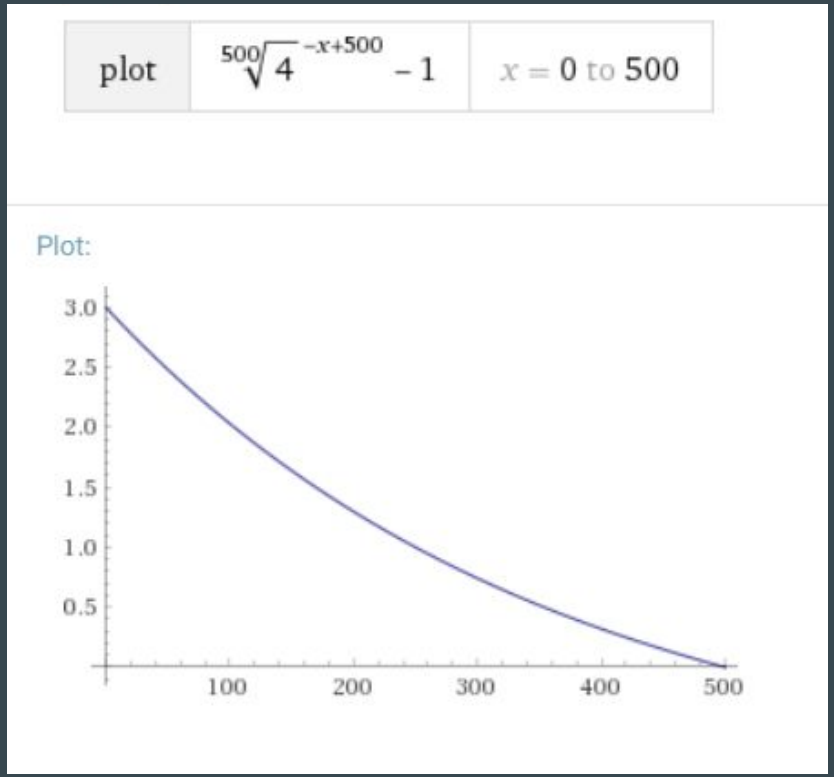
Calculate a players adjusted statistic based on sample size and recency.

http://www.sloansportsconference.com/wp-content/uploads/2016/02/1422-Baseball.pdf

Weight more recent performance heavier than older performance (back ~3 years).

Assume MLB average, deviate as sample size increases.

Statistic Modeling Weights



Weight more recent games heavier than games farther back

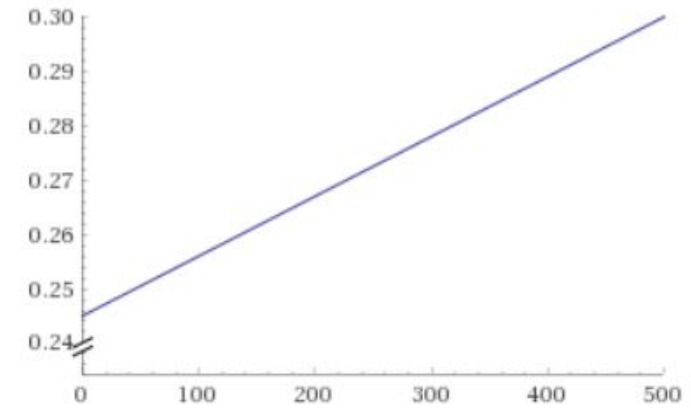
Take into account hot/cold streaks, while not ignoring underlining player skill

Statistic Modeling Weights

plot	$\frac{x}{500} \times 0.3 + \left(1 - \frac{x}{500}\right) \times 0.245$	$x = 0$ to 500

Player with .300 BA compared to .245 MLB average, x-axis is games played

Plot:



Assume a player starts at the MLB average

As their sample size increases, deviate away from the MLB average and towards their calculated average from the previous slide

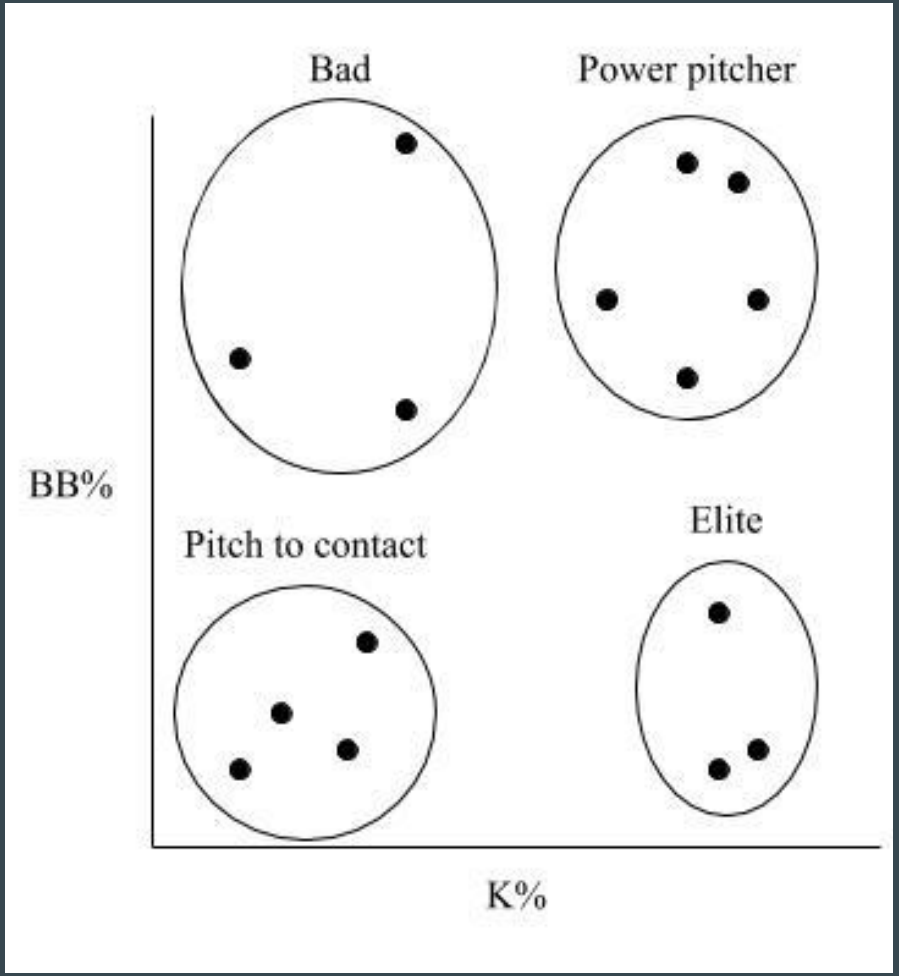
Player-Type Matchups

Player-Type Matchups

Use grouping method (method TBD) to categorize batters and pitchers (power strikeout pitchers, contact speedster batters, etc.).

Simulate player-type matchups to determine bonuses/penalties when facing other player-types (ex. high power, high strikeout batter receives .050 bonus to BA and SLG when facing low walk, low strikeout pitcher).

Take these into account when calculating optimal lineup.

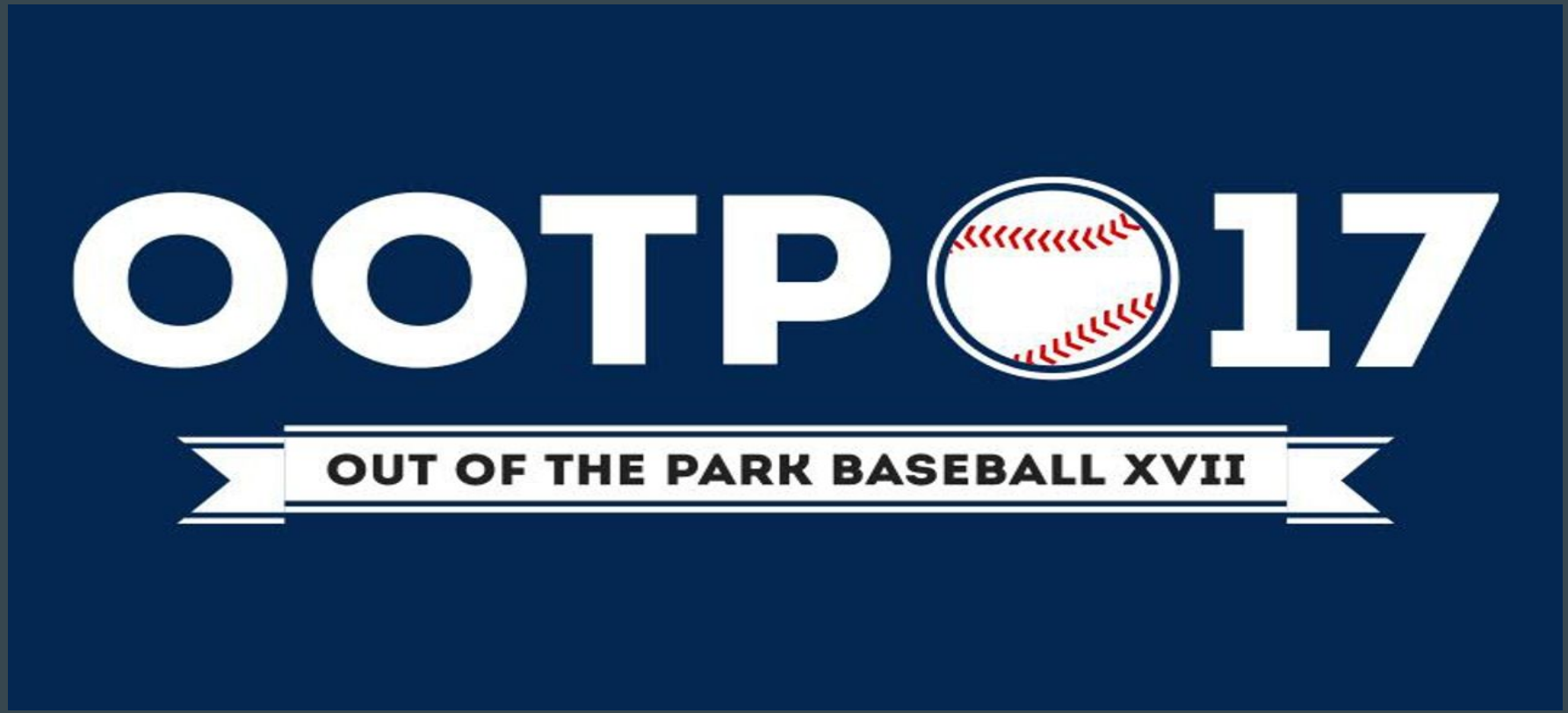


How does an “Elite” pitcher matchup against a “Contact” batter.

Theoretical examples for now.

Compare these to the expected odds as calculated via the Odds-Ratio method.

Simulation



Batting Order

Batting Order

Which spots in the batting order are most important to run production?

Which “player types” fit the best at different spots in the batting order?

How many additional runs does an optimized batting order score as compared to a “conventional” batting order?

Batting Order: Details

Simulate the NL and AL average lineup to give a baseline runs per game against league average pitching.

Simulate the same lineup with a “Mike Trout” at each lineup position to determine which position in the lineup are the most important in terms of run production (used later).

Batting Order: Details Contd.

Runs simulations with different “types” of hitters at each lineup position to determine the following (and more):

What position is a strikeout most/least costly?

What position is a HR (SLG/ISO) most/least beneficial?

What position is a walk/OBP most/least beneficial?

Is there merit to using a “second leadoff batter” in the NL?

Batting Order: Details Contd.

Use the information from the previous slide to “score” each batter in a real lineup, as well as information from “Mike Trout” simulation.

Batter A might receive a score of 87 for batting fourth, but 37 for batting first as determined by his batter profile and pitcher matchup.

Optimize the overall “score” of the lineup.

Weaknesses

Assuming correctness of OOTP 17 simulations

Defensive value

Pinch hitters

Scoring optimization as part of final product (try multiple methods to determine optimal scoring algorithm)

Proposed Timeline

- Background research, player statistic model by Thanksgiving Break
- Player-type grouping completed by Winter Break
- 1.5 months running simulations and recording results (including translating player statistics into OOTP 17 grades, do some of this over Winter Break as well)
- 1 month consolidating data into lineup optimizer code
- 1 month for writeup, testing, analysis, etc.
- 1-2 weeks leeway built in