

# Analyzing coaching decisions in college baseball using a reproducible in-game win probability model.

Dave Miller

University of Hartford

*davmiller@hartford.edu*  
*@davmiller9*

October 20, 2018



# Table of Contents

- 1 Background and Motivation
- 2 Expected Runs
- 3 Win probability model
- 4 Application of Win Probability: WPA
- 5 Player Ability and WPA

# Background and Motivation

- **Fall 2017:** Independent study in Sports Analytics working with Men's Basketball.



UNIVERSITY  
OF HARTFORD

# Background and Motivation

- **Fall 2017:** Independent study in Sports Analytics working with Men's Basketball.
- **Spring 2018:** Continued ind. study in baseball. Added to Baseball coaching staff as Director of Baseball Analytics.



UNIVERSITY  
OF HARTFORD

# University of Hartford Athletics

- Division I
- America East Conference
- Baseball, 2017: 20-30 overall, 8-13 in conference

# University of Hartford Athletics

- Division I
- America East Conference
- Baseball, 2017: 20-30 overall, 8-13 in conference

## Coaching staff's questions:

- Optimal lineup?
- When to use our relief pitchers?
- When/if should we bunt/steal/etc.?

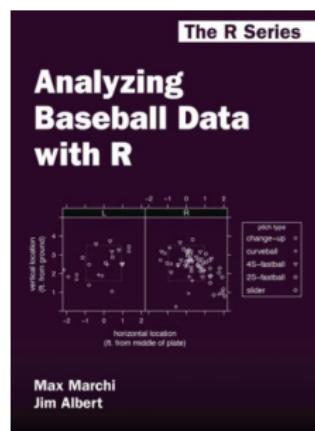
UNIVERSITY  
OF HARTFORD

# University of Hartford Athletics

- Division I
- America East Conference
- Baseball, 2017: 20-30 overall, 8-13 in conference

## Coaching staff's questions:

- Optimal lineup?
- When to use our relief pitchers?
- When/if should we bunt/steal/etc.?



# Seeking play-by-play data for college baseball

## Baseball data is widely available:

- MLB play-by-play available from Retrosheet through the `baseballr` package by Bill Petti.

# Seeking play-by-play data for college baseball

## Baseball data is widely available:

- MLB play-by-play available from Retrosheet through the `baseballr` package by Bill Petti.
- Metal bats and stark talent discrepancy (i.e., more runs) separates NCAA baseball from MLB, warranting research specific to college.

# Seeking play-by-play data for college baseball

## Baseball data is widely available:

- MLB play-by-play available from Retrosheet through the `baseballr` package by Bill Petti.
- Metal bats and stark talent discrepancy (i.e., more runs) separates NCAA baseball from MLB, warranting research specific to college.
- Play-by-play data in college baseball is limited (2014 data: @drbryan on github)

# Seeking play-by-play data for college baseball

## Baseball data is widely available:

- MLB play-by-play available from Retrosheet through the `baseballr` package by Bill Petti.
- Metal bats and stark talent discrepancy (i.e., more runs) separates NCAA baseball from MLB, warranting research specific to college.
- Play-by-play data in college baseball is limited (2014 data: @drbryan on github)

## NCAA Baseball Scraper:

- Uses the `tidyverse`, `XML`, `stringr` and `RCurl` packages in R to collect data from `stats.ncaa.com`.
- 316 NCAA Division I teams playing over 8100 games per year for 2017-2018.

# Table of Contents

- 1 Background and Motivation
- 2 Expected Runs
- 3 Win probability model
- 4 Application of Win Probability: WPA
- 5 Player Ability and WPA

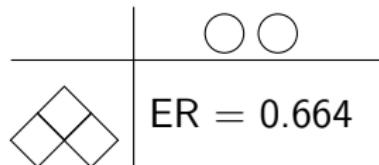
## Expected Runs

- **Expected Runs:** How many runs the average team will score the *rest of the inning* given the current base-out state?

# Expected Runs

- **Expected Runs:** How many runs the average team will score the *rest of the inning* given the current base-out state?

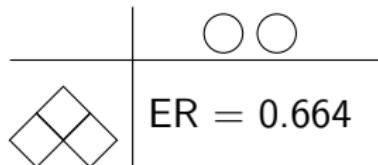
Leading off an  
inning:



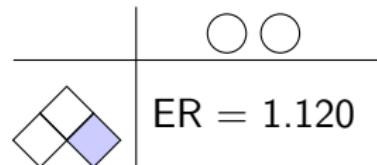
# Expected Runs

- **Expected Runs:** How many runs the average team will score the *rest of the inning* given the current base-out state?

Leading off an  
inning:



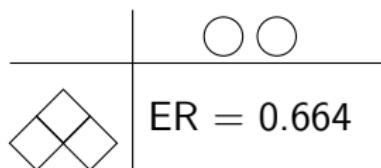
Single:



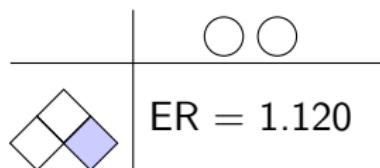
# Expected Runs

- **Expected Runs:** How many runs the average team will score the *rest of the inning* given the current base-out state?

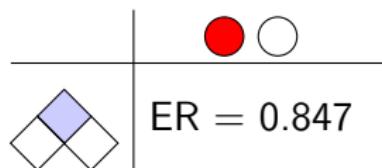
Leading off an  
inning:



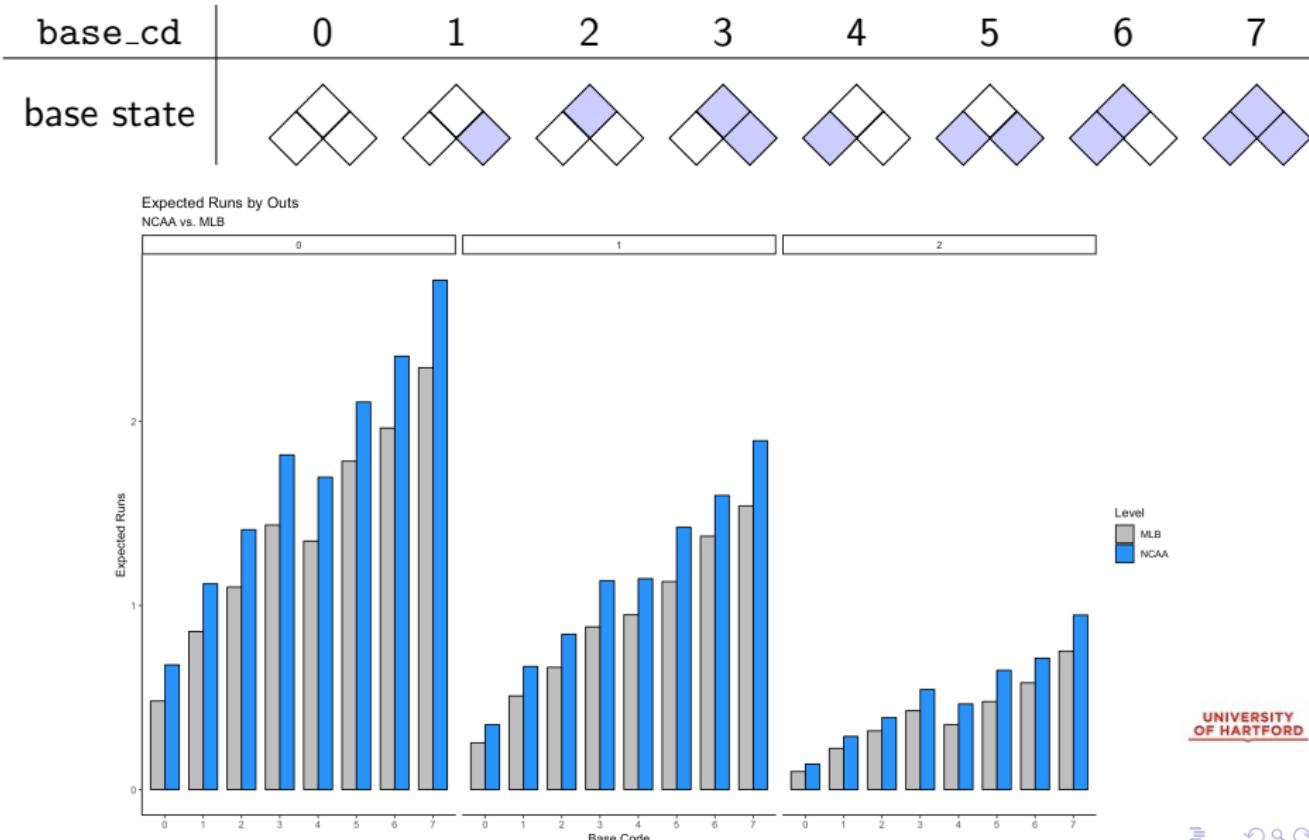
Single:



Sac Bunt:



# College Baseball Expected Runs Matrix



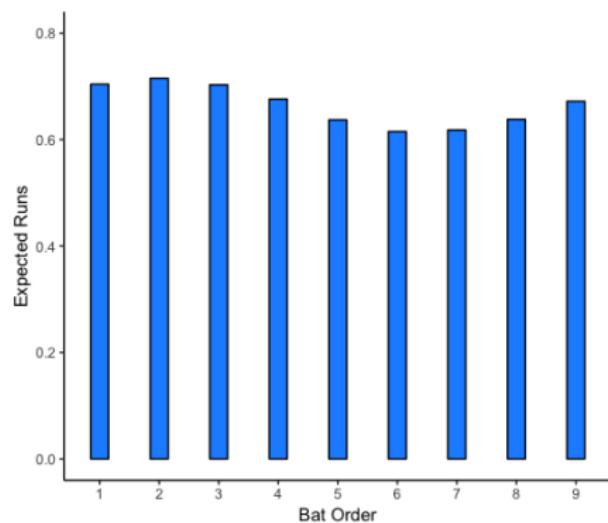
## Expected Runs: Batting Order

- Expected runs ignores the current spot in the batting order.
- Intuitively, a team can expect to score more runs when its better hitters (e.g., top of the order) are coming up.

# Expected Runs: Batting Order

- Expected runs ignores the current spot in the batting order.
- Intuitively, a team can expect to score more runs when its better hitters (e.g., top of the order) are coming up.

Figure: Expected runs in college baseball by batting order: 2017-2018



# Table of Contents

- 1 Background and Motivation
- 2 Expected Runs
- 3 Win probability model
- 4 Application of Win Probability: WPA
- 5 Player Ability and WPA

# Win Probability Model Structure

- Given the state of the game, how often does the hitting team win the game?

# Win Probability Model Structure

- Given the state of the game, how often does the hitting team win the game?

## Two models:

- Away team (Top of the inning)
- Home team (Bottom of the inning)

# Win Probability Model Structure

- Given the **state of the game**, how often does the hitting team win the game?

## Inputs:

- Inning
- Score differential
- Bases
- Outs
- Bat Order

# Win Probability Model Structure

- Given the **state of the game**, how often does the hitting team win the game?

## Inputs:

- Inning
  - Score differential
  - Bases
  - Outs
  - Bat Order
- } ⇒ Expected Runs

# Win Probability Model Structure

- Given the **state of the game**, how often does the hitting team win the game?

## Inputs:

- Inning
  - Score differential
  - Bases
  - Outs
  - Bat Order
- $\left. \begin{array}{l} \\ \\ \\ \end{array} \right\} \Rightarrow \text{Expected Runs}$

Expected Score Difference (ESD) = Score differential + Expected Runs

# Win Probability Model Structure

**Generalized Additive Model with logit link function:**

$$\log \left( \frac{p(\text{Win})}{p(\text{Loss})} \right) = s(\text{ESD}) + s(\text{inning}) + s(\text{inning}, \text{ESD})$$

where  $s$  is a smooth function.

# Win Probability Model Structure

**Generalized Additive Model with logit link function:**

$$\log \left( \frac{p(\text{Win})}{p(\text{Loss})} \right) = s(\text{ESD}) + s(\text{inning}) + s(\text{inning}, \text{ESD})$$

where  $s$  is a smooth function.

**Innings:**

- Extra innings are treated as the ninth inning.

# Win Probability Model Structure

## Generalized Additive Model with logit link function:

$$\log \left( \frac{p(\text{Win})}{p(\text{Loss})} \right) = s(\text{ESD}) + s(\text{inning}) + s(\text{inning}, \text{ESD})$$

where  $s$  is a smooth function.

### Innings:

- Extra innings are treated as the ninth inning.
- Bottom of the ninth has three outcomes:
  - Home win,
  - Home loss,
  - Extra innings.

# Win Probability Model Structure

## Generalized Additive Model with logit link function:

$$\log \left( \frac{p(\text{Win})}{p(\text{Loss})} \right) = s(\text{ESD}) + s(\text{inning}) + s(\text{inning}, \text{ESD})$$

where  $s$  is a smooth function.

### Innings:

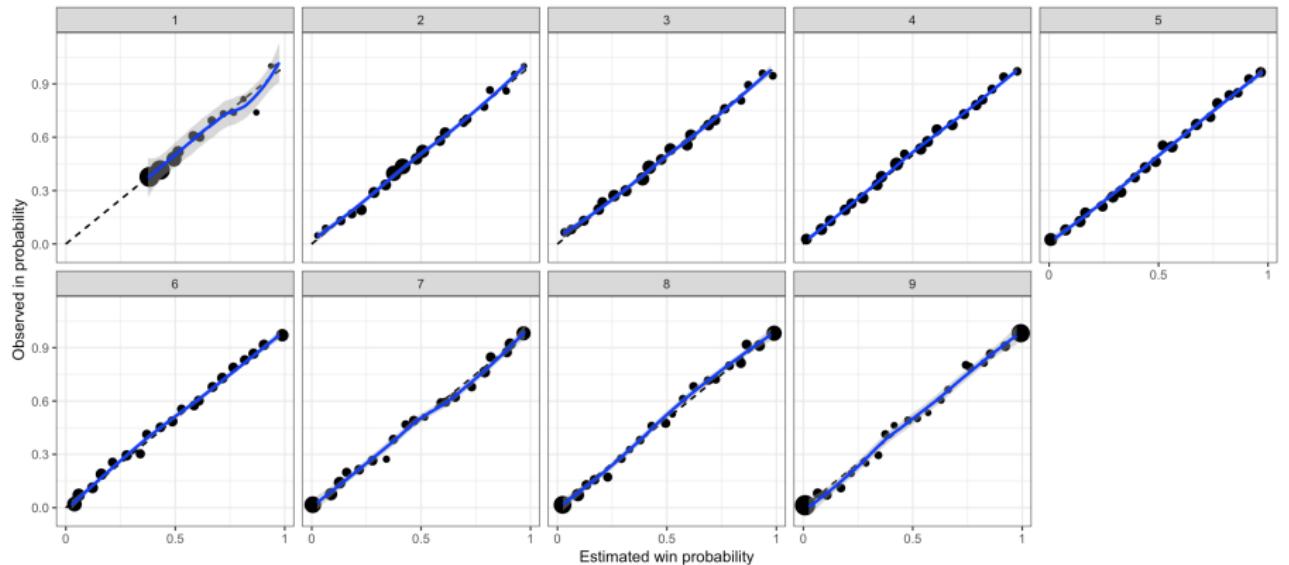
- Extra innings are treated as the ninth inning.
- Bottom of the ninth has three outcomes:

- Home win,
- Home loss,
- Extra innings.

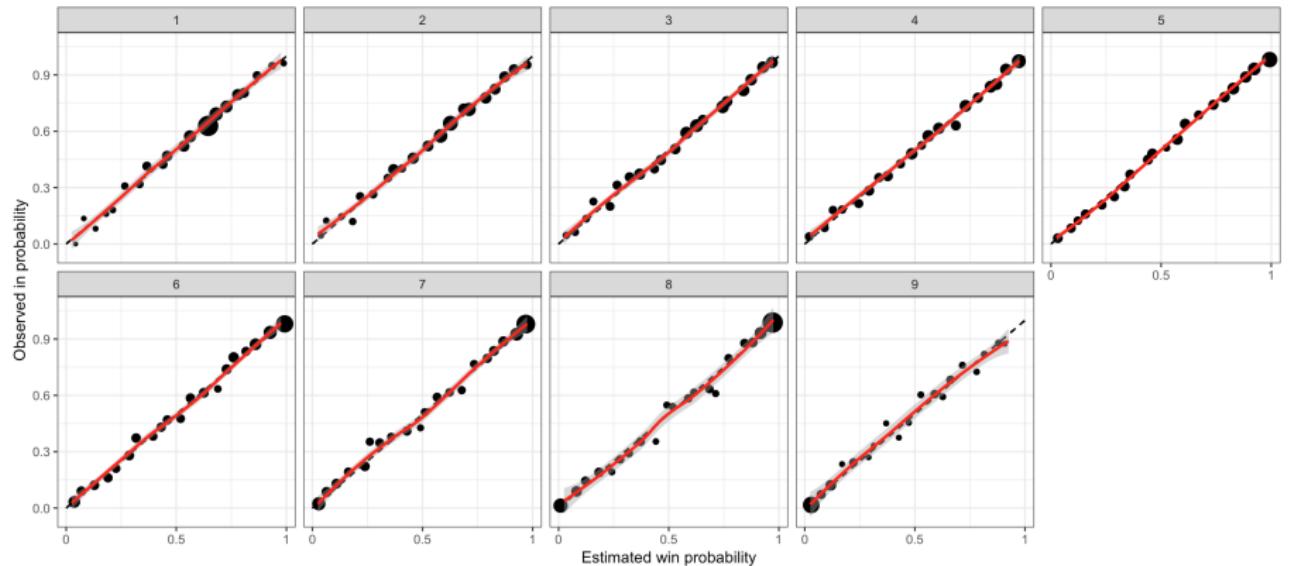
}

⇒ Multinomial Logistic Regression Model

# Model Reliability: Away



## Model Reliability: Home



# Table of Contents

- 1 Background and Motivation
- 2 Expected Runs
- 3 Win probability model
- 4 Application of Win Probability: WPA
- 5 Player Ability and WPA

# Win Probability Added

Win Probability Added (WPA) = WP<sub>after</sub> - WP<sub>before</sub>

# Win Probability Added

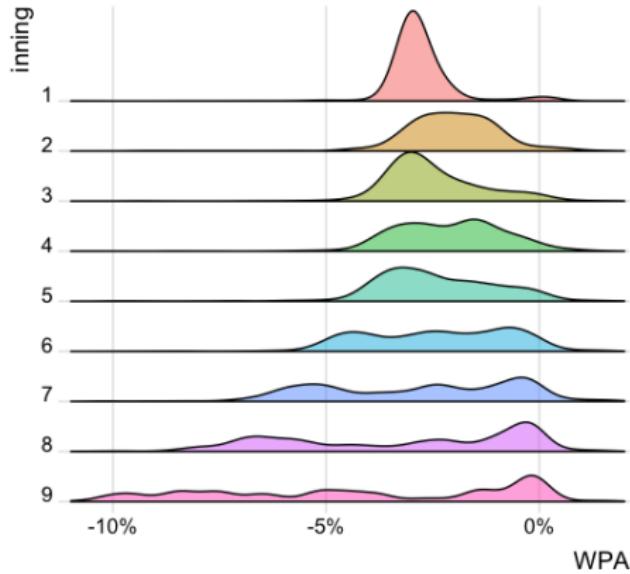
Win Probability Added (WPA) = WP<sub>after</sub> − WP<sub>before</sub>

**Bunting WPA:**

# Win Probability Added

Win Probability Added (WPA) = WP<sub>after</sub> − WP<sub>before</sub>

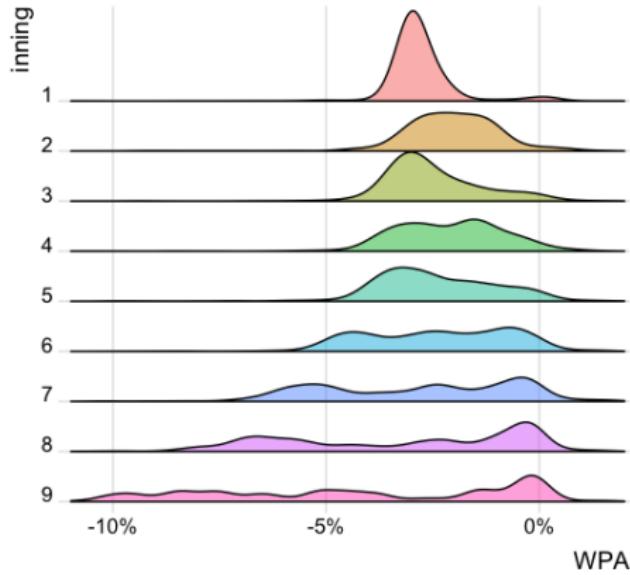
## Bunting WPA:



# Win Probability Added

Win Probability Added (WPA) = WP<sub>after</sub> - WP<sub>before</sub>

## Bunting WPA:

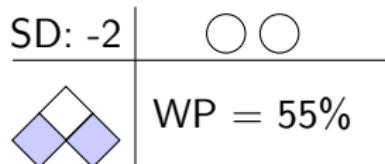


**Note:** An error occurs on 6.6% of bunt attempts (compared to 2.3% of non-bunt plays).

# WPA Example

## Situation

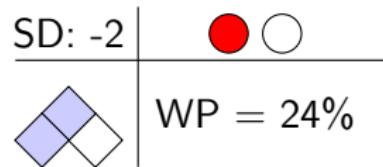
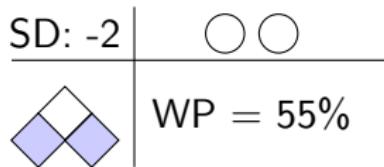
- Loyola Marymount: 6 @ Pepperdine: 4
- Bottom 9, runners on First and third, 2nd batter at bat.



# WPA Example

## Situation

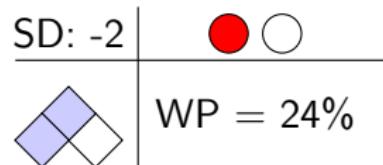
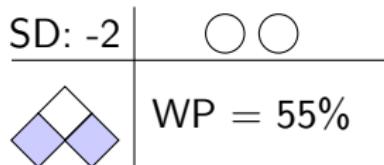
- Loyola Marymount: 6 @ Pepperdine: 4
- Bottom 9, runners on First and third, 2nd batter at bat.



# WPA Example

## Situation

- Loyola Marymount: 6 @ Pepperdine: 4
- Bottom 9, runners on First and third, 2nd batter at bat.

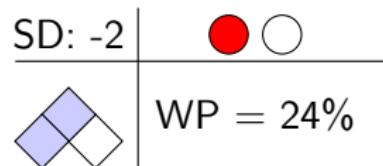
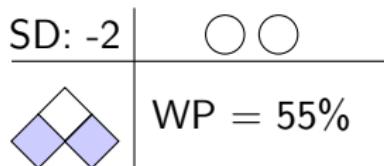


WPA = -31%.

# WPA Example

## Situation

- Loyola Marymount: 6 @ Pepperdine: 4
- Bottom 9, runners on First and third, 2nd batter at bat.



WPA = -31%.

## Analyst:

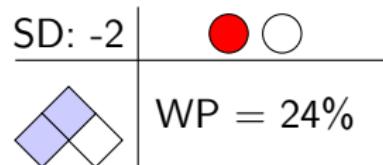
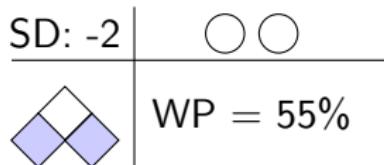


UNIVERSITY  
OF HARTFORD

# WPA Example

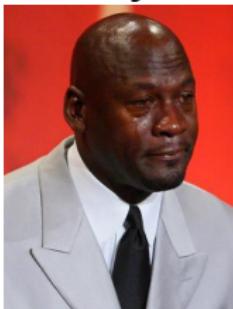
## Situation

- Loyola Marymount: 6 @ Pepperdine: 4
- Bottom 9, runners on First and third, 2nd batter at bat.



WPA = -31%.

**Analyst:**



**Coach (after walk-off HR):**



UNIVERSITY  
OF HARTFORD

# Table of Contents

- 1 Background and Motivation
- 2 Expected Runs
- 3 Win probability model
- 4 Application of Win Probability: WPA
- 5 Player Ability and WPA

# Considering Player Ability: xWPA

- Given a player's ability, can we estimate the impact they will have on the game at a particular situation?

## Considering Player Ability: xWPA

- Given a player's ability, can we estimate the impact they will have on the game at a particular situation?
- Expected Win Probability Added (xWPA) =

$$\frac{w_{BB} * BB + w_{1B} * 1B + w_{2B} * 2B + w_{3B} * 3B + w_{HR} * HR - w_O * Outs}{PA}$$

where  $w_k$  is the average WPA for each event  $k$  given the game situation.

# Considering Player Ability: xWPA

## Pepperdine 2nd batter:

- BA: 0.214, OBP: 0.335, SLG: 0.260, wOBA: 0.342.

# Considering Player Ability: xWPA

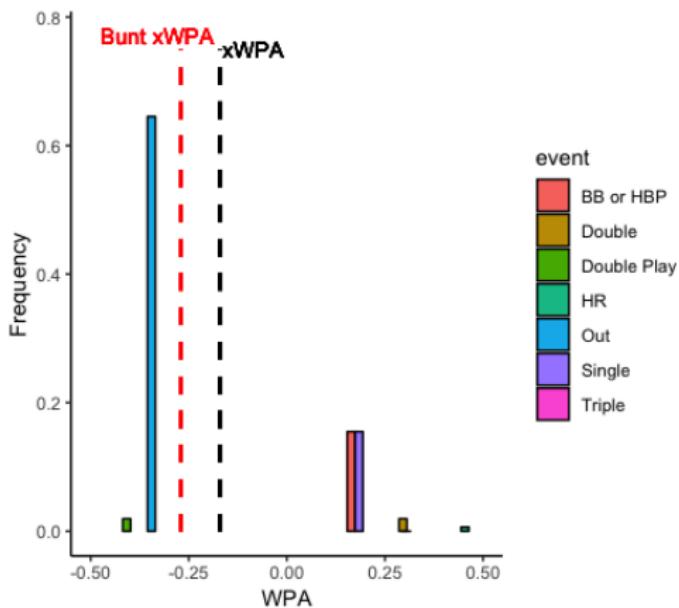
## Pepperdine 2nd batter:

- BA: 0.214, OBP: 0.335, SLG: 0.260, wOBA: 0.342.
- **xWPA: -17%,**

# Considering Player Ability: xWPA

## Pepperdine 2nd batter:

- BA: 0.214, OBP: 0.335, SLG: 0.260, wOBA: 0.342.
- **xWPA: -17%**,
- **Bunt xWPA: -27%**



# Considering Player Ability: wOBAp



Tangotiger ✅  
@tangotiger

I REALLY should generate the game-state wOBA at some point. For example:

bases are loaded, two outs, bottom of last inning, score is tied

$$\text{wOBA} = (1.0 \times \text{HB} + 1.0 \times \text{BB} + 1.0 \times 1\text{B} + 1.0 \times 2\text{B} + 1.0 \times 3\text{B} + 1.0 \times \text{HR}) / \text{PA}$$

Obviously, OBP = wOBA in this case for every single player.

UNIVERSITY  
OF HARTFORD

## Considering Player Ability: wOBAp

- Use xWPA to calculate wOBAp by shifting weights so that outs = 0:

$$wOBAp = \frac{BB + W_{1B}(1B) + W_{2B}(2B) + W_{3B}(3B) + W_{HR}(HR)}{PA}$$

where  $W_k = \frac{w_k + w_O}{w_{BB} + w_O}$ .

## Considering Player Ability: wOBAp

- Use xWPA to calculate wOBAp by shifting weights so that outs = 0:

$$wOBAp = \frac{BB + W_{1B}(1B) + W_{2B}(2B) + W_{3B}(3B) + W_{HR}(HR)}{PA}$$

where  $W_k = \frac{w_k + w_O}{w_{BB} + w_O}$ .

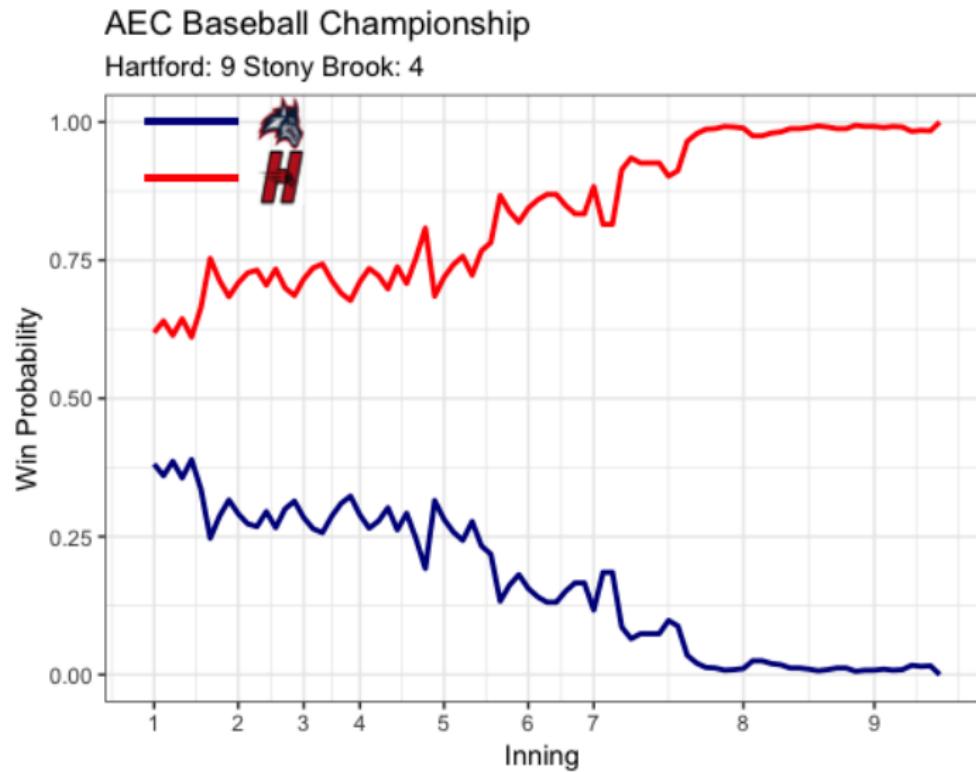
### Previous Example:

- $WP_{\text{before}} \approx 75\%$ ,
- $WP_{\text{Out}} \approx 52\%$ ,  $WP_{\text{Not Out}} = 100\%$ :  $W_k = \frac{0.25 + 0.48}{0.25 + 0.48} = 1$ :

$$wOBAp = \frac{BB + 1B + 2B + 3B + HR}{PA} = OBP$$

UNIVERSITY  
OF HARTFORD

# Conclusion



UNIVERSITY  
OF HARTFORD

# Conclusion



UNIVERSITY  
OF HARTFORD

# Thank you!

- CMSAC Organizers and Competition Committee
- Coaches Blood, Mal, Glynn and Newton from UH Baseball
- Kristina and Nicole
- Krystin, Rocky and Thor

## Contact:

- [davmiller@hartford.edu](mailto:davmiller@hartford.edu)
- @davmiller9