

# Rewarding batters for baserunner advancement

A ridge-regressed Rasch model

Scott Powers

Stanford University

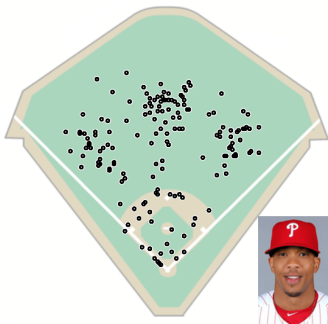
Saberseminar 2015

# Background

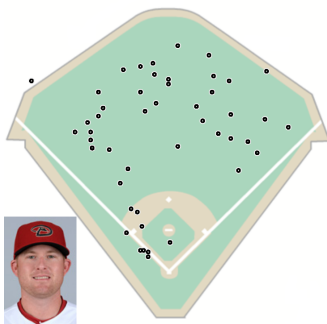
- Allen (2014): LHB singles more valuable than RHB singles
  - because if runner is on 1B, he more likely gets to 3B
  - LHBs undervalued by  $\sim \$300,000/\text{year}$  on average
  - RHBs overvalued by  $\sim \$150,000/\text{year}$  on average
- This observation has been made before
  - Pakin (1993)
  - James (1997)
- But no one quantified its value before Allen (2014)
- Today: Extend Allen's work to evaluate individual players

## Singles spray charts (2014)

Ben Revere



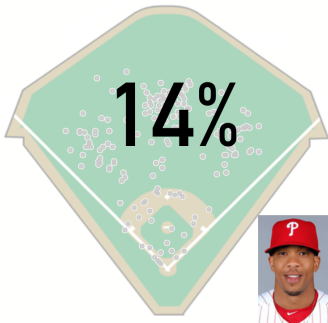
Mark Trumbo



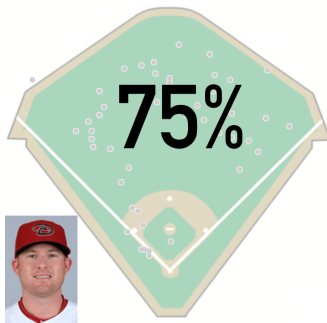
Images created with Bill Petti's Spray Chart Comparison tool  
Portraits from Baseball-Reference.com

## Singles spray charts (2014)

Ben Revere



Mark Trumbo



Images created with Bill Petti's Spray Chart Comparison tool  
Portraits from Baseball-Reference.com

## Probabilistic baserunning model

- For each single with a runner on 1B, observe:
  - $B_i$ : the batter
  - $R_i$ : the runner on 1B
  - $H_i$ : handedness of  $B_i$  (L or R)
  - $z_i$ : vector of additional covariates to control for (e.g. indicators of full count, two outs, home team)
- $Y_i = \begin{cases} 1 & \text{if } R_i \text{ safely reaches third or home} \\ 0 & \text{otherwise} \end{cases}$

Model (logistic regression):

$$P(Y_i = 1) = \frac{e^{\alpha + \beta_{B_i} + \rho_{R_i} + \theta_{H_i} + \zeta^T z_i}}{1 + e^{\alpha + \beta_{B_i} + \rho_{R_i} + \theta_{H_i} + \zeta^T z_i}} = \frac{e^{\eta_i}}{1 + e^{\eta_i}}$$

## Illustrative example

$$\alpha = -1.52$$

$$\beta_{\text{Morse}} = -0.01$$

$$\rho_{\text{Pence}} = +0.42$$

$$\theta_R = +0.00$$



$$P(Y = 1) = \frac{e^{-1.52-0.01+0.42+0.00}}{1 + e^{-1.52-0.01+0.42+0.00}} = \frac{e^{-1.11}}{1 + e^{-1.11}} = 24.7\%$$

# The dataset

- From Retrosheet play-by-play data
- All singles with a runner on 1B in 2014, excluding:
  - Plays with at least one error
  - Batted balls classified as popups
  - Plays when lead runner is thrown out
- Otherwise, anything goes
- $< 5\%$  of singles excluded
- Sample size:  $n = 8571$  singles; 641 batters; 671 runners

## Fitting the model

$$(\hat{\alpha}, \hat{\beta}, \hat{\rho}, \hat{\theta}, \hat{\zeta}) = \arg \min \left\{ - \sum_{i=1}^n \log P(y_i | \eta_i) + \lambda (||\beta||_2^2 + ||\rho||_2^2 + ||\theta||_2^2 + ||\zeta||_2^2) \right\}$$

- Solve this optimization problem with R package `glmnet` (coordinate descent)
- $\lambda$  is *regularization parameter*, chosen via cross-validation
- Benefit of regularization: regression to the mean



## Summary of results

$$\hat{\alpha} = -1.52$$

$$\hat{\theta}_L = 0.43$$

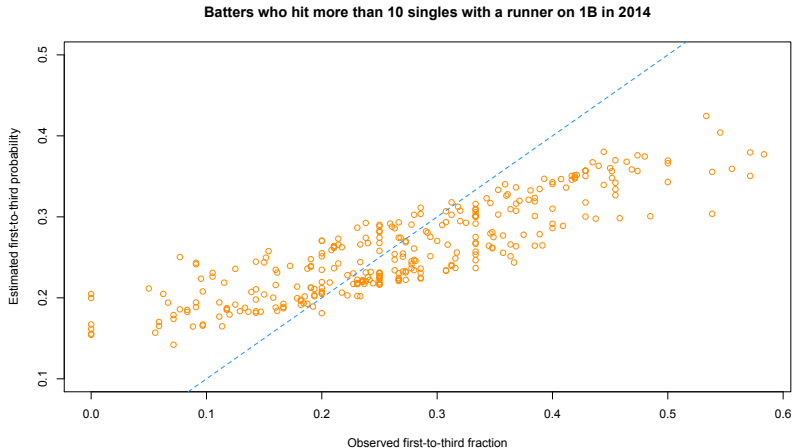
$$\hat{\zeta}_{home} = 0.13$$

$$\hat{\zeta}_{2outs} = 0.42$$

$$\hat{\zeta}_{2outs,3-2} = 1.24$$

R	$\hat{\rho}_R$
Elvis Andrus	0.73
Dexter Fowler	0.73
Rajai Davis	0.63
Jackie Bradley, Jr.	0.63
Brian Dozier	0.62
...	
Chris Carter	-0.49
Billy Butler	-0.53
Pablo Sandoval	-0.55
Starlin Castro	-0.61
Victor Martinez	-0.73

# 1<sup>st</sup>-to-3<sup>rd</sup> advancement: estimated vs. observed



## Implications for wOBA

Event	Actual Total	Theoretical breakup		
		Getting on	Moving over	Inning killer
OUT	-0.27	-0.01	-0.10	-0.16
BB	0.30	0.24	0.06	—
1B	0.46	0.25	0.21	—
2B	0.75	0.41	0.34	—
3B	1.03	0.61	0.42	—
HR	1.40	1.00	0.40	—

from tangotiger.net

$$wOBA_{naive} \propto .57 * BB + .73 * 1B + 1.02 * 2B + 1.30 * 3B + 1.67 * HR$$

## Implications for wOBA

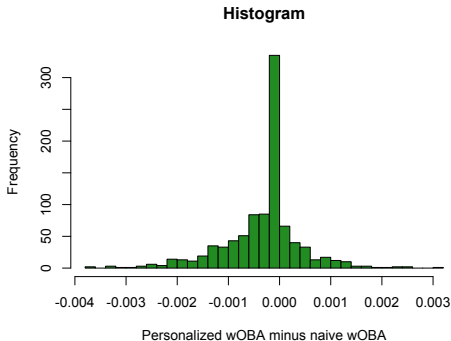
Event	Actual Total	Theoretical breakup		
		Getting on	Moving over	Inning Killer
OUT	-0.25	-0.00	-0.09	-0.16
BB	0.30	0.24	0.06	—
1B	???	0.25	.09 + .06 + 0.13	—
2B	???	0.40	.14 + 0.27	—
3B	1.01	0.59	0.42	—
HR	1.40	1.00	0.40	—

$$wOBA_{\text{Miguel Cabrera}} \propto .55*BB + .70*1B + .99*2B + 1.26*3B + 1.65*HR$$

$$wOBA_{\text{Willie Bloomquist}} \propto .55*BB + .68*1B + .97*2B + 1.26*3B + 1.65*HR$$

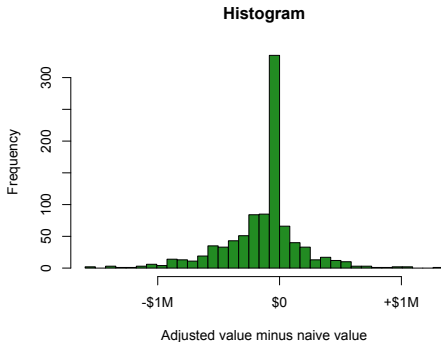
## Batters most overrated/underrated by wOBA

Player	Adjustment
James Loney	+.0032
Robinson Cano	+.0026
Eric Hosmer	+.0024
Joe Mauer	+.0024
Jed Lowrie	+.0023
...	
Rajai Davis	-.0030
Adeiny Hechavarria	-.0033
Danny Santana	-.0033
Starling Marte	-.0034
Dee Gordon	-.0037



## Batters most overrated/underrated by wOBA

Player	Adjustment
James Loney	+\$1.33M
Robinson Cano	+\$1.08M
Eric Hosmer	+\$1.03M
Joe Mauer	+\$1.00M
Jed Lowrie	+\$0.96M
...	
Rajai Davis	-\$1.26M
Adeiny Hechavarria	-\$1.37M
Danny Santana	-\$1.37M
Starling Marte	-\$1.41M
Dee Gordon	-\$1.54M



## Future work

- Consider baserunner advancement on outs
- Include identity of fielder as variable in regression
- Adjust WAR for baserunning

# Conclusions

- Identified \$1.33M in surplus value for James Loney in 2014
  - Not accounted for anywhere else, publicly
- Results on 2014 data are consistent with Allen (2014)
  - But allow for results on individual players



## Acknowledgments

- Bob McMurray for comments that led to improvements
- Retrosheet
- TL Turocy (Chadwick)
- Max Marchi and Jim Albert (Analyzing Baseball Data with R)
- Jerome Friedman, Trevor Hastie, Noah Simon and Rob Tibshirani (glmnet)

## References

- Allen D (2014). The single to right field: Why left-handed batters may be undervalued one Mike Trout Unit. *Stanford University STATS 50 Final Projects*. <http://web.stanford.edu/class/stats50/finalprojects.html>
- Dolinar S (2015). WAR diagram: position players. *FanGraphs*. Published May 15, 2015; accessed August 13, 2015.
- James B (1997). The Bill James Guide to Baseball Managers
- Pankin MD (1993). Subtle aspects of the game. SABR 23. <http://www.pankin.com/markov/sabr23.htm>
- Tango T (date unknown). How are runs really created. Accessed August 13, 2015. <http://www.tangotiger.net/runscreated.html>