Reference Point for the Baseball Players in Japan

Is the contribution of SABR metrics taken into account?

Reio TANJI

Osaka University

June.28,2018

References

- Pope & Simonsohn (2011,Association for Phychological Science) "Round Numbers as Goals : Evidence From Baseball, SAT Takers, and the Lab"
- Hakes & Sauer (2006, Jornal of Economic Perspectives) "An Economic Evaluation of the *Moneyball* Hypothesis"

Pope & Simonsohn

- Verify that **round numbers** in performance scales act as **reference points**, by examing three practical studies.
- In the first study, they found that baseball players in MLB prefer finising the season with a batting average(AVG) just above .300, to that with just below .300.
- Data : MLB player's play-by-play data from 1975 to 2008. Players with at least 200 at bat (打数) : N=8,817

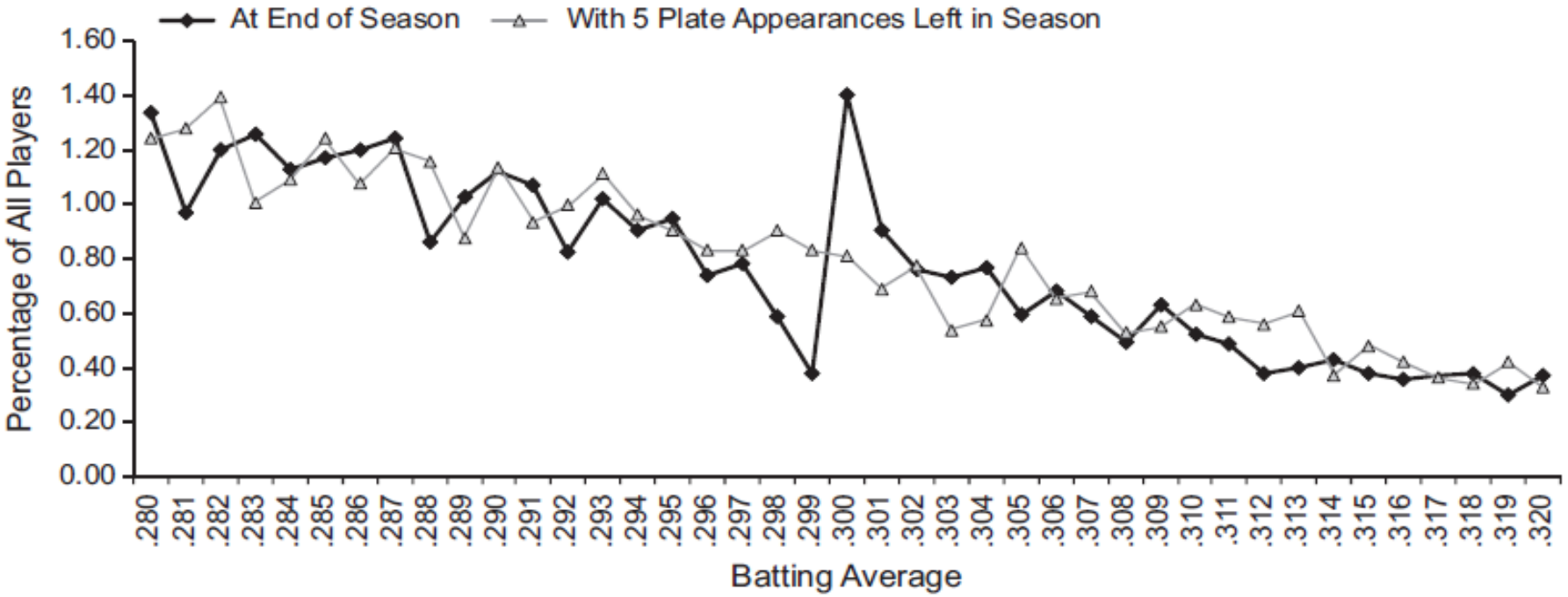
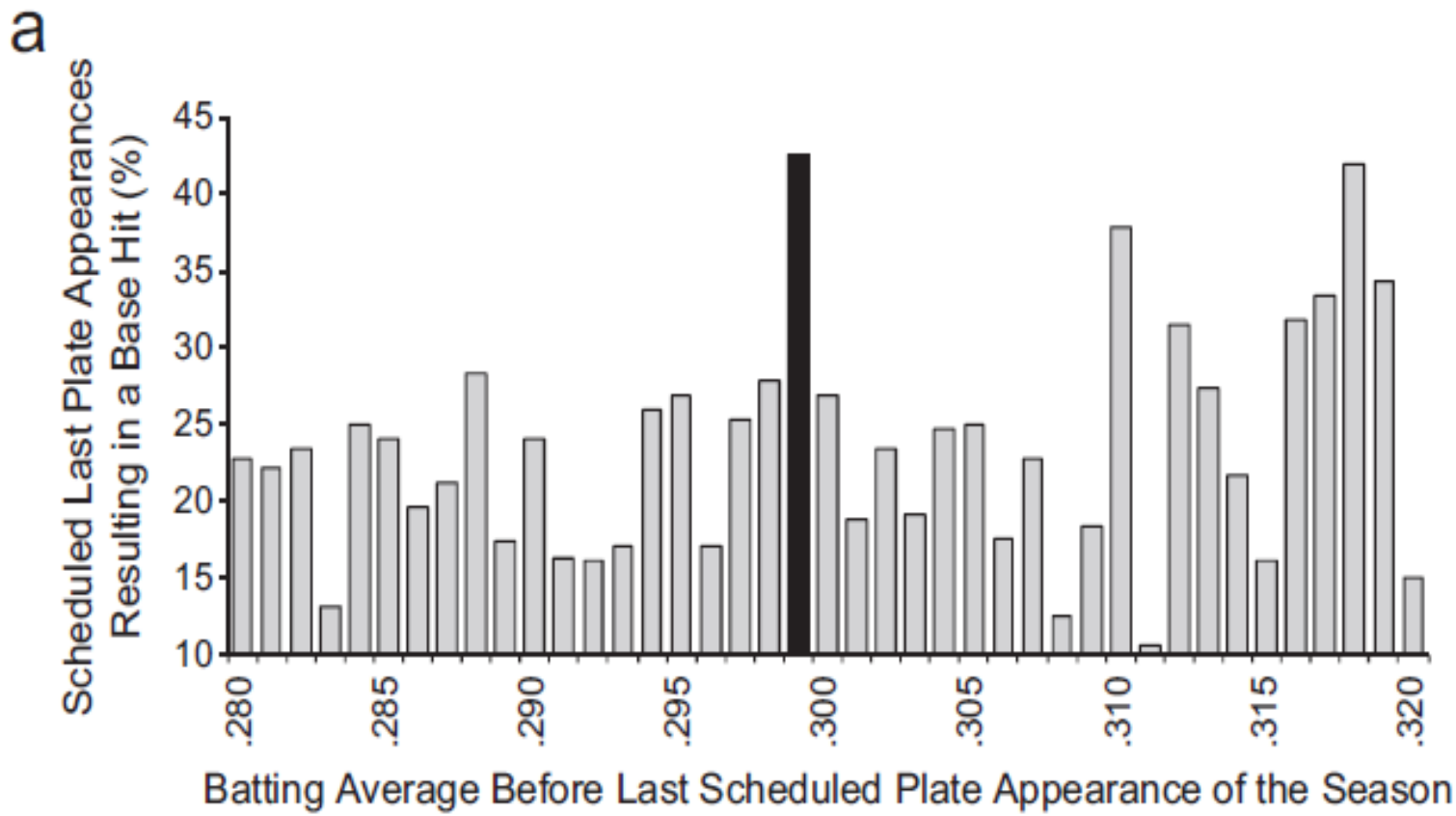
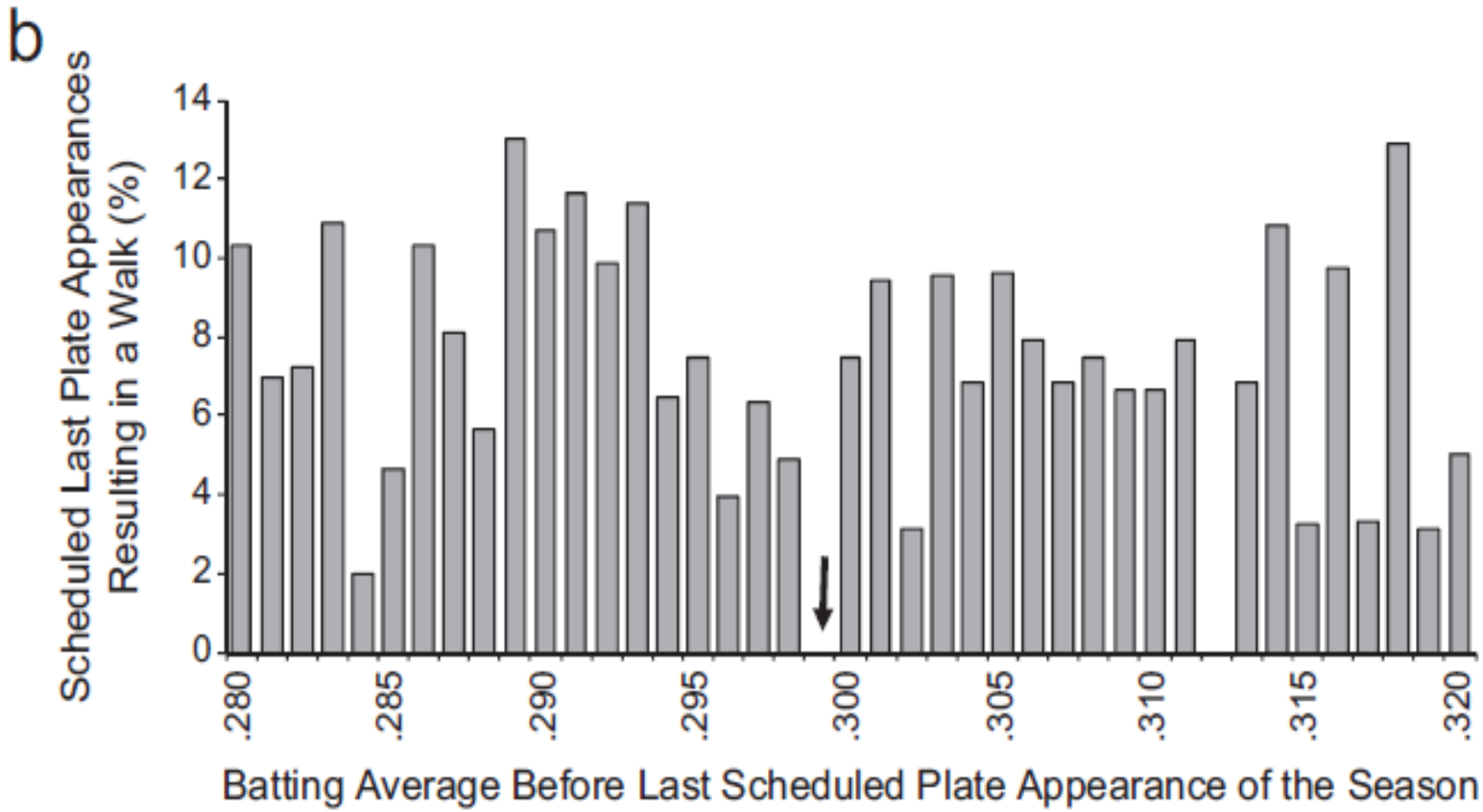


Fig. 1. Relative frequency of batting averages among Major League Baseball players between 1975 and 2008. Batting averages at the end of the baseball season and with five plate appearances left in the season are shown. The graph includes only player-seasons with at least 200 at bats.

- Players with .298 or .299 (0.97 %) $<$ with .300 or .301 (2.30 %), $Z = 7.35$, $p < .001$.
- Control distribution : when 5 plate appearances left in the season.



- Players with AVG of .299 was likely to get a base hit(43%) than overall(22.8%) at their last PA.
  $Z = 3.62, p < .001$.



- .298 or .299 players tend to walk (四球) than .300 or .301 players.
  $Z = 2.14$, $p = .032$.

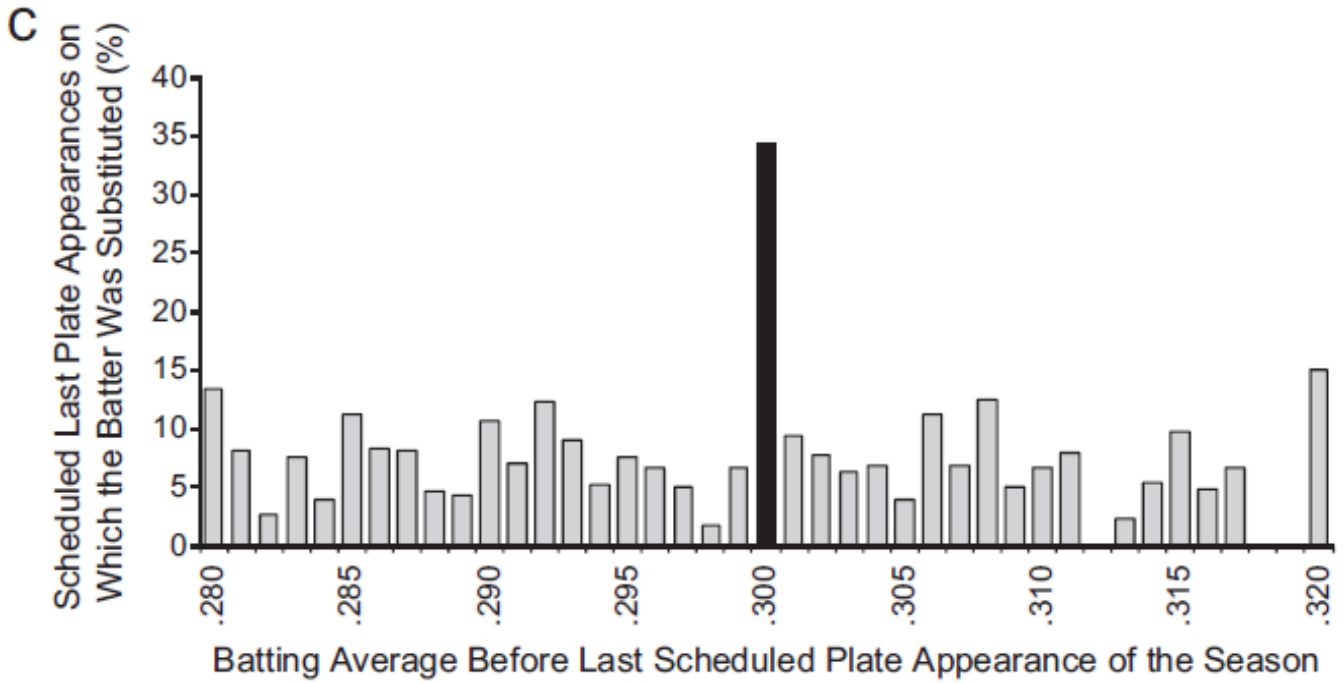


Fig. 2. Outcome of the last scheduled plate appearance of the season: percentage of plays resulting in (a) base hits, (b) walks (which cannot increase batting average), and (c) batter substitutions (pinch hitter brought in). Bars involved in tests of predictions are highlighted in black. The arrow in (b) emphasizes that not a single player with a batting average of .299 walked.

- If his AVG is just above .300, then he might end the season earlier by being substituted.
  $Z = 8.29$ and $p < .001$.

Pope & Simonsohn

- The behavior of baseball players proved the existence of the reference point of round numbers, such as batting average of .300.
- Limitations:
  There were only one relevant round number.
  Action to improve their performance took place on the last plate appearance.

Hakes & Sauer

- *"Moneyball Hypothesis"*
  : Michael Lewis's claim that the valuation of skills in MLB player's market was grossly inefficient.
- Members of the Society for American Baseball Research (In short, SABR) have studied that on-base percentage (OBP) plays more important role to consider the winning percentage than batting average.

Table 1

The Impact of On-Base and Slugging Percentage on Winning

	Model			
	1	*2*	*3*	*4*
Constant	0.508	0.612	0.502	0.500
	(0.114)	(0.073)	(0.099)	(0.005)
On-Base	3.294		2.141	2.032
	(0.221)		(0.296)	(0.183)
On-Base against	−3.317		−1.892	-2.032^{R}
	(0.196)		(0.291)	
Slugging		1.731	0.802	0.900
		(0.122)	(0.149)	(0.106)
Slugging against		−1.999	−1.005	-0.900^{R}
		(0.112)	(0.152)	
Number of observations	150	150	150	150
R^2	.825	.787	.885	.884

Hypothesis test of model 4, H^0: On-Base = Slugging
$F(1, 147) = 16.74$, p-value = 0.0001

Source: Retrosheet Game Logs, ⟨http://www.retrosheet.org⟩. The data were obtained free of charge from, and are copyrighted by, Retrosheet, 20 Sunset Rd., Newark, DE 19711.

Notes: Data are aggregate statistics for all 30 teams from 1999–2003. Coefficient estimates were obtained using ordinary least squares. Coefficients for annual 0/1 dummy variables are suppressed. Standard errors are in parentheses. Superscript "R" indicates that the coefficient was restricted to equal its counterpart in the regression. The p-value for the null hypothesis that restrictions are valid is 0.406 ($F = 0.52$).

Table 3

The Baseball Labor Market's Valuation of On-Base and Slugging Percentage

	All Years	*2000–2003*	*2000*	*2001*	*2002*	*2003*	*2004*
On-Base	1.360	0.842	1.334	−0.132	0.965	1.351	3.681
	(0.625)	(0.678)	(1.237)	(1.230)	(1.489)	(1.596)	(1.598)
Slugging	2.392	2.453	2.754	3.102	2.080	2.047	2.175
	(0.311)	(0.338)	(0.628)	(0.613)	(0.686)	(0.850)	(0.788)
Plate appearances	0.003	0.003	0.003	0.003	0.003	0.003	0.003
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Arbitration eligible	1.255	1.242	1.293	1.106	1.323	1.249	1.323
	(0.047)	(0.048)	(0.102)	(0.100)	(0.100)	(0.111)	(0.115)
Free agency	1.683	1.711	1.764	1.684	1.729	1.663	1.575
	(0.044)	(0.185)	(0.096)	(0.092)	(0.097)	(0.107)	(0.105)
Catcher dummy	0.152	0.185	0.137	0.065	0.208	0.343	0.059
	(0.056)	(0.061)	(0.124)	(0.116)	(0.122)	(0.134)	(0.133)
Infielder dummy	−0.029	−0.007	0.060	0.069	−0.087	−0.054	−0.100
	(0.040)	(0.044)	(0.087)	(0.083)	(0.086)	(0.095)	(0.098)
Intercept	10.083	10.429	10.078	10.347	10.490	10.289	9.782
	(0.170)	(0.178)	(0.360)	(0.321)	(0.358)	(0.387)	(0.414)
Observations	1736	1402	353	357	344	342	340
R^2	0.675	0.687	0.676	0.728	0.695	0.655	0.635
Value of one-standard-deviation increase (in millions of dollars)							
On-Base			0.14	0.16	0.17	0.19	0.49
Slugging			0.52	0.61	0.64	0.70	0.61

Source: Same as Table 2.

Notes: The dependent variable is ln(Salary) for year t, and performance variables are from year $t-1$. 0/1 dummies for each year are included in the pooled regressions. Standard errors in parentheses. The sample includes all players with at least 130 plate appearances during the relevant season.

Figure 1

Labor Market Returns to On-Base and Slugging Percentage Over Time

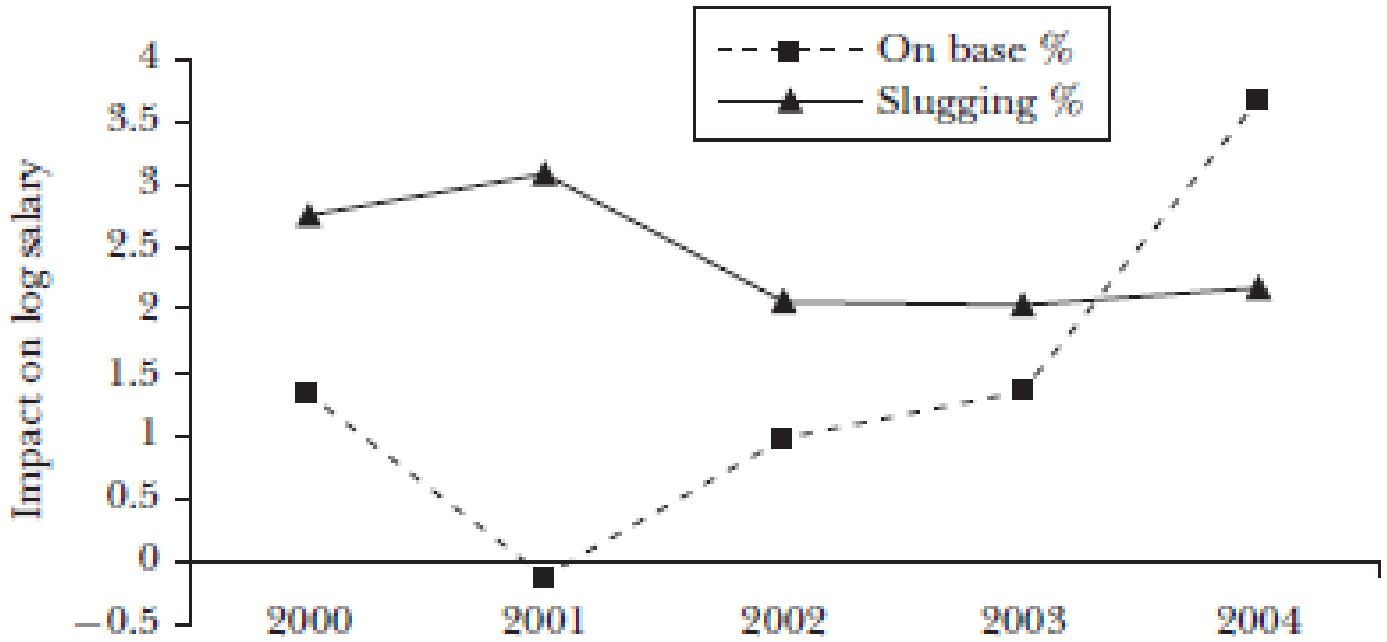
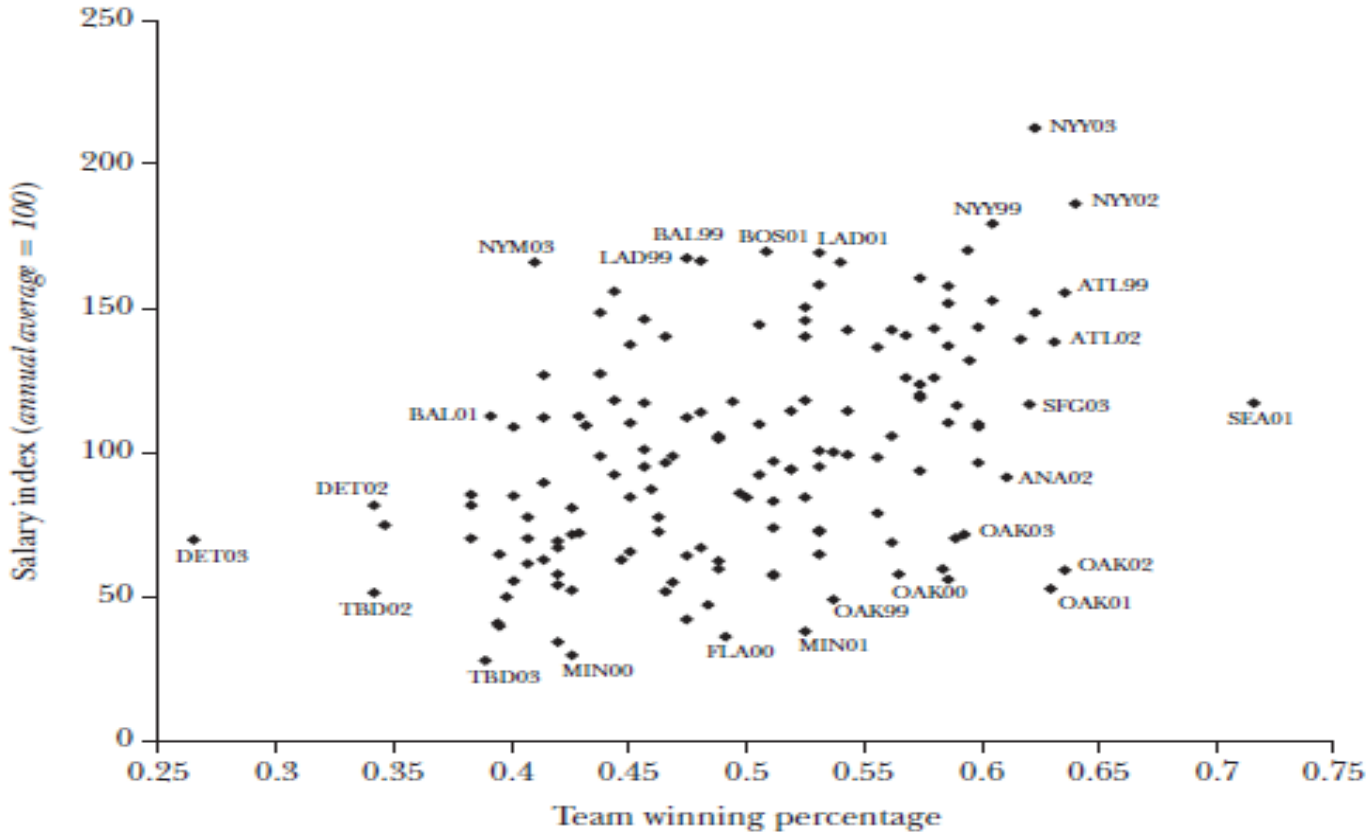


Figure 2

Frontier for Efficient Conversion of Team Salary into Team Winning Percentage, 1999–2003



Source: Won-loss records from www.baseball-reference.com. Team salaries from SABR, ⟨http://businessofbaseball.com/data.htm⟩.

Notes: Teams near the frontiers of efficient and inefficient conversion are given a team–year label, with the last two digits indicating the year. Teams near the frontiers are Atlanta (ATL), Anaheim (ANA), Baltimore (BAL), Boston (BOS), Detroit (DET), Florida (FLA), Los Angeles Dodgers (LAD), Minnesota (MIN), Oakland (OAK), New York Mets (NYM), New York Yankees (NYY), Seattle (SEA), San Francisco Giants (SFG), and Tampa Bay Devil Rays (TBD). All years for Oakland are included.

Extension

- On-base percentage may be more important than batting average, when it comes to consider raising winning percentage, and it reflects players' effort in similar way to batting average.
  $\Rightarrow$ Round numbers in on-base percentage may act as reference point, as well as batting-average : .350 or .400.

Data

- Sortable team stats of Nippon Professional Baseball (NPB), from 2008 to 2017.
  : N=120
- Indexes : Win average (勝率 : WA), Runs (得点 : R) Runs allowed (失点 : RA), Batting average (AVG), On-base percentage (出塁率 : OBP) Slugging percentage(長打率 : SLG), OPS (OBP + SLG).

Model

- Confirm that also in NPB, OBP contributes better to the Win Average than AVG or SLG.
  Applying OLS,

$$WA_i = \beta \mathbf{X}_i + RA_i + u_i$$

$$R_i = \gamma \mathbf{X}_i + v_i$$

WA_i : Win average of team i

R_i : Runs of team i

$\mathbf{X}_i$: OBP, SLG and AVG of the team i