International Journal of Statistical Sciences *ISSN 1683-5603*
Vol. 21(1), 2021, pp 95-104

Best Offensive Percentage (BOP): A Superior Way to Measure the Offensive Value of a Baseball Player Than OPS

Kenneth Ko

Graziadio Business School, Pepperdine University,
Los Angeles, California, United States
Email: kenneth.ko@pepperdine.edu

[Received October 10, 2020; Revised November 15, 2020; Accepted November 22, 2020]

Abstract

Currently, on-base plus slugging (OPS) is the preferred measure to evaluate the offensive value of baseball players. Although this is a good measure, it is not the best measure because the weights for OPS are fixed. A better measure is proposed, Best Offensive Percentage (BOP) which picks the optimal weights based on a regression model using runs scored as the dependent variable and BOP as the independent variable. Data is used from the 2008–2018 Major League Baseball seasons to determine the optimal weights. Using BOP as opposed to OPS will give baseball general managers a competitive advantage to help them to determine not only which players they should pursue but also how much they should pay them. Used appropriately, BOP will help teams select better players, win more games, and win more championships.

Keywords: Baseball, Analytics, On-base plus slugging (OPS), Regression analysis, Statistics.

AMS Classification: 62, 65.

1. Introduction

The movie Moneyball popularized the use of analytics in baseball (Merrifield, 2018; Mirsky, 2014). Billy Beane, who was the general manager of the Oakland A's from 1998–2016, realized in 2002 that on-base percentage was being undervalued by other general managers. He exploited this fact and, despite having the third lowest payroll in baseball at the time, the 2002 Oakland A's went on to have a remarkable season – winning 20 games in a row at one point and finishing with a 103-59 before losing to the Minnesota Twins in the playoffs (3 game to 2).

Theo Epstein is another strong proponent of analytics, and helped the Boston Red Sox to end the 86-year old "curse of the Bambino" and win the World Series in 2004. As if that was not enough, Epstein again used analytics to help the Chicago Cubs end an even longer, 108-year old, drought -- the longest in all major North American sports history – and win a World Series title in 2016.

Clearly, analytics plays a key role in baseball. Baseball is considered the most individual of the team sports, and this is quite true. Baseball is focused on the one-on-one matchup between the pitcher and the batter. And, even after a ball is put into play, the focus is on an individual fielder making a play. In baseball, analytics takes on great importance because it helps teams to do the two things that are essential to winning: maximizing runs scored and minimizing runs allowed.

The teams that can use analytics "better" will have more success on the field. In fact, the use of analytics in baseball is well-documented. The growth of analytics in baseball has been assessed (Shaw, 2014). The use of wins above replacement (WAR) has been examined (Baumer, et al., 2015) as applied to baseball. Furthermore, research has been done to explore how baseball-analytic principles can even applied to other industries, such as the manufacturing industry (Valerdi, 2017) and the insurance industry (Hyle, 2011).

2. A New Methodology: Best Offensive Percentage (BOP)

On-base plus slugging (OPS) is considered the "gold standard" in terms of measuring the offensive value of a baseball player. In order to understand what OPS is, we need to consider the following equations/definitions:

on-base percentage (OBP) = (H + BB + HBP) / (AB + BB + HBP + SF)

slugging percentage (SLG) = [(1B)+ (2*2B) + (3*3B) + (4*HR)] / AB

OPS = OBP + SLG

where:

H = hits, BB = bases on balls (walks), HBP = times hit by pitch, AB = at bats

SF = sacrifice flies, 1B = singles, 2B = doubles, 3B = triples, HR = home runs.

The importance of OPS is clear to even a casual follower of the game of baseball. As a case in point, the 2018 World Series champion Boston Red Sox was a historically great team that won a total of 119 games (including the playoffs), which is the third highest total of all time. Not surprisingly, the 2018 Red Sox had the highest OPS of any team in the major leagues at 0.792. Also not surprisingly, the Los Angeles Dodgers – the Red Sox World Series opponent – had the highest OPS of all National Leagues teams at 0.774.

For the data for this paper, we used final Major League Baseball team statistics from the 2008–2018 baseball seasons. One question we need to consider is how should we measure the offensive value of a baseball player? To answer this question we need to consider the ultimate goal of an offense, and that is simply to score runs. In that sense, OPS is an excellent measure of the offensive value of a baseball player because based on the aforementioned data the R^2 value of a model with runs as the dependent variable and OPS as the independent variable was an average of 0.8812 over this time period with a range of 0.7770 to 0.9349. Please see Table 2 below.

Yet, as high as the R^2 value is with OPS as the independent variable, this is definitely not the optimal model because the weights for all of the variables are fixed. So, if we allow the weights to vary to minimize the Error Sum of the Squares (ESS) between runs and estimated runs, this would certainly be a better model. So, we now introduce a new variable, Best Offensive Percentage (BOP), as an alternative to OPS where the weights are allowed to vary:

$$\text{BOP} = (w_1\text{*1B} + w_2\text{*2B} + w_3\text{*3B} + w_4\text{*HR})/\text{AB} +$$

$$(w_5\text{*BB} + \text{w6*HBP})/(\text{AB+BB+HBP+SF})$$

The weights are then chosen by using a multiple regression analysis model which minimizes the ESS. Please see Table 1 for the BOP weights for 2018.

Table 1: 2018 BOP Weights

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	1B	2B	3B	HR	BB	HBP	AB	SF	Weights	BOP	R	Est R	Squared Error
2	798	259	50	176	560	52	5460	45	1.4742	0.566	693	678.2	218.9
3	915	314	29	175	511	66	5582	43	2.6082	0.593	759	751.3	59.2
4	872	242	15	188	422	57	5507	35	1.7374	0.551	622	639.7	314.2
5	915	355	31	208	569	55	5623	48	3.8543	0.634	876	859.0	289.9
6	966	286	34	167	576	78	5624	46	1.4215	0.595	761	755.9	26.2
7	851	259	40	182	425	66	5523	32	0.9045	0.557	656	653.9	4.3
8	956	251	25	172	559	65	5532	35	-825.7391	0.585	696	730.5	1187.4
9	915	297	19	216	554	80	5595	44	2658.7140	0.614	818	805.5	155.1
10	886	280	42	210	507	51	5541	37	12	0.601	780	771.2	77.2
11	872	284	35	135	428	52	5494	40		0.543	630	616.7	177.4
12	889	278	18	205	565	61	5453	45		0.607	797	787.2	95.1
13	883	283	29	155	427	67	5505	40		0.556	638	651.4	179.5
14	837	249	23	214	514	73	5472	39		0.581	721	719.8	1.5
15	830	296	33	235	647	61	5572	39		0.622	804	829.3	639.5
16	929	222	24	128	455	73	5488	31		0.529	589	579.6	88.0
17	904	252	24	218	537	58	5542	41		0.597	754	761.2	52.4
18	874	318	22	166	534	37	5526	38		0.585	738	729.4	74.8
19	813	265	34	170	566	73	5468	42		0.565	676	676.6	0.3
20	815	269	23	267	625	62	5515	59		0.626	851	837.4	184.2
21	838	322	20	227	550	76	5579	44		0.614	813	806.1	47.9
22	813	241	30	186	582	64	5424	32		0.571	677	691.1	198.5
23	896	290	38	157	474	59	5447	52		0.576	692	706.2	200.3
24	847	250	30	162	471	31	5486	36		0.544	617	619.6	6.6
25	938	256	32	176	430	70	5513	41		0.571	677	693.3	266.3
26	906	255	30	133	448	49	5541	42		0.534	603	593.9	82.0
27	907	248	9	205	525	80	5498	48		0.587	759	733.9	629.5
28	948	274	43	150	540	101	5475	50		0.588	716	736.9	435.5
29	824	266	24	194	555	88	5453	34		0.582	737	721.1	252.9
30	783	320	16	217	499	58	5477	37		0.595	709	757.3	2335.8
31	902	284	25	191	631	59	5517	40		0.608	771	791.7	427.5
32													
33												ESS=	8707.9

From Table 1, cells I2 … I7 are weights w_1 … w_6. Cell I8 is the Intercept for the ESS formula. Cell I9 is the Slope for the ESS Formula. Finally, Cell I10 is the sum of the weights which is set to equal 12 (as a normalization factor). The number 12 is not arbitrary, as 1B generally represents 1 base, 2B generally represents 2 bases, 3B generally represents 3 bases, HR generally represents 4 bases, BB generally represents 1 base, and HBP generally represents 1 base. Summing all of these gives us 12 bases.

The formula for Estimated (EST) Runs = Intercept + Slope * BOP

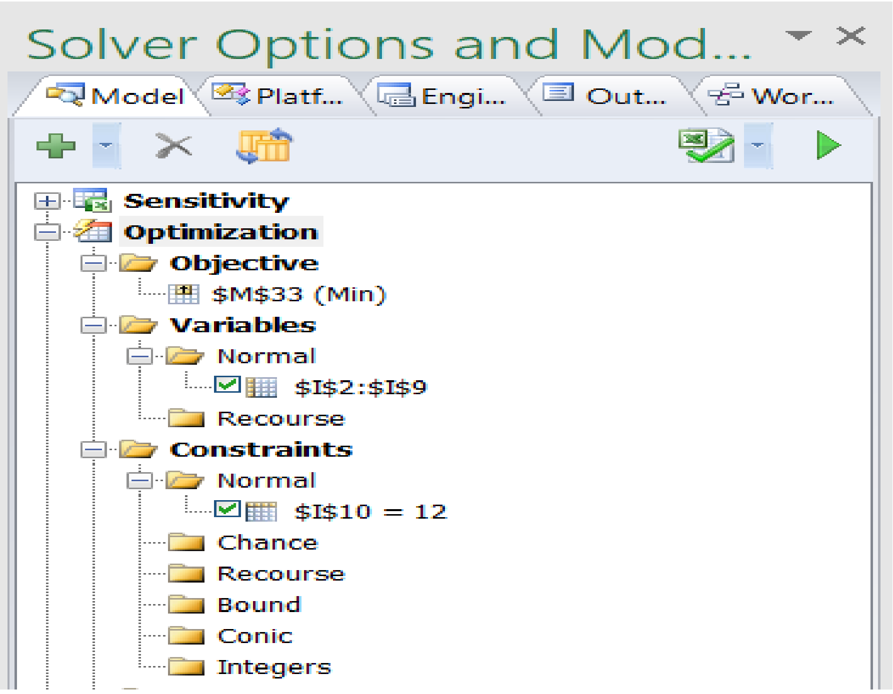

Figure 1: Inputs to the Optimization Model

From Figure 1, we see:

Objective Function: minimize M33 – the ESS.

Changing Variables: I2:I9 (I2:I7 -- w_1 … w_6; I8 – Intercept; I9 – Slope)

Constraints:

I10 (the sum of weights) = 12

We did similar modeling for years 2008-2017, and Table 2 below contains all of the BOP and R^2 values for 2008-2018.

Table 2: BOP Weights and R^2 Values

	w_1(1B)	w_2(2B)	w_3(3B)	w_4(HR)	w_5(BB)	w_6(HBP)	R^2 - OPS	R^2 – BOP
2018	1.4742	2.6082	1.7374	3.8543	1.4215	0.9045	0.9347	0.9473
2017	2.1905	2.3611	1.2029	3.7725	1.1828	1.2901	0.8729	0.9162
2016	1.3706	3.0068	2.5795	3.3240	0.5290	1.1900	0.8678	0.8920
2015	1.3990	3.7293	2.0122	4.4817	0.1817	0.1961	0.7770	0.8070
2014	1.4110	3.0310	3.7769	3.7759	0.9994	-0.9943	0.8189	0.8533
2013	1.6463	2.8609	2.1649	3.6910	0.8455	0.7913	0.8908	0.9032
2012	1.6116	2.1102	2.8708	4.3668	0.7524	0.2881	0.8966	0.9052
2011	1.4222	1.9224	3.7786	3.7543	1.0031	0.1195	0.9349	0.9476
2010	1.5128	1.6573	2.8071	3.8722	1.3966	0.7540	0.8905	0.9316
2009	1.6219	2.3347	2.5787	3.6508	0.4616	1.3524	0.9187	0.9269
2008	1.4602	2.1054	3.7007	3.4024	0.4253	0.9060	0.8906	0.9017
averages	1.5564	2.5207	2.6554	3.8133	0.8363	0.6180	0.8812	0.9029
normalized	1.0000	1.6196	1.7061	2.4501	0.5373	0.3971		

Obviously, the R^2 values for BOP exceed the R^2 values for OPS every year.

For the 2018 baseball season, 141 players had the requisite 502 plate appearances to qualify to lead statistical categories. We can break these 141 players into one of four categories based on their batting averages and slugging percentages. Please see Table 3:

Table 3: Four Categories of Hitters

	Batting Average < 0.300	Batting Average >= 0.300
Slugging Percentage < 0.500	Ordinary Players (74.47%)	Contact Hitters (7.09%)
Slugging Percentage >= 0.500	Sluggers (14.18%)	Superstars (4.26%)

Focusing on the averages from Table 2, we can see that slugging percentage undervalues singles (average BOP weight = 1.5564) and doubles (average BOP weight = 2.5207), yet it overvalues triples (average BOP weight = 2.6554) and home runs (average BOP weight = 3.8133). Furthermore, on base percentage overvalues bases on balls (average BOP weight = 0.8363) and times hit by pitch (average BOP weight = 0.6180). These facts, coupled with the other fact that general managers use OPS to measure the offensive value of a player, contribute to “sluggers” being overvalued whereas “contact hitters” are undervalued. This is extremely important because if general managers are not measuring the offensive abilities of players accurately, then this will lead to wrong decision making in terms of which players to pursue and how much to pay them.

Since “contact hitters” are generally undervalued, general managers should look for “bargains” for such players in a similar way to how Billy Beane got “bargains” for players with high on-base percentages back in 2002. This would give a general manager superior knowledge and a competitive advantage in pursuing these types of players. On the other hand, general managers should be cautious of overpaying “sluggers” because these players are generally overvalued. To illustrate this, please see Table 4 below for “a tale of 6 players.”

Table 4: Three “Contact Hitters” Versus Three “Sluggers”

Player	Batting Average	Slugging Percentage	Category	OPS	BOP	Salary
Joey Wendle	0.300	0.435	Contact Hitter	0.789	0.673	$545,000
Jose Martinez	0.305	0.457	Contact Hitter	0.821	0.673	$560,400
Whit Merrifield	0.304	0.438	Contact Hitter	0.806	0.664	$569,500
Giancarlo Stanton	0.266	0.509	Slugger	0.852	0.698	$25,000,000
Nelson Cruz	0.256	0.509	Slugger	0.850	0.688	$14,000,000
Matt Carpenter	0.257	0.523	Slugger	0.897	0.748	$13,500,000

The first three players – Joey Wendle, Jose Martinez, and Whit Merrifield -- were all “contact hitters.” The last three players – Giancarlo Stanton, Nelson Cruz, and Matt Carpenter – were all “sluggers.” Not surprisingly the “sluggers” have higher OPS values than the “contact hitters.” However, the BOPs of the “contact hitters” are almost as high as those for the “sluggers.” Also as expected, the salary of the “sluggers” is far higher than for the “contact hitters.” So, general managers overpaid the “sluggers” and underpaid the “contact hitters.” This speaks to an opportunity for general managers to use BOP to better evaluate players, find “bargains,” and pay players the “right” salaries.

Of the 141 players, 24 (17.0%) are valued correctly (same BOP ranking as OPS ranking), 57 (40.4%) are undervalued (better BOP ranking than OPS ranking), and 60 (42.6%) are overvalued (worse BOP ranking than OPS ranking). This means that the vast majority of players, 83.0%, are not ranked correctly and therefore not valued correctly. This means that there are significant opportunities for general managers to find “bargains” and avoid overpaying for players. Furthermore, virtually half of the players (49.6%) had a ranking difference of 3 or more, and nearly a third (31.9%) of the players had a rank difference of 5 or more. To see this information more specially, please see the Appendix for the top 25 players in terms of BOP for the 2018 season. We ranked these same 25 players in terms of OPS, and you will see that the rankings differences are not small.

Given the above discussion, we can see that general managers are not measuring the offensive abilities of players as accurately as they should be because they don’t have the best information on which to base their decisions. BOP provides general managers with the best information/measure to decide not only which players to pursue but also how much they should pay them. Every general manager is seeking to find the optimal group of players based on “the right measures.” By replacing OPS with BOP, general managers would have a better measure by which to accomplish the goal of maximizing the runs scored for their teams.

3. Conclusion

Now, baseball general managers can use a superior offensive statistic, BOP, as opposed to using OPS when evaluating the offensive value of players. Similar to how the 2002 Oakland A's exploited an "inefficiency in the market" where OBP was being undervalued, general managers who use BOP will be in a better position than their counterparts who use OPS. In future research, we could examine the right amount to pay players based on BOP. In valuing players "more correctly," BOP will allow general managers to maximize the production of the money that they spend on players, which should lead to more wins and more World Series championships!

References

[1] Baumer, B. S., Jensen, S. T. and Matthews, G. J. (2015). openWAR: An open source system for evaluating overall player performance in major league baseball, Journal of Quantitative Analysis in Sports, 11(2), 69-84.

[2] Hyle, R. R. (2012). Take Me Out To The Ballgame: Insurance And Baseball Analytics, National Underwriter / P&C, 115(12), 5.

[3] Merrifield, B. (2018). Taking lessons from the analytical champs, Supply House Times, 61(7), 50.

[4] Mirsky, S. (2014). Numbers Game, Scientific American, 310(4), 88. Newsweek Global, 2014, "Field of Streams, 163(5), 1-5.

[5] Shaw, S. K. (2014). The Sabermetric Revolution: Assessing the Growth of Analytics in Baseball, Library Journal, 139(3), 109

[6] Valerdi, R. (2017). What we can learn from baseball analytics, ISE: Industrial & Systems Engineering at Work, 49(8), 28.

Appendix: BOP Ranking Versus OPS Ranking for Top 25 Players

Player	BOP	OPS	BOP Rank	OPS Rank	Rank Difference
Trout, M	0.909	1.088	1	1	0
Betts, M	0.894	1.078	2	2	0
Martinez, J	0.844	1.031	3	3	0
Yelich, C	0.813	1.000	4	4	0
Ramirez, J	0.777	0.939	5	5	0
Bregman, A	0.772	0.926	6	7	1
Arenado, N	0.766	0.935	7	6	-1
Goldschmidt, P	0.758	0.922	8	8	0
Harper, B	0.749	0.889	9	16	7
Rendon, A	0.749	0.909	10	10	0
Carpenter, M	0.748	0.897	11	12	1
Machado, M	0.738	0.905	12	11	-1
Story, T	0.738	0.914	13	9	-4
Freeman, F	0.737	0.892	14	14	0
Nimmo, B	0.731	0.886	15	17	2
Aguilar, J	0.729	0.89	16	15	-1
Suarez, E	0.728	0.892	17	13	-4
Bogaerts, X	0.727	0.883	18	18	0
Lindor, F	0.711	0.871	19	21	2
Votto, J	0.708	0.837	20	35	15
Davis, K	0.707	0.874	21	20	-1
Haniger, M	0.706	0.859	22	25	3
Chapman, M	0.705	0.864	23	23	0
Hoskins, R	0.704	0.85	24	30	6
Baez, J	0.701	0.881	25	19	-6