James Ingram

May 8 2019

Does BB/K For Hitters Actually Matter?

Abstract:

In this paper, I will be analyzing Major League Baseball team statistics for walk-to-strikeout ratio for hitters (denoted BB/K), the runs that the team scores per game (denoted R/G), and the team's winning percentage denoted (W %), and the team's slugging percentage (SLG) to see if there is a correlation between BB/K and run production, BB/K and W %, and to see how BB/K and SLG have interacted over time. Overall, I have found in the analysis of these statistics featured below that BB/K has decreased over time while SLG has increased, but that neither of these facts have affected W % nor R/G. These findings may signify a stylistic change to the game of baseball, not affecting the amount of runs scored by teams or their ability to win.

Background:

Throughout my lifetime being around baseball, it has been understood by many around the game that if a team walks more and strikes out less, then that will lead to better offensive production and more wins in the long run. This notion intuitively makes sense, as strikeouts are statistically the least productive outcome possible of any given plate appearance that only results in one out, and that high walk totals will allow a team to reach base with more regularity, and thus score more runs. However, as sabermetrics and statistical analysis have become more

prevalent in baseball, many have speculated whether this notion really leads to better offensive production.

In the article “Special Ks” written by Joe Posnanski of NBC Sports, he highlights how strikeout totals have risen in the MLB, and how it should not be expected to go down any time soon. He cites the the rise in strikeouts is a “one way road”, meaning, “Good pitchers get more strikeouts than weaker pitchers”, and that, “Good hitters (historically) ALSO strikeout more than weaker hitters,” (Posnanski). Although Posnanski claims that better pitchers and better hitters are leading to increased strikeout totals, he doesn’t necessarily claim that this is a bad thing. In fact he claims that these inflated strikeout totals are due to increased competition between both the pitcher and hitter. Thus, a gradual decrease of the once revered BB/K may not be a sign that the game is in decline, but that the style of play is shifting.

Baseball has seen many changes that have affected competition, creating different “eras” of baseball. The first notable era would be the “Deadball Era”, in which foul balls first started being counted as strikes, baseballs were not swapped out until they got lost (and fresh balls go much farther than heavy, dirty ones), and pitchers were allowed to deface the ball to make it harder for the batter to hit it. There have been many such eras in baseball where the style of play drastically changed from previous decades, and it is very possible that the decline in strikeouts, as we be shown by the decrease in BB/K in my data, will show that the style of play in baseball is shifting once again. Using Matlab, I will analyze a data set of team statistics for every MLB team since the founding of the American League in 1901 to see if there is a correlation between a team’s BB/K and their winning percentage, and I will also be looking at how R/G, BB/K, and slugging percentage (SLG) has changed over time. Slugging percentage is the measure of total

bases a player collects per at bat, a single being one base, a double two, a triple three, and a homer run four. A higher slugging percentage does not necessarily indicate a player reaches base often, but that he does a lot of damage while at the plate. It will be interesting to see that if as BB/K decreases as anticipated, if SLG increases. This would signify a large change in the style of play across the league over time.

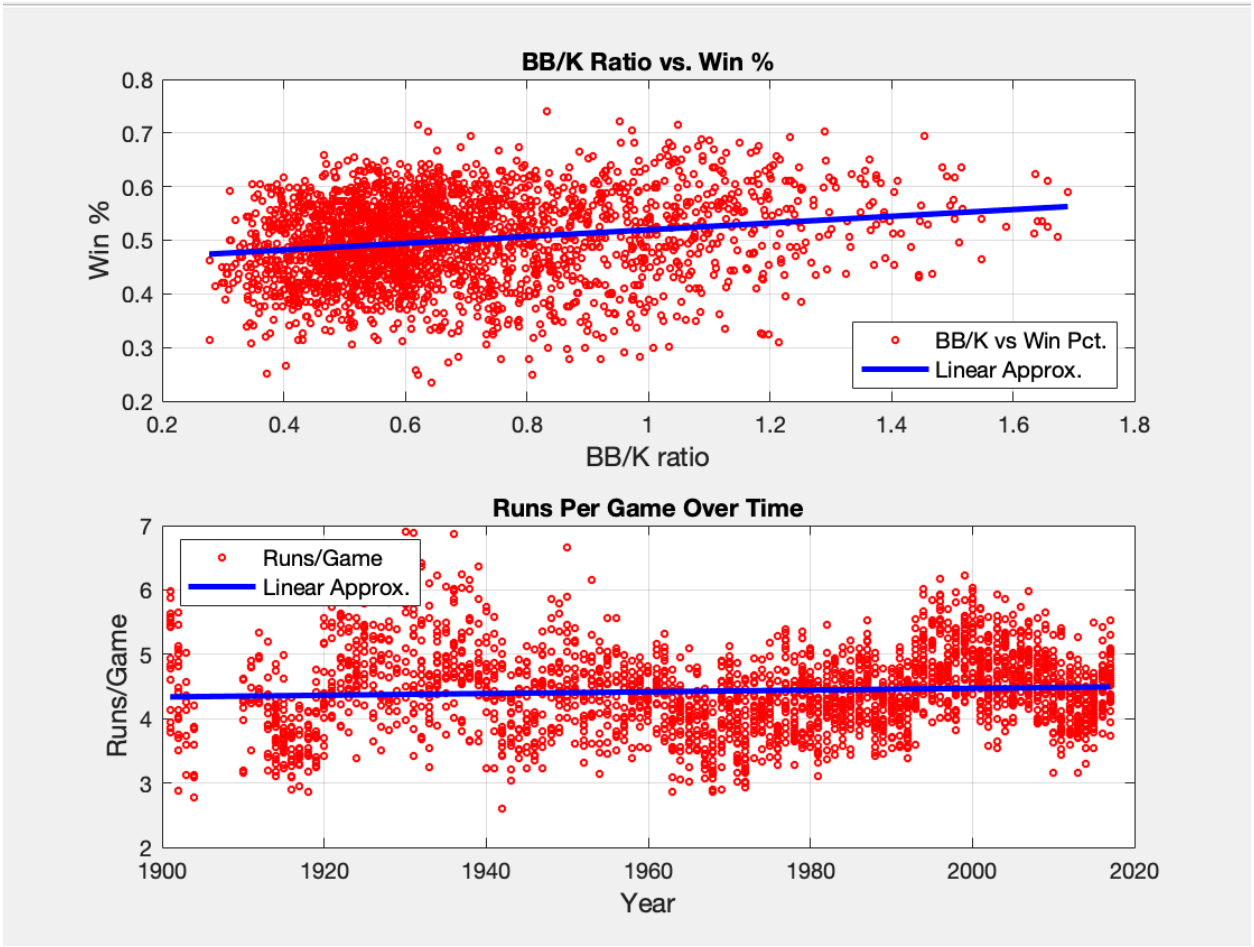
Materials and Methods:

My dataset for the MLB team statistics from 1901-2017 comes from Lahman's Baseball Database at SeanLahman.com. The comma-delimited dataset entitled Teams.csv, contained the teams statistics for ever Major League Baseball team from 1871-2017, including seasons from before the inaugural season of the American League in 1901. Using Terminal, I shortened the dataset to only include seasons since 1901 via the tail function in Unix, and I saved it in the file MLBTeams.csv.

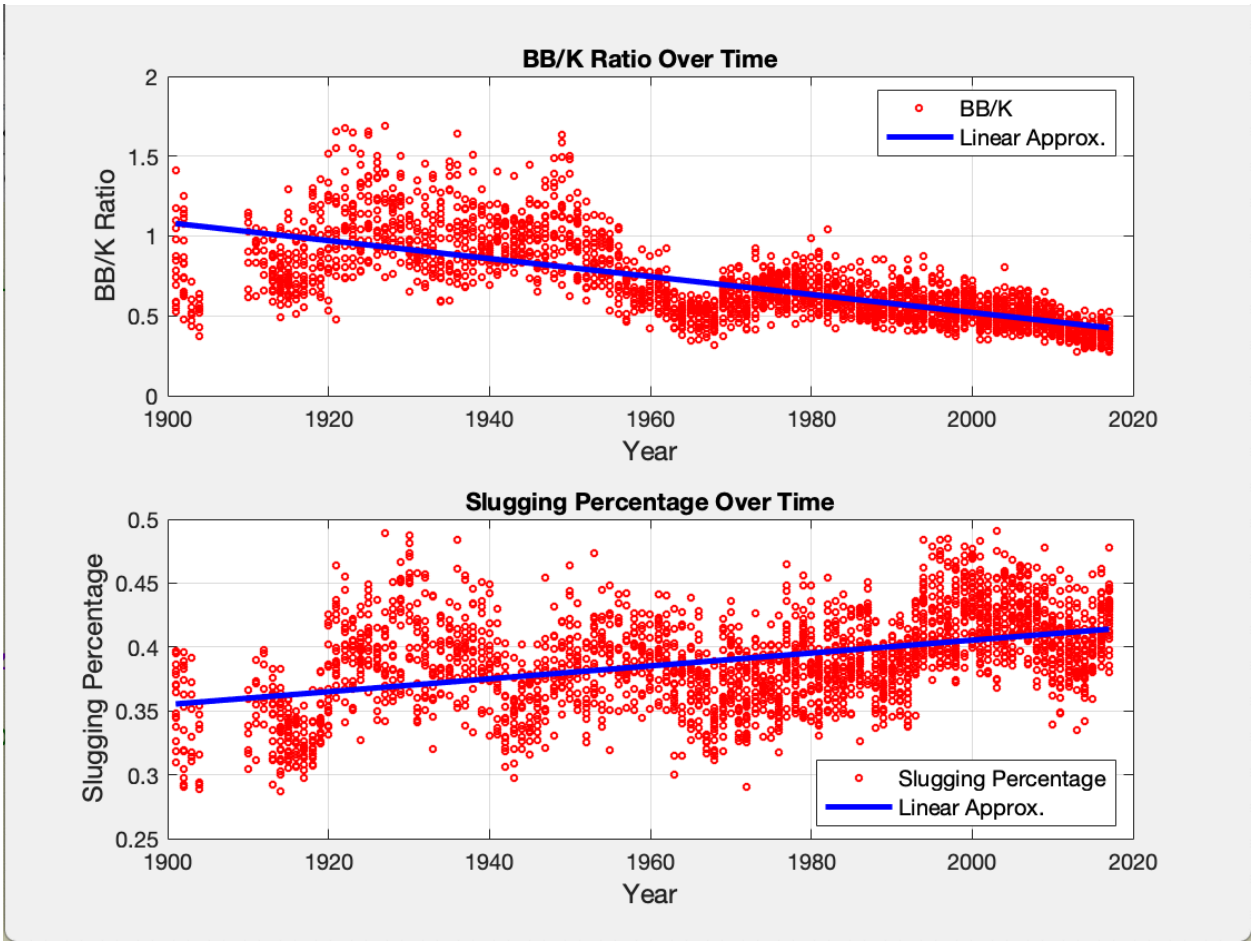
In Matlab, I then read MLBTeams.csv into Matlab, saving the data in the cell array "C". I then extracted data for each team's Wins (W), Losses (L), Year (year), Walks (BB), Strikeouts (K), Games Played (G), Runs (R), At Bats (AB), Hits (H), Doubles (Doubles), Triples (Triples), Home Runs (HR), and Singles (Singles). I then calculated some variables to be used in plots: Runs Per Game (R/G = R/G), Win Percentage (W_pct = W/(W+L)), BB per K (BB_per_K = BB/K), and Slugging Percentage (SLG = (Singles + 2*Doubles + 3*Triples + 4*Home Runs)/(AB)). There was a discrepancy in the dataset, and K were not counted for some teams, which returned NaN values for BB/K. To correct this, I made a logical vector "x" with all of the years with NaN values, and adjusted each variable using this logical vector.

I then made 2 figures with 2 subplots each. The first figure has BB/K vs Win % scatterplot and fit line, and R/G vs Time scatterplot and fit line. The second figure has BB/K over time scatterplot and fit line, and SLG over time scatterplot and fit line. All of the fit lines are to a line (first-degree polynomial).

Analysis:



The first subplot has BB/K on the x-axis and Win % on the y-axis. The first-degree polynomial (linear) fit is in blue (R^2 = .0395) and the data points are in red. The second subplot is Year on the x-axis and R/G on the y-axis, with a fit line in blue (R^2 = .0047) and the data points in red.



The first subplot is BB/K ratio vs the Year, with the data in red and the linear fit line in blue (R^2 = .5112). The second subplot is SLG vs Year, with the data in red and the fit line in blue (R^2 = .2009).

Discussion:

It is clear from analysis of the data that BB/K has decreased over time, as discussed in the background. Not only is the fit for BB/K the strongest of all the plots analyzed (R^2 = .5112), but if you will notice, after 1960, there was only one team to have a BB/K over 1 (meaning the team walked more than they struck out). This is significant compared to 1901-1960, in which there

were 315 teams that had a BB/K over 1. This signifies a drastic change in the style of play from that era to the current day, as teams now strikeout more and walk less than they did in earlier decades.

Moreover, another thing this is clear from the data is that the decrease in BB/K has not affected run production over time (R/G), nor has it affected a team's ability to win. The R/G over time model had the weakest fit out of all of the graphs (R^2 = .0047), meaning that there is no trend for how R/G in the MLB has changed over time. It is important to note though that the R/G data does appear to oscillate over time, having increased run production around 1930, a decrease around 1970, and another increase around 2000. There is also a very weak correlation between BB/K and Win % (R^2 = .0395), meaning that there is not data here to support that BB/K greatly affects Win %.

Another interesting takeaway from the analysis is that SLG has increased over time as BB/K has decreased. To clarify, SLG (slugging percentage), is the amount of bases that a team collects per at bat, with one base being a single and 4 bases being a homerun. A higher SLG for a team equates to a team that hits more extra-base hits (hits that garner more bases than a single), and with a moderately strong correlation (R^2 = .2009), it appears that teams are slugging better now than they did in the past. It's possible when discussing the style of play from yesteryear to now that teams are focusing more hitting more extra-base hits, and sacrificing their BB/K as a result. The fact that there was no trend found for R/G over time, this change in the style of play in which teams slug more and sacrifice their BB/K is merely a stylistic change in the game as there is no significant data to show that either are better for run production.

Conclusion:

From the data set analyzed, which featured the team statistics from 1901-2017 (omitting some teams in which data was not available), it was found that there is no correlation between BB/K and the ability of a team to win or produce runs. However, SLG has increased as BB/K has decreased over time, which indicates that there has been a change to the style of play as time has passed, in which teams nowadays try to slug more, and thus sacrifice their BB/K. This analysis does leave many questions and new avenues of research, as it would not make much sense for teams to change the way the play the game this drastically for no apparent reason. Research should considered into why teams have sacrificed their BB/K to slug more, and into if a return to that early style of play with higher BB/K would help or hurt offensive production.

References:

"Special Ks." *NBC SportsWorld*, 19 Apr. 2016, sportsworld.nbcsports.com/baseball-strikeouts-more-than-ever/.

http://www.seanlahman.com/baseball-archive/statistics/