What is the Correlation between Pitcher Spin Rates and other Peripheral Baseball Statistics?

Michael Avlonitis

Texas A&M University

What is the Correlation between Pitcher Spin Rates and other Peripheral Baseball Statistics?

Ever since the turn of the century, the world of baseball has gone through a transformation. The game has turned to find new innovations in order to find the diamond in the rough[[1]](#footnote-1). This change in culture is made possible through now widely popular machine learning.[[2]](#footnote-2) Machine Learning is the science of getting computers to learn and act like humans do, and improve their learning over time in autonomous fashion, by feeding them data and information in the form of observations and real-world interaction[[3]](#footnote-3). In an age where many teams are trying to be competitive, but some owners provide teams with more funding than others, Major League Baseball (MLB) Front Offices have allowed experts in math and statistics to develop new ways to identify talent, giving them a higher probability for success. Free public databases, such as baseballsavant.com or baseball-reference.com, allow anyone to access some of the most advanced or new statistics that are driving decision making in MLB. It is important that Baseball data analysis focuses on outcomes that are directly related on things that specific player can control in order to get a true sense of his ability.In this study, I am going to be doing analysis on spin rate, in order to determine what kind of relationship it has with Earned Run Average, and the quality of contact opponents make.

**Earned Run Average**

Naturally, one would think that in order to have a higher chance at winning games, a team would want as many players as possible with low ERAs. However, this statistic isn’t quite indicative of a pitcher’s performance, because it takes into account many extraneous variables that are out of the control of the pitcher. This is because the pitcher isn’t necessarily solely responsible for how many runs he allows. As a result of the ERA formula being so simple, it does not completely cover the scope of the game. Ballpark dimensions, fielder ability, and weather can all contribute to the outcome of the baseball game, shouldn’t be used to dictate the skill of a pitcher. I’m arguing that since ERA has many variables that go into it, perhaps baseball analysts should take a deeper look at spin rate as a measure of a pitcher’s effectiveness. This paper will utilize spin rate as an alternate way to evaluate pitcher performance throughout the season.

**Spin Rate**

Spin rate is a variable that describes the speed of pitch spin in units of revolutions per minute (rpm). Spin rate is a variable that the pitcher has full control of. He doesn’t have to rely on other people to do their job in order to have a good spin rate. Spin is important in that it’s what allows the ball to move in the air and deceive the batter. A specific example of the impact that spin rate has on a pitcher’s effectiveness is Trevor Bauer. In August of 2019, Bauer’s fastball had about 2300 rpm. Then suddenly in September, it jumped up to 2700. The following season it was still around this mark, and even though his pitches were about .5 MPH slower, he was putting up his best numbers of his career[[4]](#footnote-4). In 2020, Trevor Bauer won the National League Cy Young Award, and he has made it publicly known that he attributes his success to his rise in spin rate.

An analysis on spin rate is important because it could allow teams to find pitchers who may be perceived as flawed due to a high ERA. There is no salary cap in Major League Baseball (MLB), which leads to a disparity in spending among teams. This forces teams who don’t spend as much money to look for players who may not cost as much money on the open market. For example, Player A may have a high ERA, but he pitched on a team that fielded a poor defense behind him. This is certainly something that would inflate his ERA compared to someone who had a good defense behind him. This isn’t something that he could control, and therefore he shouldn’t be penalized. When Team B does a deeper dive in his statistics, they see that his spin rate was in the 90th percentile of the all MLB pitchers. This could imply that he has the ability to make hitters swing and miss, and if he is aided by Team B’s elite defense, he could eventually be considered by the media as a reclamation project.

**Current Study**

I am going to be doing analysis on spin rate, in order to determine what kind of relationship it has with Earned Run Average, and the quality of contact opponents make.

This study will conduct data analysis on pitchers in 2020 in an attempt to explain how a pitcher’s ERA may not be indicative of his overall performance due to the fact that the simple ERA formula doesn’t explain the myriad of factors that goes into pitcher success. Pitcher spin rate is effective because it will explain some of the phenomena like the one that Trevor Bauer had last year. The ultimate goal is to see how does spin rate relate to these key outcomes, and to show how inducing poor contact due to spin rate will translate to a more effective pitcher with a lower ERA. I predict that variables such as strikeout percentage, whiff percentage, and swing percentage will be highly correlated with spin rate, which will be a key indicator for an effective pitcher. In order to determine the degree to which pitcher spin rate predicts ERA, series of statistical tests will be conducted to address the following research questions.

1)What is the relationship between spin rate and Earned Run Average?

2)How does a players average spin rate correspond with the batters’ quality of contact?

# Methodology

## **Data and Procedures**

In order to conduct this research, a custom data set was created using a baseball database called Baseball Savant[[5]](#footnote-5) . The data was downloaded as a .csv file and uploaded to R version 1.2.5033 for data analysis. This dataset included all 469 MLB pitchers who threw at least one pitch in the 2020 season. Baseball Savant gets the data from high resolution optical cameras with radar tracking in order to measure movements and locations of not only the ball as it moves, but also players. These cameras are set up across all 30 MLB stadiums uniformly so that this data is readily available for everyone to use. All data methods and graphs will be conducted using R packages base and Hmisc.

## Instrumentation

**Name**

The first two variables are self-explanatory, they are the players first and last name. Each player will have all of their subsequent statistics listed in the row that follows their first and last name. The next column states that all the statistics in this dataset pertains to the year 2020.

**Strikeout Percentage**

This is how many batters he strikes out divided by how many batters he faces. This is important to show that some people strike out batters more effectively, and the spin on their pitches may be the cause of that.

**Earned Run Average**

This statistic explains how many runs a pitcher is predicted to allow if he were to pitch all 9 innings of a game. This is the most popular statistics to define a pitcher’s effectiveness.

**Exit Velocity and Barrel Percentage**

This is how fast the ball was going when it came off the bat in miles per hour, and barrel percentage is how often the batter hit the ball. That is, the batter hit the ball with an exit velocity of at least 98 MPH, and a launch angle between 26-30 degrees. A good barrel percentage for a batter starts at 10%.

**Whiff Percentage**

This is a proportion of how often the opposing batter swings and misses at their pitches. This should have a high correlation with strikeout percentage, as it directly relates to a pitcher’s ability to fool the batter.

**Swing Percentage**

This is how often the opposing batters swing at the pitcher’s pitches. This variable will relate with the whiff percentage and give us an idea of how well opposing batters are able to make contact with their pitches.

**Spin Rate**

The next three variables are the pitcher’s average spin rate converted to revolutions per second. The spin rates are divided into three different pitch categories: Fastballs, Breaking Balls, and Off-speed Pitches. A fastball can be classified as a 2 seam, 4 seam, cutter, or sinker. A breaking ball is either a curveball or slider. Finally, an off-speed pitch is a change-up. Some of the boxes for that particular player may be empty, and this is because they don’t throw that specific type of pitch. These are the variables that I am going to use to answer the above research questions.

**Plan of Analysis**

In order to address the first research question (What is the relationship between spin rate and Earned Run Average?), a regression analysis will be performed with strikeout %, whiff %, barrel%, swing%, exit velocity, and the three different types of spin regressed on ERA**.** Next, the second research question will be addressed (How does a players average spin rate correspond with the batters’ quality of contact?). I am going to find the correlation between pitcher ERA and other peripheral statistics like spin rate on different pitch types, swing and miss percentage, and the quality of contact that a batter makes. I will be doing regression analysis and graphing some of my results.

Results

As an introduction to the data, summary statistics are listed in Table 1 with regards to the nine variables listed above:

**Table 1 (Summary Statistics)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | *M* | *SD* | *Skew[X]* | *Kurt[X]* |
| Strikeout % | 23.95 | 6.94 | 0.6 | 0.63 |
| Earned Run Average | 4.45 | 2.39 | 1.6 | 5.76 |
| Exit Velocity (MPH) | 88.4 | 2.16 | -0.16 | 0.14 |
| Barrel % | 7.46 | 3.53 | 0.71 | 0.77 |
| Whiff % | 26.83 | 6.17 | 0.52 | 0.64 |
| Swing % | 46.16 | 3.82 | 0.16 | 0.67 |
| Fastball Spin Rate | 2269.66 | 168.39 | 0.005 | 0.1 |
| Breaking Ball Spin Rate | 2449.38 | 248.11 | -3.51 | 20.48 |
| Offspeed Spin Rate | 1744.95 | 271.19 | -1.29 | 0.4 |

Table 1 includes the mean, standard deviation, skewness, and kurtosis for the data representing the 2020 MLB season. This will allow for discussions on the validity of the data. The only variable that can be considered highly skewed is ERA (the response variable). A deeper look into its distribution would give tell us why that is.

**Figure 1 (Distribution of ERA)**

Chart, histogram

Description automatically generated

# As listed above the mean Earned Run Average is 4.45. This is illustrated in the graph, not surprisingly, there are players who have an ERA>10. These people are driving up the mean and creating a distribution that appears to be right skewed. This is normal during the course of an MLB season, and their data is still of importance.

**Figure 2 (Correlation Matrix)**

A correlation matrix will allow for comparisons across different variables. Table

Description automatically generated with low confidence

Here, there are notable moderately strong correlations. As stated above, the spin rate variables were transformed to revolutions per second, instead of revolutions per minute. Some notable correlations include strikeout percentage with ERA fastball spin rate, and off-speed spin rate, fastball spin rate with whiff percentage, along with ERA and the two measures of quality of opponent contact. An important note is there isn’t many notable correlations with breaking ball spin rate. In the summary statistics above there is a large skewness and kurtosis value for breaking ball spin rate. After doing an investigation into the density curve, there is are 7 who simply don’t have a breaking pitch. These include curveballs, sliders, and others. This coupled with the high spin rate of players who do have these pitches causes high skew values and low correlation. There are even more players who don’t have an off-speed pitch. However, this isn’t as big of an issue because of the low spin rate of the players who do have an off-speed pitch. Removing all 80 players who don’t have all three pitch types would be an issue, because it wouldn’t be indicative of the entire MLB.

## **What is the relationship between spin rate and Earned Run Average?**

In order to do a deeper dive into the correlations between some of the moderately correlated variables, linear regression models will be fitted for the data.

**Table 2 (Linear Regression with ERA as the Criterion)**

## Table Description automatically generated

The two highly significant P-values here are Strikeout Percentage and Barrels. For the 2020 data, a 15% rise in Strikeout Percentage leads to the ERA going down by 1, say from 4.5 to 3.5. Similarly, a 27% rise in Opposing Barrel Percentage leads to the ERA going up by 1. As mentioned above, Barrel% is a metric of how well the opposing batter hit the ball well. When the batter gets a barrel, these are the instances most likely to be home runs. When a pitcher has a high barrel percentage, he clearly isn’t deceiving the batter with either his movement, location, or both. These two values are important because they directly pertain to both of our research questions. Although spin rate was not a significant predictor of ERA, it may be important to note that spin rate may be indicative of a high strikeout percentage, or a low barrel percentage. Linear regression models will now be run for both, with strikeout percentage and barrel percentage as the response variables respectively.

## **How does a players average spin rate correspond with the batters’ quality of contact?**

**Table 3 (Linear Regression with Strikeout Percentage as the Criterion)**

Table

Description automatically generated

Here, the output says that all three different pitch types are significantly correlated with Strikeout Percentage. This indicates that if a pitcher is able to attain a certain spin rate, he will be more likely to strike batters out. According to the output, a 1.04 rise in revolutions per second (RPS) will lead to a 1% rise in strikeout percentage. The opposite is seen with both breaking pitches and off-speed pitches. A -.14 decrease in (RPS) will lead to the strikeout percentage going up by one, and a -.13 decrease in RPS will lead to the strikeout percentage going up by one, respectively. According to the correlation matrix, there is a strong (r=.80) relationship between strikeout percentage and whiff percentage. Similarly, a regression output with whiff rate as the response variable will show significant p-values with all three different types of spin5.

**Table 4 (Linear Regression with Barrels as the Criterion)**

# Table Description automatically generated

Next, we have the regression output with Barrel Percentage as the response variable. Here, there isn’t quite any correlation between barrel percentage and the three different pitch types’ spin rates. This could indicate that there isn’t a direct correlation between spin rate and barrel percentage, however it is possible there is a relationship with other variables not listed in the study. A deeper analysis would have to be conducted.

# Discussion

# It was surprising at first to see that there was no significant relationship between spin rate and ERA. However, there was an interesting indirect relationship. In mathematical logic, this can be referred to as transitivity. There is a moderate relationship between ERA and strikeout percentage, mainly because of the numerous variables that create a player’s overall ERA. We then see that there is a significant relationship between Strikeout Percentage, Whiff Percentage and each spin rate variable.

In the regression output with Strikeout Percentage as the response, we see a different relationship between fastball spin, off-speed spin, and breaking spin. This result is explained in data analysis done by a company called Rapsodo. Rapsodo is a leader in measuring baseball and golf data by using smaller personal cameras in order to track different metrics related to the respective sports. They say that for pitchers to have an ideal spin rate range, they want to have a high spin rate with high spin efficiency on their fastballs, and lower spin rate with varying spin efficiency on their slower pitches. In order for a pitcher to be successful in the middle ranges, their pitching release needs to be unique in order to add another level of deception[[6]](#footnote-6). A low RPM pitch that is down in the zone is ideal, as well as a high RPM pitch up in the zone. This is because of how to spin affects the balls interaction with gravity[[7]](#footnote-7). In order for a fastball to have as little pressure from gravity as possible, it has to be able to spin as quickly as possible in order to create the most effective vertical movement. Inversely, a lower RPM pitch would be able to allow gravity to force it to drop down in the zone. If an off-speed pitch has higher or even middle level RPM, it starts to flatten out, and that creates an easy to read, slow, flat pitch that a hitter is likely to barrel up.

**Limitations**

There are some notable limitations within this study. First and foremost, a similar study that encompasses data over multiple years could both increase the validity and show trends. Also, there are many people in the 2020 dataset that are missing certain pitch types, which increases the skewness of the distribution. Without these values, there could be even higher correlations or lower P-values.

**Conclusion**

This study is valuable for various reasons. First of all, as mentioned before, it can be a way to judge pitcher success based of individual factors that pitcher’s can control. ERA has long been considered flawed[[8]](#footnote-8). It is important that new statistics are introduced in order to have the highest probability of winning. Player development personnel and coaches can look at spin rate data in order to help out players at youth and high school levels. This will help them develop themselves into pitchers with ideal spin tendencies. Similarly, we can compare pitchers without certain pitch types, as in the data set, and see how their stats would compare with players who do have those pitch types.

In reference to Trevor Bauer, as mentioned above, from 2019-2020, he saw his barrel percentage go down by 2.3%, his strikeout percentage went up by an astonishing 10.2%, and his ERA went down by 2.19. This is a huge jump, and according to this study, it can partially by explained by him being in the 100th Percentile for Fastball spin and the 92nd Percentile for curve spin. All while being in only the 24th Percentile for velocity of pitches thrown. Everyone is dazzled by throwing the ball 100 MPH+, however pitchers like Bauer prove that velocity isn’t the driving factor in pitching success, its spin rate.

References

Daniels, Seth, and Brian Page. “Rapsodo Pitching Certification .” Lecture, April 11, 2021.

Faggella, Daniel. “What Is Machine Learning? - An Informed Definition.” Emerj. Emerj, February 26, 2020. https://emerj.com/ai-glossary-terms/what-is-machine-learning/.

Howell, Ben. “Modeling Plate Discipline from the KBO to MLB.” *SABR Virtual Analytics Conference*. Lecture, March 14, 2021.

Ishii, Tatsuya. “Using Machine Learning Algorithms to Identify Undervalued ...” Accessed June 7, 2021. http://cs229.stanford.edu/proj2016/report/Ishii-UsingMachineLearningAlgorithmsToIdentifyUndervaluedBaseballPlayers-report.pdf.

Koseler, Kaan, and Matthew Stephan. “Machine Learning Applications in Baseball: A Systematic ...” Accessed June 7, 2021. http://www.users.miamioh.edu/stephamd/papers/Baseball\_Machine\_Learning.pdf.

Lewis, Michael. *Moneyball: the Art of Winning an Unfair Game*. New York: W.W. Norton, 2013.

Marks, Isaac. “Earned Run Average Is a Bad Statistic and Should Be Retired.” Medium. RO Baseball, August 16, 2016. https://medium.com/ro-baseball/earned-run-average-is-a-bad-statistic-and-should-be-retired-97e4c9feaf5f.

“MLB Stats, Scores, History, & Records.” Baseball. Accessed June 7, 2021. https://www.baseball-reference.com/.

Moore, Jack. “The Most Convoluted Statistic: ERA.” FanGraphs Baseball. Accessed June 7, 2021. https://blogs.fangraphs.com/the-most-convoluted-statistic-era/.

Nagami, Tomoyuki, Takatoshi Higuchi, and Kazuyuki Kanosue. “How Baseball Spin Influences the Performance of a Pitcher.” *The Journal of Physical Fitness and Sports Medicine* 2, no. 1 (2013): 63–68. https://doi.org/10.7600/jpfsm.2.63.

Rowan , Nate. “Whiff Effect: Are Pitchers More Likely to Repeat the Previous Pitch After A Swing-And-Miss? .” *SABR Virtual Analytics Conference*. Lecture, March 14, 2021.

“Statcast Custom Leaderboards.” baseballsavant.com. Accessed June 7, 2021. https://baseballsavant.mlb.com/leaderboard/custom?year=2020&type=pitcher&filter=&sort=1&sortDir=asc&min=50&selections=p\_k\_percent%2Cp\_era%2Cexit\_velocity\_avg%2Cbarrel\_batted\_rate%2Cwhiff\_percent%2Cswing\_percent%2Cfastball\_avg\_spin%2Cbreaking\_avg\_spin%2Coffspeed\_avg\_spin%2C&chart=false&x=p\_k\_percent&y=p\_k\_percent&r=no&chartType=beeswarm.

“Trevor Bauer Should Just Admit It.” Defector, September 24, 2020. https://defector.com/trevor-bauer-should-just-admit-it/.

“What Is Gold Glove Award In Baseball? Definition & Meaning.” SportsLingo, November 13, 2017. https://www.sportslingo.com/sports-glossary/g/gold-glove-award/.

**Table 5 (Linear Regression with Whiff as the Criterion)** Table

Description automatically generated

Linear regression output with whiff rate as the response

1. Michael Lewis, *Moneyball*(New York: W.W. Norton & Co, 2003) [↑](#footnote-ref-1)
2. Koseler, and Stephan, Machine Learning Applications in Baseball: A Systematic Literature Review [↑](#footnote-ref-2)
3. Emerj, “What is Machine Learning? – An Informed Definition” [↑](#footnote-ref-3)
4. Defector, “Trevor Bauer Should Just Admit It” [↑](#footnote-ref-4)
5. Baseball Savant, “Statcast Custom Leaderboards” [↑](#footnote-ref-5)
6. Page, Daniels, “Rapsodo Pitching Certification” [↑](#footnote-ref-6)
7. Tomoyuki, Higuchi, Kanosue, “How Baseball Spin Influences the Performance of a Pitcher” [↑](#footnote-ref-7)
8. Marks, “Earned Run Average Is a Bad Statistics and Should be Retired” [↑](#footnote-ref-8)