# An Analysis of Defensive Shifts in the MLB, 2016-2019

By Alex Veroulis and Haley Rao

Bruin Sports Analytics, 2020

## Contents



## Abstract

In the ever-changing arms race of information in today's MLB, one phenomenon integral to the modern game is the defensive shift, an alignment that aims to position defenders in the most frequent hitting areas of their opponents. Although this strategy was unheard of even 10 years ago, it has been quite effective in getting batters out in recent seasons. However, teams have begun to adjust offensively against the shift over the past few years, which may have diminished the initial effectiveness of the defensive strategy. To examine this effectiveness, we analyzed team defensive statistics over the past four seasons to see which variables were most correlated to shift usage using exhaustive stepwise regression; based on

the chosen variables, we created a shift effectiveness metric and compared team averages across the four seasons. In addition to analyzing team strategy, we compared hitters' performances during plate appearances without the shift and plate appearances with the shift to see how they responded to the defensive strategy. We identified important hitting variables that changed with a shift in play using random forest and logistic regression models that classified observations in terms of whether or not they were plate appearances with the shift. Through further examination of why some hitters beat the shift while other hitters suffer, we concluded that teams should employ a defensive shift based on a hitter's skill rather than their tendency to hit the ball in a certain direction. Future studies could examine how the shift has impacted the game over a longer course of time, as we would have more team and hitter data to work with.

# Introduction

The shift is a relatively new development in baseball, as it has only been around for a few seasons. Specifically, the shift is a defensive alignment where players are strategically placed in fielding positions where hitters most frequently put the ball in play. Today it is used to turn ground balls and line drives that otherwise would have been base hits into outs. Historically, the shift dates back to the 1940s, when player-manager Lou Boudreau told his players to change their infield configuration against Ted Williams, an incredible batter who once finished the season with a .406 batting average in 1941 (the last time anyone has finished above .400). Manager Joe Maddon is considered one of the pioneers of the modern infield shift, reviving the strategy as early as the mid-2000s (Paine). At first, the shift was only used against hitters who would frequently pull the ball to the other side of the field: right-handed hitters who would often hit the ball to the left side of the field, and left-handed hitters who would 'pull' the ball to the right side of the field. These players were only a small fraction of batters; in recent years, however, some teams have begun employing the shift on almost half their at-bats (Baseball Savant). What used to be an uncommon strategy is now almost an integral part of the game. With an increased frequency in its use, is the shift still as effective as it originally was?

Baseball Info Solutions has previously attempted to quantify shift effectiveness by creating their own statistic: SRS, or shift runs saved. A high SRS indicates that a team's use of the shift is saving them from runs scored by their opponent. They observed that teams that shift the most actually have some of the lowest SRS values, pointing to the idea that the overuse of the defensive strategy has decreased its benefits (Miller). The use of the shift was also previously considered in an ESPN article that used the difference between hitting statistics like BABIP and their expected values to evaluate the shift's effectiveness in the long run. A lower difference between the observed batting statistic and its expected value is somewhat correlated with a decrease in shift usage, suggesting the two may be related (Miller). Furthermore, a collaboration between Major League Baseball and Amazon Web Services conducted a similar analysis as they looked at fielding configurations to create a shift impact metric. To create the metric, they employed a model that predicted the probability of a base hit and then subtracted expected batting average from the odds (Karimi). These pieces of literature inspired our study of the shift, but failed to categorically explain the overall effectiveness of the shift as a defensive strategy. Our research is geared towards determining whether the shift is as effective as it is perceived to be, especially when it is used so frequently. We took a look at a more comprehensive set of variables than most other research to examine the shift and we came up with metrics of our own that are unique to this paper, such as the shift effectiveness metric.

There was also a glaring lack of analysis that looked at how hitters specifically respond to the shift. We wanted to examine whether hitters were changing their approach at the plate when met with an infield shift by looking at a multitude of batting statistics including (but not limited to) batting average on balls in play (BABIP), slugging percentage (SLG), weighted on base average (wOBA), and launch angle. If these statistics changed significantly during plate appearances with a shift in play, it could give us some insight into what players are doing to overcome the different infield configuration. It is also apparent throughout Major League Baseball's recent history that some players consistently perform well against the shift and 'beat' it, while others continually suffer at the plate when hitting against the shift. We wanted to examine the reasons why some hitters perform better than they usually do against the shift, while other hitters perform worse than normal when the shift is in play.

## Methods

To start, we acquired data from a number of sources, including Fangraphs, Baseball Reference, and Baseball Savant to conduct our analysis. We used Fangraphs data for each year from 2016-2019, which included pitching statistics for each team such as strikeouts per nine innings, batting average on balls in play (BABIP), earned run average (ERA), and wins above replacement (WAR), among others. As for Baseball Savant, we found data on this site to analyze the frequencies of teams shifting. Specifically, we obtained the proportions that teams shifted against left-handed hitters, against right-handed hitters, and overall from 2016-2019. This same site included more team shift data, including the amount of balls in play allowed, as well as the expected and actual values of key statistics like batting average, slugging percentage, and weighted on-base average allowed against opposing hitters. We then created new variables for the differences between the actual and expected values of these variables for consolidation purposes. Additionally, from Baseball Savant, we obtained player batting statistics from 2016-2019 for both instances when there was a shift in play and when there was not. We limited the data to players who had 50 plate appearances in both categories. The data from both Fangraphs and Baseball Savant were exported as csv files and then imported into R for analysis.

To clean the Fangraphs data, we began by removing the actual and expected fielding independent pitching statistics (FIP and FIP%, respectively), as these do not consider shifts, making them irrelevant to our research. We also removed the Losses variable and the Games Started variable, as these were linearly dependent with some other variables. The Baseball Savant and Fangraphs datasets were combined after each was arranged by team name alphabetically, so that the rows of the columns would match. We then standardized the predictors by taking each observation in each column, subtracting the mean of the column, and subsequently dividing by the standard deviation of the column. This standardization allowed for each variable to be equally considered in the model. For the player-level Baseball Savant data, the datasets from each year were combined into one and a year variable was added to indicate the specific season. The shift dataset and non-shift dataset were also combined and a shift indicator variable was created; a 1 indicates the observations are player batting statistics with a shift in play and a 0 indicates the player batting statistics when a shift was not in play. Extraneous variables were removed, including the distance each fielder was from the plate, total pitches, and pitch percent. The final player data set consisted of 2500 observations and 16 variables.

To examine the player data, we decided to create classification models to predict whether there was a shift or not based on the batter's hitting statistics. This would allow us to clearly see which metrics change the most when a shift is in play. Before creating the models, we performed synthetic data generation using the ROSE package in R, which uses a smoothed bootstrap approach to generate a synthetic data set that is the same size as the original data set but with more balanced classification. In the original data set, there were 1888 observations without a shift and only 612 observations with a shift; after we oversampled from the shift observations, we ended up with a data set that had 1312 observations without a shift and 1188 observations with a shift. We first performed logistic regression in R to predict whether there was a shift or not based on the player's batting statistics. We also created a random forest model using the randomForest package in R. After using these two models to determine which batting statistics changed the most against a shift, we pivoted to examining why some players beat the shift while others perform consistently worse when the defensive strategy is employed. Since previous research used BABIP as an indication of player success against the shift, we used the difference in BABIP when there was a shift and when there was not to separate the players into three factions: those who perform better, about the same, and worse against the shift. Next, we performed many two sample t-tests to investigate the difference in batting statistics between the three groups to see why some players' BABIP increases against the shift while others' decreases.

# Results

For some context on shift usage in the past few years, we decided to take a look at general trends over the past few seasons in the MLB, from 2016-2019. With respect to the visuals below, Figure 1 represents the league-wide shift proportion averages in three categories: the total proportion of shifts, the proportion of shifts against left-handed batters, and the proportion of shifts against right-handed batters. We notice that, while there was a dip in shift usage from 2016 to 2017, the shift has become increasingly popular with steady increases in both 2018 and 2019. Also, shift usage against both lefties and righties follows the overall proportion of shifts, as we would expect. As for Figure 2, we created actual minus expected statistics for each year, meaning that we took an actual statistic (like batting average) and subtract its expected version (expected batting average) from it to get a difference. We plotted differences for batting average, slugging percentage, and weighted on base average for 2016 to 2019. Evidently, while batting average difference remained fairly constant, the slugging percentage and weighted on-base average differences seemed to follow the trend lines for shift usage in the other visual. This would suggest that these offensive statistics outperform their expected values when shift usage increases, which hints at the phenomenon that shifts may not be as effective as many experts would think.

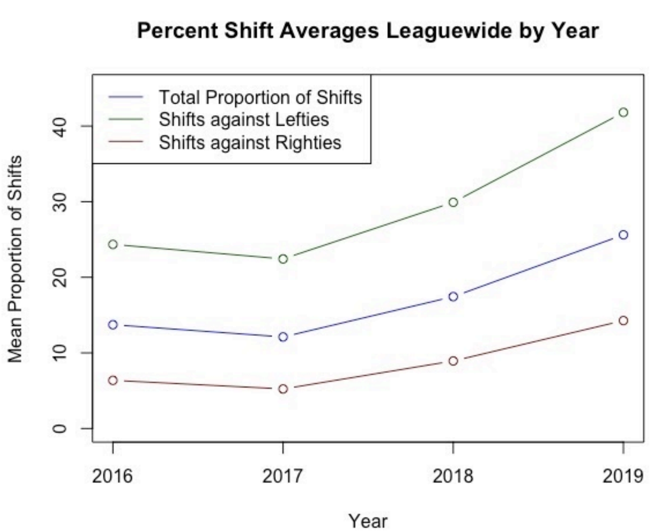


*Figure 1*

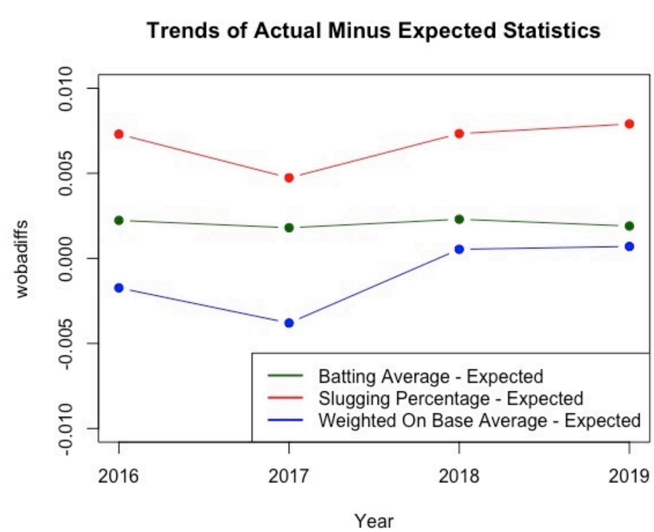


*Figure 2*

On a more granular level, we compared the shift rates of teams in 2018 and 2019 and found that even a year's time could lead to significant changes in how teams strategize defensively. In the histograms below, Figure 3 represents the distribution of shift rates for the MLB in 2018, while Figure 4 shows the distribution for 2019. The red lines are the league-wide averages for shift percentage, and the blue lines represent the maximum shift frequency of all teams in the league. Clearly, teams in 2019 shifted much more often than in 2018, with an 8% increase in the average team's shift usage. Also, the 2019 Los Angeles Dodgers shifted the most that year at an astronomical 50.6% clip, which was quite higher than the highest proportion for 2018 (37.6% for the Houston Astros). Lastly, only a few teams in 2018 had a shift rate above 30%, but about a third of the league had shift percentages at least that high in 2019.

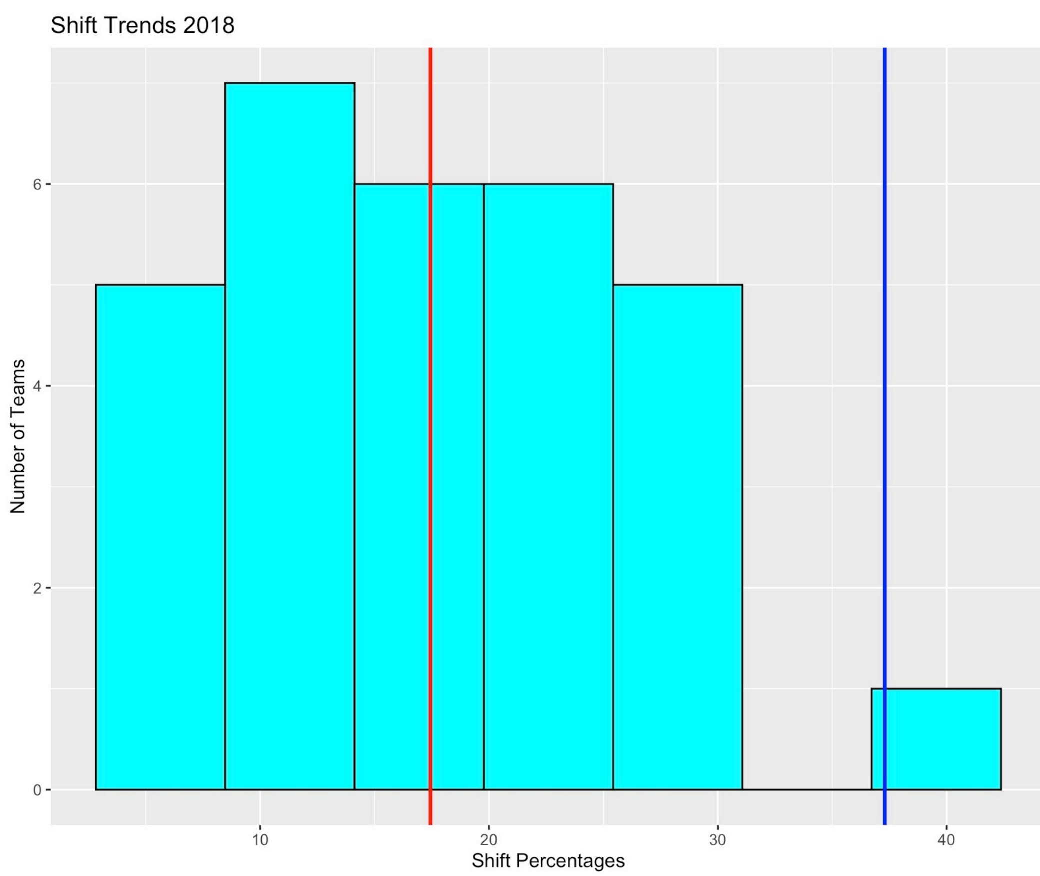


*Figure 3*

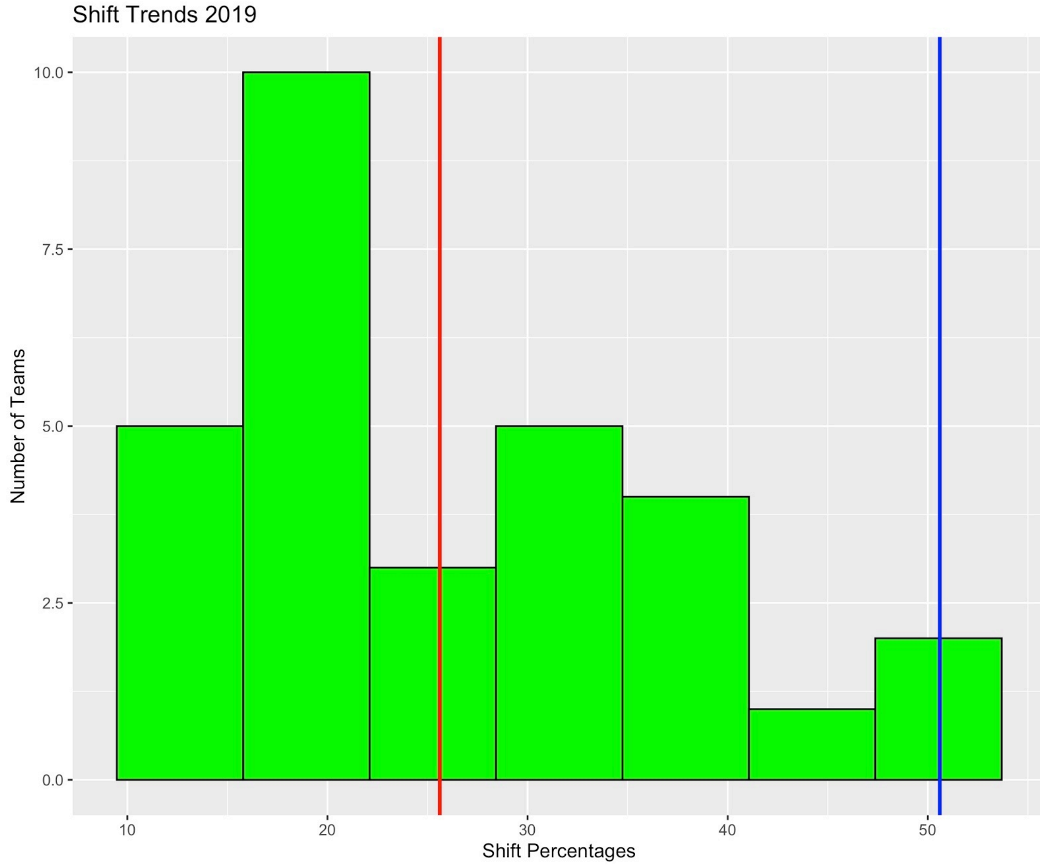


*Figure 4*

For the models that predicted shift effectiveness, we went through each individual season from 2016 to 2019 and looked at which variables were the best predictors of shift rates for the teams. To perform variable selection for these models, we used exhaustive stepwise regression; 17 predictors were incorporated in the stepwise regression, and the optimal number of predictors was determined by the minimum adjusted R-squared value for each model with a certain number of predictors. The optimal size model was then selected for each year. As a side note, all rate variables were on a per 9-inning basis (the duration of a game), all variables were standardized, and significant variables were those that had a p-value of 0.1 or lower.

For the 2016 season, the optimal model had six variables, which was comprised of Strikeout Rate, Walk Rate, Balls in Play, the difference between batting average and expected batting average (xBAdiff), the difference between slugging percentage and expected slugging percentage (xSlgdiff), and the difference between weighted on-base average and expected weighted on-base average (xwOBAdiff). After squaring the Strikeout Rate, Walk Rate, and Balls in Play variables, the model yielded an adjusted R-squared value of 0.5517. Of the six variables in the model, the significant variables included Balls in Play, xBAdiff, xSlgdiff, and xwOBAdiff.

Moving on, the 2017 model was optimized with only five variables, which were Saves, Games Pitched, Strikeout Rate, Walk Rate, and Ground Ball Rate. In terms of transformations, we squared Saves and Strikeout Rate, which yielded an adjusted R-squared value of 0.3617. For this model, Saves, Games Pitched, Walk Rate, and Strikeout Rate squared.

As for the 2018 model, the optimal number of variables was 8, more than any of the other models. The 8 variables included Games Pitched, Strikeout Rate, Batting Average on Balls in Play (BABIP), Ground Ball Rate, Earned Runs Average (ERA), xwOBAdiff, xBAdiff, and xSlgdiff. With squared transformations to BABIP and ERA, the model had an adjusted R-squared value of 0.5002. The significant variables for this model were Strikeout Rate, BABIP, ERA, xBAdiff, xSlgdiff, and xwOBAdiff.

For the 2019 model, the 5-variable model was the optimal one. Variables chosen were Home Run Allowed Rate (HR), BABIP, Runners Left on Base Rate (LOB), ERA, and xBAdiff. The only transformations performed on the model were the squaring of HR, BABIP, and LOB, and this model yielded an adjusted R-squared of 0.4625. The significant variables were BABIP squared, HR, ERA, and LOB squared.

As for the composite model, we used data from all seasons (2016-2019), and we determined that the optimal model size was 7 variables. The variables chosen were Wins, Saves, Games Pitched, Strikeout Rate, BABIP, LOB, and xSlgdiff. Even though no transformations were necessary, the adjusted R-squared for the model was quite low in comparison to the individual season models, at only 0.14. Significant variables for this composite model were Games Pitched, BABIP, Strikeout Rate, and xSlgdiff.

Another composite model we tried was a ridge regression with all of the variables, with a specific focus on the slopes of BABIP, xBAdiff, xSlgdiff, and xwOBAdiff, as these showed up most frequently in the other models and were mentioned as good indicators of shift usage in our background research. With an optimized lambda of 0.63 for this ridge regression model, we obtained an R-squared (unadjusted) of 0.1545, which was not much of an improvement from the original composite model. The slopes for the variables we deemed important are in Table 1; only BABIP had a negative slope.

Ridge Regression Slopes

| Variable Name | Slope Value |
|---|---|
| BABIP | -0.11886 |
| BA_minus_xBA | 0.00620 |
| SLG_minus_xSLG | 0.1095 |
| wOBA_minus_xwOBA | 0.08524 |

*Table 1*

In addition to the models above, we implemented a scoring system for shift effectiveness using the four important variables: BABIP, xBAdiff, xSlgdiff, and xwOBAdiff. The purpose of this scoring system was to see how impactful shifts were in the MLB over time (specifically from the 2016-2019 seasons). To calculate the shift effectiveness score for each team, we added these four variables and multiplied the sum by -1. We multiplied by -1 because higher values for each individual variable would suggest that a team's defensive strategies were not effective, so the multiplication inverts these high values to become negative values in the scoring system. So, in this system, higher scores represent better defensive tactics and shift effectiveness. Below are distributions of the scores for each individual year from 2016 to 2019 (Figures 5 through 8), as well as the composite scores of all years (Figure 9).

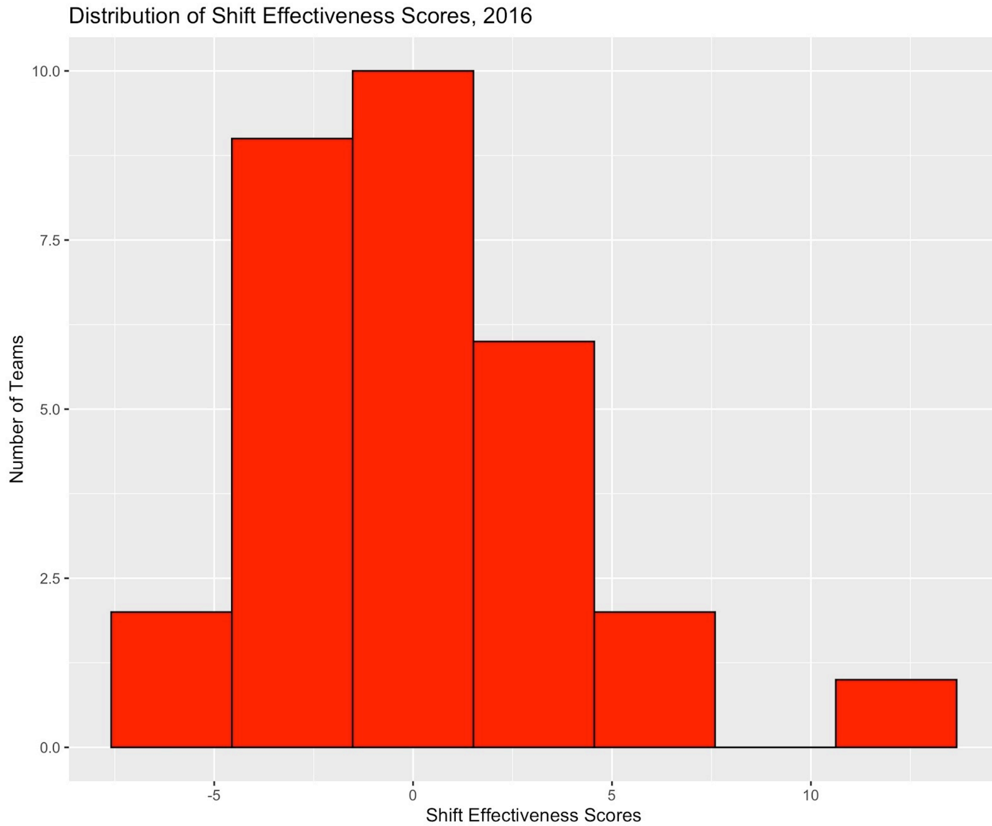


*Figure 5*

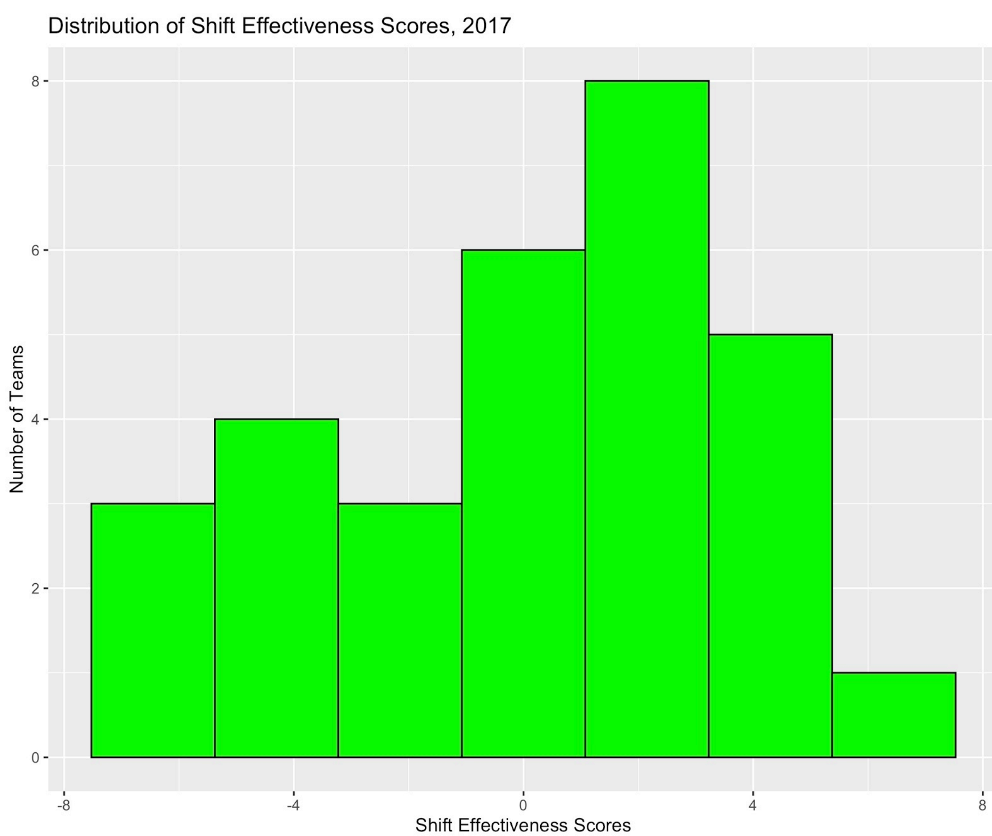


*Figure 6*

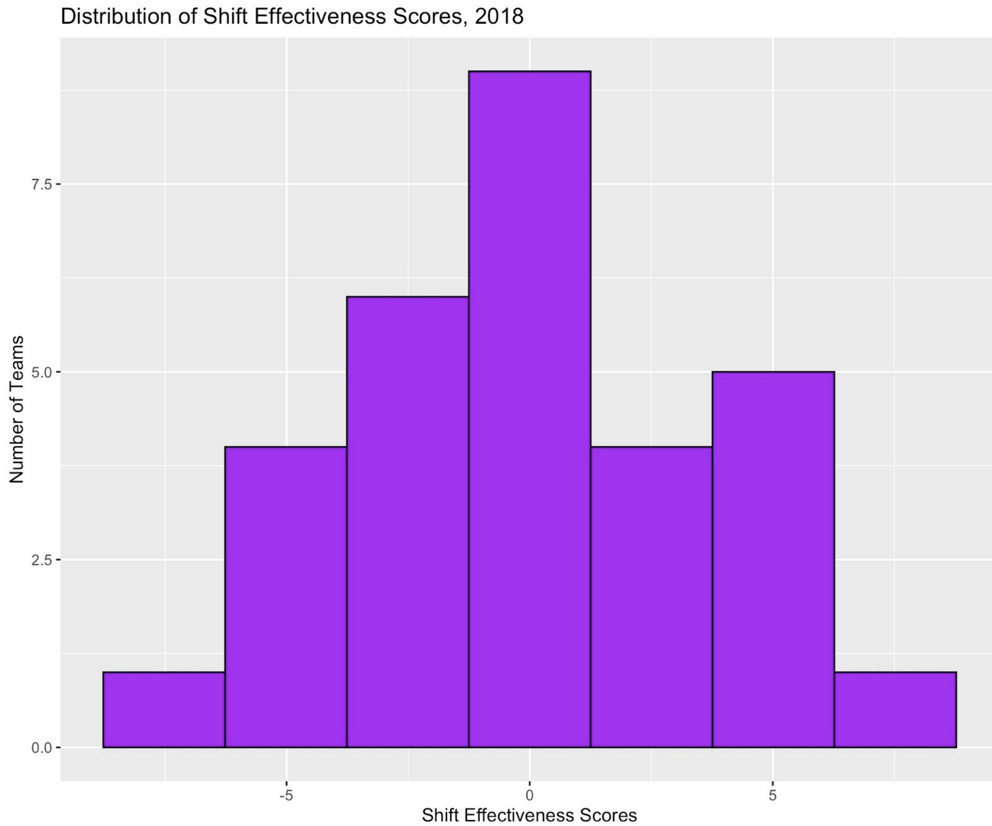


*Figure 7*

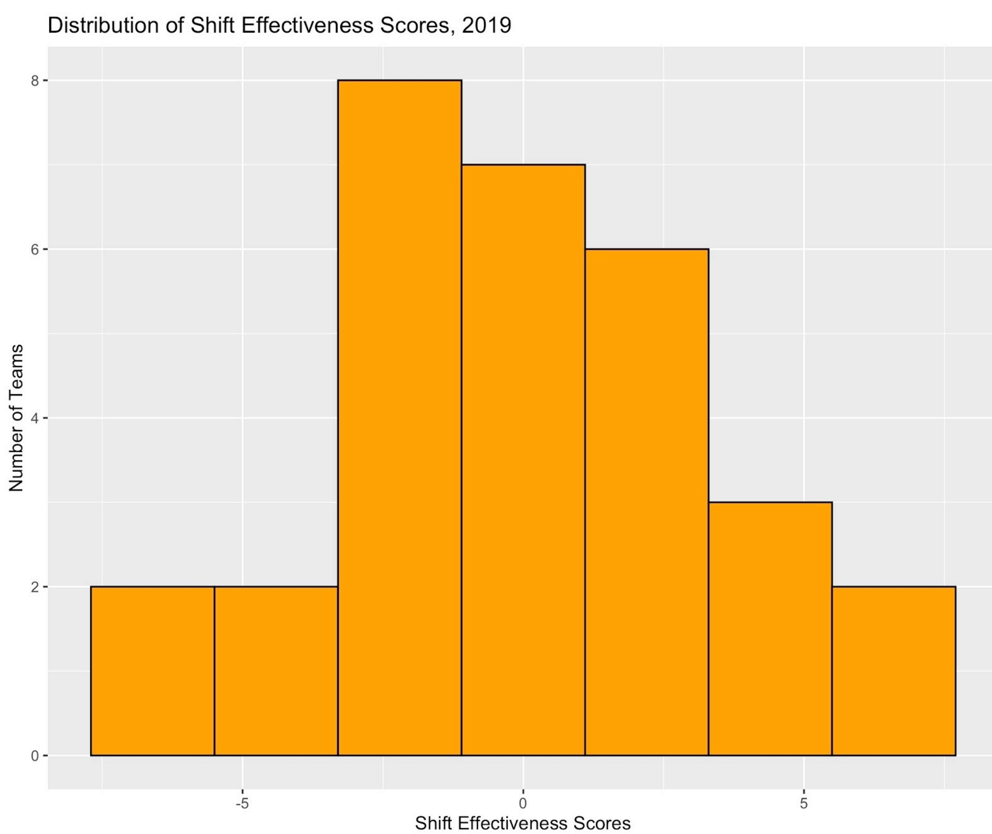


*Figure 8*

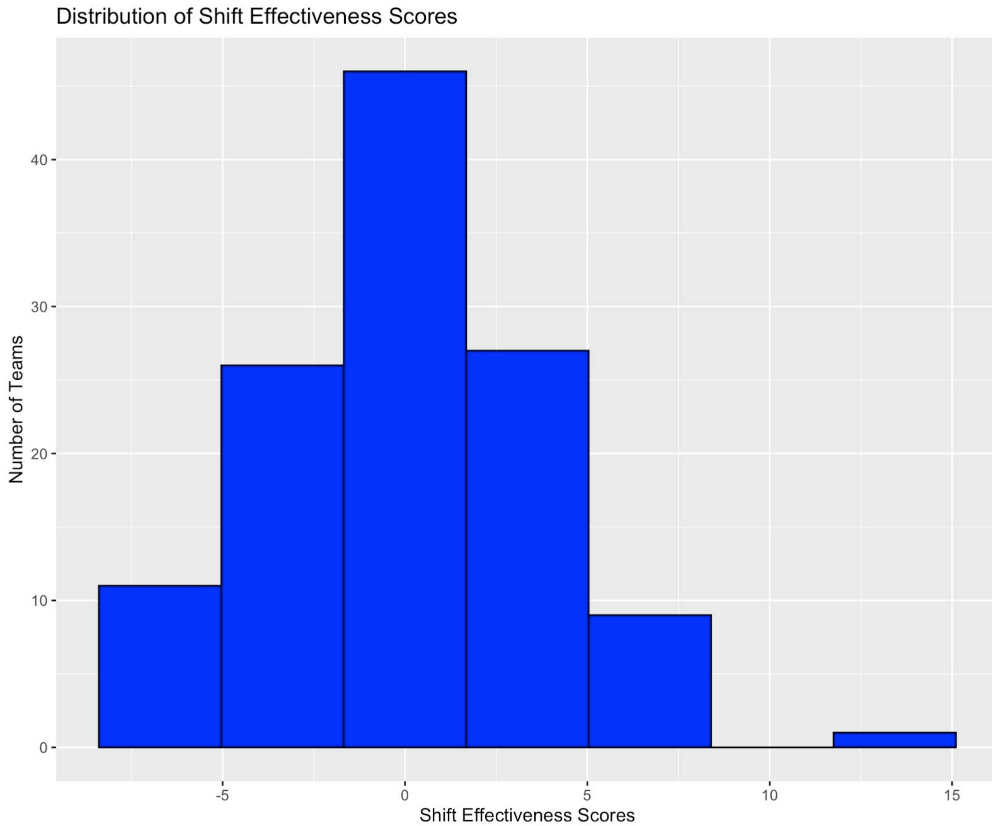


*Figure 9*

For the player level data, two models were created to predict whether or not an observation (a given player's batting statistics) occurred with or without a shift in play. A logistic regression model was trained on the data created through synthetic data generation and tested on the actual data obtained from Baseball Savant.

| | Actual Value | | |
|---|---|---|---|
| Predicted Value | | No Shift | Shift |
| | No Shift | 1407 | 181 |
| | Shift | 481 | 431 |

*Table 2*

When tested on the actual data, the model had a misclassification rate of 0.26. Out of 2,500 observations, the model classified 481 no shift observations as being with a shift and 181 with shift observations as being without a shift; there were far more false positives than false negatives. The results of this model give insight into the specific batting statistics that change the most when the shift is employed.

| | **shift** | |
|---|---|---|
| *Predictors* | *Odds Ratios* | *p* |
| ba | 0.03 | **0.005** |
| iso | 12.63 | **0.001** |
| babip | 0.04 | **<0.001** |
| slg | 1.37 | 0.601 |
| woba | 0.99 | 0.996 |
| xwoba | 132.41 | **<0.001** |
| xba | 8.73 | 0.145 |
| launch_speed | 1.10 | **<0.001** |
| launch_angle | 1.08 | **<0.001** |
| spin_rate | 0.99 | **<0.001** |
| velocity | 0.71 | **<0.001** |
| effective_speed | 0.86 | **0.031** |
| eff_min_vel | 0.93 | 0.756 |
| release_extension | 0.30 | **0.037** |
| year | 1.21 | **<0.001** |
| Observations | 2500 | |
| $R^2$ Tjur | 0.190 | |

*Table 3*

The most significant variables in the logistic model included BABIP, xwOBA, launch speed, launch angle, spin rate, velocity, and year. Each of the variables had a p-value of less than 0.001, indicating that they were extremely significant in predicting the presence of a shift.

A random forest model was also trained to predict whether an observation was with a shift in play or not. Similar to the logistic regression model, the random forest was trained on the synthetically generated data and tested on actual observations. When training the random forest model, eight variables were available for splitting at each node and the importance of each predictor was estimated and reported.

| | | Actual Value | |
|---|---|---|---|
| Predicted Value | | No Shift | Shift |
| | No Shift | 1619 | 244 |
| | Shift | 274 | 363 |

*Table 4*

This model had a misclassification rate of 0.207, an improvement on the logistic model. Out of 2,500 observations, the model predicted 244 false negatives and 274 false positives. This is a much more even distribution of errors compared to the logistic model.

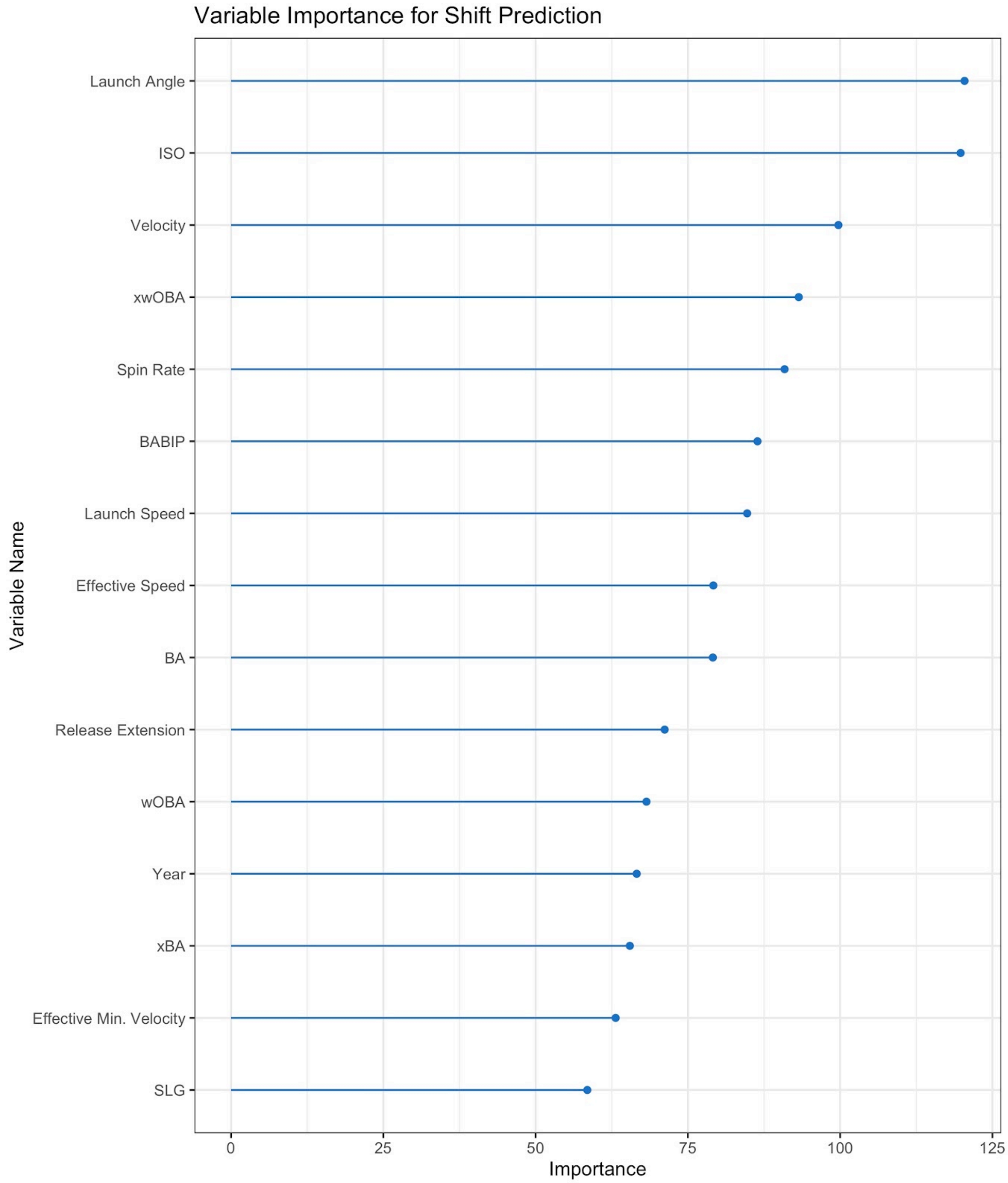


*Figure 10*

The six most important variables in the random forest were launch angle, ISO, velocity, xwOBA, spin rate, and BABIP. The logistic and random forest models generally agreed on the most important variables for predicting the presence of a shift. We will further explore these variables in our analysis of why some players beat the shift while others suffer.

To compare the players' performance against the shift, we created three equal groups of 136 players based on their change in BABIP with and without a shift. A *good* player against the shift would typically increase their BABIP with a shift in play, an *average* player's BABIP would not change much with or without a shift, and a *bad* player's BABIP would typically decrease when there was a shift. For this analysis we focused only on players from 2019. To test the difference in performance between the three groups, we performed many two sample t-tests in an effort to find out if there was a significant difference in means between the good, average, and bad players against the shift.

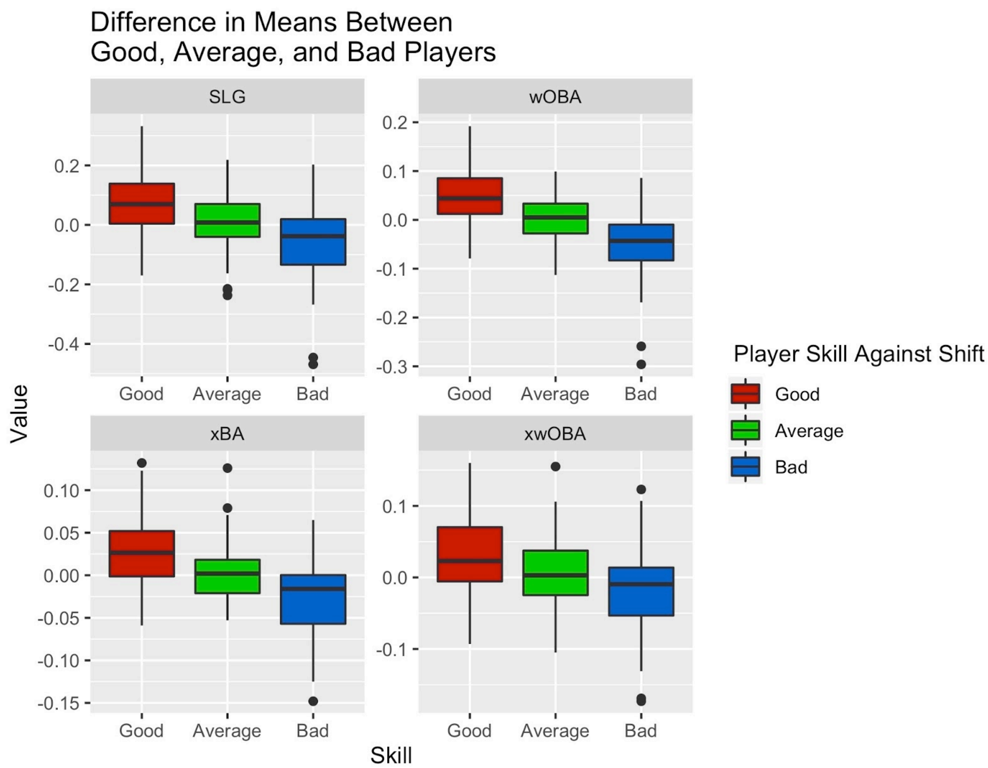


*Figure 11*

We initially examined hitting technique variables like launch angle, launch speed, effective speed, and velocity. For each technique variable we took the players' difference between when a shift was present and when a shift was not present. We then compared this difference across the three groups. For the technique variables, no significant difference was found between the three groups; the groups all seem to have similar hitting techniques with and without the shift. However, there was a significant difference when it came to power hitting statistics like SLG, wOBA, xBA, and xwOBA. Across each of these variables, the *good* players hit for significantly more power than the *average* players, and the *average* players were significantly more powerful against the shift than the *bad* players.

## Conclusion

For the models predicting shift usage, the standard composite model with 7 variables had a fairly low R-squared value of 0.14, and the significant variables were not all consistent with the significant variables of the individual year models. Based on these factors, we deduced that variables associated with the shift may depend from season to season; this is not at all surprising when considering the volatile nature of the game, where dynamic changes in the fundamental game of baseball can take place over startlingly short intervals. A great example of this was the contextual analysis of shift usage in the MLB, where shifts were used less in 2017 than in 2016 but subsequently became more popular in both 2018 and 2019. As for the ridge regression model, we looked at four slopes in particular despite integrating all variables into the model itself. The important variables were BABIP, xBAdiff, xSlgdiff, and xwOBAdiff, which we determined by seeing their prevalence in prior research and their significance in some of the models. We looked at the slopes of these variables and noticed that all but BABIP had positive slopes. Now, if the shift were an extremely effective defensive tactic, one would think that all of these slopes would be negative, as offensive statistics should generally be worse than expected with good defense. However, since three of the four important slopes were negative in the ridge regression model, we concluded that the shift may not be as definitively effective as some experts propose. Lastly, one challenge with these models was the presence of only 30 teams in the MLB, yielding a relatively small sample size to use, which perhaps explains the low R-squared values. If we continued with the project, we would perhaps look at more seasons of data to increase this sample size; even then we would only have a limited number of seasons in which the shift is even used.

Looking at the shift effectiveness scoring system, we calculated these scores for all teams for each year from 2016 to 2019. The first two histograms in the shift effectiveness section are score distributions for 2016 (Figure 5) and 2017 (Figure 6). In 2016, the scores were skewed to the right, with the majority of teams having a negative shift effectiveness score. The majority of teams had scores in between -3 and 3, and there were only four notable outliers, with two scores greater than 5 and two scores less than -5. The highest score, 12.7, comes from the 2016 World Series Champions, the Chicago Cubs; this was the highest recorded shift effectiveness score out of any team we examined in this project. As for 2017, there were many more positive scores than in the year prior, with the majority of teams earning positive scores. The distribution in 2017 is not skewed like 2016's distribution, as there is a more even spread of scores; 7 scores were below -3, 16 scores were from -3 to 3, and 7 scores were above 3. Moving on to the next row of visuals, the 2018 (Figure 7) and 2019 (Figure 8) seasons had quite different results from 2016 and 2017 in their own right. For 2018, the distribution looked more like a bell curve than previous years, with a fair amount of teams concentrated near a score of 0. In fact, 19 of the 30 teams had scores in between -2 and 2, so fewer teams had extreme scores this season; there were slightly more teams with positive scores than negative scores overall, but seven of these positive scores were less than 1. The last individual season was 2019, where a lot of scores were concentrated around 0 as well. However, the distribution of 2019 was not like the curve from 2018; instead, there were slightly more negative scores than positive ones, with 13 positive scores and 17 negative scores. Of the 17 negative scores, 8 of them were greater than -2, so there were not too many extreme scores, similar to 2018. Finally, the composite histogram (Figure 9) reveals a clear bell curve even more rigidly defined than the 2018 distribution; this makes sense because the composite data should level out any imbalances from a particular year. Comparing score distributions from year to year, one could notice that 2017 seemed to have the highest number of positive scores out of any season, which would suggest that defensive strategies had the most success during this time. However, this was also the year where shift usage slightly dipped, which may be counterintuitive for proponents of the shift. Going along with this, 2018 and 2019 were years in which usage of the shift increased in the MLB, but shift effectiveness scores did not show marked improvement, so we cannot conclude that the shift is a significantly better defensive strategy than non-shifts.

For the player level analysis, the logistic and random forest models generally agreed on the variables that change the most when there is a defensive shift: launch angle, ISO, velocity, xwOBA, spin rate, BABIP, and launch speed. The logistic regression indicated several significant relationships between predictors and the presence of a shift. A higher xwOBA is 132 times more likely to occur with a shift in play, indicating that there is a higher expected number of bases hit when there is a shift in play. A higher ISO is 12.63 times more likely to occur when there is a shift in play, again indicating that players are more likely to hit for power and get more bases when there is a shift. Additionally, an increase in launch angle and launch speed is 1.1 and 1.08 times (respectively) more likely to occur when there is a shift in play, indicating that when there is a shift, batters are more likely to hit the ball higher and harder. On the other hand, an increased BABIP is .04 times less likely to occur when there is a shift, indicating that typically BABIP decreases when there is a shift. Based on the logistic model, while players are more likely to hit for power and get additional bases when there is a shift, their overall batting average tends to go down.

The two sample t-tests performed to test the difference between the *good*, *average*, and *bad* players against the shift in 2019 showed significant differences between the groups when it came to the power hitting statistics SLG, wOBA, xwOBA, and xBA. The statistics xBA and xwOBA measure the skill of a player by looking at comparable balls in terms of exit velocity, launch speed, and launch angle; they essentially remove defense from the equation. Since defensive performance does not influence these variables, the statistical significance of the t-tests indicates that it is the skill of the players that causes a significant increase in these hitting metrics, which the shift is not the cause for. Simply put, the *good* hitters will always hit better than the *average* and *bad* hitters whether there is a shift or not. Similarly, the *average* hitters will always hit better than the *bad* hitters regardless of the presence of a shift. Since players who beat the shift are just skilled batters who typically get more bases than average, perhaps teams should focus on shifting against the *bad* batters rather than batters who tend to hit the ball in a certain direction. This would be a novel approach to implementing the shift compared to how teams currently use the defensive strategy. We are not concluding that the shift is an ineffective strategy; rather, we are suggesting that it may be more effective when used against poorly skilled players. If teams did start using the shift in this newly proposed manner, it would take multiple years to see if the effectiveness of the shift increases.

Overall, our analysis was tremendously limited by the lack of data available for the shift. This defensive strategy has only been used for a handful of years compared to how long baseball has been played historically. This analysis would improve tremendously in the coming years as more shift data is collected. In the future, we are also interested in looking at how pitchers may change their pitching approach when a shift is in play.

# References


Karimi, Hussein, et al. “Calculating New Stats in Major League Baseball with Amazon SageMaker.” *Amazon*, Greenhaven Press/Gale, 16 Oct. 2019, aws.amazon.com/blogs/machine-learning/calculating-new-stats-in-major-league-baseball-with-amazon-sagemaker/.

*Major League Team Stats*. *"2016-2019." Pitchers Dashboard, FanGraphs Baseball.* www.fangraphs.com.

Miller, Sam. “MLB Myth-Buster: The Shift Isn't Curbing Runs; It's Creating Them.” *ESPN*, ESPN Internet Ventures, 7 Aug. 2018, www.espn.com/mlb/story/_/id/24265798/mlb-myth-buster-shift-curbing-runs-creating-them.

“MLB Team Fielder Positioning.” *2016-2019*. Baseball Savant. baseballsavant.mlb.com.

Paine, Neil. “Why Baseball Revived A 60-Year-Old Strategy Designed To Stop Ted Williams.” *FiveThirtyEight*, FiveThirtyEight, 13 Oct. 2016, fivethirtyeight.com/features/ahead-of-their-time-why-baseball-revived-a-60-year-old-strategy-designed-to-stop-ted-williams/.