**Fielder Independent Pitching: Identifying Problems and Formulating Improvements**

*Introduction:*

Since 1912, Major League baseball has kept track of earned run average (ERA), along with other statistics such as wins and strikeouts, as a measure of evaluating pitcher performance. Earned runs are the number of runs scored by the other team against the pitcher, not including any runs allowed due to fielding errors. ERA is the average of earned runs allowed per nine innings pitched, calculated as:

ERA is a relatively strong proxy for pitcher ability and is proven to be much more informative than other commonly evaluated statistics such as wins. While ERA has been heavily used for much of baseball’s long history, there is a large amount of inherent statistical error from basing the calculation on earned runs. Errors are subjective to the scorer and are confounded by fielder ability. For example, if an above-average fielder makes a fault on an incredibly difficult play, it will often count as an error, whereas an average fielder would not attempt such a play. Since an error is given, that run does not count against the pitcher. Additionally, worse fielders may allow baserunners to take extra bases, increasing the likelihood of that runner scoring. These among many other reasons causes ERA to be directly dependent on defensive quality.

As baseball has grown more statistically advanced over the last several decades, different methods have surfaced in attempting to calculate a proxy for true pitching success. One of the most commonly used methods of which is called Fielder Independent Pitching (FIP). FIP takes advantage of baseball’s three true outcomes: base-on-balls (BB), strikeouts (K), and home runs (HR). Each of these are considered true outcomes as there is thought to be no little-to-no fielding variability within the recording of these results. FIP is calculated by:

Where C represents a constant centering FIP around the league average ERA. By definition, FIP attempts to eliminate luck and provide information regarding true outcomes. However, if we wish to generalize this statistic further, there is a noticeable source of error regarding the use of home runs. Since dimensions vary between ballparks, flyballs considered home runs at one park may not clear the wall at another. As a solution to this problem, expected Fielder Independent Pitching (xFIP) utilizes the quantity of flyballs a pitcher allows and multiplies this by the league average home run per flyball rate, as demonstrated by:

It is worth noting that the estimation for expected home runs is not related to power whatsoever, but solely based on how often the pitcher allows the opposition to hit the ball in the air.

While FIP and xFIP convey much more information than ERA, both ignore the potential of a ball being put in play without going over the wall. During the 2018 season, home runs only made up 5.2% of all non-strikeout at-bats, meaning 94.2% of balls hit were completely disregarded by these estimators for pitcher quality. Throughout the past 5 seasons (not including 2021), home runs only corresponded to roughly 40% of scoring. To closely match ERA, FIP overutilizes home runs allowed, potentially resulting in added error.

Additionally, both FIP and xFIP have no method of estimating pitching accuracy, otherwise known as command. Traditionally, there is no way to accurately measure command through basic outcome statistics. However, in 2017, Baseball Prospectus created the statistic Called Strikes Above Average (CSAA), which they define as, “the additional called strikes outside the reference zone that are credited to the pitcher after accounting for catcher, umpire, pitch type, etc.” Essentially, the calculation utilizes called strike probability to provide a measure of how well a pitcher can consistently hit difficult spots, proxying command.

Through this project, we attempt to improve on Fielder Independent Pitching by modeling earned run average using available sabermetrics to estimate initially ignored effects. As a byproduct, our model will improve on xFIP’s estimate of expected home runs allowed.

*Data collection*:

We obtained much of our data from the 2018 and 2019 data from FanGraphs’ online pitching database. Tables were merged from the “Standard”, “Advanced”, and “Statcast” tabs. Variables kept include home runs, exit velocity, hard-hit rate, flyball rate, strikeouts, base-on-balls, total batters faced, hit-by-pitches, launch angle, FIP, xFIP, team, and ERA.

In addition, we utilized the 2018 and 2019 data from Baseball Prospectus’ “Plate Discipline” tab for its Called Strikes Above Average (CSAA) statistic.

Only pitchers throwing more than 100 innings over the duration of the season included in our calculations. This is to reduce noise and increase confidence in our outcomes. No outliers were removed from the dataset. To obtain the rate statistics for strikeouts, walks, and homeruns, we divided the season total by the number of total batters faced.

*Descriptive Statistics:*

Our sample contains 125 pitchers for the 2018 season and 118 pitchers for the 2019 season, with an overlap of 76 pitchers throwing over 100 innings in both seasons. For each variables collected, the averages by year are displayed below:

**Table 1**

|  |  |  |
| --- | --- | --- |
| Variable | 2018 Mean (SD) | 2018 Mean (SD) |
| ERA | 4.049 (0.931) | 4.287 (0.921) |
| FIP | 4.061 (0.77) | 4.351 (0.808) |
| xFIP | 4.04 (0.661) | 4.383 (0.692) |
| Strikeout% | 22.4% (5.0) | 22.9% (5.4) |
| Walk% | 8.6% (2.4) | 7.4% (1.8) |
| CSAA | 0.001 (0.008) | 0.001 (0.008) |
| HR% | 3.1% (0.9) | 3.6% (0.9) |
| Exit Velocity | 88.379 (1.125) | 88.67 (1.225) |
| Launch Angle | 12.214 (4.087) | 12.71 (3.845) |
| Hard-Hit% | 35.3% (4.786) | 38.1% (3.98) |
| Flyball% | 35.2% (6.24) | 35.7% (5.924) |
| League |  |  |
| American | 44% | 41% |
| National | 48% | 47% |
| Both | 8% | 13% |

From Table 1, we can see that there is a general increase in offense between 2018 and 2019. The leaguewide ERA increases from 4.049 to 4.287 between 2018 and 2019, as well as its estimators FIP and xFIP. Home run rate, exit velocity, launch angle, hard-hit rate, and flyball rate all increase as well. This suggests that, on average, hitters had more success in 2019 than in 2018.

Observing our main response variable, the distribution of ERA over the 2018 season is as follows:

Chart, histogram

Description automatically generated

The distribution of ERA appears to be approximately normal. As a descriptive overview, we can illustrate the predictors’ relationships with the response and one another through the 2018 Pearson correlation heatmap in Figure 2.

A picture containing treemap chart

Description automatically generated

Figure 2 supports the strength of FIP and xFIP as ERA estimators, as both are strongly positively correlated with ERA. Additionally, looking at the top right of the figure, it is evident that flyball rate is strongly correlated with launch angle. This makes intuitive sense, as both variables represent similar results. However, flyball rate could still maintain importance as it may provide the model with unique information regarding result frequency. Similarly, exit velocity and hard-hit rate are strongly positively correlated but can relay slightly different information.

From Figure 2, we can see that CSAA is relatively uncorrelated with the majority of the other variables in our model, suggesting that it can provide added information regarding pitching command. The top 10 CSAA leaderboard from 2018 is as follows:

|  |  |  |
| --- | --- | --- |
| Name | CSAA | ERA |
| Jon Lester | 0.020 | 3.32 |
| Blake Snell | 0.019 | 1.89 |
| Marco Estrada | 0.018 | 5.64 |
| Jordan Zimmermann | 0.017 | 4.52 |
| Wei-Yin Chen | 0.016 | 4.79 |
| Kyle Gibson | 0.015 | 3.62 |
| Masahiro Tanaka | 0.014 | 3.75 |
| Patrick Corbin | 0.014 | 3.15 |
| Tyson Ross | 0.014 | 4.15 |
| CC Sabathia | 0.013 | 3.65 |

The names on the leaderboard support the claim that CSAA is a proxy for control, as the majority of pitchers on the list (excluding 2018 Cy Young Award winner Blake Snell) are not known for having particularly strong pitch movement, but generally supplement this by inducing soft contact through pinpoint accuracy. The relationship between CSAA and ERA has a correlation of -0.28, as shown by Figure 3.

Chart, scatter chart

Description automatically generated

Note the point with an extreme negative value of CSAA. This observation is James Paxton of the Seattle Mariners, who pitched to a 3.76 ERA in 2018.

Since we model expected home run rate in the process of modeling ERA, we need to observe the distribution of home run rate.

Chart, histogram

Description automatically generated

From Figure 4, we see that the distribution of home run rate is approximately normal. It is important to note that home run rate is bounded by 0.

*Inferential Results:*

In an attempt to maintain interpretability, we utilize an ordinary least squares model to predict both expected home run rate and ERA. Both models will be trained using the 2018 data and tested using the 2019 data. For the expected home run rate model, the predictors we consider are exit velocity, launch angle, flyball rate, and hard-hit rate. The model output is as follows:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Coefficients:** | **Std. Error** | **t statistic** | **p-value** |
| **Intercept** | -0.14506 | 0.058335 | -2.487 | 0.01427 |
| **Exit Velocity** | 0.001346 | 0.000727 | 1.851 | 0.0666 |
| **Launch Angle** | -0.00131 | 0.000583 | -2.254 | 0.026 |
| **Hard-hit Rate** | 0.000481 | 0.000173 | 2.775 | 0.00641 |
| **Flyball Rate** | 0.001586 | 0.000376 | 4.213 | 4.90\*10^-5 |

Therefore, the formula used to calculate expected home run rate (xHR%), is:

For the 2018 training data, xHR% has a Pearson correlation of 0.684 with observed home run rate. On the 2019 testing data, we observe a correlation of 0.597. This drop can likely be attributed to a change in league environment. As discussed previously, offensive factors increased in 2019.In future analysis, it would be wise to include a term centering xHR% on the league average. For reference, over the 2019 season, the xFIP estimate for expected home runs has a correlation of 0.451 with observed home run rate. To evaluate the fit of the model, we can observe Figure 5.

Chart, scatter chart

Description automatically generated

The residuals appear to be distributed normally and the model does not violate the homoscedasticity assumption. The leftmost observation represents Noah Syndergaard, who gave up 9 home runs while facing 644 batters.

Utilizing this model to calculate expected home runs, we can build our improved version of Fielder Independent Pitching, titled iFIP. In total, we will build our model on the predictors strikeout rate, walk rate, xHR%, exit velocity, league (a binary variable with one representing the American League), and CSAA. The model summary is as follows:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Coefficients: | Std. Error | t statistic | p-value |
| Intercept | -15.1954 | 5.13961 | -2.957 | 0.003758 |
| K rate | -10.0006 | 1.12718 | -8.872 | 9.19\*10^-15 |
| BB rate | 6.35934 | 2.24806 | 2.829 | 0.005492 |
| xHR rate | 23.01426 | 10.8266 | 2.126 | 0.035612 |
| Exit Velocity | 0.22855 | 0.05966 | 3.831 | 0.000206 |
| League | 0.09108 | 0.11255 | 0.809 | 0.41997 |
| CSAA | -15.1558 | 7.23288 | -2.095 | 0.038274 |

Every predictor in the model is significant at the 0.05 level of significance except league. While the coefficient may not be significantly different from zero, we are motivated to leave it in the model. Our prior knowledge of the American League containing a designated hitter validates the slight difference in ERA observed by the model. The formula for iFIP can be written out as follows:

On the 2018 training data, iFIP has a strongly positive correlation of 0.7738 with ERA.

Chart, scatter chart

Description automatically generated

Using the metric to predict ERA in 2019, we see a reduced correlation of 0.6423. While there is a noticeable decrease between years, it is important to remember, as with xHR%, there is no centering term based upon league average ERA. It is important to keep this in mind as we compare iFIP to FIP.

In comparing the metrics, we look at four separate results. The first two columns of the table below measure accuracy by observing the statistic’s relationship with ERA. The last two columns measure consistency, or how predictive it is with next year’s results.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Statistic | 2018 Correlation with ERA | 2019 Correlation with ERA | 2018 Correlation with 2019 ERA | 2018 Correlation with Itself in 2019 |
| FIP | 0.7811 | 0.7797 | 0.3641 | 0.4433 |
| xFIP | 0.6589 | 0.6251 | 0.4077 | 0.6572 |
| iFIP | 0.7738 | 0.6423 | 0.3496 | 0.6776 |

Out of the three measures, xFIP appears to be the least accurate in comparison to ERA, which makes intuitive sense. Since FIP uses home runs, which are guaranteed to factor into ERA, it will be more predictive of ERA. However, while FIP is the most accurate, it lacks consistency on a year-to-year basis. iFIP provides a middle ground between FIP and xFIP. It attains more accuracy than xFIP in 2018 and 2019 (despite being uncalibrated) while still not utilizing home runs as a factor, and it is more consistent on a year-to-year basis than FIP. For the column comparing the 2018 statistic with ERA, it is important to emphasize that noise in ERA can make these values rather unstable. 2018 ERA has a 0.2195 correlation with 2019 ERA.

*Conclusion:*

From the model creating iFIP, it is interesting that the ratio between the coefficients for strikeout rate and walk rate is approximately the inverted ratio between the FIP coefficients. For iFIP, the strikeout coefficient divided by the walk coefficient is -1.57 (about -3/2) compared to the -0.66 (-2/3) ratio from FIP. Additionally, using iFIP, the ratio of the strikeout coefficient is much higher compared to the home run coefficient than it is in FIP (0.4345 to 0.1538).

As a brief post-hoc analysis, we built a model to predict ERA using xFIP, CSAA, and exit velocity. Interestingly enough, all three predictors have coefficients significantly different from zero.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Estimate | Std. Error | t stat | p-value |
| Intercept | -25.7083 | 4.35488 | -5.903 | 3.34E-08 |
| xFIP | 0.79266 | 0.0839 | 9.448 | 3.33E-16 |
| CSAA | -15.6932 | 7.12389 | -2.203 | 0.0295 |
| Exit Velocity | 0.30067 | 0.04985 | 6.032 | 1.82E-08 |

This further demonstrates that xFIP alone can be heavily improved in its estimation of ERA.

In further analysis, iFIP should be improved by adding a constant centralizing the mean prediction to the current season’s mean ERA. We can extend this idea to the xHR% term within the calculation for iFIP. Another limitation we faced in analysis was that, from our dataset, there was no way to determine if a player’s league(s) if he played for multiple teams throughout the season. As a result, players who played for multiple teams are grouped with the National League.

Takeaways from this project include a new statistic that teams and fans can utilize to identify successful pitchers. Additionally, it opens the door for more research and improvements by demonstrating that FIP and xFIP can lack important information that is readily available.