**Hackathon: Switching Pitcher’s Roles**

As mentioned by the prompt’s background, “MLB pitching staffs are becoming increasingly amorphous in nature, and the types of roles available to pitchers are more fluid than ever.” With pitchers having a very different role than the “traditional player”, we had been tasked with finding characteristics that make each player successful at their given role, and finding two players that would be well suited to switch to a different role. In response to this, we developed a method of fitting multiple models for starting pitchers and relievers to figure out if they would be best fit for the rotation or bullpen.

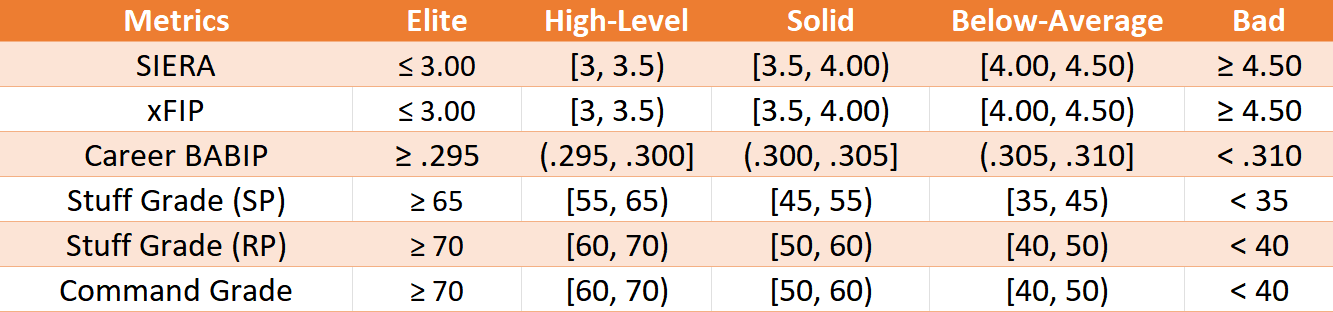
First, using the pitchers that appeared in the play-by-play data given to us, we classified them on a ranking scale with a number from zero to four with zero being a “poor-quality” pitcher, and four being an “elite-level” pitcher. Each bracket was broken down using five different metrics: SIERA, xFIP, Career Average BABIP, Stuff, and Location. ****

Table 1: Pitcher Classifications

Our methodology deviates from multiple places: SIERA and xFIP are both derived from Fangraphs estimators and taken on a number between their excellent and great category. BABIP was also pulled from Fangraphs, as Piper Slowinski writes “the average BABIP for pitchers is also about .300,” and that baseball “won’t see many pitchers outside the .290 to .310.” Therefore, we adjusted for our “elite” and “poor” to be inclusive to those numbers, and each category in between was evenly distributed.As for the stuff and command grade’s, using PitchingBot’s 20-80 grading scale, each player was distributed along the scale. It should be noted that while command grades remained consistent for relief pitchers and starting pitchers, stuff grades needed to be weighted differently. Relievers generally showcase and grade out with better stuff metrics, and therefore needed to be graded harsher.

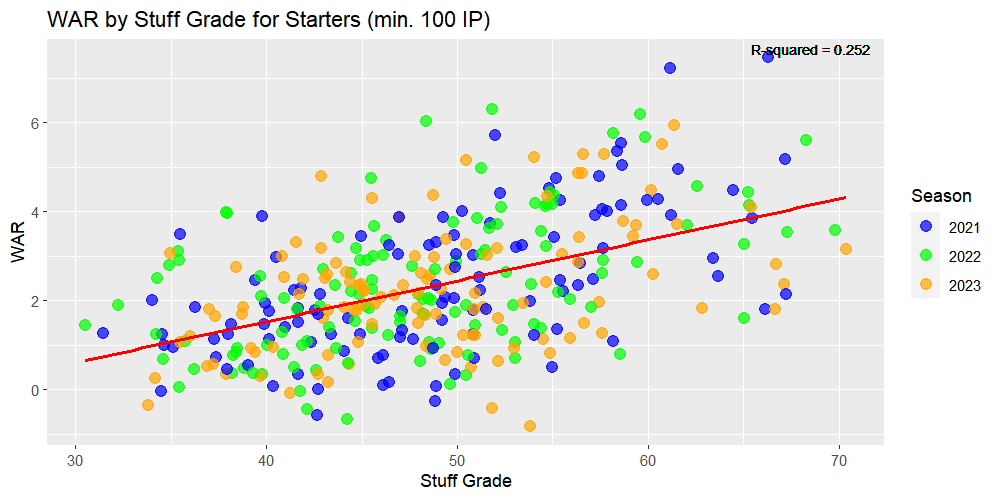
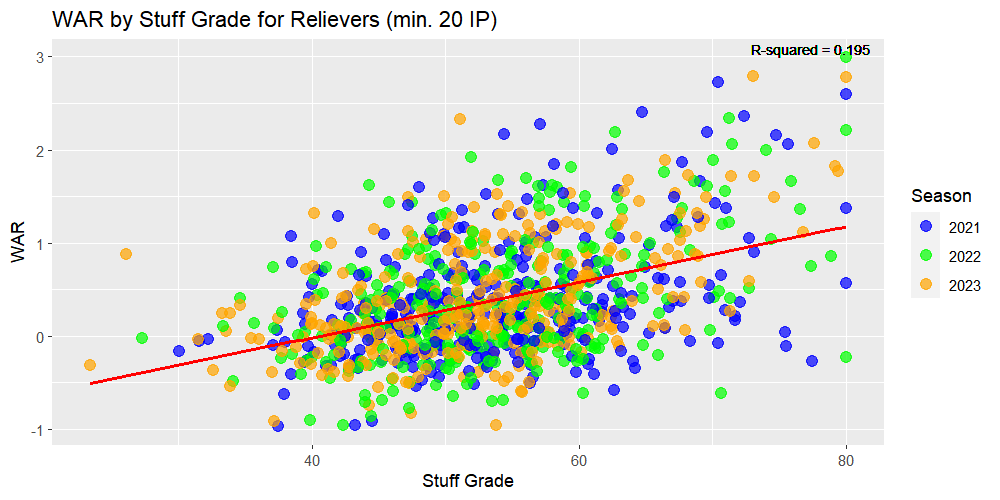


Figure 1: WAR vs Stuff Grade for SPs

 Figure 2: WAR vs Stuff Grade for RPs

As can be seen from the graphs above, only one starter across all three years had a 70+ Stuff Grade, while the relievers had seven different pitchers that graded with an 80. Each player’s score then gets averaged out and judged based on their final score and rounded up if need be. For example, if a pitcher has two elite attributes, one solid attribute, and two bad attributes, they would grade out as a two, or a “solid pitcher”. After running code to grade every pitcher, we then omitted all pitchers from the data that graded as “high-level” or “elite” as these pitchers are already established enough in their roles, and a change to a different role would not make intuitive sense. Lastly, all pitchers were then divided up simply between starters and relievers, where starters were defined by people whose appearances were at least 60% starts so as to eliminate swingmen and openers.

Starting with the relievers, it was presupposed that when switching roles appropriate resources and time were delegated to transforming the role of the pitcher, and results will be based as if a reliever was already built up as a starter. This also means we have to take feasibility into account. For example, if our model told us Craig Kimbrel should move to the rotation, intuitively, we wouldn’t decide to move him because no team would. Kimbrel is 35 now, has never been a starter in the MLB, and only throws two pitches. This is the case with most relievers, which is why we built a XGBoost model that checks the capability of each reliever to be in the rotation based on the stuff, command, and size of pitch mix of the guys in the rotation. The average starting pitcher’s Stuff Grade was nine percent lower than the average reliever, so we multiplied all of the relievers’ Stuff Grade’s by 0.91. Using this adjusted Stuff Grade, we then plugged this into our model with Command Grades. These results returned a 57.3% accuracy within the exact tier of player (bad, below-average, solid, etc.) , and 96.1% accuracy within one tier.

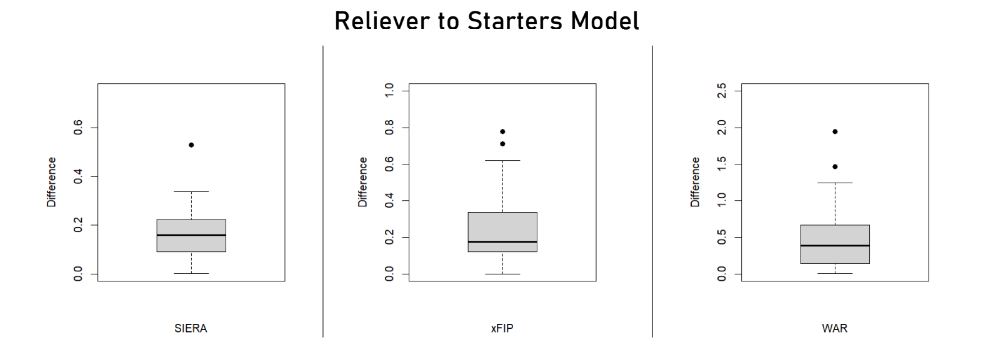


Figure 3: Model Accuracy for SP

We then omitted any pitchers with 3 or less pitches, as starters need to have deception, and it is unfair to assume someone can add a pitch or dominate with a limited repertoire as very few have the capability of doing so. We also exercised discretion by excluding pitchers who had been moved out of the rotation in the majors or high minors due to extenuating reasons, such as injury or performance issues. For example, Beau Brieske of the Tigers was moved to the 15-Day and 60-Day IL twice each in 2022, which made Detroit move him to the bullpen for the upcoming season. Our model suggested that Brieske could make the transition back to a starter role, but our model doesn’t account for underlying injury history.

However, two names that our model did return were two lefties: A.J. Puk and Tim Herrin. In 2023, Puk appeared in 58 games, all out of the bullpen, to the tune of a 3.97 ERA and a 0.9 fWAR with Miami. However, in our model, Puk is projected to have 2.86 WAR, 3.76 SIERA, and 3.79 xFIP.Last year, he finished 31 games with 15 saves and a 62.5 save percentage, indicating that he is in the wrong role. After a trade, Puk was thrust into a setup/closer type role, but he grades out with four different good pitches, the worst of which is his 4-Seam which has a .220 xBA and .298 xWOBA, and all of his pitches having better expected numbers. The other pitcher we were able to spotlight was Tim Herrin. Only 26 years old last year, Herrin threw 27.2 IP in the majors with the Guardians at a 5.53 ERA and just a 0.1 fWAR and was optioned five different times. Herrin couldn’t catch much of a break, as his xERA sat 83 points lower than his actual, and three of his four pitchers had a lower xWOBA than true WOBA. While a higher BABIP can indicate quality of contact, .361 is unsustainable. Regression to the mean could’ve possibly kept him on the roster, which is a strange way to talk about a guy who projected in our model for 2.53 fWAR and 4.07 SIERA as a starter. In essence, Puk’s value is more than tripled, and Herrin would be more than 25 times as useful.

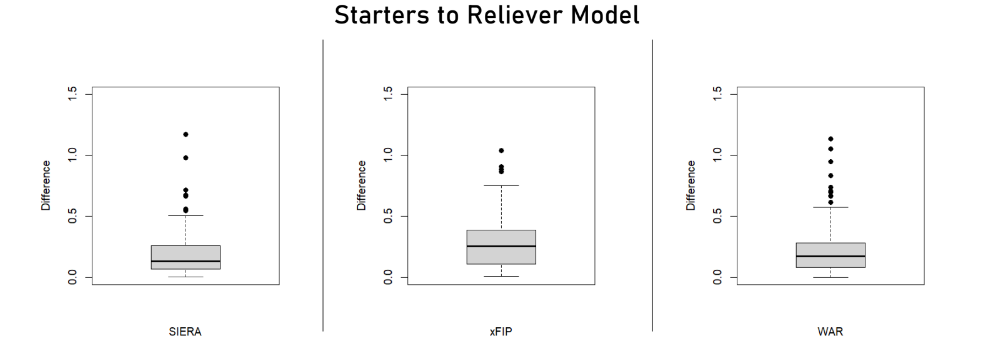
Moving onto the starters switching into relievers, we did not identify anyone of note. Nobody saw a significant increase in value in our model when transitioning from starter to reliever, and therefore we couldn’t make any justifiable predictions for moving a starter to the bullpen. This does make some sense though due to smaller sample size among relievers in innings, and the fact that there isn’t much to suggest that a specific player couldn’t make a transition to a bullpen role, a decrease in accuracy solidifies our findings. 

Figure 4: Model Accuracy for RP

All code used for modeling and visualizations are in the attached code file below.

References

Grove, Cameron. “PitchingBot Pitch Modeling Primer.” *Sabermetrics Library*, library.fangraphs.com/pitching/pitchingbot-pitch-modeling-primer/. Accessed 5 Feb. 2024.

Slowinski, Piper. “BABIP.” *Sabermetrics Library*, library.fangraphs.com/pitching/babip/. Accessed 5 Feb. 2024.

Slowinski, Piper. “SIERA.” *Sabermetrics Library*, library.fangraphs.com/pitching/siera/. Accessed 5 Feb. 2024.

Slowinski, Piper. “XFIP.” *Sabermetrics Library*, library.fangraphs.com/pitching/xfip/. Accessed 5 Feb. 2024.