Senior Project Thesis Draft

Tate Evans

Advisor: Professor Myers

Introduction

In baseball, a pitcher is defined by several things, their build, the way they move in their pitching motion, their signature pitch, but most importantly by their repertoire - the pitch type, movement, speed, and usage of each pitch that they throw. All of these factors work together to influence the effectiveness of a pitcher. In the modern age of baseball, a repertoire is not just the group of pitch names the pitcher utilizes; advanced technology allows for the precise measurement of each pitch's usage, movement, and velocity, allowing for an extremely in depth and quantified way to describe a pitcher.

Over recent years, these metrics have become increasingly important. A specific pitcher's 'stuff', the quality of his pitches based on their velocity, movement, and other metrics, in combination with the ability to locate the ball, has proven to be the strongest predictor of a pitcher's performance (fangraphs link). Baseball analysts have coined the term *Stuff+* to describe how effective an individual pitch is compared to pitches of the same type within a single value (higher Stuff+ typically comes from more movement or velocity); equivalent metrics have been made for location, *Location+*, and the combination of the two, *Pitching+.*

Given the benefit that better 'stuff' is likely to bring a pitcher, we expect to see changes in a pitchers' repertoire year-to-year as they chase that increased performance. And add on that hitters will quickly adapt to a pitcher's repertoire if there is not significant change. Everything in the world of baseball is pushing pitchers to change in pursuit of better performance and larger contracts. But change does not always come easy; every pitcher has certain mental and physical biases they have developed throughout their career that limit the extent and effectiveness to which changes can be made. Abandoning an already successful repertoire in

search for even more *Stuff+* might leave a pitcher in no man's land come the start of the season. The same could happen from a pitcher trying to build a completely new repertoire and their body not being able to adapt.

These ideas are the motivation behind the main question of this paper - how does repertoire change relate to performance change for MLB pitchers? Do larger changes tend to lead to better results, or are there risks and inconsistencies that come into play? Repertoire change is quantified using Earth Mover's Distance calculated from the velocity, horizontal movement, and vertical movement of every pitch in two consecutive years. Performance is measured through advanced statistics that try to account solely for a pitcher's ability, including FIP, SIERA, xERA, and K-BB%.

Background

To analyze the relationship between a pitcher's repertoire change and performance change, it's necessary to use metrics that most effectively quantify those things. A major difficulty with trying to measure performance for pitchers is that most statistics are outcome-based and partly rely on the defense's ability to prevent runs. This can create noise within traditional stats, something more advanced stats attempt to solve by only capturing a pitcher's ability. One popular metric is *Fielding Independent Pitching (FIP)*, which uses the outcomes most controllable by a pitcher to evaluate performance. The formula for FIP is:

$$FIP = \frac{13 * HR + 3 * (BB + HBP) - 2 * K}{IP} + constant$$

SIERA is another step up; it accounts for flyball and groundball rate as well in a very complex formula that is more descriptive than FIP. A final metric is xERA, which accounts for contact quality (measured as a combination of exit velocity and launch angle) as its measure of performance; this attempts to take out the luck that comes with outcome based results in baseball.

There have been many attempted methods to quantify a repertoire's variation or year-over-year change. A simple method (2 patrick brennan, 3 fangraphs) is to capture the randomness of a repertoire by calculating the entropy of a repertoire using each pitch's usage percentage. In information theory, entropy is calculated as:

$$H = -\sum_i p_i \, log(p_i)$$

Where p_i is the usage percentage of each pitch type *i* a pitcher throws.This metric gives higher entropies to repertoire's that have more even pitch usages across a larger amount of pitches. While this is an easily understandable metric, it does not contain a lot of information when examining change from one year to the next. It does not consider each pitch's movement and velocity, only the usage; because of this, it would categorize usage differences from a fastball to a cutter (two similar pitches in movement and velocity) and fastball to curveball (possibly the two most different pitches in baseball, being an objectively larger change) as the same. A similar metric that captures usage variation is gini index, which comes with the same shortcomings as entropy. These metrics fail to capture movement differences in the same pitch and the magnitude of change between different pictures, which are crucial to understanding a pitcher's true repertoire and their attempt at becoming a better pitcher. Likely the most sufficient metric for this analysis is Earth Mover's Distance; this metric quantifies the 'work' that is needed to reshape one distribution into another. This metric can be calculated using any feature space we

wish, but we stick to pitch velocity, horizontal movement, and vertical movement. In this article, the authors describe the strengths of this metric when it comes to analyzing repertoire changes, and they also apply several other steps in the data cleaning process to ensure more accurate data. The metric does an excellent job in finding the most extreme similarities and differences between pitchers, however I will use it to dive deeper into the impact of repertoire change on individual performance.

A recent paper looks at a similar idea by measuring variation in pitch sequencing using entropy-based metrics rather than year-to-year repertoire change. Shannon entropy and other measures are calculated from pitch sequences and compared to ERA and wins, in an effort to relate the variation of pitch sequencing to performance. Only small correlations were found using usage-based descriptors, which supports the usage of a metric that accounts for pitch characteristics, such as EMD.

Sources for background:

1) Stuff+, location+, pitching+ - Fangraphs

- Predictiveness of stuff,location,pitching+
- Description of each term and how powerful they are

2) Using pitch type and location entropy as descriptor or repertoire

- Describes one way to quantify/describe a repertoire
    - Only looks at usages, not type, movement, etc not the greatest

3) Pitch Mix Variation

- Attempt at describing total variation in a pitcher's repertoire
- Entropy, Gini index, custom metric

4) Baseball Prospectus: EMD/Wasserstein Metric

- Discussion of previous works that compare differences in pitches of same type
- Describes EMD - 'work' needed to move distribution to the new distribution (using movement and velocity profiles)
- Whitening transform to account for differences in variances/correlation of speed, horizontal,vertical movement
- Compare only against RHB to RHB and then combine
- Looks at specific year-to-year changes and differences in ERA
    - Extend this to look at advanced pitching metrics and larger sample size
    - Younger people change more, looking to find something that keeps them in the league