RSGC
Royal St. George's College

The Young Researcher

2023 Volume 7 | Issue 1

Inefficiencies in the Baseball Labour Market: An Analysis of Roster Construction Strategies of the Cleveland Guardians and the Tampa Bay Rays for 2022 MLB Season

Will Houslander

Recommended Citation
Houslander, W. (2023). Inefficiencies in the baseball labour market: An analysis of roster construction strategies of the Cleveland Guardians and the Tampa Bay Rays for 2022 MLB season. *The Young Researcher*, 7(1), 40-63. http://www.theyoungresearcher.com/papers/houslander.pdf

ISSN: 2560-9815 (Print) 2560-9823 (Online) Journal homepage: http://www.theyoungresearcher.com

Inefficiencies in the Baseball Labour Market: An Analysis of Roster Construction Strategies of the Cleveland Guardians and the Tampa Bay Rays for 2022 MLB Season

Will Houslander

Abstract: In this study, a three-step method was used to assess one main question: How have the Cleveland Guardians and the Tampa Bay Rays found playoff success despite their low payrolls? First, this study found the correlation between 29 different baseball statistics and MLB teams winning percentages during the 2022 season. The correlation was found using Pearson's correlation coefficient, specifically looking for statistics with strong correlations with r-values > 0.6 or > -0.6. Ten statistics were found to have a strong correlation to winning percentage in 2022. Next, a log-linear regression analysis was used to analyze which of the strongly correlated metrics were the most undervalued in the baseball labour market. Lastly, step three discovered whether the Rays and the Guardians value the undervalued statistics identified in step two. The results of this study indicate that both the Guardians and the Rays found success through their strong pitching staffs.

Keywords: baseball, baseball analytics, sabermetrics, market inefficiencies, low payroll, payroll strategies

Introduction

Michael Lewis' 2003 book *Moneyball* details how low-payroll teams like the Oakland Athletics could succeed despite their low payroll (Lewis, 2003). Billy Beane and the Oakland Athletics front office found success by exploiting inefficiencies in player evaluation in the MLB. They discovered that other MLB teams were undervaluing players who possessed a great ability to get on base but struggled to attain a high batting average (AVG) and slugging percentage (SLG), which were prominent statistics in player evaluation in the early 2000s (Brown, et al., 2017). Oakland found great success using these tactics, and for that reason, all teams, no matter the payroll, adopted Oakland's Moneyball tactics by 2019 (University of Toronto, 2019). With all teams quickly increasing their usage of Moneyball tactics, low-payroll teams such as the Oakland Athletics lost their competitive advantage, thus leading to high-payroll teams having vast success, as since 2004 all World Series winners have had an opening day payroll in the top half of the league in the year they won (Glaser, 2019). However, since 2016 both the Cleveland Guardians and the Tampa Bay Rays have made the playoffs at least three times in a row, all while keeping a payroll in the bottom half of the league (Baseball Reference, 2022). Out of 30 major league teams, Cleveland's payroll ranked in the bottom eight in three of their five playoff runs (Cot's Contracts, 2022). Meanwhile, Tampa Bay's payroll was even lower in all four of their postseason appearances. They consistently ranked in the bottom seven (Cot's Contracts, 2022).

Historically, researchers have directed their studies at discovering different ways that Michael Lewis'

Moneyball revolutionized the way MLB teams build their rosters. Caporale et al. (2013) discovered that *Moneyball* did not affect the teams' draft philosophy. Additionally, Chang and Zenilman (2013) studied the change in free-agent salaries since *Moneyball* and found that teams started paying players more based on Wins Above Replacement (WAR).

However, there has been little research on market inefficiencies in today's baseball market and if successful low-payroll teams like the Cleveland Guardians and the Tampa Bay Rays are exploiting these inefficiencies when constructing their rosters. Through a 3-step quantitative analysis, this study will aim to discover what statistics correlate the most to winning percentage, what market inefficiencies are present in today's baseball labour market, and whether the Guardians and the Rays leverage these inefficiencies. A three-step method was put in place to complete this study. The method will aim to answer the research question: What inefficiencies are there in today's baseball labour market and did the Cleveland Guardians and the Tampa Bay Rays value these statistics when constructing their rosters for the 2022 MLB season?

This study will help major league teams, agents, and writers know what statistics to look for when evaluating players. Additionally, this study will research ways that struggling organizations can still succeed with a low payroll, and will discover the proper way to utilize teams' limited money supply.

Literature Review

Correlation Between Statistics and Winning

A regression analysis conducted by Houser (2005) examined what baseball statistics correlated the most with a team's percentage of wins. The regression analysis included seven different statistics as explanatory variables. Just as Houser (2005) hypothesized, OBP along with WHIP had the highest correlation to a team's winning percentage. Furthermore, Frerker (2013) used Pearson's correlation coefficient to determine the correlation between a teams' winning percentage in the 2012 MLB season and their strikeout-to-walk ratio (K/BB), on-base percentage (OBP), and home runs. Frerker (2013) discovered a positive strong correlation between a team's K/BB ratio and its winning percentage as demonstrated by its 0.68 r-value. Both OBP and home runs had a r-value < 0.5, indicating that both have a moderately weak positive correlation. Step 1 of this study's method is based on Freker's (2013) method; however, this study provides insight into more than three statistics. Similar to Frerker (2013), Fullerton et al. (2014) tested the importance of earned run average (ERA), errors per game, and on-base plus slugging (OPS) on a team's success in his econometric analysis of the 2013 MLB season. Fullerton et al. (2014) discovered that the most successful teams in the 2013 season were teams that found a solid balance between pitching, hitting, and fielding (Fullerton et al., 2014).

MLB Market Inefficiencies

Lewis' (2003) *Moneyball* provided awareness of market inefficiencies that were present in the baseball labour market and how the Oakland Athletics exploited those inefficiencies. Following the successful reception of *Moneyball,* Hakes and Sauer (2006) conducted a linear regression analysis to confirm Lewis's (2003) hypothesis that OBP was undervalued in the baseball labour market. This study successfully validates Lewis's (2003) hypothesis as it found that from 2000-2004, on-base percentage was undervalued; however, in the years after *Moneyball* was released, the market abruptly corrected itself and valued OPS appropriately (Hakes and Sauer, 2006). Hakes and Sauer (2006) provided an insightful analysis into how the baseball industry responded to the Moneyball hypothesis; however, it failed to account for the fact that players with less service time make less money, even though they may provide more value for teams. Unlike Hakes and Sauer (2006), this study included players' service time as a control variable in regression analysis to ensure valid results. Baumer and Zimbalist (2014) expanded upon Hakes and Sauer's (2006) study and tested whether the changes in the labour market valuations due to Moneyball tactics would be permanent. Baumer and Zimbalist (2014) discovered that the baseball labour market's immediate response to *Moneyball* was unsustainable. Baumer and Zimbalist (2014) mark *Moneyball* as the catalyst in accelerating the process of making high on-base players more valuable. However, in the end, the immediate responses in the baseball labour market were just a part of the modest, long-term increase in high-OBP players in the MLB. Lastly, research done by Duquette et al. (2019) provides a modern-day view of the state of *Moneyball* 15 years after the book was published. Since the release of *Moneyball* in 2003, all major league baseball teams have shifted their payroll to align with its analytics (Duquette et al., 2019). Duquette et al. (2019) also discovered that the ability to draw walks may still be undervalued in the present-day labour market.

Moneyball's Impact on Baseball Operations

Since Michael Lewis' book was published in 2003, many studies have been done on how Moneyball tactics have impacted the way MLB front offices construct their teams. A study by Chang and Zenilman (2013) measured the impact of *Moneyball* on free agent salaries, and whether teams had started to award players based on different statistics. The focus of the study was on three specific years of free agency, 2000, 2005, and 2011. These years represent years before, immediately following, and eight years after the book was written. Chang and Zenilman (2013) discovered that front offices shifted their focus from players with great physical traits whom they valued highly pre-*Moneyball* to players who accumulate high WAR and OBP. In contrast, Caporale and Collier (2012) analyzed whether *Moneyball* affected a team's draft strategy. In Lewis's (2003) book, he advises that teams should focus more on drafting college players over high school players as they have more success in the MLB. Caporale and Collier's (2012) study challenges this theory, finding that the performance of college players in MLB is no better than the performance of high school players. Additionally, Caporale and Collier (2012) discovered that there had been no significant change in the team's draft strategy following *Moneyball*. After adopting the original Moneyball tactics, all major league teams transitioned to developing their own Moneyball theories, as evidenced by research conducted by Barrella (2018). Since 2012 the league has seen a 15% increase in the average number of relievers used per game. Barella (2018) concluded that this rise is a result of teams realizing they can pay relievers much less than starters, even though they provide the same value as starters just on fewer innings.

Gap Analysis

Current research on how certain baseball metrics correlate with a team's winning percentage has provided insight into the relationship between certain statistics and winning percentage, but it is not without limitations. The two main limitations are that many papers have relied on outdated and flawed metrics, and many have used a limited variety of metrics. For instance, Houser's (2005) study focused on how flawed statistics like OBP, batting average, and fielding percentage correlate to a team's winning percentage. Due to flaws in these major statistics, the results of Houser's research are significantly hindered. This study aims to use newer and more accurate statistics such as ISO (isolated power) and K/BB (strikeout-to-walk rate) that better represent a player's value and are expected to yield more valid results. In addition, most studies have failed to use a variety of statistics when examining the correlation between team performance in certain metrics and a team's winning percentage. Many studies including Frerker (2013), and Fullerton et al. (2014) only looked into how three statistics correlate to winning. In contrast, this study examines how a team's performance in 29 different metrics correlates with their winning percentage in the 2022 MLB season. I used domain knowledge to select these 29 metrics (listed in Appendix B) as they fully encompass players' performance.

Current research on market inefficiencies in the MLB labor market has increased understanding of how certain statistics are undervalued. However, these studies have not examined whether teams are taking advantage of undervalued statistics. For example, Duquette et al. (2019) found that the ability to draw walks was undervalued, but did not investigate whether any teams, specifically low-payroll teams were exploiting this inefficiency. This study aims to address these limitations by examining whether the Cleveland Guardians and the Tampa Bay Rays have successfully capitalized on market inefficiencies to achieve more postseason success. The findings will provide valuable insights for struggling low-payroll teams seeking to improve their roster construction strategies.

Method

A comprehensive and in-depth study is needed to evaluate the market inefficiencies present in today's baseball labour market. This study consisted of three carefully designed steps that provided valuable insights into the current state of the baseball market and its inefficiencies.

Step 1:

Step one builds on previous research and helps better understand what factors contribute the most to a team's success. Previous research, such as Frerker's (2013) study on strikeout to walk ratio (K/BB), on-base-percentage (OBP) and home runs, has shown that certain skills contribute more to a team's success. To build on Frerker's (2013) research, I collected data from the 2022 MLB season and used Pearson's correlation coefficient to determine the strength of the relationship between a team's performance in each metric and their winning percentage. Pearson's correlation coefficient generates an r-value that ranges from -1 (perfect negative correlation) to 1 (perfect positive correlation). From here I was able to identify which statistics are most important to predicting success and help teams understand what skills to focus on when building their rosters. This method is the most suitable for step one because it provides teams with an easy means of fully understanding how different statistics correlate to winning. By using domain knowledge, I collected 29 different statistics that provide a comprehensive view of a player's performance and are most relevant when attempting to predict team success. A full list of the statistics used with definitions can be found in Appendix B.

To fully understand how these statistics correlate with teams winning percentages, I collected team performance for all 29 statistics and their winning percentage in 2022. The strategies and tactics used by teams to score and prevent runs tend to remain consistent from season to season, making it sufficient to evaluate the relationship between baseball metrics and winning percentages using data from a single season. In this study, I have chosen to focus on data from the 2022 MLB season to provide a comprehensive and up-to-date analysis. The winning percentage data was collected from MLB.com, while the team performance statistics were collected from Fangraphs. Once all the data was collected, it was exported from Fangraphs and sorted into Google Sheets. Google Sheets was used to ensure that all the data collected is stored safely. The results for this step are presented primarily on graphs showing the correlation between the two variables. The results can be found in the "Findings Section" and "Appendix A."

Step 2:

Step two identified undervalued statistics in the current baseball labour market. To do this, I used a log-linear regression model similar to the one used by Baumer and Zimbalist (2014) in their study on the relationship between player salaries and performance metrics. My log-linear regression model used player salaries as the dependent variable, and all of the statistics used in step one as the independent variables. However, unlike step one, all statistics were individual player statistics from the 2022 MLB season. In total, three different log-linear regression models were run: one for position players, one for starting pitchers, and one for relief pitchers. Only arbitration-eligible players or free-agent-eligible players were included in the model, as they are the only players with some control over their salary. To create the regression model, I first collected data on player salaries from a reliable source, Cots Contracts (2022). Then, I collected individual player data on all of the statistics in step one from Fangraphs. When all the data was collected, I conducted a log-linear regression model using the following equations:

Position Player

ln(Salary) = β0 + β1·PA + β2·G + β3·BB/K + β4·BB% + β5·K% + β6·ISO + β7·Contact% + β8·Hard% + β9·LD% + β10·Barrel% + β11·GB% + β12·IFFB% + β13·Soft% + β14·DRS + β15·UZR + β16·OAA + β17·Catcher + β18·Infielder + β19·Outfielder + β20·ArbEligible + β21·Free Agent

Pitchers

ln(Salary) = β0 + β1·G + β2·IP + β3·BB% + β4·K% + β5·K/BB + β6·K-BB% + β7·SwStr% + β8·CSW% + β9·HARD% + β10·Soft% + β11·LD% + β12·Barrel% + β13·IFFB% + β14·GB% + β15·O-Swing%+ β16·Contact% + β17·ArbEligible + β18·FreeAgent β represents the coefficient value. This indicates the fractional increase in a player›s expected salary if the statistic corresponding to the particular β value increases by one unit while all other statistics remain constant. This regression analysis technique is found in papers such as Baumer and Zimablist (2014) and Hakes and Sauer (2006). Similar to Baumer and Zimbalist (2014), I used control variables to attempt to include all possible factors that go into a player›s salary. Variables were coded as 1 if they fell within the assigned category, and 0 if they did not. The control variables include:

1. ArbEligible - Is the player arbitration eligible? Their service time is between three and six years.
2. FreeAgent - Has this player ever been a free agent? Their service time is above six years.
3. Catcher - Is this player a catcher? Catchers are typically valued differently because fielding is more important than hitting for catchers.
4. Infielder - Is this player an infielder?
5. Outfielder - Is this player an outfielder?

Control variables 3-5 were only used in the position players regression model.

The control variable "DesignatedHitter" is not included to ensure there is not a problem with multicollinearity. Multicollinearity occurs when independent variables are correlated. Excluding a fourth category of "DesignatedHitter" avoids having correlation of all four control variables: for example, if a player was not an "Outfielder" they would have to fall into one of the other three categories. However, by excluding "DesignatedHitter," there is a possibility that a player falls into none of the three control variables. The results of this section showed how an increase of one statistical unit affects a player's expected salary. This information will help guide the team's decision-making regarding player acquisitions. The results of the regression models are presented through tables that can be found in the "Findings" section.

Step 3

Lastly, step three of my method aimed to determine whether successful low-payroll teams, such as the Cleveland Guardians and the Tampa Bay Rays, value the undervalued statistics identified in step two. To accomplish this, I collected data on the performance of these teams across the 29 different statistics used in step one. I then analyzed whether Cleveland and Tampa Bay prioritized the undervalued statistics by examining which statistics they ranked in the top ten in 2022. If any of their top ten statistics aligned with the undervalued statistics, it would suggest that these teams were leveraging market inefficiencies in the baseball labour market to construct competitive rosters. Step three of this study sheds light on the strategies employed by successful low-payroll teams, providing insights that can inform how teams should approach the construction of their roster.

Findings

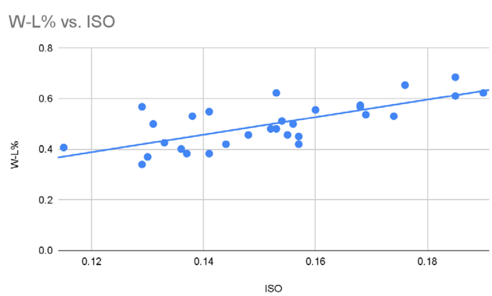
Step 1

Definitions of all statistics can be found in Appendix B

Position Players

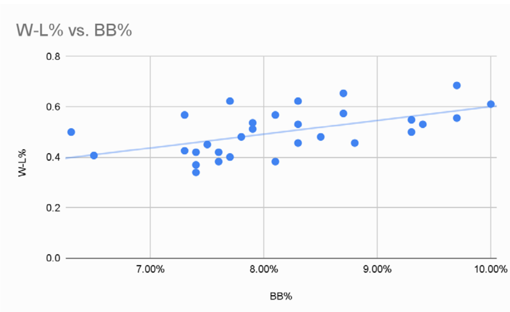
The analysis of MLB data from the 2022 season revealed that three position player statistics are strongly correlated with winning percentage. All three of the statistics had an r-value of greater than 0.6 or -0.6 which aligns with the definition I followed.

ISO (r =0.72)
BB/K (r =0.69)
DRS (r = -0.66)



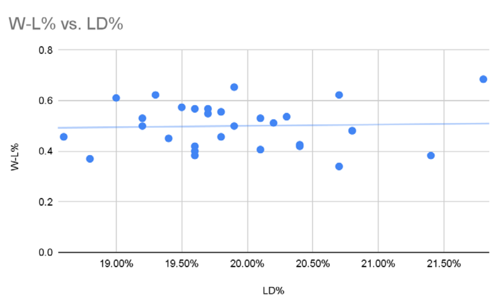
Additionally, I discovered that several other position players' statistics had a moderately strong relationship to teams winning percentages in the 2022 MLB season. These statistics include:

BB% (r = 0.56)
Barrel% (r = 0.55)
GB% (r = -0.50)
OAA (r = 0.48)
HARD% (r = 0.44)
K% (r = -0.41)



Lastly, there were six statistics that I determined had a weak correlation to a team's winning percentage in 2022, as they all had an r-value of less than 0.4 or -0.4. These statistics include:

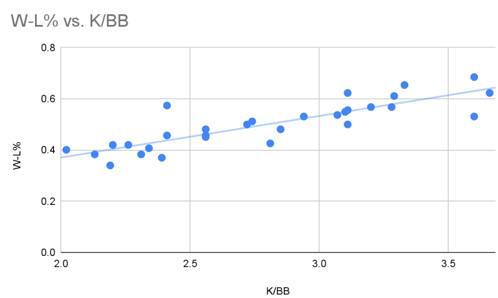
Soft% (r = -0.33)
Contact% (r = 0.30)
Bsr (r = 0.21)
UZR (r = 0.16)
IFFB% (r = -0.1)
LD% (r = 0.043)



Pitchers

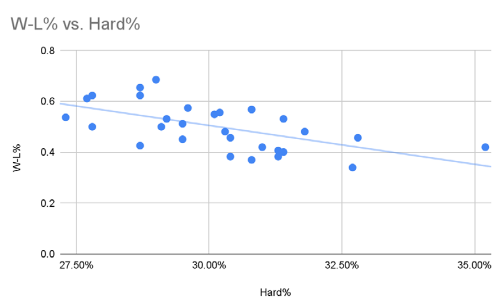
I found that seven different pitching statistics strongly correlated to teams winning percentages in the 2022 MLB season. These seven statistics are:

K/BB (r = 0.85)
K-BB% (r = 0.84)
K% (r = 0.74)
BB% (r = -0.72)
SwStr% (r = 0.67)
O-swing% (r = 0.64)
CSW% (r = 0.64)



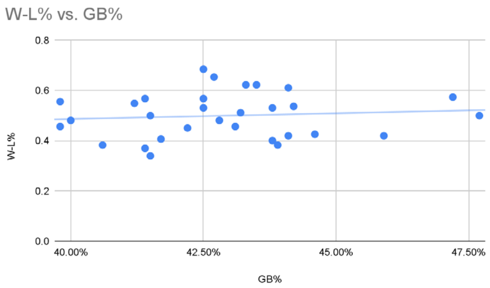
Next, I discovered that five pitching statistics exhibited a moderately strong correlation to teams' winning percentages in the 2022 MLB season. Those statistics were:

Soft% (r = 0.59)
Contact% (r = 0.58)
HARD% (r = -0.58)
Barrel% (r = -0.54)
LD% (r = -0.44)



Finally, only two statistics showed a weak correlation to teams' winning percentages in the 2022 MLB season. These two statistics were:

IFFB% (r = 0.34)
GB% (r = 0.098)



Appendix A provides graphs showing the correlations between each statistic and winning percentage.

Position Players

When looking into how position player statistics are valued in the MLB labour market, my log-linear regression model had an r^2 value of 0.61. Table 1 highlights how the position player statistics with a strong correlation to winning percentage are valued in the 2022 market.

Starting Pitchers

When looking into how starting pitcher's statistics are valued in the MLB labour market, the log-linear regression model had an r^2 value of 0.41. Table 2 shows how the starting pitching statistics with a strong correlation to winning percentage are valued in the 2022 market.

Relief Pitchers

When looking into how relief pitcher's statistics are valued in the MLB labour market, my log-linear regression model had an r^2 value of 0.16. Table 3 found how the relief pitcher's statistics with a strong correlation to winning percentage are valued in the 2022 MLB labour market.

<u>Step 2:</u>
Definitions of all terms mentioned are in Appendix C

	BB/K	ISO	DRS
Coefficient Value	0.75	0.01	-0.99
Coefficient s.error	0.49	0.001	1.47

Table 1 - How Position Player Statistics are Valued in the Baseball Labour Market

	K/BB	K-BB%	K%	BB%	SwStr%	O-Swing%	CSW%
Coefficient Value	12.99	4.35	-1.34	0.93	1.86	0.009	1.38
Coefficient s.error	7.72	5.02	3.78	2.27	5.24	0.008	3.33

Table 2 - How Starting Pitcher Statistics are Valued in the Baseball Labour Market

	K/BB	K-BB%	K%	BB%	SwStr%	O-Swing%	CSW%
Coefficient Value	-4.74	-0.38	-5.15	-1.24	1.44	0.02	1.75
Coefficient s.error	4.66	2.48	2.35	1.29	3.03	0.01	2.27

Table 3 - How Relief Pitcher Statistics are Valued in the Baseball Labour Market

Step 3:
Tampa Bay Rays 2022 Performance:

Position Player Performance:

In 2022, the Rays were not in the top ten of any of the 15 position player metrics that I used to analyze their performance.

Plate Discipline/Vision	Performance	League Rank
K%	23.20%	19th
BB%	8.30%	12th
BB/K	0.36	15th
O-Swing%	32.50%	13th
Contact%	75.40%	23rd
Batted Ball Data		
LD%	20.10%	11th
GB%	44.40%	23rd
IFFB%	35.50%	21st
Soft%	16.30%	13th
Hard%	28.90%	25th
Barrel%	6.10%	26th
Advanced Stats		
ISO	0.138	22nd
BSR	4	11th
DRS	15	14th
OAA	2	15th

Figure 1 - Tampa Bay Rays Position Player Performance in 2022 Season

Starting Pitchers Performance

In the 2022 MLB season, the Tampa Bay Rays starting pitching staff was in the top ten in a staggering ten of the 14 pitching metrics I reviewed. A yellow highlight represents a top 10 finish in that metric.

Advanced	Performance	League Rank
BB%	5.60%	1st
K%	23.60%	10th
K-BB%	18%	4th
K/BB	4.23	1st
Batted Ball Data		
GB%	42.70%	14th
LD%	19.80%	9th
IFFB%	37.50%	14th
Hard%	29.90%	11th
Soft%	15.90%	18th
Barrel%	7.10%	8th
Put Away Stuff		
O-Swing%	36.40%	1st
SwStr%	12.70%	1st
Contact%	74.80%	4th
CSW%	29.10%	3rd

Figure 2 - Tampa Bay Rays Starting Pitchers Performance in 2022

Relief Pitchers Performance

The relief pitchers for the Tampa Bay Rays ranked in the top ten of 7/14 pitching metrics in the 2022 season.

Advanced	Performance	League Rank
BB%	7.50%	1st
K%	23.10%	18th
K-BB%	15.70%	8th
K/BB	3.1	5th
Batted Ball Data		
GB%	42.30%	23rd
LD%	19.40%	13th
IFFB%	12.80%	2nd
Hard%	28.30%	11th
Soft%	17.70%	9th
Barrel%	7.60%	25th
Put Away Stuff		
O-Swing%	34.70%	3rd
SwStr%	11.90%	14th
Contact%	75.70%	21st
CSW%	28.70%	10th

Figure 3 - Tampa Bay Rays Relief Pitchers Performance in 2022

Cleveland Guardians 2022 Performance

Position Player Performance

In the 2022 MLB season the Guardians excelled in 6/15 position player statistics; however, they were well below average in almost all of the 11 remaining statistics.

Plate Discipline/Vision	Performance	League Rank
K%	18.20%	1st
BB%	7.30%	28th
BB/K	0.4	9th
O-Swing%	34.20%	25th
Contact%	80.80%	1st
Batted Ball Data		
LD%	19.70%	18th
GB%	44.40%	22nd
IFFB%	11.20%	24th
Soft%	18.10%	27th
Hard%	26.30%	29th
Barrel%	4.90%	30th
ISO	0.129	28th
BSR	13.2	4th
DRS	79	3rd
OAA	19	6th

Figure 4 - Cleveland Guardians Position Player Performance in 2022

Starting Pitching Performance

In the 2022 MLB season, the Cleveland Guardians had an average starting pitching staff as they finished in the top ten in 3/14 pitching metrics.

Advanced	Performance	League Rank
BB%	6.50%	7th
K%	21.30%	17th
K-BB%	14.80%	13th
K/BB	3.28	9th
Batted Ball Data		
GB%	40.30%	24th
LD%	20.50%	24th
IFFB%	39.20%	20th
Hard%	32.00%	22nd
Soft%	15.00%	26th
Barrel%	8.70%	25th
Put Away Stuff		
O-Swing%	34.40%	5th
SwStr%	11.20%	14th
Contact%	77.30%	14th
CSW%	26.60%	20th

Figure 5 - Cleveland Guardians Starting Pitchers Performance in 2022

Relief Pitchers Performance

A huge reason for the Cleveland Guardians success in 2022 was because of their strong relief pitchers' performance. They ranked in the top ten of 11/14 pitching metrics.

Advanced	Performance	League Rank
BB%	8.60%	9th
K%	26.40%	6th
K-BB%	17.80%	6th
K/BB	3.08	6th
Batted Ball Data		
GB%	46.40%	4th
LD%	18.60%	5th
IFFB%	10.90%	11th
Hard%	28.70%	14th
Soft%	16.00%	7th
Barrel%	7.30%	21st
Put Away Stuff		
O-Swing%	33.70%	5th
SwStr%	13.00%	4th
Contact%	73.70%	6th
CSW%	28.70%	8th

Figure 6 - Cleveland Guardians Relief Pitchers Performance in 2022

Discussion

The main takeaways from my findings highlight five key items of importance:

1. The importance of power and plate discipline on team success;
2. The importance of pitchers limiting walks and getting swings-and-misses on team success;
3. Batted ball data is not important for team success;
4. Power and swing-and-miss abilities are undervalued;
5. Tampa Bay and Cleveland found success through strong pitching.

Importance of Power and Plate Discipline on Team Success

The strong positive correlation of r= 0.69 between a team's BB/K and their winning percentage indicates the importance of batters' plate discipline on the team's success. A high BB/K ratio indicates that a team's hitters are patient and can wait for the pitches they can hit or take walks when necessary. In addition, teams with a high BB/K have an easier time creating runs as they are less reliant on getting hits to drive in the runs. Additionally, the strong positive correlation of r= 0.72 between a team's ISO and their winning percentage also demonstrates the importance of extra-base power on their success. Teams who fail to generate enough extra-base power tend to struggle to generate quick and easy runs, whereas teams with a high ISO do not have to rely as much on stringing singles together to generate enough runs to win. My findings reveal a stark contrast to Murray's (2022) results, which indicate a moderate correlation between team success and ISO, and a weak correlation to BB/K. In contrast, my research demonstrates a strong correlation between both ISO and BB/K and team success. Overall, the strong positive correlations observed between batters' BB/K and a team's success, as well as between a team's ISO and their winning percentage, underscore the crucial role that patient, power-hitting batters play in the successful outcomes of MLB teams.

Importance of Pitchers Limiting Walks and Getting Swings-and-Misses on Team Success

The strong positive correlation of r= 0.67 and r= 0.64 between a team's performance in SwStr% and O-swing% exemplifies the importance of generating swings-and-misses on an MLB team's success. A high SwStr% and O-swing% suggest that pitchers are inducing more swings-and-misses from opposing batters, which is a sign of strong pitch quality and can lead to more strikeouts. The strong correlation highlights the importance of effective pitching strategies that prioritize high quality pitches to get swings and misses. Next, my study revealed a strong positive correlation between a team's winning percentage and their performance in K/BB, with a coefficient value of r= 0.85. A high K/BB ratio is indicative of a pitcher's ability to control their pitches effectively, limiting walks while recording a high number of strikeouts. Consequently, teams with pitchers who possess these skills are more likely to achieve success. This study's findings align with those of Frerker (2013) and Murray (2022), demonstrating a consistent and robust association between a team's K/BB performance and their success. The consistency across different years reinforces my initial hypothesis that the strategies and tactics employed by baseball teams to score and prevent runs tend to remain stable over time.

Batted Ball Data is not Important to Team Success

For both batters and pitchers batted ball data has very little importance in determining team success. This is evident as batted ball metrics such as GB%, HARD%, and Soft% all had r-values <0.6 for both batters and pitchers. The low r-values in all batted ball data for batters suggest that teams still can succeed even if they hit more ground balls, or do not hit the ball as hard as the best teams. For pitchers, the low r-values in all the batted ball data imply that while a team's performance in batted ball metrics is important to consider, it is not the only factor that should be taken into account when assessing a team's overall performance. Other variables such as opponents' skill, injuries, and fielding also play a significant role in determining a team's success. Once again, my study's findings align with those of Murray's (2022) study as they also found a moderate or weak correlation between all batted ball statistics including GB%, LD%, and HARD% on teams W%. As a result of the similarity of mine and Murray's (2022) findings, it can certainly be concluded that a team's batted ball data is not a good indicator of success.

Power and Generating Swing and Miss is Undervalued

Even though ISO showed the strongest correlation to teams winning percentage in 2022, it was also deemed the most undervalued statistic in the current MLB labour market. A one-unit increase in ISO only results in a 0.01% increase in a player's expected salary. Additionally, ISO's s.error of 0.001 would mean that the estimated coefficient for ISO has a high degree of precision. A low s.error suggests that I can comfortably conclude that ISO is the most undervalued position player statistic that is strongly correlated to the teams' winning percentages. The results of this analysis suggest that teams should place a greater emphasis on acquiring players with a high ISO, as they may be able to acquire players with high power production at a lower cost. Next, for both starting pitchers and relief pitchers, O-swing% was deemed to be the most undervalued metric in the baseball labour market. For starting pitchers, a one-unit increase in O-swing% resulted in only a 0.009% increase in expected salary. Additionally, the s.error for this coefficient was 0.008, indicating that the coefficient estimate is reliable. For relief pitchers, a one-unit increase in O-swing% resulted in a decrease of a player's expected salary by 0.02%. In addition, an s.error of only 0.01 suggests that I can trust these results. However, I may not be able to trust the results for the remainder of the log-linear regression for relief pitchers as it only has an R^2 value of 0.16. A low R^2 value indicates that many factors are missing from this regression analysis that are important in the valuation of relief pitchers. Overall, this analysis suggests that both ISO and O-swing% are undervalued statistics in the MLB labour market, and acquiring players with high values in these statistics may be a cost-effective strategy for teams. Additionally, while the estimated coefficients for these statistics have a high degree of precision and reliability, it is important to note that other factors may also be important in evaluating player performance and determining player value.

Tampa Bay and Cleveland Found Success Through Strong Pitching

In the 2022 MLB season, both the Cleveland Guardians and the Tampa Bay Rays found incredible success with their strong pitching staffs despite low payrolls (Baseball Reference, 2022). At this time, both teams' payrolls were in the bottom eight of 30 (Spotrac, 2022). Tampa Bay placed a strong focus on their starting pitching as evident through their ten, top-ten finishes in the fourteen pitching metrics reviewed. Interestingly, the Rays did show signs of taking advantage of the market inefficiencies found in step two as they ranked first in O-swing% for starting pitchers. The Rays' strong performance in O-swing% suggests that they may have capitalized on the market inefficiency to acquire starting pitchers, with great ability to get swing-and-miss for a low price. This strategy is apparent in the Rays 2022 off-season signing of Corey Kluber for $8 million. Before signing with the Rays, Kluber ranked in the 80th percentile in O-Swing% (Baseball Savant, 2022). And in 2022, with the Rays, he ranked in the 98th percentile in O-Swing% (Baseball Savant, 2022). While other teams focused on Kluber's flaws such as his low fastball velocity, the Rays may have been focusing more on his high chase rate. Next, in the 2022 season, the Cleveland Guardians focused heavily on having a strong bullpen. This is apparent through eleven top-ten finishes in the 14 pitching metrics observed. Similar to the Rays, the Cleveland Guardians also seemed to emphasize leveraging market inefficiencies as both their starters and their relievers ranked 5th in O-swing%. Alongside achieving pitching success, Tampa Bay and Cleveland have avoided signing players based on batted ball data. The batted ball variables for these two teams do not just rank outside the top ten in the league; they are very near the bottom. Efficient hiring not only means signing players with undervalued characteristics, it also requires avoiding players who are more expensive because of overvalued characteristics like their batted ball data. It can be concluded that low-payroll teams should prioritize pitching over offense when constructing their rosters as both Cleveland and Tampa Bay have consistently had success doing so. Overall, the success of the Cleveland Guardians and the Tampa Bay Rays despite their low payrolls in the 2022 MLB season suggests that there may be market inefficiencies in the baseball labour market that can be leveraged by struggling low-payroll teams to acquire talented players for a lower cost. This also indicates that low-payroll teams should prioritize pitching over offense when constructing their rosters.

Conclusion

Limitations

To evaluate the reliability of my findings, it is important to state the limitations of the study. Two main limitations exist: only one season of data was used, and I am unaware of the thought process of Tampa Bay's and Cleveland's front offices. For step two of my method, I used only one year of player statistics due to having limited time to complete this study. This limited me from knowing whether the undervalued statistics I found in the baseball labour market were consistent from previous years. Secondly, without actually consulting with people from the front offices of either Tampa Bay or Cleveland, I cannot conclude whether or not they were leveraging the market inefficiencies I discovered in step two of my method. I was only able to hypothesize based on my findings.

Implications

MLB teams, agents, and writers can all use this paper's findings to better understand how to properly evaluate players. I show that statistics such as ISO, BB/K, and DRS for position players and K/BB, O-swing%, and SwStr% for pitchers correlate the strongest to a team's success. Therefore, instead of using more traditional statistics such as BA, RBIs, or ERA to evaluate players, teams should focus more on players' performance in the statistics which correlate most strongly to winning. Additionally, my findings have helped discover further insight into how low-payroll teams can succeed in modern baseball. Struggling low-payroll teams like Pittsburgh, Cincinnati, and Baltimore are made aware of strategies that successful low-payroll teams like Cleveland and Tampa Bay have used to find success. One of these strategies includes focusing heavily on pitching, instead of focusing heavily on offence. This is because 7/10 statistics strongly correlated to winning percentage were pitching statistics. Additionally, both the Guardians and the Rays have found the bulk of their success through strong pitching staffs. In the future, low payroll teams will also now know that the best way to gain an advantage in team building is by searching for market inefficiencies and undervalued statistics in the labour market. Moreover, all MLB teams are now aware of the market inefficiencies that are present in the current labour market. This information will allow the market to correct itself and grant current undervalued players the opportunity to get paid fairly.

Future Direction

Future studies should use data from more than one season when testing for market inefficiencies as Baumer and Zimablist (2014) did in their study. This will allow readers to understand how the labour market fluctuates, and if any statistics have consistently been undervalued. Furthermore, if possible, future studies could conduct interviews with front-office members from successful low-payroll teams to get added certainty about their team-building strategies.

References

Barella, N. (2018) Moneyball II: A new idea for MLB roster construction. https://www.linkedin.com/pulse/moneyball-ii-new-idea-mlb-roster-construction-nicholas-barella/

Baseball-Reference. (2022). World Series and MLB playoffs. https://www.baseball-reference.com/postseason/

Baumer, B., & Zimbalist, A. (2014). Quantifying market inefficiencies in the baseball players' market. Eastern Economic Journal, 40(4), 488-498. https://www.researchgate.net/publication/265647013_Quantifying_Market_Inefficiencies_in_the_Baseball_Players'_Market

Baseball Savant. (2022). Corey Kluber's statcast, visuals, and advanced metrics. https://baseballsavant.mlb.com/savant-player/corey-kluber-446372?stats=statcast-r-pitching-mlb

Brown, D. T., Link, C. R., & Rubin, S. L. (2017). Moneyball after 10 years: How have Major League Baseball salaries adjusted? *Journal of Sports Economics*, 18(8), 771-786. https://doi.org/10.1177/1527002515609665

Caporale, T., Collier, T. (2013). Scouts versus Stats: The impact of Moneyball on the Major League Baseball draft. https://www.researchgate.net/publication/254230987_Scouts_versus_Stats_The_impact_of_Moneyball_on_the_Major_League_Baseball_draft

Chang, J., Zenilman, J. (2013). A study of sabermetrics in Major League Baseball: The impact of Moneyball on free agent salaries. *Unpublished dissertation, Washington University in St. Louis.* https://olinblog.wustl.edu/wp-content/uploads/AStudyofSabermetricsinMajorLeagueBaseball.pdf

Cot's Contracts. (2022). Opening Day Salaries 2000 – 2022. https://legacy.baseballprospectus.com/compensation/cots/

Duquette, C. M., Cebula, R. J., & Mixon Jr, F. G. (2019). Major league baseball's Moneyball at age 15: a re-appraisal. Applied Economics, 51(52), 5694-5700. https://www.tandfonline.com/doi/full/10.1080/00036846.2019.1617399

Fangraphs. (2014). Complete list (Offense). https://library.fangraphs.com/offense/offensive-statistics-list/

Fangraphs (2014). Complete list (Pitching). https://library.fangraphs.com/pitching/complete-list-pitching/

Frerker, M. (2013). Measuring correlation between statistics and wins in Major League Baseball. https://courses.cs.washington.edu/courses/cse140/13wi/projects/frerkerm-report.pdf

Fullerton, S. L., Fullerton Jr, T. M., & Walke, A. G. (2014). An econometric analysis of the 2013 major league baseball season. https://www.aabri.com/manuscripts/141841.pd

Glaser, K. (2019). World Series champion opening day payroll ranks in the Wild Card era. Baseball America. https://www.baseballamerica.com/stories/world-series-champion-opening-day-payroll-ranks-in-the-wild-card-era/

Hakes, J. K., & Sauer, R. D. (2006). An economic evaluation of the Moneyball hypothesis. Journal of Economic Perspectives, 20(3), 173-186. https://www.aeaweb.org/articles?id=10.1257/jep.20.3.173

Hamilton, D. F., Ghert, M., & Simpson, A. H. R. W. (2015). Interpreting regression models in clinical outcome studies. Bone & Joint Research, 4(9), 152-153. https://boneandjoint.org.uk/article/10.1302/2046-3758.49.2000571

Houser, A. (2005). Which baseball statistic is the most important when determining team success. The Park Place Economist, 13(1), 29-36. https://digitalcommons.iwu.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1247&context=parkplace#:~:text=Teams%20with%20better%20offensive%20statistics,statistic%20in%20determining%20team%20wins

Lewis, M. (2003). Moneyball: The art of winning an unfair game.

Murray, B. (2022). MLB winning percentage breakdown: which statistics help teams win more games? *Samford University.* https://www.samford.edu/sports-analytics/fans/2022/MLB-Winning-Percentage-Breakdown-Which-Statistics-Help-Teams-Win-More-Games#:~:text=He%20claimed%20runs%20are%20the,allow%20more%20or%20less%20runs.

Siegel, A., & Wagner, M. (2022). Standard error of estimate. *Science Direct.* https://www.sciencedirect.com/topics/mathematics/standard-error-of-estimate

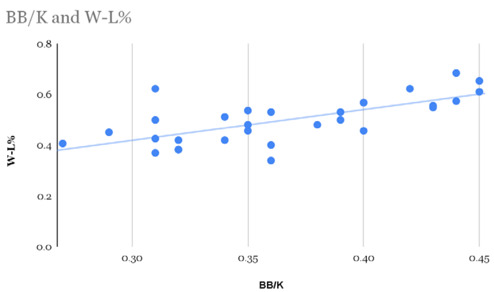
Spotrac. (2022). MLB team payroll tracker. https://www.spotrac.com/mlb/payroll/2022/

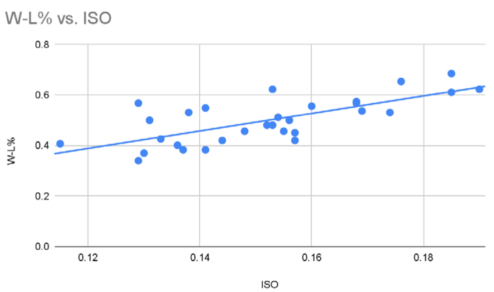
University of Toronto, Rotman School of Management. (2019). Moneyball advantage peters out once everyone's doing it. Science Daily. www.sciencedaily.com/releases/2019/04/190408161654.htm

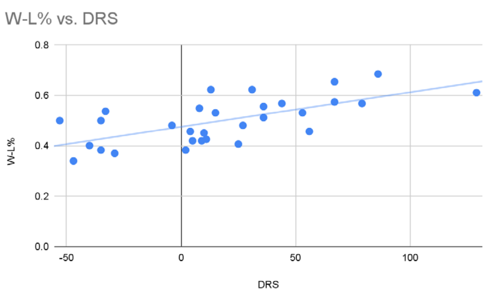
Appendix A

Position Players:

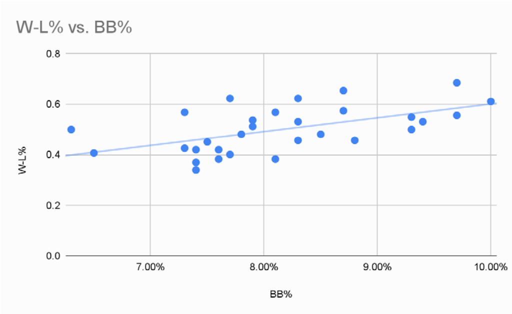
Strong Correlation Graphs:

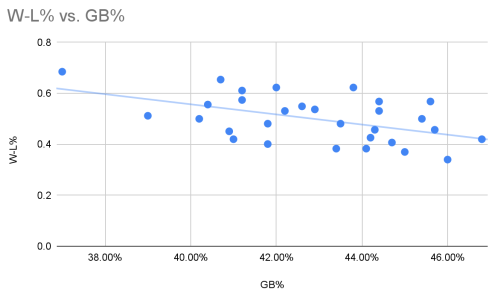


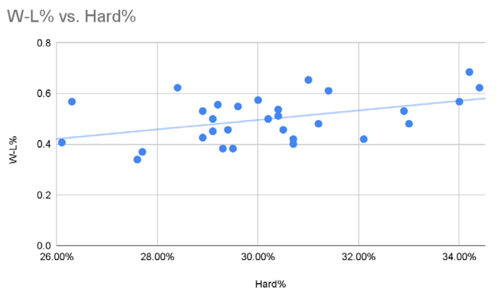




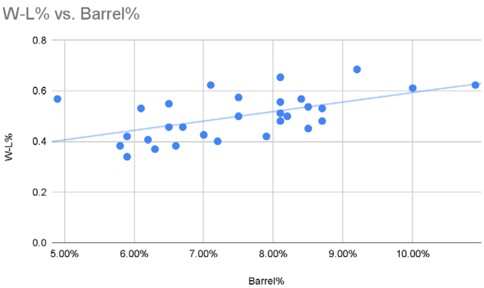
Moderately Strong Correlation Graphs

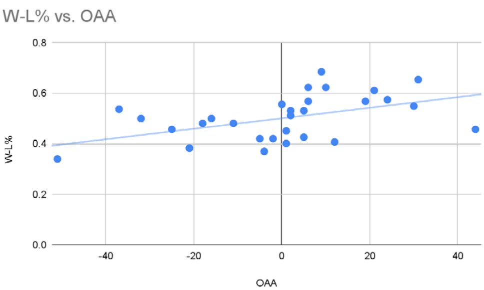


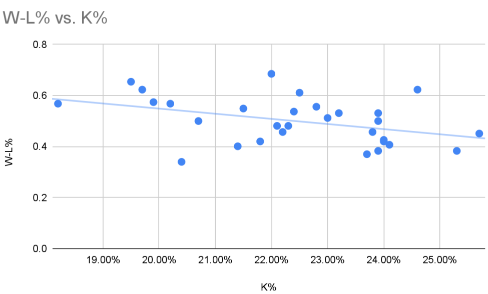


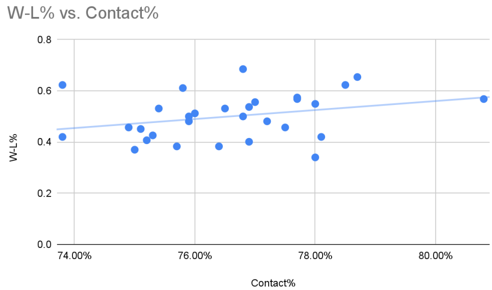


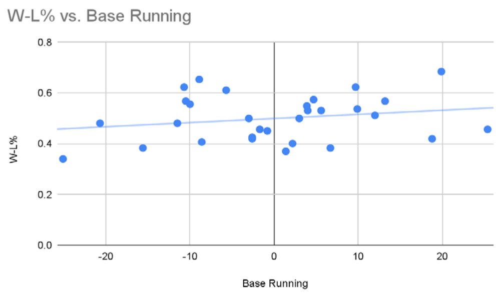
Weak Correlation Graphs

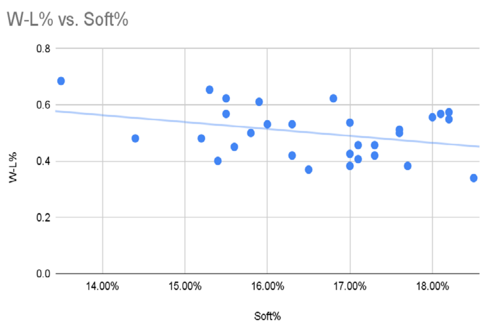


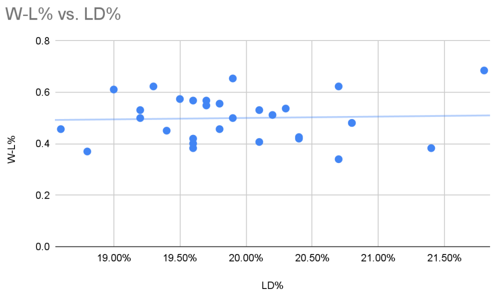


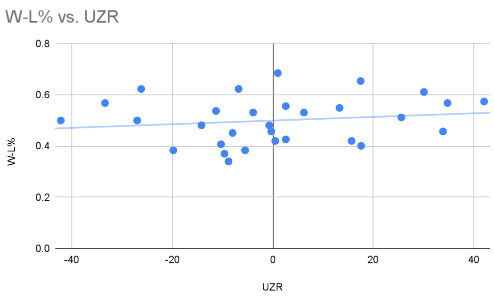


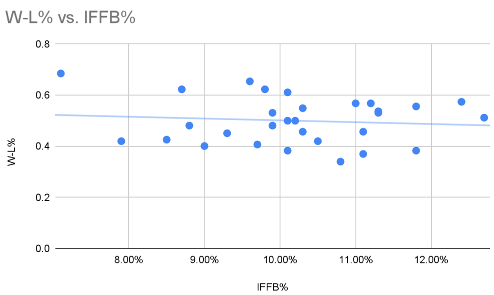






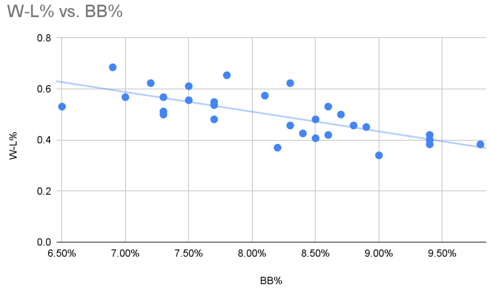


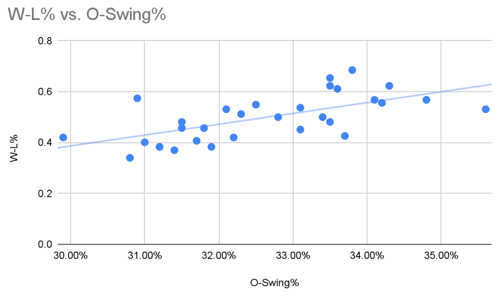


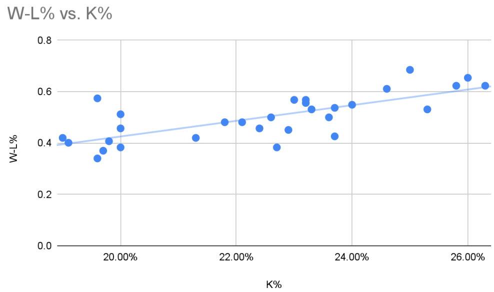


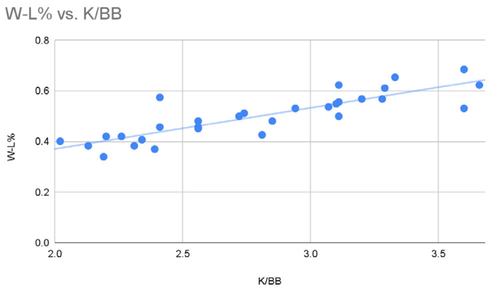
Starting Pitchers

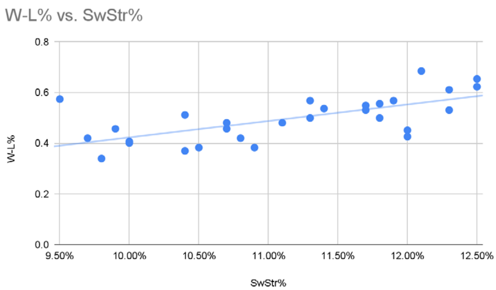
Strong Correlation

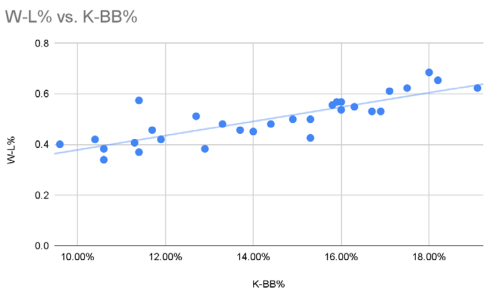


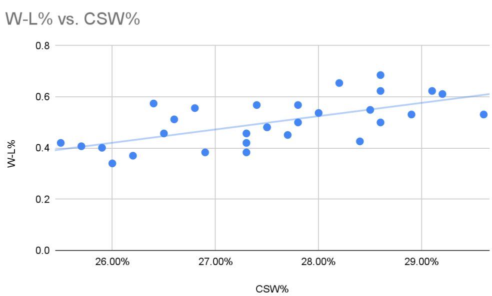




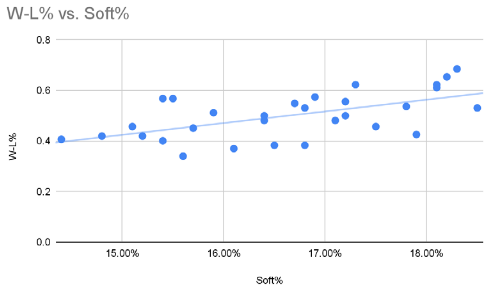


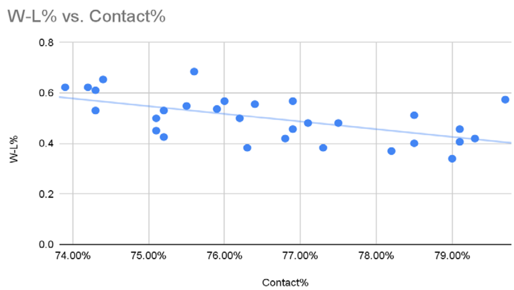


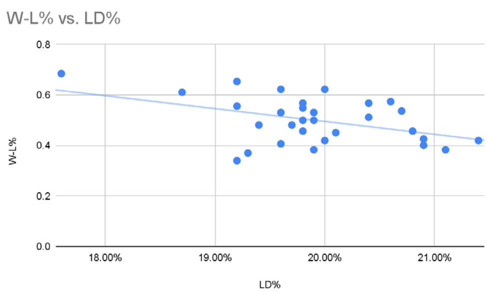


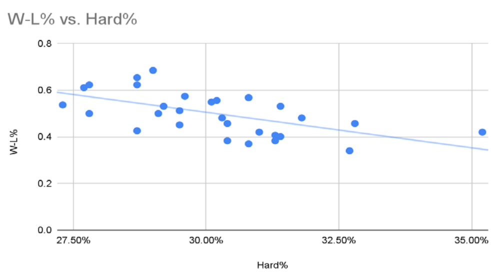


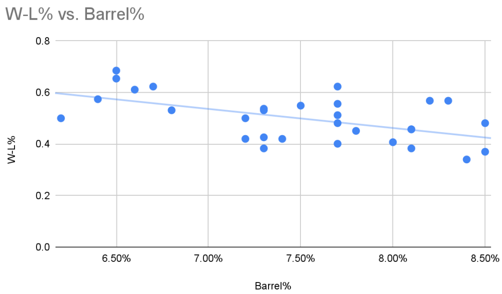
Moderately Weak Correlation



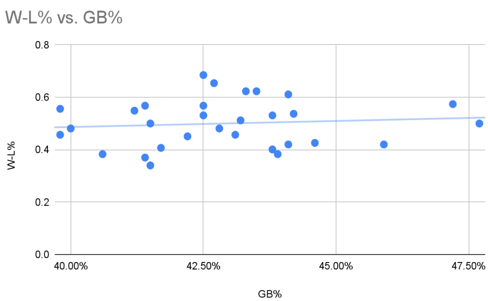


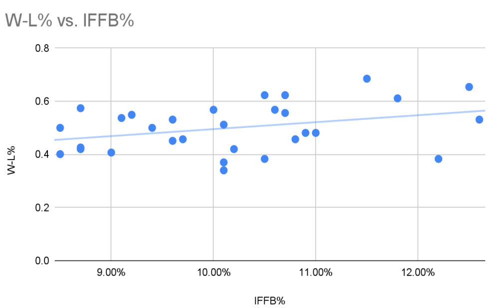






Weak Correlation





Appendix B

Definition of Variables (and their desired value)

Position Players:
Plate Discipline/Vision
1. K% (total strikeouts/plate appearances) – Lower K% is desirable.
2. BB% (total walks/plate appearances) – Higher BB% is desirable.
3. BB/K (total walks/total strikeouts) – Higher BB/K is desirable.

Batted Ball Data
1. LD% (total line drives/balls in play) – Higher LD% is desirable.
2. GB% (total ground balls/balls in play) – Lower GB% is desirable.
3. IFFB% (total infield fly balls/fly balls) – Lower IFFB% is desirable.
4. Soft% (percentage of soft-hit balls) – Lower Soft% is desirable.
5. HARD% (percentage of balls hit 95 mph +) – Higher HARD% is desirable.
6. Barrel% (total balls hit 98 mph +/balls in play) – Higher Barrel% is desirable.

Advanced Statistics
1. ISO (SLG – AVG) – Higher ISO is desirable.
2. Bsr (baserunning runs above average) – Higher Bsr is desirable.
3. OAA (outs above average) – Higher OAA is desirable.
4. DRS (defensive runs saved) – Higher DRS is desirable.
5. UZR (defensive runs above average) – Higher UZR is desirable.

Pitchers:
Advanced Statistics
1. BB% (walks/batters faced) – Lower BB% is desirable.
2. K% (total strikeouts/batters faced) – Higher K% is desirable.
3. K/BB (total strikeouts/total walks) – Higher K/BB is desirable.
4. K-BB% (K% - BB%) – Higher K-BB% is desirable.

Batted Ball Data
1. LD% (total line drives/total balls in play) – Lower LD% is desirable.
2. GB% (total ground balls/total balls in play) – Higher GB% is desirable.
3. IFFB% (total infield fly balls/total balls in play) – Higher IFFB% is desirable.
4. Soft% (percentage of soft-hit batted balls against) – Higher Soft% is desirable.
5. HARD% (percentage of balls hit 95 mph + against) – Lower HARD% is desirable.
6. Barrel% (total balls hit 98 mph+/total balls in play) – Lower Barrel% is desirable.

Put Away Stuff
1. O-Swing% (swings at pitches outside the zone /pitches outside the zone) – Higher O-Swing% is desirable.
2. SwStr% (swings and misses/total pitches) – Higher SwStr% is desirable.
3. Contact% (number of pitches on which contact was made/swings) – Lower Contact% is desirable.
4. CSW% (called strikes + swing strikes/total pitches) – A higher CSW% is desirable

All definitions are from Fangraphs (2014)

Appendix C

Definitions for Step 2 of Method

R^2: r^2 is a percentage value that measures how well a linear regression model fits the observed data (Hamilton, 2015). **Example:** an r^2 of 0.61 means that approximately 61% of the variation in the dependent variable (Salary) is accounted for by the independent variables (All of the statistics) in the model. The remaining 39% of the variation is still unexplained and may be due to other factors that were not included in the model.

Coefficient Value: The coefficient value represents the percentage change in the dependent variable (salary) with an increase in one unit of the independent variable (all the statistics). **Example:** If a player had a salary of $5,000,000 and increased their BB/K by one unit (up 0.01) their expected salary would increase by $37,500.

Coefficient s.error: The coefficient s.error represents how much the coefficient value varies when running the model again with different datasets (Siegel & Wagner, 2022)