BUSINESS 620

Allegheny College

Meadville, Pennsylvania 16335

Impact of Data Analytics in Major League Baseball



Alejandro Samayoa

4/7/2025

Impact of Data Analytics in Major League Baseball

by

Alejandro Samayoa

Submitted to The Department of Business and Economics

Project Advisor: Professor Tomas Nonnenmacher

Second Reader: Professor Tim Bianco

Date: 4/7/2025

I hereby recognize and pledge to fulfill my responsibilities as defined in the Honor Code and to maintain the integrity of both myself and the College as a whole.

Signature:

Alejandro Samayoa

Acknowledgements

I would like to thank my parents John and Tracy for always supporting and pushing me to be better all my life. I would not be here today without them. I would also like to thank my grandparents, Fred and Dianne, who have always been there for me whenever I was in need. They have taught me so much about myself and I am so grateful to have them in my life. Finally, I would like to thank Professors Nonnenmacher and Bianco for their help with this research paper. They were always there for me when I needed help and always could find time to meet and answer any questions I had, you two were truly the best and I could not have picked two better professors to work with.

Table of Contents

Acknowledgements…….. iii
List of Figures……2
List of Tables…….3
Summary of AI Use by Author……..4
Abstract……..5
CHAPTER 1: Introduction…….6
CHAPTER 2: Literature Review……... 9
CHAPTER 3: Theories & Expectations……...18
CHAPTER 4: Data & Methodology……26
CHAPTER 5: Results……...30
CHAPTER 6: Conclusion……41
Bibliography……43

List of Figures

Figure 3.1. Percentage (%) of Teams Adopted by Year……22

List of Tables

Table 5.1. Win percentage as a function of TeamSalary……...31

Table 5.2. Win percentage as a function of TeamSalary and DA……..32

Table 5.3. Win percentage as a function of TeamSalary, DA, and TeamSalary * DA…….33

Table 5.4. Win Percentage as a function of TeamWARHitting, TeamWARPitching, and TeamOBP……...35

Table 5.5. Win Percentage as a function of BattingAverage, and RunsBattedIn……35

Table 5.6. Win Percentage as a function of TeamWARHitting, TeamWARPitching, TeamOBP, war_hit_trend, war_pitch_trend, obp_trend, and year_centered……...37

Summary of AI Use by Author

During this study there was little use of AI. Other than Figure 3.1 as well as some grammatical changes, everything else was done by hand.

Abstract

The introduction of data analytics (Sabermetrics) and the concept of *Moneyball: The Art of Winning an Unfair Game* transformed baseball by providing new insights into player performance data, which impacted team decision making and management. Data analytics transformed more than just baseball. It is widely used in all sports today as well as businesses to help make strategic decisions and provide advantages so they can thrive. This research examines whether teams, particularly those in smaller markets with lower budgets, that use sabermetrics continue to maintain a competitive advantage, allowing them to compete with larger market teams. It also explores whether teams that adopt Sabermetrics achieve better on field performance relative to their payroll. My findings support that Sabermetrics initially provided early adopters with a competitive advantage, but that advantage could not be sustained over time. Additionally, I found that certain Sabermetric statistics were more strongly correlated with team success than traditional statistics. Analytics have become essential across all sports, but they are especially significant in baseball. This research is important because it can shed light on the long-term efficiencies of data analytics and its impact on decision making and management within the sports and businesses industries.

CHAPTER 1: Introduction

What is data analytics? Why is it so important when it comes to decision making? Data analytics is the widespread use of data, descriptive and predictive models, quantitative and statistical analysis, and fact-based management to advise actions and decisions (Davenport 2006). In today's world, run by technology and constant change, businesses need to stay ahead of the curve in order to compete.

One area that has seen a dramatic increase in data analytics is the world of sports. Over the past 3 decades, Major League Baseball (MLB) has experienced a major technological shift inspired by Sabermetrics which was created in 1980 by Bill James, who was a renowned baseball statistician, writer, and historian. James defined Sabermetrics as "The search for objective knowledge about baseball" and it has served as the main empirical analysis of baseball statistics. The first use and application of data analytics was for data driven decision making in 1997 when Billy Beane and the Oakland Athletics baseball team began using Sabermetrics to evaluate players' overall value to build a roster to better compete within the MLB with a small, limited payroll.

The practice gained popularity league-wide in 2003 with the release of the best-selling book *Moneyball: The Art of Winning an Unfair Game* written by Michael Lewis followed by the 2011 Oscar-winning movie *Moneyball* starring Brad Pitt. Johnson & Wales University (2024) states teams now use data analytics for recruiting players, coaching and developing players, strategic optimization of performance, enhancing fan experience and engagement, and redefining sponsorships and revenue generation.

Even though Sabermetrics is arguably the most famous use of data analytics in sports, the NBA has its own version of data analytics like SportsVU which is a camera tracking system following the movements of players and the ball while on the court. The NFL has also taken a profound liking to data analytics with their partnership with Catapult which is a sports analytics technology company that provides athlete monitoring to help prevent injury and keep athletes operating at the top of their game. Data analytics have changed the way businesses around the world operate.

Previous research has found that Sabermetrics provided teams who first adopted it with great success like the Oakland Athletics from 1997-2003. While the studies have highlighted early success, there is still a need to assess whether the advantages Sabermetrics provided could stretch across the entire league and still be as impactful over time.

In this research paper I answer this main research question "How does the application of Sabermetrics affect team performance and payroll efficiency in Major League Baseball?" by exploring how data analytics, more specifically Sabermetrics, impacted team performance and payroll efficiency in Major League Baseball and if that effect, if any, was sustained over time. To test this, I ran multiple regression analyses comparing teams winning percentages relative to their payroll, the year they adopted a data analytics program, their team wins above replacement (WAR) for hitting and pitching, team on base percentage (OBP), as well as teams batting averages (BA) and runs batted in (RBIs).

My main finding suggests that Sabermetrics initially provided the teams that adopted it early with an advantage, but that advantage was not sustained over time. Instead, team salary proved to have more of an impact on how much a team won, suggesting that the more money a team had to spend provided them with more games won. Additionally, I found that certain

Sabermetric statistics, particularly wins above replacement (WAR) were more strongly correlated with team success than traditional statistics such as team batting average and team runs batted in.

Understanding how data analytics has affected team performance and payroll efficiency is critical, as it highlights the growing role of data analytics in shaping outcomes and economic sustainability in competitive business markets. In a world shaped around technology and constant adoption using data analytics to gain a competitive advantage is necessary for businesses to thrive and compete in their given markets.

CHAPTER 2: Literature Review

This chapter examines the existing literature surrounding data analytics and Sabermetrics. When answering the research question, there are two key areas to explore: team performance and resource allocation to improve payroll efficiency. Utilizing relevant literature, I will explore the following questions: How has Sabermetrics led to on-field performance and success, and to what extent has it helped teams to optimize their payroll expenditures in management and strategic decisions? The chapter further explores the brief history of Sabermetrics, its challenges and the broader implications, and the future direction of data analytics within baseball, the wider world of sports, and in business.

The chapter first addresses the history and evolution of Sabermetrics, establishing its historical context and its core principles. The second section examines how Sabermetrics affected team performance and influenced management's decision making. The third section shifts to look at the effects of Sabermetrics on payroll efficiency, analyzing how well teams allocate their resources and spend their money on their players. The fourth section considers the future of Sabermetrics, more specifically, what it can lead to and its broader implications.

History of data analytics and Sabermetrics

Sabermetrics and statistics in baseball have a rich history. While many people who know baseball attribute Sabermetrics and data analytics to Billy Beane and the Oakland A's, that's not where it began (Syracuse University 2024). In 1859, a British-born journalist, Henry Chadwick, familiar with the game of Cricket, became fascinated with baseball and began to write about it (Schiff, 2008). Chadwick is credited with transforming how the game was recorded and analyzed, creating the first-ever box score. Chadwick went on to develop the framework for

today's modern scoring systems of hits, home runs and total bases including the backwards K to indicate a looking strikeout[1] which is still one of the most important stats for pitchers today (Pesca 2009). Later in 1872, Hervie Alden Dobson, a fan of Chadwick's, wrote a note proposing a new measure of batting "hits per time at bat," which eventually became batting average, one of the most important batting statistics (Syracuse University 2024). In 1947, the Brooklyn Dodgers hired Allen Roth, who became the first full-time statistician working for a major league baseball team (Syracuse University 2024). He went on to perform calculations for on-base percentage, performance in certain balls and strikes counts, and batting average with runners in scoring position (Syracuse University 2024). Finally in 1971, the Society for American Baseball Research formed in Cooperstown, NY promoting the history of baseball and statistics (Syracuse University 2024).

In 1980, an analyst named Bill James defined it as "the search for objective knowledge about baseball" and coined it "Sabermetrics" in honor of the Society of American Baseball Research (SABR) of which James was one of the co-founders (Kelly 2019). The problem, however, was that front offices wanted nothing to do with Sabermetrics. They were comfortable with what they knew - the old traditional scouting methods. It wasn't until 1997 that Billy Beane and the Oakland Athletics showed teams the power of Sabermetrics.

Due to an owner's orchestrated reduction in payroll, Billy Beane, the General Manager (GM) of A's, had to figure out a way to not only put a roster together, but build one that could compete. Beane and his team adopted Sabermetrics to put together a competitive team roster of undervalued players. They looked at undervalued player metrics like on-base percentage rather

[1] Batter is called out on a third strike without swinging at that pitch

than the traditional batting averages and RBIs. Beane and the A's achieved exceptional success leveraging these new metrics (Lewis 2003) that led to the publication of *Moneyball* in 2003. It didn't take long for the book to become hugely successful, selling over a million copies. It was made into a movie in 2011 which was nominated for 6 Academy Awards including Best Picture and won many film critic awards. More broadly, the book and film launched the adoption and popularity of Sabermetrics within businesses, sports, and government agencies (Elitzer 2018).

Impact of Sabermetrics on Team Performance

Ever since the beginning of the game, front office personnel had always relied on traditional scouting methods when it came to player evaluations and signings (Weller 2020). Scouts would travel all around the U.S. evaluating players, seeing if they passed the "eye test" and then they would report back to the front offices who would draft or sign a player based on "gut feelings" and information shared by their Scout (Brocklebank 2018). Scouting is still widely used today, but its role has evolved, as many flaws in traditional scouting methods have been uncovered (Weller, 2020).

Following its adoption, Sabermetrics was first only used for players who were already in the MLB. Even though there are new metrics that can help evaluate a player, there is barely any literature that can help predict how players will turn out after they are drafted. Kilanowski & Moloney (2024) found a way to utilize Sabermetrics to assess and predict how amateur players will perform. Focusing on the Cape Cod Summer league, known for being the best summer league in the U.S., they developed their own metric called "cWAR" which is the same as the metric WAR, but it focuses specifically on college players. They found that cWAR effectively

compared the top college players' performance on an even scale, highlighting the potential for it to be a valuable tool for MLB teams to use when it comes to the draft. Teams can now sign and draft players who might be overlooked and not pay them as much, then develop them to help teams win.

Sabermetrics has proven to be a useful tool in helping teams win. Elitzur (2018) provides an in-depth analysis by looking at the teams who successfully adopted and implemented Sabermetrics before, during, and after the publishing of "*Moneyball: The Art of Winning an Unfair Game."* Teams such as the A's, Houston Astros, and Tampa Bay Rays were the first teams to win more games while spending less money. This type of approach was coined the *Moneyball* approach and teams were called *Moneyball* teams as they focused on key metrics like wins above replacement (WAR) and on base percentage (OBP) when making strategic managerial decisions (Elitzer 2018). Year by year, this led to more teams adopting a data analytics program to help them succeed.

Sabermetrics can also be used as an extremely powerful predictive tool. Beneventano et al. (2012), found that certain metrics are better run-related predictors than the traditional statistics used in the past and therefore were a key factor in helping teams win or lose. Later, Brogan (2023) evaluates what metrics help the most when predicting what teams will make the postseason. He found that runs created (RC) and wins above replacement (WAR), one of the biggest Sabermetrics, had the most significance when it came to predicting success. The impact of Sabermetrics and the power of data analytics is clearly seen in the St. Louis Cardinals 2009 draft culminating in a winning 2013 season and the Boston Red Sox's 2004 World Series win which came just two years after hiring "Mr. Sabermetrics", Bill James (Sumeraj 2014).

Economic Impact of Sabermetrics on Payroll Efficiency

Sabermetrics was first adopted as a way for small-market teams with limited payrolls to compete against large-market teams with bigger budgets. There is no salary cap in the MLB so large market teams can spend as much money as they want creating a division between the two market types. Hakes and Sauer (2006) provide an economic analysis to *Moneyball* and found that the market was heavily undervaluing Sabermetrics. Furthermore, they used economic tools to test Lewis's main theory from *Moneyball* and found he was correct. They found that during the time period in *Moneyball,* batters' salaries didn't correctly reflect how much their skills helped their team win. This inefficiency was large enough for the Oakland Athletics to use it to their favor and create a competitive advantage over other teams.

In 1997, the A's revealed these inefficiencies in player evaluations noting that on base percentage (OBP), a key Sabermetric, had a stronger correlation with winning percentage than both batting average (BA) and slugging percentage (SLG), two traditionally valued metrics. The world saw this firsthand when the A's in 2002 let go of former MVP Jason Giambi, star outfielder Johnny Damon, and stud closer Jason Isringhausen and instead, Billy Beane, and the A's, used their reduced payroll to sign cheaper, undervalued players into their roster spots, which eventually led to them making the playoffs. More recently, Brown et al. (2015), analyzed team payrolls, current salaries, and player evaluations to determine future salaries. They found that even in the post *Moneyball* era, metrics like on base percentage (OBP) continue to play a huge role in determining a player's salary and their signing potential. Data analytics and Sabermetrics have forever changed the game of baseball. "Gut feeling" no longer solely determines how teams scout players, evaluate talent, value players, and manage their rosters.

Challenges and Limitations of Sabermetrics

Though past results show the success of implementing Sabermetrics in improving team performance and payroll efficiency, several issues have also been identified. Much of the literature I have found that has supported the positive impact of data analytics has also found that the advantage it once brought cannot be sustained. A team's challenge to sustain a competitive advantage through innovation has been well documented across various business settings. Barney (1991) talks about understanding sources of a sustained competitive advantage and the link between a business's resources and sustaining a competitive advantage. He looks at four specific indicators that a firm resource must have to sustain a competitive advantage: It must be valuable, rare among the field of competition, imperfectly imitable, and there cannot be equivalent substitutes.

Sabermetrics in the beginning had those four indicators. When Sabermetrics was first adopted it was extremely valuable, as the jumps in the Athletics' records indicate. It was also rare, as it had never really been seen or used before in any sport, and not many teams considered using technology like it. To this day, there really are no substitutes to Sabermetrics, but it has vastly been revolutionized and upgraded, branching out far beyond just some specific metrics. However, Sabermetrics was not imperfectly imitable and could easily be duplicated or copied. Before *Moneyball,* only the Athletics and Padres adopted Sabermetrics, and that small sample size provided a competitive advantage for a time. It was only after *Moneyball* was published that the duplication process began, leading to a league-wide adoption over time which diminished the advantage Sabermetrics once provided.

In the *pre-Moneyball* era from 1997-2002, Sabermetrics brought a competitive advantage to the few teams that adopted it. Elitzur (2015) studied whether the *Moneyball* effect was still relevant after the book was published in 2003. He found a positive correlation between higher spending and team performance, which suggested that Sabermetrics no longer provided a competitive advantage. However, this did not mean that teams were not using Sabermetrics, instead they were combining their higher spending with Sabermetrics to give the best possible outcome. After 2003, only the Tampa Bay Rays showed an advantage in their roster according to their payroll. Freeman (2019) finds a similar result, but their study looked at Sabermetrics effects on winning percentage, attendance, and sustained success over time. They found that teams that implemented Sabermetrics before *Moneyball* was published had a sustained competitive advantage the first couple years after their adoption over non-Sabermetric teams. However, after the book was published this advantage started to disappear as more teams adopted some type of data analytics program. This result makes sense as it follows Barney's argument as Sabermetrics was not rare, nor valuable anymore since more teams were adopting it regardless of whether they were small or large market teams. Today, all 30 teams in the MLB have some sort of an analytics program (Elitzur 2018) utilizing them in similar but different ways.

Broader Implications and Future Directions

Data analytics and Sabermetrics were literal game-changers when they were first introduced into the MLB. Today, Sabermetrics has expanded and become much larger than what Bill James thought it could ever be. Mizels et al. (2022), studies teams in the MLB that have new metrics like Statcast and Trackman which are advanced, machine-learning algorithms that provide insights into how hard players hit the ball, the angle their bat is swung, and the spin rate pitchers have with certain pitches. These advancements have given teams more data to analyze

that can impact their in-game management, strategies and how they evaluate players' performance. These new machines are credited with increasing fan engagement as every team posts how hard and far a ball was hit or how hard the pitcher threw so fans can feel even more connected to the game. The next predicted step for data analytics is to look at how to prevent injuries. With these new technologies, teams and trainers can analyze how athletes' bodies move and work so if there is anything wrong, the coaches and training staff can act beforehand, preventing any career ending injuries or setbacks (like Tommy John) from ever happening.

Baseball is not the only sport that is up to date with data analytics and technology. All sports have adopted some sort of analytics in varying ways. If you watch an NFL game and it is 4th down and the team is going for it or a team is going for a two-point conversion, they are doing it because they are learning from analytics. When announcers analyze the plays, they credit analytics and how it has shaped the game into what it is today. Television networks promote analytics by displaying a small graphic image with the percentage of probability the team has in order to get the conversion and its chances of succeeding or failing. These new algorithms are changing the way sports are played, and we are seeing a new, more aggressive approach to game strategies that put teams in the best position to win.

Auto racing is another example of a sport adopting the use of analytics. Car development teams analyze past results to make their cars faster, have better turn radius, and be more fuel efficient and ultimately safer (Schoborg 2023). Even though data analytics has had a major effect in most sports, that is not the only industry where it is being implemented. Wolfe et al. (2007) explores using the *Moneyball* method on a broader basis, helping people understand the book and take in more than just lessons in baseball.

Wolfe points out that the book also offers important management lessons when it comes to business. He highlights different themes, and proposes *Moneyball* lessons for managers including: leadership, innovation, overcoming resistance to change, and creating and sustaining a competitive advantage. He also investigates how the "Moneyball" approach can be related to fields in finance, academia, and film. Managers need to be innovative and creative to get an edge over their competition. In a world run by technology and constant change, sticking to outdated methods no longer works - businesses that don't adapt will fail (Cole 2022).

In conclusion, Sabermetrics and data analytics have radically changed and shaped Major League Baseball and enhanced team performance and payroll efficiency. Sabermetrics has come a long way since it was first introduced in the 1997 season. Teams can now optimize their resource allocation, better evaluate players and find undervalued talent to compete. Despite that, the rapid widespread adoption of analytics after *Moneyball* has diluted its once dominant competitive advantage and created an extremely competitive environment demanding constant innovation and change. The future of data analytics is endless. It has already been shown to impact other sports and the business world. As technology continues to innovate and upgrade, so will data analytics. This chapter reviewed the literature on data analytics and Sabermetrics and lays the groundwork for my hypotheses which we will further explore in the next chapter about the theory and analytical framework.

CHAPTER 3: Theories & Expectations

Major League Baseball operates in an extremely competitive environment where even though there is no salary cap, teams must still strategically use their financial resources to achieve success. In this chapter, I explore two economic principle theories: resource allocation using strategic decision making and diffusion of new technologies. I will look at how these theories connect with Sabermetrics and how they influenced teams' success and their payroll efficiency. Through the use of these economic principles, I formulated three hypotheses: *H1:* teams that use Sabermetrics will have a higher performance relative to their payroll, *H2:* specific Sabermetrics like on base percentage (OBP) and wins above replacement (WAR) will correlate more strongly with team success than traditional metrics, and *H3:* adoption of Sabermetrics cannot provide a sustained competitive advantage over time. Throughout this chapter, I will explain how these hypotheses are consistent with the two economic principle theories and how they will help describe and support the data collected in later chapters.

Theory

While there is no salary cap in Major League Baseball, managers must carefully and strategically manage their payroll to achieve the most on-field success. This challenge is tied to the issue of resource allocation efficiency, which refers to how effectively teams can distribute their available resources across projects and other operations. Sabermetrics and data analytics provide a structured approach to help teams make smarter, more effective choices.

MLB teams can only spend as much money as their ownership allows. Large-market teams like the Boston Red Sox, New York Yankees, and Los Angeles Dodgers have capitalized on their owners' willingness to pay top dollar to acquire top talent. In comparison, small-market teams such as the Oakland Athletics and San Diego Padres have faced financial limits and had to find more creative ways to build competitive teams.

To help address spending disparities between large and small market teams, the MLB introduced the Competitive Balance Tax (Luxury Tax) in 1997. This tax penalizes teams that exceed a predetermined payroll limit with increasing rates for consecutive violations. For instance, first-time offenders pay a 20% tax on the overage amount, while a second consecutive violation incurs a 30% tax, and so on. The tax calculation includes the average annual value (AAV) of all contracts on a team's 40-man roster, plus additional player benefits, with final amounts determined at the end of each season (MLB.com).

Small market teams still struggled after the tax was instituted. General Managers like Billy Beane had to think more strategically and more like a finance manager within a business. To maximize his team's resource efficiency, Beane had to strategically evaluate the return on investment (ROI) for every player and payroll dollar. He used Sabermetrics to find the most undervalued players who could help his team win. Stats like WAR and OBP were used to measure players' effectiveness to their teams' success relative to their "cost." The Athletics were first to utilize metrics to generate more wins per dollar spent. He optimized allocating his resources in order to make the most strategic decisions to help his team win.

Sabermetrics was highly effective for the Athletics because they were among the first teams to use it, giving them a strategic advantage. It wasn't until Moneyball was published in 2003 that Sabermetrics became well known and no longer provided a sustained competitive

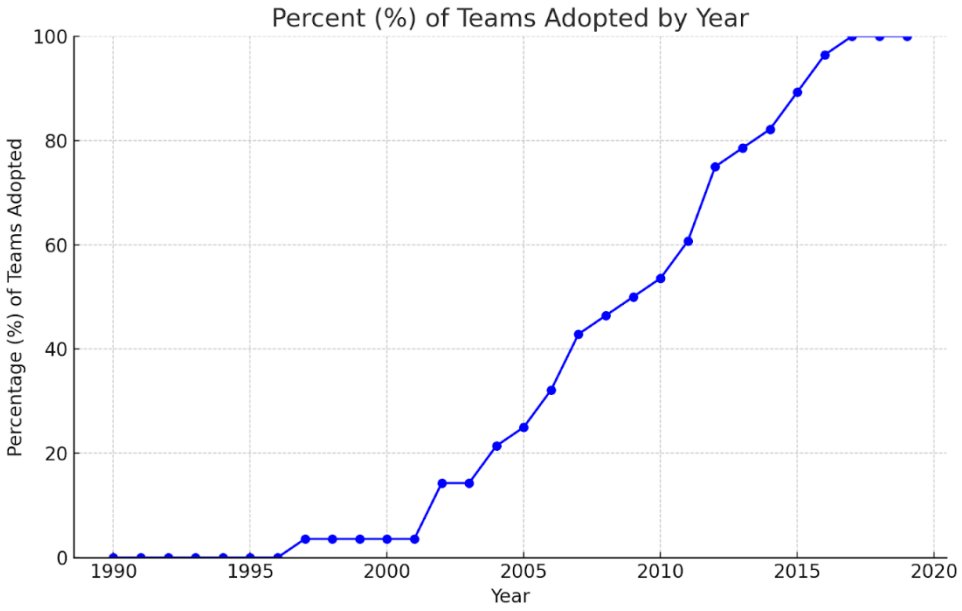
advantage. One by one, teams started to adopt their own form of a data analytics program. Sabermetrics was no longer unique to just small market teams. The league wide adoption can be understood through the theory of diffusion of new technology. When a new technology or product materializes and is successful, it starts a process where the number of people who use the technology or product increases over time (Stoneman & Battisti, 2011).

Sometimes, the process has a faster rate of increase and other times it can be slower but, in the end, there is a marginal increase in the use of a new technology or product. The adoption process can be faster or slower depending on high or low barriers to adoption. High barriers typically result in a longer time to adopt because they often involve legal issues like patents, copyrights, or trademarks, which firms must navigate through before adopting this new technology. Radio frequency identification (RFID) technology in retail and supply chain management is a prime example of high barriers. Early RFID technology was patented, therefore requiring firms to pay very high license fees to use it. Semiconductor tags and scanners, two RFID patented components, created a monopoly-like control, limiting widespread use and adoption. As a result, RFID technology was too expensive for medium and small sized retailers to adopt and implement. However, large retailers like Walmart and Amazon could afford RFIDs and were some of the first to adopt it. Similar to early adoption of Sabermetrics in the MLB, this adoption provided them with a significant competitive advantage, enabling them to manage their inventories more efficiently and effectively. When RFID technology became cheaper and patents expired, it became the industry standard for all retailers, leading to widespread adoption.

Similarly, reflecting how after *Moneyball* was published, more teams began to adopt Sabermetrics as they saw the strategic advantages it brought to team performance and payroll efficiency. In the case of Sabermetrics, adoption was a slower process even though Sabermetrics

was a non-excludable good with low barriers for adoption. Typically, technology with low barriers for adoption has a faster rate of adoption since firms can implement the technology as soon as they become aware of it, not having to deal with navigating legal or regulatory restrictions. However, the adoption of Sabermetrics proved to be an exception. Beane and other early adopters didn't have any type of patent or exclusive rights on Sabermetrics; any team could adopt it whenever they were aware of it. The reason teams didn't adopt Sabermetrics when the Athletics did was because they simply were not aware of Sabermetrics, it was something no teams ever considered. It wasn't until *Moneyball* was published that teams were aware of Sabermetrics, leading to its eventual league-wide adoption, consequently, diminishing the competitive advantage it once brought. Teams like the Diamondbacks were slower to adopt, waiting until 2017 to make the change, but it took other teams like the Red Sox, Yankees, and Cardinals only one to two years after *Moneyball* was published to adopt. However, today, all 30 teams have some sort of data analytics program. Figure 3.1 shows the percentage of teams that adopted Sabermetrics over time.

Figure 3.1



As previously mentioned, the widespread use of data analytics made Sabermetrics the standard practice rather than a unique advantage. Geroski (2000) provides a framework to help us understand how the diffusion of new technologies through different markets typically follows an S-curve. He explains how in a very competitive market due to low barriers of adoption, increased knowledge of new technology leads to a higher usage and adoption therefore leading to a weakened effect of the advantages the technology once brought to the firms that first adopted it. This diminishing effect also explains why there are fast and slow rates of adoption of new technologies since firms who take longer to adopt can look at the effects it had on other firms and if they see positive results they will incorporate faster and vice versa if the results are not positive. This helps explain why some teams like the Arizona Diamondbacks took longer to create a data analytics program than others.

Hypotheses and Expectations

This research paper is guided by three main hypotheses. The following sections detail each hypothesis and explain how I expect the data to reflect these predictions.

H1: Teams that use Sabermetrics will have a higher performance relative to their payroll.

H2: Specific Sabermetrics like on base percentage (OBP) and wins above replacement (WAR) will have a stronger correlation with team success than traditional metrics like batting average and runs batted in (RBIs).

H3: Adoption of Sabermetrics cannot provide a sustained competitive advantage over time.

First, I hypothesize that teams that utilize sabermetrics will achieve greater on field success relative to their payroll compared to teams that do not. This hypothesis is supported by and based on the premise that a more data driven approach allows teams to allocate their resources more strategically and effectively which was the first theory we talked about in this chapter. By using Sabermetrics to pinpoint underrated players, teams could build optimal rosters that allowed them to maximize wins per dollar spent. When I look at the empirical data in chapter 5, I will expect this hypothesis to hold true. Therefore, I would count on teams who employed Sabermetrics to have higher win percentages relative to their payroll expenditures, fortifying the idea that Sabermetrics improves resource efficiency.

My second hypothesis is that certain metrics within Sabermetrics like Wins Above Replacement (WAR) and On Base Percentage (OBP) have stronger correlations with team success than the more traditional metrics like batting average and runs batted in. This hypothesis is supported by the second theory written about in this chapter, of strategic decision making. Managers could look at those stats and strategically decide if they want to sign that player or not and use them effectively to give their team the best chance to win. Since OBP highlights a player's ability to get on base, giving his team a scoring opportunity, and WAR assesses a player's overall impact to helping his team win, we should see these two metrics be more predictive of team performance. Through some statistical modeling, we anticipate Sabermetrics to have a stronger relationship with winning, further supporting the strategic and overall effectiveness of data analytical methods.

My third and final hypothesis keys in on the growth of data analytical practices across the entire MLB. It proposes that as more teams adopt Sabermetrics and incorporate their own type of data analytics programs, the early competitive advantages Sabermetrics once brought to the teams that adopted it first will start to diminish. The Athletics, the team showcased in *Moneyball*, saw a tremendous increase in team performance when Billy Beane adopted Sabermetrics and used it to build their roster. However, once *Moneyball* was published and the information was available for everyone to see and use, that increase in performance disappeared. This hypothesis is further aligned with the theory of diffusion of new technologies which is the third and final theory.

To conclude, in this chapter I discussed some economic principal theories that aligned with the adoption of data analytics in the MLB and how it provided a framework to better my understanding of its effect on team performance and payroll efficiency. I also mentioned my

three main hypotheses that we expect to be true and proved in the next chapter. In that chapter, I will look at the empirical data collected for this research paper and talk about some statistical tests that will be run in chapter 5.

CHAPTER 4: Data & Methodology

This chapter outlines the data I collected, along with the methodological framework used to look at the effects of Sabermetrics on team performance and payroll efficiency in the MLB. Specifically, this section describes the variables I selected to use in my statistical analysis, the reasoning behind their inclusion, and the sources from which I found them. Additionally, this chapter introduces the types of statistical analyses I will use to test my hypotheses.

Data Variables

When looking at what data to extract that would help answer the three hypotheses in the previous chapter and the central question to this study: how does the application of sabermetrics affect team performance and payroll efficiency in Major League Baseball, I carefully chose and explored 7 main variables in a given time frame from 1997-2019: win percentage (WinPct), team salary (TeamSalary), team WAR hitting (TeamWARH), team WAR pitching (TeamWARP), team batting average (BA), team runs batted in (RBIs), and the year each team adopted a data analytics program (DA) crowning them as a "Sabermetric team".

I chose these specific variables as I believe they would be the best fit to test my three hypotheses. Teams' winning percentage, salary, and date of adoption will be used to test my first hypothesis: *H1:* Teams that use Sabermetrics will have a higher performance relative to their payroll. Win percentage, Team WAR hitting, team WAR pitching, team batting average, and team RBIs will be used to test my second hypothesis *H2:* Specific Sabermetrics like on base percentage (OBP) and wins above replacement (WAR) will have a stronger correlation with team success than traditional metrics like batting average and runs batted in (RBIs). Winning

percentage, team WAR hitting, team WAR pitching, team OBP, trend interaction variables, and a case study on the Oakland A's will be used to test my third hypothesis *H3:* Adoption of Sabermetrics cannot provide a sustained competitive advantage over time. I chose this specific time period 1997-2019 because we can see the change and effects Sabermetrics brought in the first year it was adopted (1997) all the way through a couple years after the last team adopted Sabermetrics (2017). However, in 1997 there were only 28 teams in the MLB as the Arizona Diamondbacks and Tampa Bay Rays were not established organizations yet but would become official teams in 1998. This had no real effect on the statistical tests that were tested and observed.

All the data used in this study was extracted from several different sources, including Baseball-Reference.com, The Baseball Cube, the paper *Data Analytics Effects in Major League Baseball* (Elitzur 2018), and FanGraphs. The dependent variable chosen for this study is win percentage. When finding teams winning percentages, I looked at Baseball-Reference.com to find all 30 teams wins and losses throughout 1997-2019 and took the teams wins and divided it by their losses then multiplied it by 100 to find their exact win percentage for each year. I also used Baseball-Reference.com to find each team's yearly batting average and runs batted in. For team salary I examined each team's opening day payroll per year from 1997-2019 using the Baseball Cube. For the two variables: team WAR hitting, and team WAR pitching. I looked at FanGraphs for each team per year. Lastly and most importantly using the paper *Data Analytics Effects in Major League Baseball* (Elitzur 2018), I was able to look at when each team adopted some sort of data analytics program.

Methodology

To test my hypotheses, I used Stata to run multiple different cross-sectional time-series regressions also known as "panel data regressions." I chose to run this type of regression because it is used to estimate linear regression models when looking at panel data, which means data that has both cross-sectional and time-series components. In my study all 30 MLB teams are my cross-sectional components while my chosen time range 1997-2019 are my time-series components. I also will talk about a case study examining the first team to adopt Sabermetrics to see its initial competitive advantage it provided before other teams started adopting data analytics in 2004. The two teams being the Oakland Athletics and San Diego Padres in 1997-2002.

Panel Regressions Equations

To test my hypotheses multiple different panel regressions were run. In this section we will go over the equations used for each regression and what table each equation corresponds with. As I previously mentioned, winning percentage is the dependent variable while team salary, team WAR hitting, team WAR pitching, team batting average, team runs batted in, and date of adoption were my independent variables. In the next chapter I use these regressions to test my three hypotheses.

Table 5.1: $WinPct_{it} = \beta_1 TeamSalary + \alpha_i + \lambda_{it} + \epsilon_{it}$

Table 5.2: $WinPct_{it} = \beta_1 TeamSalary + \beta_2 DA_{it} + \alpha_i + \lambda_{it} + \epsilon_{it}$

Table 5.3: $WinPct_{it} = \beta_1 TeamSalary + \beta_2 DA_{it} + \beta_3 (TeamSalary * DA_{it} + \alpha_i + \lambda_{it} + \epsilon_{it}$

Table 5.4: $WinPct_{it} = \beta_1 TeamWARH_{it} + \beta_2 TeamWARP_{it} + \beta_3 TeamOBP + \alpha_i + \lambda_{it} + \epsilon_{it}$

Table 5.5: $WinPct_{it} = \beta_1 BattingAverege_{it} + \beta_2 RunsBattedIn_{it} + \alpha_i + \lambda_{it} + \epsilon_{it}$

Table 5.6: $WinPct_{it} = \beta_1 TeamWARH_{it} + \beta_2 TeamWARP + \beta_3 TeamOBP +$

$\beta_4 Hitwartrend_{it} + \beta_5 Pitchwartrend + \beta_6 Obptrend_{it} + \beta_7 Yearcentered + \alpha_i + \lambda_{it} + \epsilon_{it}$

In all of these equations, win percentage is the dependent variable being tested. Beta (β) refers to the regression coefficients, which determine how much of the dependent variable is expected to change when the independent variable changes by one unit. Alpha (α) represents team- fixed effects and captures all unobserved, team-specific traits that stay constant over time. Lambda (ʎ) is a term that represents time-fixed effects, capturing yearly specific effects that equally influence every team. Epsilon (∈) is the error term, representing the random variation in the dependent variable that is not explained by the model.

In conclusion, this chapter looked at the specific data and variables I found and will be using in my statistical analysis. I also mentioned the methods I used to run my different panel regressions and what each regressions equation is. In chapter 5, I will run those regressions, examine the results, and compare them with my hypotheses and my initial thoughts on what the results would be.

CHAPTER 5: Results

This chapter presents the regression analyses and the results. I will explain key results from each regression and test them with my hypotheses to see if they were correct. Each table will include all the independent variables, fixed effects, number of observations, and r-squared values.

Fixed Effects and R-squared values

Although r-squared (Overall) is the main r-squared value that will tell us how much variation in the dependent variable is explained by the model, there are two other important r-squared values that I included. R-squared (Within) represents how much of the variation in the dependent variable within each unit is demonstrated by the model. R-squared (Between) measures how much of the variation in the dependent variable between different units is explained by the model. I included two fixed effects in my regressions: time and team. Team fixed effects control for differences between teams that may not change over time. This ensures how payroll and performance change within each team over time, instead of comparing different teams to one another. Year fixed effects account for things that may affect any team in a given year. This effect makes sure that the results are not influenced by any major changes that happened across the league.

Table 5.1: Win Percentage as a function of TeamSalary

	(1)	(2)
TeamSalary	0.00313***[2]	0.0002113
DA	-	-
TeamSalary * DA	-	-
Time fixed effect?	No	Yes
Team fixed effect?	No	Yes
Number of observations	684	684
R-squared (Within)	0.0146	0.0146
R-squared (Between)	0.5270	0.5270
R-squared (Overall)	0.0781	0.0781

[2] *** $p < 0.01$

Table 5.2: Win Percentage as a function of TeamSalary and DA

	(3)	(4)
TeamSalary	0.000398***	0.00028**[3]
DA	-.014503	-.0100349
TeamSalary * DA	-	-
Time fixed effect?	No	Yes
Team fixed effect?	No	Yes
Number of observations	684	684
R-squared (Within)	0.0178	0.0178
R-squared (Between)	0.5011	0.5021
R-squared (Overall)	0.0907	0.0907

[3] ** $p < 0.05$

Table 5.3: Win Percentage as a function of TeamSalary, DA, and TeamSalary * DA

	(5)	**(6)**
TeamSalary	0.000448***	0.000358**
DA	-.006839	0.0014781
TeamSalary * DA	-0.0000838	-0.000134
Time fixed effect?	No	Yes
Team fixed effect?	No	Yes
Number of observations	684	684
R-squared (Within)	0.0189	0.0193
R-squared (Between)	0.4927	0.4795
R-squared (Overall)	0.0900	0.0868

In tables 5.1, 5.2, and 5.3 there is a consistent theme with TeamSalary being the only independent variable that has any statistical significance with the dependent variable winning percentage as you can see every coefficient for TeamSalary besides one in Table 5.1 are significant with at least 95% confidence. In all three tables, TeamSalary also has a positive coefficient. For example, in the first regression run in table 5.1, which includes no fixed effects, every $1 million increase in team payroll is associated with a 0.000313 increase in win percentage. When applied over a full season, a team can expect to have an additional 0.05 wins. This was expected since it would make sense that the teams who spent more money would win more games. Even though TeamSalary helps contribute to wins, that value is very small, so

teams still have to optimize how they spend their money to get the most wins per dollar. However, even though TeamSalary is positively statistically significant, the within and overall r-squared values are low indicating that the model does not have much explanatory power and TeamSalary alone does not predict a lot of a team's success. The other two independent variables, date of adoption (DA) and the interaction variable (TeamSalary * DA) are not statistically significant in any regression that was run indicating they did not have any impact on team performance.

Table 5.4: Win Percentage as a function of TeamWARHitting, TeamWARPitching, and TeamOBP

	(1)	**(2)**
TeamWARHitting	0.0052766***[4]	0.0052521***
TeamWARPitching	0.0069312***	0.0069188***
TeamOBP	0.2015383	0.1761284
Time fixed effect?	No	Yes
Team fixed effect?	No	Yes
Number of observations	688	688
R-squared (Within)	0.7739	0.7740
R-squared (Between)	0.9181	0.9176
R-squared (Overall)	0.8043	0.8042

Table 5.5: Win Percentage as a function of BattingAverage (BA), and RunsBattedIn (RBI)

	(1)	**(2)**
BattingAverage (BA)	-0.3393183	-0.2432972
RunsBattedIn (RBI)	0.0005231***	0.0005105***
Time fixed effect?	No	Yes
Team fixed effect?	No	Yes
Number of observations	688	688
R-squared (Within)	0.2930	0.2931
R-squared (Between)	0.3845	0.3795
R-squared (Overall)	0.3120	0.3111

[4] *** $p < 0.01$

In Table 5.4, TeamWARHitting and TeamWARPitching are the only positive statistically significant variables when explaining Win Percentage. In regression 1, TeamWARHitting has a coefficient of 0.0052766 and a p-value of 0.000, meaning it has 99% confidence level. This means that for every additional point of a team's Wins Bove Replacement, that team's winning percentage increases by 0.0052766. Both TeamWARHitting and TeamWARPitching are positive and significant in both regression models, with and without fixed effects, and the r-squared values for each regression are all high. This suggests that both variables are consistently significant predictors of team winning percentages. TeamOBP, however, is not significant in either of the two regressions, implying no reliable impact on win percentage.

Additionally in Table 5.5, RunsBattedIn (RBI) is the only positive significant variable. However, the coefficients are smaller than TeamWARHitting and TeamWARPitching, and the r-squared values are all low. This implies that RBI does not serve as much as a consistent significant predictor of a team's win percentage as WAR, nor does it explain nearly as much variation in win percentage. BattingAverage surprisingly is negative and not statistically significant in either regression, meaning it has no impact on win percentage.

Table 5.6: Win Percentage as a function of TeamWARHitting, TeamWARPitching, TeamOBP, war_hit_trend, war_pitch_trend, obp_trend, and year_centered

	(1)
TeamWARHitting	0.0044892***[5]
TeamWARPitching	0.0073808***
TeamOBP	0.6672114***
War_hit_trend	0.000132***
War_pitch_trend	-0.0000814**[6]
Obp_trend	-0.0606551***
Year_centered	0.189248***
Number of observations	688
R-squared	0.8083
Adjusted R-squared	0.3184

Table 5.6 tests whether Sabermetrics could sustain the once provided competitive advantage it once brought to the teams that adopted data analytics before the publishing of *Moneyball: The Art of Winning an Unfair Game.* Before running this regression, I generated a year centered variable based on the year 2002, as well as interaction variables like War_hit_trend (TeamWARHitting * year centered) to see the change in impact of Sabermetric statistics over time. Generating these variables allowed me to read the results of the trend variables more clearly, and they were tied to a meaningful reference point being 2002, the year before

[5] *** $p < 0.01$
[6] ** $p < 0.05$

Moneyball was published. The results would be strictly after 2002 not including the 1997-2001 seasons. All the independent variables are statistically significant, but not all are positive. The one exception in this regression is TeamWARHitting, which is positive and significant on its own and its trend interaction variable as well. This means a team's WAR for hitting is the only variable to continue to show a stable or slightly increasing impact on team win percentage over time. In contrast, the other two variables, WAR pitching and OBP showed the opposite and declined overtime. The r-squared value is very high at 0.8102, indicating this model explains a lot of the variation in team winning percentages.

Results with hypotheses

Based on the results from all the regressions run my first hypothesis, *H1:* Teams that use Sabermetrics will have a higher performance relative to their payroll was partially supported. In Tables 5.1-5.3, TeamSalary was the only independent variable to show any significance, and it had a very weak relationship, and the coefficients were very small and didn't explain much of the variation in the models due to its consistently low r-squared values. I also ran an additional regression testing TeamSalary and TeamWARHitting and TeamWARPitching to see if team salaries and the year they adopted data analytics impacted their team WAR and to see if teams found an advantage by allocating their resources decisively to find undervalued players at an undervalued price that helped them win. TeamSalary did show a consistent positive significant relationship with both TeamWARHitting and TeamWARPitching in the random effect models indicating salary plays a role in team performance. Unfortunately, in all fixed effect models, TeamSalary lost all its significance, suggesting that spending alone does not ensure higher team performance. Date of adoption (DA) and the interaction term TeamSalary * DA was mostly

insignificant, meaning that teams adopting Sabermetrics earlier did not necessarily achieve a sustained performance boost in relation to payroll.

My second hypothesis *H2:* Specific Sabermetrics like on base percentage (OBP) and wins above replacement (WAR) will have a stronger correlation with team success than traditional metrics like batting average and runs batted in (RBIs), was strongly supported by the results from Tables 5.4 & 5.5. In Table 5.4 where I examined the relationship between Sabermetric statistics and win percentage, I found that in both the fixed effects and not fixed effects models WAR for hitting and pitching were highly significant predictors of team's winning percentages. The coefficients for both WAR variables were consistently positive and statistically significant at the 99% confidence level, the r-squared values were also very high specifying that an increase in either variable is associated with an increase in a team's winning percentage. In contrast, in Table 5.5 when running the same type of regression using more traditional statistics like batting average and runs batted in the results (RBIs) were almost opposite. Batting average failed to show any significance, and its coefficient was negative. RBIs, however, did have positive significant coefficients but the r-squared values were very low indicating very little explanatory power. Overall, these results reinforce that Sabermetric statistics are more correlated with team success than traditional statistics.

When looking at my third hypothesis, *H3:* Adoption of Sabermetrics cannot provide a sustained competitive advantage over time, the results strongly supported this. In Table 5.6, team WAR for hitting and pitching, along with team OBP, are positive and significant, indicating a positive relationship with team success. However, the predictive power of War pitching and OBP over time has significantly decreased, suggesting that the competitive advantage these metrics once offered has diminished as more teams adopted data analytics. WAR hitting is an outlier, as

it still has a stable increase on team success, meaning some Sabermetric statistics remain valuable to teams to win more games, but it's once desired and valuable advantage has greatly diminished. To further explore my third hypothesis, I conducted a case study on the Oakland Athletics, one of the first teams to adopt Sabermetrics. From the 1997-2002 seasons, before Sabermetrics was introduced to the world, the Athletics increased their winning percentage significantly every year and made it to the American League Divisional Series (ALDS)[7] in 2000, 2001, 2002, and 2003, although they lost every series. However, after 2003 when *Moneyball: The Art of Winning an Unfair Game* was published and Sabermetrics became mainstream, the Athletics maintained a steady winning percentage slightly over 50% for a couple years before seeing a sharp decline. They did not make it back to the ALDS until 2018. This further supports my third hypothesis: that Sabermetrics could not provide a sustained competitive advantage over time.

In this chapter, I explored the specific results I was able to gain through running my regression analysis and I was able to compare the results to my hypotheses. In the next chapter, I will be wrapping up this research paper by talking about my main findings, some of the limitations I encountered, and the bright future of data analytics, more specifically Sabermetrics.

[7] Best of 5 series in the MLB to determine which two teams will make American League Championship

Chapter 6: Conclusion

Summary:

In this study, I examined how the application of Sabermetrics affected team performance and payroll efficiency in Major League Baseball. My main findings suggest that while Sabermetrics initially increased team performance, it did not have a significant effect on payroll efficiency. Instead, Team salaries had a stronger influence, indicating that the financial resources a team had, the more wins they are likely to win. I also found that certain Sabermetric statistics, such as wins above replacement, were more strongly correlated with team success than traditional statistics like batting average and runs batted in. Finally, the competitive advantage Sabermetrics brought to early adopters, like the Athletics from 1997-2002, could not be sustained over time.

Limitations:

During this study, I encountered two limitations when it came to collecting data. First, there is no website or source that provides a list of when all 30 MLB teams adopted some form of data analytics program, as a result I relied on the data from Elitzur (2018), *Data Analytics Effects in Major League Baseball.* Elitzur had to use unconventional methods such as inspecting multiple LinkedIn pages, trying to find when each team adopted Sabermetrics. The second limitation I faced was that, as of today, all 30 MLB teams have a data analytics program. Over time, after 2003 when *Moneyball* was published, it was difficult to see which teams relied on or used Sabermetrics heavily to try and gain an advantage.

Future Directions

Data analytics and Sabermetrics have forever changed the game of baseball. Management decisions, scouting, and allocating payroll were some of the first and biggest aspects of the game that changed. In today's landscape, new technology continues to emerge, such as the recently developed "Torpedo bat" by Aaron Leanhardt, a former MIT physicist. This bat leverages data analytics and statistics to analyze where batters most frequently hit the ball on the bat. Using this data, the bat is designed to put more of the barrel on those contact points based off the individual batter's tendencies, therefore helping batters hit the ball on the barrel more often. This is just one current example of how data analytics continues to revolutionize the world of sports. Data analytics has expanded far beyond just Sabermetrics. Data analytics is transforming not only the world of sports but the world of business as well. In today's world of constant adaptation and learning are essential to staying competitive. In sports, data analytics is still just getting started. There is still so much we don't know about what data analytics can accomplish. The possibilities are limitless.

Bibliography

Beneventano, P., Berger, P. D., & Weinberg, B. D. (2012, June). *Predicting run production and run prevention in baseball*. Predicting Run Production and Run Prevention in Baseball: The Impact of Sabermetrics. http://www.ijbhtnet.com/journals/Vol_2_No_4_June_2012/7.pdf

Brocklebank, J. (2018, November 13). *The truth about Sabermetrics*. Samford University. https://www.samford.edu/sports-analytics/fans/2018/The-Truth-about-Sabermetrics

Brogan, C. (2023). Logistic regression analysis on the value of statistics in Major League Baseball (Bachelor's thesis, Walsh University). Division of Math and Science, Walsh University.

Brown, D. T., Link, C. R., & Rubin, S. L. (2015). *Moneyball after 10 years: How have Major League Baseball salaries adjusted?* Journal of Sports Economics.

Cole, B. M. (2022, April 14). *Innovate or die: How a lack of innovation can cause business failure*. Forbes. https://www.forbes.com/sites/biancamillercole/2019/01/10/innovate-or-die-how-a-lack-of-innovation-can-cause-business-failure/

Davenport, T. H. (2006, January). *Competing on Analytics*. Harvard Business Review. https://cs.brown.edu/courses/cs295-11/competing.pdf

Dickinson, M. (2023). *X-Ray Vision: Finding success through Sabermetrics in Major League ...* Finding success through Sabermetrics in Major League Baseball. https://scholarsarchive.jwu.edu/cgi/viewcontent.cgi?article=1068&context=student_scholarship

Elitzur, R. (2018). *Data analytics effects in Major League Baseball*. Rotman School of Management, University of Toronto.

Elitzur, R. (2015, July 8). *Is Moneyball a myth? perceived advantages to sabermetrics vs reality*. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2627721

Fangraphs baseball. (n.d.). https://www.fangraphs.com/

Freeman, L. A. (2019). *How long before performance improves*. The Impact of Analytics in Professional Baseball: How Long Before Performance Improves. http://jitm.ubalt.edu/XXX-2/article3.pdf

Geroski, P. A. (2000, April 21). *Models of Technology Diffusion*. Research Policy. https://www.sciencedirect.com/science/article/pii/S004873339900092X

Hakes, J. K., & Sauer, R. D. (2006). *An economic evaluation of the Moneyball hypothesis.* Journal of Economic Perspectives.

Kilanowski, H., & Moloney, T. (2024). Sabermetrics by the sea: Evaluating college players with the Cape Cod Baseball League. *Journal of Sports Analytics*, *10*(1), 67–75. https://doi.org/10.3233/jsa-240771

Mizels, J., Erickson, B., & Chalmers, P. (2022, August). *Current state of data and analytics research in baseball*. Current reviews in musculoskeletal medicine. https://pmc.ncbi.nlm.nih.gov/articles/PMC9276858/

MLB stats, scores, History, & Records. Baseball. (n.d.). https://www.baseball-reference.com/

Pesca, M. (2009, July 30). *The Man Who Made Baseball's box score a hit*. NPR. https://www.npr.org/2009/07/30/106891539/the-man-who-made-baseballs-box-score-a-hit

Schiff, A. (2021, April 16). *Admin*. Society for American Baseball Research. https://sabr.org/journal/article/henry-chadwick-the-father-of-baseball-was-a-sportswriter/

Schiff, A. (2021, April 16). *Admin*. Society for American Baseball Research. https://sabr.org/journal/article/henry-chadwick-the-father-of-baseball-was-a-sportswriter/

Schoborg, C. (2023). *Football by the numbers: A look into sports analytics currently used in the National Football League*. STARS. https://stars.library.ucf.edu/etd2020/1762/

Stoneman, P., & Battisti, G. (2010). *Chapter 17 - The Diffusion of New Technology* . Science Direct. https://www.sciencedirect.com/science/article/abs/pii/S0169721810020010

Sumeraj, J. (2014). *Changing baseball forever - digitalcommons@cod*. Changing Baseball Forever. https://dc.cod.edu/cgi/viewcontent.cgi?article=1510&context=essai

The baseball cube - data/statistics research. The Baseball Cube - Data/Statistics Research. (n.d.). https://www.thebaseballcube.com/

University, S. (2021, May 19). *Sabermetrics: Baseball Analytics and the science of winning*. SYR. https://onlinegrad.syracuse.edu/blog/sabermetrics-baseball-analytics-the-science-of-winning-accessible/

University, Wales. J. (2024, June 4). *How sports analytics enhances performance & efficiency: JWU CPS*. JWU College of Professional Studies. https://online.jwu.edu/blog/how-sports-analytics-enhances-efficiency-and-performance/

Weller, E. (2020, May 9). *The Data Revolution: An Examination of the use of Scouting and Analytics in Major League Baseball Front offices*. eGrove. https://egrove.olemiss.edu/hon_thesis/1306/

Wolfe, R. (2007, November). (PDF) Moneyball: A Business Perspective. International Journal of Sport Finance. https://www.researchgate.net/publication/299052147_Moneyball_A_business_perspective