Spring 2020

Investigating Discrimination in Major League Baseball: Not So Perfect Competition

Esteban Rivera

Investigating Discrimination in Major League Baseball: Not So Perfect Competition

Thesis Submitted to Levy Economics Institute of Bard College

by Esteban Rivera

Annandale‐on‐Hudson, New York May 2020

Acknowledgements

First, I would like to thank Martha Tepepa and the rest of the Spring 2019 Economics 508 class for supporting me in my decision to pursue this issue. Without your support, I simply would not have investigated this, leaving these players with one less person fighting for them. Next, I am beyond grateful for the help from my thesis advisor, Fernando Rios-Avila. His critiques and teachings made the work in this thesis possible. To Jan Kregel and my fellow 2020 M.S. peers, thank you for your valued input throughout this journey.

I would also like to thank all of the faculty members at the Levy Economics Institute for pushing me to become the best possible student and learner that I can be. Special thanks go to my coaches and teammates for their support in these last few years on and off the field.

Lastly, I want to take this opportunity to thank my family and girlfriend. Emily, thank you for listening to my rants on this topic, and many others. You have been my rock through all of this. Raven and David, you have challenged me my entire life, and it pushed me to be the best version of myself. Mom and dad, I owe you for all of my accomplishments in the past, present and future. I love you all.

Plagiarism Statement

I have written this project using my own words and ideas, except otherwise indicated. I have subsequently attributed each word, idea, figure and table which is not my own to their respective authors. I am aware that paraphrasing is plagiarism unless the source is duly acknowledged. I understand that the incorporation of material from other works without acknowledgment will be treated as plagiarism. I have read and understand the Levy Economics Institute of Bard College statement on plagiarism and academic honesty as well as the relevant pages in the Student Handbook.

Esteban Rivera May 19th, 2020

Abstract

In recent years, there has been a growing trend in Major League Baseball. Elite Latino players are signing long-term extensions early in their careers, years before they reach free agency. This dissertation investigates wage differentials between races among players with over six years of service time in the MLB. First, some of the reasons why Latino players are pressured into accepting extensions early in their career are presented. Then, I estimate several models controlling for different types of performance. The statistic Wins Above Replacement measures all aspects of a player's performance, making it an ideal tool for labor market studies. The results from the model show convincing evidence of the presence of discrimination against top-earning Latino hitters and starting pitchers. The models suggest that from the years of 2017-2019, Latino players experienced a wage penalty in comparison to white-American players.

Keywords: Discrimination; Major League Baseball; Latinos; Wins Above Replacement; Performance; Quantile Regression; Minor League Baseball; Labor Economics

JEL Classifications: C22; D31; J15; J24; J52; Z2; Z22

TABLE OF CONTENTS

I: List of Figures

INTRODUCTION

Over the past decade, Major League Baseball has experienced an analytical revolution. What once was a game of instincts and reactions, has now become calculated and technical. Just 20 years ago the game had an entirely different identity. Teams often used what we call in baseball "the eye test." In order to define what players were the best or were bound for greatness, scouts trusted their eyes above anything else. Now, the game is reliant on data. Biased opinions are dwindling away, and now data rules. Statistical analysis and profit-seeking behavior are a key part of that analytical revolution.

Gone are the days of teams giving away masses of millions to players without profound evidence that this player will continue to perform for years to come. The league is obsessed with getting the very most out of a player, and not paying a dime over what they are worth. It is a business and emotions have taken a back seat. Owners do not want to pay an additional fee for a player that does not bring everything to the table, while players and agents want to make sure they get what they deserve.

This has created a divide in the league in several ways. Since the last collective bargaining agreement was agreed upon, tensions have risen between players and owners. Players have begun to notice that a power struggle exists from the beginning of their careers, all the way to the end. Most minor league players live below the poverty line (Williams 2019). This is due to MLB lobbying Congress to exempt minor leaguers from the United States' minimum wage laws. On top of that, they play under awful conditions, with little to no support from the major league owners.

This struggle is worse Latino players. In this paper, you will see the many ways in which Latino players disadvantaged in comparison to other races, mainly white-American players. From the minor leagues through the major leagues, Latino players face the greatest struggles. It begins with their careers as amateurs. It takes a lot of sacrifice to pursue a life as a professional baseball player in countries like the Dominican Republic, where the majority of international players in MLB come from. Once they are signed by an MLB team and enter the minor leagues, things do not get any easier. After that, acclimating to the big league is no easy feat.

Much of the literature on labor dynamics in MLB does not focus on the qualitative aspect, which is why it is discussed in detail in this paper. That being said, the literature heavily focuses on quantitative analysis in the sport. This is due to the availability of performance statistics. In U.S. labor market studies, A combination of education, age, and other characteristics are used to define a worker's performance (O'Neill 1990). In sports, and particularly baseball, several statistics provide more accurate measures of a player's performance and value relative to other players. This makes sports performance measures a nice analytical tool in economics for controlling for skill. For this reason, we can more accurately define whether a phenomenon like discrimination exists in MLB (Holmes 2011, Palmer and King 2006, Singh et al. 2003). The literature focuses on free agency because in theory, it is an environment where there is perfect competition. Players can be bid on for any number of years and dollars by any of the 30 teams in the sport.

With labor tensions rising and collective bargaining dynamics changing, there is a need for a reinvestigation of discrimination in the sport. Latino players make up half the players in the minors, and a third of the players in the majors. They have become a significant reason for the success of the league and change in labor relations. The crux of the paper comes in the analysis of wage differentials amongst Latino and white players using quantile regression. I will show that there is a need to re-open the literature, due to the convincing evidence of discrimination against top-tier Latino players.

CHAPTER I: LITERATURE REVIEW ON WAGE DIFFERENTIALS AND PAY DISCRIMINATION IN THE UNITED STATES

The research on theories of discrimination starts with Gary Becker (1971). One of the aspects of Becker's theory states that discriminating firms are lower-profit than non-discriminating firms. Firm's unwillingness to pay for equally productive minorities would leave them paying higher wages for less productive white workers. In due time, they would be driven out of the market, resulting in the disappearance of discrimination in the long run (Heckman, Lazear and Murphy 2018). Following Becker (1971), other economists investigated other theories of discrimination within the market and out of the market. Research on pre- and post-market discrimination surveys the disadvantages minorities face outside of typical market negotiations. Bertrand and Mullainathan (2004) examined the disadvantages minority workers face in the job application process and Darity Jr. and Mason (1998) investigate the class dynamic created by concentration of white men in higher-up positions. After a survey of discrimination within and out of the labor market, this overview will focus on Hispanic/Latino workers. It will highlight the role of factors such as physical appearance and immigration in explaining the differences in wages. These differences occur within the Hispanic/Latino population and relative to the white workers as well. It then describes the current conditions facing Hispanic labor in the United States.

Section I: Economics Of Wage Discrimination In the United States

As Charles and Guryan (2008) suggest, the expansion of the study of discrimination in economics was initiated by the publication of Becker's (1971), *The Economics of Discrimination*. Becker (1971) analyzed wage gaps between different racial groups and how discrimination from employers, employees, and customers could contribute to these gaps. His model construction emphasizes the role of preferences. Employers, employees, and customers all have preferences for what they want to hire, sell or purchase. The racial/gender composition of a workplace could influence a worker's decision to quit or not (Lang and Spitzer 2020). In other words, if a worker has is prejudiced against a certain race, he or she may prefer to not work with members of an opposing race. Similarly, an employer may not be prejudiced themselves, but their customers are (Lang and Spitzer 2010). If the employer suspects their business would suffer from hiring a worker from a certain race, then they might not be inclined to hire this worker. People could prefer to have as little contact with minorities as possible because of racial prejudice. Becker (1957) and others investigated the extent of the role of racial prejudice and taste-based discrimination in explaining wage gaps.

Ultimately, Becker finds that discriminating employers would be driven out of the market in the long-run, given perfect competition. Firms would not survive in a market where they are consistently reaping lower profits than firms which do not discriminate. Arrow (1972) criticized the approach, noting that it concludes that discrimination does not exist, which eliminates the phenomenon it intended to explain. Arrow (1972) investigated if Becker's (1957) conclusions would hold up in an environment with imperfect competition and found if there are nonconvexities or informational problems, then discrimination would persist. Charles and Guryan (2007) showed that racial distaste and wage gaps could survive in perfect competition as well. If prejudicial tastes are mobile across different roles within the labor market, then discriminating employers could earn roles as employees after their firm is driven out of the labor market and continue to discriminate in other ways.

Several empirical studies that have been undertaken investigating the claims put forth by Becker (1957) and Arrow (1972). Charles and Guryan's (2008) empirical assessment of Becker's preference model focuses on the role of racial prejudice in explaining wage gaps between whites and minorities. In particular, they explore the cross-sectional aspect of Becker's analysis. By exploring the extent of racial prejudice in different regions across time, they conclude that racial wage gaps are more largely affected by the level of prejudice of the marginal person, opposed to the average person. When controlling for skill of black workers across different regions, they find that areas with a larger share of black workers are more likely to be sorted to prejudiced employers. Their results suggest that about a quarter of the unconditional wage gap could be attributed to racial prejudice. This means that without controlling for any skills or other variables, race explains a quarter of why black workers earn less than white workers, but this suggests that other factors that come into play when explaining the wage gap, among them statistical discrimination and human capital differences.

The theory of statistical discrimination explores the idea that employers use easily observable variables that are correlated with productivity, such as race or education, to make hiring decisions (Altonji and Pierret 2001). People, including employers, are left to make decisions based off their own experiences and limited information. Without knowing everything about a prospective employee, employers rely on their experiences and whatever observable information they might have. One of these observable characteristics is race. If an employer perceives that a particular race is less productive than another on an average basis, then this could affect his or her willingness to hire somebody. Altonji and Pierret (2001) find that race plays a significant role in initial wages upon entry into a market, but that in the long-run productivity and performance are more significant in explaining wages. They also observe that employers learn over time. Meaning that they their perception of a group may change over time and that they use this information to make decisions. They state that their investigation of statistical discrimination could be useful in areas where there are hard-to-observe characteristics, such as productivity.

More recently Lang and Manove (2011) investigated wage differentials between black and white workers, controlling for education and cognitive ability. They analyze the difference between employers' prejudice and the impression produced by membership of a particular group, formally known as statistical discrimination. An emerging consensus in the literature had pointed towards most of the origin in wage differentials being due to pre-market discrimination, but their results attribute more explanatory power to statistical discrimination and prejudice. When controlling for ability, they found that blacks were more likely to have higher education. In other words, black workers with equal ability to white workers had higher levels of education. They indicate this could come about because black workers have greater incentive to signal their productivity.

Their findings show significant evidence that black workers with the same education, cognitive tests scores, and familial background still experienced an 11% penalty relative to their white counterparts. Specifically, they estimate a black worker would receive the pay of an identical white worker with one less year of education. In prior work done by Neal and Johnson (1996), they explicitly reject controlling for education when estimating wage differentials because they claim it is endogenous. However, Lang and Manove (2011) argue that since black workers have higher education levels than white workers with equal test scores, not controlling for education understates the impact of discrimination. When they control for ability, black and white workers earn about the same, but when controlling for education and ability, black workers earn less. Not including education creates an omitted variable bias. Black workers were also much more likely to be unemployed (Lang and Lehmann 2012), even when controlling for human capital. Lang and Manove (2011) showed that in addition to pre-market factors, statistical discrimination and racial prejudice also play a role in explaining the wage gaps between black and white workers.

In addition to prejudice and statistical discrimination, differences in human capital have been found to explain the wage gap between whites and minorities (O'Neill 1990). Although, O'Neill (1990) focuses on black and white men. Between 1940 and 1980, black workers had a sharp increase in their educational attainment which was accompanied by an increase in earnings. The result was a convergence in the racial wage gap between black and white men in those four decades. However, through the 1980s the black-white wage gap increased again. O'Neill (1990) proposes that this could be due to the restructuring of the job market. Businesses sought more highly educated skilled workers, which benefitted better educated white males.

Neal and Johnson (1996) conducted seminal work on the role of skill gaps in explaining differentials Skill gaps can be due to a multitude of variables such as, lower quality of schooling, lower quality of teachers, regional differences, and other factors. The question is whether minorities face a disadvantage in acquiring the same resources as white workers in preparation for their entry to the labor market. Using cognitive test scores as a measure for skill, they estimate why pay gaps come about between black and white workers. They estimated that gaps in test scores between races are more significant predictors of this gap, than discrimination. Even though discrimination did play a role in their findings, they interpret their results as black children facing obstacles to acquiring productive skill at a young age. Lang and Manove's (2011) critique of this study proved that there is a need to control for education when estimating pay gaps. However, other pre-market factors have been explored in estimating the disadvantages faced by minorities.

For example, a study investigated the role that a job candidate's name played when receiving a call back from a job application (Bertrand and Mullainathan 2004). They found evidence suggesting black workers received significantly fewer call-backs than their white counterparts with similar skills. If the worker had a culturally African-American name, they would find it harder to overcome this hurdle even with improved skills and credentials. Thus highly-skilled workers with culturally African-American names receive the same barrier to work as lower-skilled workers. Darity Jr. and Mason (1998) note that once a worker obtains a job, the racial-gender composition of these roles directly affects worker bargaining power. By excluding minority workers from higher-paying/higher-power jobs, an earnings ceiling evolves, and minorities have less opportunity to rise in the work force. This creates a cycle of discrimination that is reinforced by a class separation.

Section II: Economics Of Wage Discrimination In the United States For Hispanic/Latinos

In the 1970's, Hispanic men earned slightly less than black workers and significantly less than white workers (Reimers 1983). While the experience of discrimination varied across different ethnicities of Hispanic men, they all earned significantly less than white men with a penalty range of 12%-36%. Reimers (1983) suspected that the reasons for the wage differentials were due to variables such as age/experience, education, geographic location, language barriers and discrimination. For Puerto Rican men, discrimination was estimated to be responsible for 18% of the gap, the greatest among all Hispanic groups. Discrimination accounts for only a 6% wage difference than white men, the smallest among all Hispanic groups. Reimers (1983) also investigated if Hispanic men suffered from more labor market discrimination than black workers.

Using a Oaxaca-Blinder decomposition she found that Hispanic men suffer from greater differences in human capital than black workers among certain groups. In terms of gaps, black men fall between Puerto Rican men and what Reimers (1983) labels as "other Hispanic" men. This would put the black worker gap between 23% and 33%. She suspects a few factors could explain why discrimination and gaps vary across Hispanic groups. For Puerto Ricans, it could be due to their concentration in northeastern cities. These cities have a higher cost of living and more white workers. If they experience discrimination at a higher rate, then it is more likely that it has a higher impact on wages. Hispanic workers of all origins experience a wage gap compared to their white counterparts, but in the 1980s the gap was even larger than that between white and black workers. One theme throughout Reimers' (1983) study was that Cuban Americans are at a significantly greater advantage than other Hispanic workers in the United States.

George Borjas (1982) further explored why Cuban Americans experience this advantage in comparison to other Hispanics. The time of arrival in the United States is strongly associated with the wages that are observed by a given sub-group of Hispanic workers (Borjas 1982). However, in the 1980s, Cuban American wages were significantly higher than other Hispanic groups. Borjas (1982) finds Cubans were more economically advanced because their access to US schooling. The group that immigrated to the US from Cuba in the 1960s was mainly composed of upper middle-class workers who were more likely to have higher levels of education. Another reason as to why Cubans progressed faster than non-Cuban Hispanics could be to lower "return mobility" (Borjas 1982). For example, Puerto Rican and Mexicans were more likely to return to their native homelands than exiled Cubans. Borjas (1982) argued that Cubans had more incentive to "adapt" to the US labor market than any other Hispanic immigrants.

Sanchez-Soto (2018) investigates the differential returns on education between non-Hispanic whites and four groups of Hispanics: Puerto Ricans, Mexicans, Cubans and other Latin Americans. These are exactly the same groups Borjas (1982) investigated. While Borjas (1982) concluded that Cubans invested more in U.S. education during the latter half of the 20th century, Sanchez-Soto (2018) provides updated evidence showing how these groups' investment in education has evolved in the 21st century. The study concludes that all Hispanics have lower returns on higher education than white workers with the same abilities. Borjas (1982) found this to be the case for all groups except Cubans, so Cuban differentials converged over time to that of other Hispanic subgroups. Hispanics also have lower returns to post-graduate and graduate education, which makes the associated costs higher for their group because they are often underpaid (gap between 18%-20%) even when obtaining a job in line with their skills (Sanchez-Soto 2018).

For women, Mexican and Puerto Ricans have the highest returns for low levels of education. Meaning in the short run, Mexican and Puerto Rican women are more likely to work because the gains of pursuing a higher education are not large enough to compensate for time spent in school. Sanchez-Soto's findings are consistent with Borjas' (1982) when it comes to time spent in the U.S. for foreign born Hispanics. Cubans experience an increase in wages when spending more time in the U.S., but Mexicans and other Hispanics do not. Comparing literature from the U.S. labor market for blacks versus Hispanics, we see that both experience wage discrimination, but that the returns to education are lower for Hispanics than black workers. This makes it important to consider how pre-market discrimination affects Cubans and other Hispanic immigrants in comparison to black workers.

Similar to the experiences of black workers and black children, Hispanic children start at a disadvantage in comparison to their white counterparts (Carneiro 2005). Carneiro (2005) refers to this disadvantage as a cognitive deficit, but this does not mean that Hispanic children are not as intelligent as white children. It means that have reduced access to schooling and are inherently less likely to attend school. In fact, Hispanic children are less likely to attend school than both black and white children (Carneiro 2005). This leads to Hispanic children having lower test scores than black children. Carneiro (2005) also points out that when controlling for test scores and schooling, there is no evidence suggesting a significant wage gap for Hispanics in comparison to whites. This means that Carneiro's (2005) results point towards pre-market discrimination as the main source of discrimination, unlike studies such as Darity Jr. and Mason (1998). With this, he concludes that social policy should be designed to target younger minorities because by the time they are of working age, they are at an extreme disadvantage due to their lack of development. In addition to Carneiro (2005), other studies have explored the relationship between the physical appearance of Hispanics and the discrimination they face.

Hispanic people can come from several different types of origins. Hispanics can be black, white, Native American or other ethnicities. The dynamic between black and white workers in the US is similar to that of white-Hispanic and non-white Hispanics in the United States (Telles and Murouia 1988). In addition to the wage differentials between all Hispanics and white workers, darker-skinned Hispanics tend to face a sort of double-exploitation. This type of analysis is known as within-group discrimination. Telles and Murouia (1998) looked at this dynamic by examining a group of Mexican-Americans who had groups with three different phenotypic features; light, medium and dark skin tone. They found that each group experienced lower wages than their white counterparts with equal skills and experiences. However, their results also highlighted a significant difference between lighter and darker Mexican-Americans. The lighter, more European looking worker had a significant premium compared to the darker worker. Although this analysis focused on Mexican-Americans, other Hispanic groups suffer from similar dynamics because of the variance in the physical features of Hispanic people.

In addition to examining wage differentials for Hispanics based on physical features, differences based on immigrant status have become relevant as well. This is due to the differences in wages within immigrant groups and between immigrants and United States born Hispanics (Rivera-Batiz 1999). With millions of illegal immigrants in the United States, the analysis of the conditions they face could provide insight into wage differentials for certain Hispanic groups. Rivera-Batiz's (1999) analysis of illegal Mexican immigrants in the United States displays that illegal immigrants take significant pay cuts to work in the United States. In addition to discrimination, illegal immigrants suffer from this pay cut because of lower educational attainment, lower English proficiency, and fewer years spent as residents in the United States.

Rivera-Batiz (1999) estimated that Mexican illegal immigrants earn over 40% less than documented Mexican immigrants. He also concludes that human capital differences do not makeup as much of the differentials as prior research had suggested. Mainly because undocumented workers experience extreme levels of discrimination, relative to legal residents. Illegal status plays a significant role in the earnings differentials; when undocumented workers become legalized, their earnings immediately increase.

Several studies point out the significance of wage differentials between Hispanic and white workers. The key indicators of the wage gap are discrimination, education, age, etc., but none of these studies examine what Hispanic people experience in their everyday lives. Findling (2017) investigated this issue in a survey that asked questions like; do you feel that you experience discrimination in health care, workplace, and several other settings.

Using a logistic regression, she attempted to analyze the odds that a white or Hispanic worker experiences discrimination. As was expected, discrimination occurs in several different areas for Hispanics in comparison to whites. First, Hispanics reported significantly higher levels of discrimination in health care and other social institutions, regardless of socioeconomic status. Additionally, Hispanics reported that they avoid seeking health care due to anticipated discrimination.

The results also showed that education was not sufficient to offset discrimination in any domain (Findling 2017). For example, Latinos with higher education reported higher levels of discrimination in their jobs, police interactions, health care, and college. On the other hand, whites with higher socioeconomic status reported the lowest level of discrimination of any sub-group in the sample. She states she is unsure of the cause of this result, but other research suggested that it could be to lack of exposure, lack of connections and other social disadvantages (Mora and Dávila 2018). The difference in earnings between Hispanics and whites with higher education is alarming because it is directly related to the lack of representation of Latinos in decision-making roles. Without a seat at the table, it has been and always will be difficult for Latinos to overcome these hardships.

Section III: The United States Labor Market Today For Hispanic/Latino Workers

This section presents an overview of the variables that affect wage differentials experienced by Hispanic and Latino workers in the U.S. labor market. To start, Hispanic men have earned about 15% less than white men in recent decades, adjusted for human capital and other individual characteristics (Mora and Dávila 2018). Mora and Dávila (2018) provide an in-depth analysis of the Hispanic-white wage gap today and how factors Hispanics have faced recently might differ from years past. First, they address is the growth of Hispanics in the U.S. and their expanding participation in the labor force. In recently discussed papers, the composition of the workforce was different because Hispanics did not play as great a role as they do today.

Taking a look at age and experience we see the average age of a Hispanic worker, both males and females are consistently younger than the average white worker. In 2017, the average age of a Hispanic male worker was 38.6 years and 41.9 years for a white male worker (Mora and Dávila). This lines up with previous evidence stating Puerto Rican and Mexican workers are younger than white workers on average. When a white worker has more experience, it means they have higher salaries, and can get raises and promotions faster as well.

In terms of education, Hispanic men and women have both made improvements in recent decades, but still trail far behind white workers. From 2000-2017, the average share of men with a bachelor's degree or higher was 36.6% for white men and 12.7% for Hispanic men (Mora and Dávila 2018), an improvement from 9.4% between 1979-1999. For Hispanic women, the share increased from 11.8% to 19.3%. Gonzalez (2013) suggests that Hispanics have lower levels of educational attainment because they have an obligation to their families to be an economic contributor at an early age. He also states a key barrier to Hispanics achieving higher rates of education is their English languages literacy rates. Furthermore, Hispanic men with college degrees face a wider wage gap than those with less education. For Hispanic women, the wage gap is the same across all levels of education. This confirms the literature that there are additional costs to Hispanic workers pursuing higher education. Oreopoulos and Petronijevic (2013) researched the question of positive returns to education for all workers in general and found that returns vary across schools and majors but did not make any specifications on race or gender.

Hispanic and black workers are catching up with white workers as far as education is concerned. For men, the total college enrollment rates in 2015 were as follows; 39% White, 34% Black and 33% Hispanic (Gilette 2017). White men have stagnated since 2005, but black and Hispanic men made jumps of 6% and 12%, respectively. However, these statistics are not at all reflected when we observe the racial and ethnic composition of administrators in higher education. In 2016, these positions were composed of 86% white workers, 7% black and 3% Hispanic/Latino (Bischel and McChesney 2017). This dynamic almost exactly mirrors the private industry where senior-level executives are white 87% of the time. Again, this confirms the idea from Darity Jr. and Mason (1998) that certain racial/gender groups can dominate certain positions for a multitude of years and create a split between workers and executives.

We cannot expect the numbers to be anywhere close to the composition of the college enrollment rates. These rates are as of 2015 and the pool of candidates must consist of those workers with a master's degree. Understanding the racial composition of this group will give a better idea of how these outcomes come about. The racial/ethnic composition appear as; 66% White, 14% Black and 9% Hispanic (Bischel and McChesney 2017). This issue has been prevalent in the U.S. for years, yet there have been minimal gains made by minority groups. In 2005, 77% of students with master's degrees were white, so it did decrease over 10 years, but that has yet to come across in the composition of higher education administrators.

The Hispanic sub-group breakdown of college enrollment in two- or four-year colleges and universities is dominated by the same four groups as prior literature suggests but in slightly different percentages. In 2014, it was; 46% Cuban, 41% Dominican, 33% Puerto Rican and 33% Mexican (Gilette 2017). Although Puerto Rican workers are included in the Hispanic sub-groups, it should be noted that they differ from other Hispanics/Latinos in the U.S. since they are considered U.S. resident citizens. Workers who come from the island to the mainland are not usually considered immigrants. Immigrants makeup a large portion of the Hispanic workers in the U.S. and are typically paid less than non-immigrant workers. Immigrants make up less than 1% of stateside Puerto Rican workers and about 48.7% of Hispanic workers in the US are considered immigrants while only 4.3% of non-Hispanic white workers are considered immigrants (Mora and Dávila 2018).

The density of immigrants in the group of Hispanic workers heavily inflates the wage gap in comparison to white workers (Mora and Dávila 2018). The variance in wage gaps among Hispanic sub-groups supports this idea. Of Puerto Rican, Mexican and Cuban male workers, the wage gaps in 2016 in comparison to white male workers from largest to smallest was; Cuban, Mexican then Puerto Ricans. Looking at the percentages of immigrant workers in those same groups, there are intuitive results. In 2017, 61.9% of Cuban-American workers were immigrants followed by Mexican-American workers at 47.5% then Puerto Ricans at less than 1%.

However, the most recent decade signaled a shift in the composition of immigrants in the United States. With fewer arrivals from Latin America from 2008 to 2018, the share of Latino immigrant workers who lived in the U.S. for fewer than 10 years fell from 32% to 17% (Kochhar 2019). In the same period, the share of immigrant workers with over 20 years spent in the U.S. from 36% to 50%. Kochhar (2019) points out that this shift resulted in an increase in personal incomes of Hispanic immigrants. However, Kochhar (2019) also finds that new arrival Latino immigrants reported higher incomes in 2017 and 2007. One reason for this is because they have higher levels of education than in the past.

Puerto Rican men in the US face higher unemployment rates than all other Hispanic men. The unemployment rate of Puerto Rican men and women is higher on the mainland than on the island. Although they experience advantages in terms of wages, they still face disadvantages elsewhere. Findling (2017) pointed out that few differences exist in perceived discrimination for foreign-born and US-born Hispanics after controlling for several sociodemographic factors, so it seems as if these reasons are more institutional or due to markets, rather than Puerto Ricans being discriminated against for their physical appearances. As was previously suggested, this group typically lives in northeastern cities in the United States in which many workers are highly educated or skilled.

CHAPTER II: LITERATURE REVIEW OF DISCRIMINATION IN MAJOR LEAGUE BASEBALL

Much of the literature on discrimination in the labor market of Major League Baseball suggests there is little, if any, wage discrimination against Hispanics or minorities (Palmer 2006). In a game where performance and skills can be accurately defined, this is not such a surprise. Essentially, general managers and decision-makers have perfect information on a player's history of performance. If we assume that teams want to do whatever they can to win, then a player who enters free agency should earn the pay he deserves according to his past performance. Baseball provides economists with a unique case study because the availability of performance data is easily accessible to the public. There is information on individual characteristics, salary data, and traditional and sabermetric[1] statistics. Using this, many studies have been conducted to explore the phenomenon of market discrimination in baseball. Similar to the US labor market, there has also been work that investigates pre-market and non-market discrimination. This section provides an overview of the literature and a brief critique of how these studies could be improved.

Section I: Discrimination In The Labor Market

[1] Sabermetric statistics are a scientific exploration of player performance in baseball using statistical methods

The process of evaluating discrimination in baseball is similar to that of the U.S. labor market, but the key difference is the skills of players are more evident and available due to performance statistics. Performance statistics can demonstrate whether a player is below average, average or above average in every aspect of the game. By using performance statistics and comparing salaries by race, it is possible to explore the existence of discrimination by owners and general managers when paying players in Major League Baseball. MLB players are under team control for the first six seasons of their careers; the first three are under their rookie contract and in the next three they are eligible for arbitration. In arbitration, a player receives increases in pay based on your performance and other statistics. This is meant to protect both the team and the player (Palmer and King 2006). After those six years, the player is eligible to be a free agent and be bid on by all teams for their service for as many years as both sides agree. By analyzing players with over six years of experience in the major leagues, we can analyze a group of players whose salaries is based on their performance (Palmer and King 2006). While this thesis explores the conditions faced by minor leaguers in their lead-up to the major leagues, the quantitative analysis is strictly on those players with over six years of experience in the major leagues.

By analyzing the entire group containing white, black and Hispanic players, conclusions can be drawn on pay discrimination by controlling for race. In a sample of players who were major leaguers in the 2001 MLB season, it was concluded that there was no presence of discrimination in the MLB (Palmer and King 2006). However, just as the US labor market, analysis can also be conducted which looks at different levels of economic status. In this sample of players from 2001, Palmer and King (2006) manually composed groups of players from three income brackets, less than $2 million salary (low), between $2 million and $7 million (middle), and above $7 million (high). In the entire sample there are 171 white players, 110 Hispanic players and 81 black players, but they do not provide details of each bracket. Using a reduction of errors sum of squares test with logged salaries and a dummy race variable, they estimated the effects of their performance variables and personal characteristics on the salaries on the entire sample, then on each salary group. They found that discrimination does exist for both Hispanic and black "low" income group players. Palmer and King's (2006) interpretation was that fans and owners are more interested in winning games, so they must pay the best players what they deserve or pay the price of losing. In theory, it should be easier for owners to discriminate against players whose performance is easily replaceable because they have little to no effect on winning.

With a similar goal, Holmes (2011) investigated the presence of salary discrimination in Major League Baseball but by using quantile regressions and a larger set of players from additional seasons. Palmer and King (2006) method of manually choosing salary brackets proved to be effective but introduces a type of endogeneity problem that was not accounted for. In contrast, the quantile regression allows the identification of discrimination size by looking at the whole wage distribution. Quantile regression allows you to identify if players experience discrimination when comparing them to other players, assuming they have similar levels of performance and characteristics, and are in the same relative position, or quantile, among their peers.

Holmes (2011) only analyzed position players who entered free agency, and leaves pitchers out of the analysis like Palmer and King (2006). In baseball, there are pitchers and hitters. The hitters are often referred to as position players because they all play a position in the field. Ultimately, he also finds little evidence of salary discrimination in Major League Baseball. The one area where he does find statistically significant conclusions is in the bottom quantiles of the salary distribution. He found a 16% premium for free agent white position players relative to black players. He also finds a statistically significant 5% penalty for Latinos. A possible reason for the differences in their conclusions is the use of conditional quantile regression and removal of position players with between four to six years of experience. In the context of Holmes (2011) analysis, Hispanics experience an 11% premium relative to black players and 5% penalty relative to white players in the bottom quantiles of earners.

Singh, Sack and Dick (2003) investigate whether salary discrimination exists among players who have yet to reach free agency. They investigated the group of players with one to six years of experience through hierarchical regression and Ordinary Least Squares. Hierarchical regression can yield estimates of association that are more precise than conventional OLS estimates because it can stabilize imprecise estimates of regression model parameters (Richardson 2015). They concluded that there is no existence of discrimination among those who have yet to reach free agency, or those who have already been bid on in free agency (Singh, Sack and Dick 2003). Both the former and latter conclusions are in line with Holmes' (2011) results, although he only analyzed those who have already hit free agency. The key controls in this analysis were position, experience and team's salary cap, but in terms of performance, they only used offensive statistics. This could create a bias on the sample because it will more heavily weigh offense-first players.

From the results, they concluded that race was a significant predictor of position, but not of salary. However, they also concluded that position was a significant predictor of salary, so race seems to play an indirect role despite the lack of statistical significance. Experience was also a key predictor of salary, but this is expected because of the way contracts in the MLB are structured according to service time. There are automatic temporal wage increases so the relationship is linear. This points to the idea of pre-market discrimination. Singh, Sack and Dick (2003) do not entertain this, but other work has been done in baseball research to explore it.

Bellemore (2001) investigates if discrimination exists in minority players promotion from the minor leagues to the major leagues, specifically black and Hispanic players. A player often must go through at least three levels of the minor leagues before they receive a promotion to the majors. When that happens and the player becomes a full-time major leaguer, they receive the minimum salary for a season which is now $563,000 but as recently as 2009 had been $400,000.[2] However, upon entry to the MLB players get a significant pay jump. Players also get raises for being promoted within the minor leagues. For these reasons, promotion is huge because of the financial security associated with it.

The variables chosen to indicate performance in Bellemore's (2001) study were Batting Average, Runs Batted In (RBI) and Home Runs. An RBI is a batted ball event that allows a runner on base to score. It is a counting statistic that indicates how many runners have scored when a specific player was up to bat. The data for this study was taken from mainly the 1990s. The rules of negotiation and labor dynamics were extremely different from what they are in today's game. Therefore, this analysis reflects the institutions of the time.

The main conclusions are consistent with expectations and more reflect labor pre-market outcomes in the US labor market. Black players were about 8% less likely to be promoted than their white counterparts and Hispanic players were about 7% less likely. However, during the expansion years when more jobs opened up, this advantage disappears. An expansion year can be defined as a season when another team is added to the league. Since an MLB team is added,

2 Baseball Reference. 2020. Minimum Salary. Available at: https://www.baseball-reference.com/bullpen/Minimum_salary

there affiliated minor league teams must be added as well, leading to hundreds of jobs becoming available. Essentially, Bellemore (2001) assumes that the pool of productive players was not large enough to allow for discrimination when these jobs became available. This tells us that minority players benefitted from expansion because they were no longer marginalized. The one caveat here is there has not been an expansion year in the MLB in over a decade and there are currently no negotiations or rumors which point towards this possibility.

For the time, this paper was effective in bringing an issue to the forefront. It provides significant results that point towards the existence of pre-market discrimination in the MLB. Bellemore (2001) may have introduced a bias when using RBI's as a variable since players with better offensive teammates who have a higher probability of being on base and able to score create a higher probability of achieving an RBI. Therefore, a player may get an advantage because he plays for an elite team.

Bellemore's (2001) research and conclusions address a serious issue in MLB and it was just a mere prelude to what has happened in the MLB in the last 20 years. There has been a huge spike of Latino players in the MLB. If Bellemore's conclusions persist when updated to today's players, Latino players could be at a huge disadvantage. The demographics of the majors and minors already suggest Latino players do not get the same opportunity, but if this proves statistically significant, then it could be a way for Latino players to get themselves representation in CBA negotiations. Also, is it possible this discrimination that was seen in the 1990s is a reason why there has been such a huge depletion of black players in the MLB? In other words, black players may see greater opportunity elsewhere since there was a struggle to be promoted to the major leagues. Since the 1990s, the percentage of black players in the MLB has declined. It could be for a multitude of reasons, and it would be in line with expectations for the struggle for promotion to be one of them.

Breunig et al. (2012) studies the effect of wage dispersion on team performance finding that teams with a higher dispersion of wages are worse off than those with an equal dispersion. In other words, he looks to investigate the relationship between wage inequality within teams and their effect on winning percentage. From a baseball perspective, he queries the impact of a team's dispersion of talent on the chances of winning. Using a contest success function, he makes winning a probabilistic function of a team's effort, and the opponent's effort. Optimal individual effort is a function of wage dispersion, a parameter which determines if wage dispersion has a positive effect on effort and the team's probability of winning.

The results identify a significant negative relationship between wage inequality and winning percentage. Therefore, teams with a larger dispersion of wages are less likely to win under his theoretical model. A handful of conclusions can be drawn from his statistical findings. First, it is possible to construct a model in which wage dispersion, through the use of something like a Gini coefficient, matters in the determination of a given variable. The second is that wage data can be used to predict outcomes in the MLB on both a game and season-long level.

From this, they say there can be several alternative interpretations depending on the loftiness of the assumptions. By fully accepting the assumptions, larger wage dispersion leads to lower individual effort and therefore lower team performance. They draw this conclusion because teams with lower ability have a lower marginal benefit of exuding full effort when their win probability is already low. While effort is on itself an intangible concept, performance can be measured with a high degree of accuracy. Assuming that players with lower ability give less effort is an idea that cannot be calculated. Promotion from the minor leagues to the major leagues comes with great benefit. The costs of losing this opportunity are high in terms of potential salary. Players with lower abilities are more at-risk of losing huge amounts of dollars because they can lose their job at any point and are not guaranteed long-term security. The shares of Latinos in the minors is greater than that of the majors. This does not necessarily mean that Latinos give less effort. If this was the case, then Bellemore's (2001) conclusion of why these players are less likely to receive promotion is easily explained. For these reasons, they soften the assumptions.

They assume that all players give 100% effort, and interpret the results as follows; teams with a more equitable distribution of ability and wages have a higher probability of winning (Breunig et al. 2012). Therefore, if teams have $10 million to spend, it makes more sense to sign two good players ($6 million and $4 million), opposed to one great player and one bad player ($9 million and $1 million). When they ease the assumptions, they come to a more conservative, and realistic, conclusion. Baseball requires more than one player to perform well in the quest to victory. To win a game, a team needs good performances from hitters and fielders, starting pitchers, and relief pitchers. If a team can manage to keep a group of elite and good players for an affordable price, then their chances of winning increase. As Breunig et al. (2012) suggest, teams with a larger dispersion of salary, are less likely to win in the short and long-term. Perhaps this is why the trend of signing young, elite Latino players while they are cheaper has grown so quickly.

The last interpretation explores the idea that player salaries are not meant to optimize team winning, but rather to optimize team revenue and sales. By having a high-profile player who makes way more in salary than any other player, there could be a boost in revenue, ticket sales, etc. This has been largely neglected in the analysis of sports, but it makes sense if we assume that teams wish to make more revenue than they do to win as many games as possible. Fans of sports teams generally think that teams want to win, but sports are a business. Nielsen's 2013 Sports Media Report[3] states 83% of baseball fans who used a television to watch baseball games were white. If owners are attempting to satisfy their fanbase, it could make sense to invest into a player who resembles their fans. In baseball, this is how customer discrimination takes form. Customer discrimination could also explain why owners are less likely to promote minority players with equal performance of white players.

Other work has been done on the effects of race in baseball, outside of players' wages. It takes a certain skill or ability to be a manager in MLB. A manager is the leader of the players, and the rest of the coaching staff. As U.S. labor market research suggested, decision-making/managerial roles could be dominated by certain races and genders. This dynamic persists in Major League Baseball (Volz 2012). Even when controlling for variables like performance, education, coaching, minor league managing, and playing experience, it was found that between 1975 to 2008, former Hispanic players were over 65% less likely to be a major league manager than former white players. Former black players were 74% less likely than former white players (Volz 2012). Volz (2012) found a key reason for this was Hispanic and black players are much more likely to play positions that do not supply managerial positions in the major leagues. The density of certain races at a particular position is a key reason as why Hispanics face discrimination. Without many Hispanic and black players in managerial roles for MLB teams, it could make it harder for these players to easily assimilate into a team's culture. This introduces the idea of employee/co-worker discrimination in MLB. Since most players in

[3] Chang, Alvin, 2013. This is Why Baseball is so White. Vox. Available at: https://www.vox.com/2016/10/27/13416798/cubs-dodgers-baseball-white-diverse

the league are white, they could prefer a white manager, even if a minority manager is just as qualified.

In addition to pre-market and out of market discrimination, there are analyses of exit-discrimination when it comes to the racial composition of players in the National Baseball Hall of Fame (Jewell, Brown and Miles 2002). They tested the idea that discrimination could extend to voting for player membership in the hall of fame. The consensus in prior research had pointed to discrimination against black and Latin American players eligible to enter the hall of fame. Jewell, Brown and Miles (2002) controlled for all types of career statistics, then concluded minority players who were both Latin American and black have been discriminated against in hall of fame, but not other Latin Americans or black players. A similar result was observed on the investigation of wage differentials between Mexican-Americans of different physical complexion. This may not disincentivize Latin American players at all, but the hall of fame is the highest honor for any baseball player. Not receiving the recognition deserved is consequential.

In general, research in baseball tends to trail behind US labor market research. Jackson and Pradhan (2019) spelled out how MLB front offices have implemented predictive analysis in their decision making, whether that be in business or sport-related decision making. The goal was to see how this type of analysis can see-saw between explorative and exploitative. On the baseball side of decision making, teams now develop projection systems that can estimate a player's performance in the short and long-term. When deciding on whether to sign a free agent, teams use these models to determine the expected future performance of a player. This same method can be used for deciding whether to extend a player's contract before they become a free agent, or when to use a player in a trade. Projection systems have become a staple of baseball analysis.

None of discussed literature has been updated to highlight the strides made in the analysis of the performance of players. Comprehensive statistics have been developed which assess a player's ability to create or save runs, irrespective of his teammate's performance. When it comes down to it, teams want players who dominate the game on both offense and defense. The literature's use mainly offensive output created bias in a few of the studies. For example, Singh et al. (2003) only focused on offensive statistics, so it is natural to question how his results might change if he included offense and defense. Holmes (2011) and Palmer and

King (2006) used defensive merit as well. However, they used publicly recognized awards that do not always coincide with the best statistical defenders at each position. There is significant human bias when awards for each season are chosen because there is no statistical standard. Those vote upon the awards do not rely on statistical data which measures defensive performance. Instead, the voters can support whichever player they personally think is the best defender

By molding the quantitative techniques of Holmes (2011) with modern sabermetric statistics, this analysis will serve a few purposes. First, it will solidify if sabermetric statistics are significant in the estimating salaries more so or the same than traditional statistics. Next, it will clarify whether the trends that were observed in Holmes (2011) and Palmer and King (2006) persist under the dynamics of the current collective bargaining agreement. Pitchers will be incorporated into the analysis as well. WAR is available for all types of players, so the methodology can be applied across the entirety of positions.

In addition, I analyze the conditions faced by Latinos leading up to the major leagues. This is relevant to the quantitative analysis because it provides context on the motivation to exploring wages for Latino players. As Bellemore (2001) shows, Latinos and other minorities face entry-discrimination into the major leagues. This topic will be investigated further. The next section presents context about baseball's labor market, the conditions Latinos face in their march to the major leagues, and the improvements made to the analysis of current and future performance in baseball. Context is important in understanding why this analysis is relevant to the current events that MLB faces.

CHAPTER III: INFORMATION ON BASEBALL TODAY AND THE PATHWAY TO THE MAJOR LEAGUES

Every one of the analyses of both the U.S. labor market research and MLB labor market research pre-date the most recent collective bargaining agreement between the Major League Baseball Players Association and the team owners. The conclusions from the literature review will shape how I form my argument, but there is a key addition. First, I will need to know if the arguments made by Holmes (2011) and Palmer and King (2006) persist under today's institutions and second, does the use of a more accurate set of performance statistics make any difference to the analysis. But since my quantitative analysis will be similar to Holmes (2011) and Palmer and King (2006) in the sense that it is specific to only a certain subgroup, there could still be evidence that suggests some sort of discrimination elsewhere, such as for amateurs or minor leaguers. For now, I will restrict the quantitative analysis to those who enter the free-agent market in the major leagues, but the conditions they face leading up to that could influence why their outcomes come about.

Section I: Alternative Outlets

Although there is not any published work investigating all types of discrimination in baseball in one grand study, there are other outlets that have looked into it. For example, Matt Swartz (2014) used the Wins Above Replacement (WAR) framework to investigate wage differentials across different races and player-types in MLB. WAR is a comprehensive performance statistic proposed by Tango (2006) in his pivotal analysis of several important baseball topics. Equation A3-A14 in the appendix chapter discuss WAR in detail.

WAR attempts to combine a player's entire contribution to run creation by assigning run values to his performance on both offense and defense. It then normalizes across the entire league and converts run creation to a win total. This win total indicates how many more, or fewer, wins a player has contributed to his team's success. A 0 WAR player is one whose performance can be easily replaced from the highest level of the minor leagues, triple-A, or from available free agents. The replacement level player is worth the minimum salary in the major leagues. A common misconception in the use of WAR is that a 0 WAR player is an average player. This is not the case. An average player is usually worth about 2 WAR. An average major league player is not the same as a replacement level player (0 WAR). More will be discussed on WAR later in the analysis. For further knowledge, read the creator himself, Tango's (2006) explanation.[4]

In baseball, understanding contract structures is necessary if we want to carry out any labor market research. After a player enters the league and attains permanently full-time status, they begin their team-controlled contract. The first part is the three-year rookie deal at minimum salary with automatic pay increases, in which the magnitude is random. In the next three-year

[4] Tango, Tom. "How to Calculate WAR."
http://www.insidethebook.com/ee/index.php/site/comments/how_to_calculate_war/

period remuneration is determined by arbitration. Arbitration is a period in which a player receives significant raises based on his performance, position and other variables. It is the first time where there is formal negotiation between the team and the player. In rare cases, a player can have four years of arbitration and only two years of a rookie deal. In the first two years of arbitration, there is a lot of negotiating which is meant to protect both sides, but usually favors the team. In the last year of arbitration, the player is supposed to earn what both sides expect him to earn as a free agent. Swartz (2014) finds evidence of discrimination against Latino players who sign extensions years before they reach free agency, when controlling for their WAR.

Teams have the option of offering their own players extensions before they hit free agency. Extensions buy-out a player's rights for the negotiated amount of time. In other words, this could be to buy-out their pre-free agent years, some free agent years, or both. Swartz (2014) split his sample of pre-free agent players into two groups; those who signed extensions within a year of free agency and those who signed earlier than a year before. Every player in the sample was at least a replacement level player (0 WAR). Swartz (2014) used a $/WAR framework to compare players. The $/WAR framework uses the average value of a player's contract and divides it by the total WAR of a player in a given season. The contract values were adjusted for inflation to control for teams paying a different $/WAR in different seasons.

For white players, the cost per WAR is nearly identical between both groups. However, for Latino players the number drastically changes in a negative direction for players who sign earlier than a year before free agency. For those Latinos who sign within a year, the cost is $6.8 million per WAR and for those who sign over a year away it is $4.9 million per WAR. In addition, for white players who sign over a year away from free agency, the cost is $7.9 million per WAR. Latino players take a severe discount relative to white players. Swartz (2014) does not adapt his data to deal with players on different points of the WAR distribution. However, anecdotal evidence suggests that teams are only willing to commit money to elite performing players. There is less risk in committing long-term money to a guaranteed elite player. Otherwise, teams could lose millions when they had no reason to commit to the risk.

Although I will not analyze players who sign extensions before free agency, it is important to understand how the WAR framework can be adapted and used as a comparison tool in general. For example, when looking at a player with a 6 WAR versus a 3.5 WAR, it is easily distinguishable which is the more valuable player. The same goes for a 3.5 WAR player versus a 1 WAR player. It is a method in which we know which players are elite, good, average and replaceable. Furthermore, there could be player who are below replacement level, or have negative WAR totals. This is quite rare, but it means that this player was worse than the minimum level free agent or minor league replacement. It is rare because teams almost always do not tolerate a player who performs awfully when there are several other options. A player with a negative WAR has a such a high sunk cost, because every team has access to replacement level (0 WAR) players.

However, the outcomes in the baseball labor market research do not line-up with Swartz' findings for pre-market players. That is one of the motivating factors for conducting the analysis on those with over six years of service time using WAR as one of the main performance statistics. There are a number of reasons why Latino players might accept these severely discounted contract extensions before free agency. There is an issue with asymmetric information regarding present and future performance of players, giving teams an advantage over both players and opposing teams in the future. Teams know a lot about their own players. They know where they come from, what they were like as a teenager, what their work ethic is, etc. They can create projections for the expectations of the player's future performance based on a plethora of information they gathered over the years. If the team is well aware that the player came from a very underprivileged background, they are presented with an opportunity to extract surplus out of the player. That is, get them to accept a deal below what they expect them to be worth in the future.

It is hard for a 20-23-year-old youth to refuse the certainty of millions of dollars today. Especially when they never had significant financial security. Teams take the chance on a player having a career-ending injury or a fall-off in performance. However, Swartz (2014) has shown this has not been the case thus far. Teams have earned a bargain on these contracts. The sacrifice of potential future earnings has been too great for the players and it has saved teams millions. You might ask what the teams have to lose in this scenario. The answer is not much. Teams have become keen to taking these so-called risks because they continue to pay off. Teams are relying on their projection systems and so-far they have yet to fail. As each season comes, more and more young Latino players sign extensions very early in their careers.

Section II: How Analytics In Baseball Can Be Used For Exploitation

Jackson and Pradhan (2019) describe how MLB front offices have implemented predictive analysis in their decision making, whether that be in business or sport-related decision making. The goal was to see how this type of analysis can see-saw between explorative and exploitative, just as seen in Swartz' (2014) work. On the baseball side of decision making, teams now develop projection systems that can estimate a player's performance in the short and long-term. When deciding on whether to sign a free agent, teams use these models to determine what they may expect from a player in their future performance, like with young Latino players. This same method can be used for deciding whether to extend a player's contract before they become a free agent, or when you trade for a player. Projection systems have become a staple of baseball analysis and possibly the key way in which teams decide on who to target for an early extension, and who not to target.

For example, there is a well-known public projection system in Major League Baseball called the Szymborski Projection System (ZiPS).[5] This system uses growth and decline curves based on player type to find trends in player performance. Using those trends and prior performance, a projection can be made for a player or team. The system uses the past four years of performance and weights the most recent seasons more heavily. However, teams use variations of these types of systems when factoring in their own projections of players. Jayal, McRobert, Oatley and O'Donoghue (2018) discussed the many ways machine learning and data analysis are used to make decisions in sports today.

The question that Jackson and Pradhan (2019) attempt to answer is how teams can use data to oscillate between exploration and exploitation. This question serves as an incentive to investigate why players may sign contracts that do not necessarily go in-line with their projection systems. In other words, it is quite simple to conduct an experiment that tries to find out how much a player is worth depending on their performance or what they can be worth in the future. Using statistics like WAR allows us to get an accurate estimate of how many wins a player was worth in a given season. Applying that and their salary data, these systems can help estimate if the player was under-valued, over-valued, or performed to expectations.

When a player like Ozzie Albies signs a deal for $35 million over seven years, the question of whether these projection systems can be used for exploitation is brought to the

[5] Szymborski, D. 2020. ZiPS. FanGraphs. Available at: https://www.fangraphs.com/projections.aspx?pos=all&stats=bat&type=zips

forefront. Albies' contract has two team-friendly club options attached that can bring his final value to nine years for $45 million. ZiPS tells us that over the lifetime of Ables' nine-year extension he is projected to be worth $282 million, according to WAR and other statistics. In a more conservative estimate, he is expected to be worth $153 million. Even being grossly conservative in a projection system that is already inherently conservative, it can be estimated that the Atlanta Braves will accrue over $100 million in surplus for Albies' performance. This makes the analysis I conduct even more prevalent. Although I will analyze players who have hit free agency, this makes it obvious that the research on discrimination in baseball is multi-faceted. Players who have more than six years of service time have already entered free agency or signed an extension which covers some of their free agent years. If a player signed an extension early in their career and has over six years of service, they will be included in my analysis. The idea is that players from different ethnicities may experience different levels of pay, like Swartz finds in his analysis.

Section III: Labor Relations And Why Latino Players Are More Likely To Face Wage Discrimination

Major League Baseball players have always been the losers in labor negotiations. Although the league has been around since the early 20th century, the first collective bargaining agreement between the Major League Baseball Players Association (MLBPA) and the owners was not signed until the 1968 (Williams 2019). To ensure that players never entertained leaving the league for a competitor, the Supreme Court ruled in 1922 that the league was exempt from the antitrust laws of the United States (Williams 2019). In other words, if you wanted to play professional baseball, you could only work for MLB.

The players have been subject to other uncommon regulations. Today, a player is bound to the team that drafted or signed him as an amateur for six years, and in rare cases, seven years. However, this was not always the case. Until 1975, players were subject to the reverse clause. The reverse clause was agreed to in secrecy in 1879 in the yearly owners meeting. The clause was only beneficial to the owners. It ignored players' rights for almost a decade. Their wages were suppressed because the clause bound them to their team for their entire careers. Led by union activist and economist, Marvin Miller, the MLBPA finally negotiated arbitration into players' contracts in 1970, allowing them to negotiate pay increases and resolve grievances (Williams 2019).

Only in 1975, did the players and Miller successfully abolish the reverse clause and guarantee player's the right to free agency. Williams (2019) explains that in the economic explosion of the post-WWII economy, the average worker experienced a salary raise of $2,500 while the MLB worker only experienced an increase in minimum player salary of $1,000 ($6,000 to $7,000) from 1950 to 1965. Since the dawn of the league, the players have been the losers in labor negotiations. Today, MLB is still exempt from the country's anti-trust laws, making it impossible for the league to have equal rights as the rest of U.S. workers. Miller was sought out by the players of the league because of his expertise in union relations. Despite his work, he still described MLB players has the most exploited group of workers he had ever seen (Williams 2019). The league's most valuable commodity is its players. Without them, there would be no league or labor sport empire.

For decades, the league was able to exploit all of its players. Eventually the players came together and unionized, letting them gain workers' rights that were long-stripped of. Today, exploitation has taken on a new face. For years Latino players have claimed they are the product of prejudice and a tougher set of standards during their lives in MLB (Vargas 2000). Vargas (2000) was the President of the Venezuelan Baseball Players Association for over a decade. He represented all Venezuelan players in the MLBPA. Among prejudice, Latino players say they deal with isolation from family members and lack of career opportunities after baseball. The players have trouble obtaining visas for their family members to come to the states. As we learned earlier, it very unlikely a Latino player becomes a coach or manager. This sentiment is felt among the players when they retire. Vargas (2000) states MLB teams have systematically denied Latinos these opportunities because of a "boatload" mentality. He means, when the teams are finished with the players, they ship them away and do not care what comes about their lives.

The Latino Experience as an Amateur and Minor Leaguer

Puerto Rican baseball players have always experienced different circumstances than their fellow Latino players. Since Puerto Rican players are citizens of the U.S., they are not subject to the same set of rules as international players. Attending schools in the U.S. brings them a serious advantage because they can participate in the MLB draft instead of becoming an international free agent. Players from the U.S., Canada and Puerto Rico are eligible for the MLB First-Year Player Draft. The island's status as a commonwealth is a benefit for its players. It is an interesting case study to see if being part of the MLB draft helps Puerto Rican players when it comes time to enter free agency. Since they have already reaped the benefits of larger signing bonuses in the draft than international free agency, they start their careers with an advantage relative to their fellow Hispanic players.

The reason why players who enter the draft are at an advantage is due to the signing bonuses. Top tier talent in the draft could have signing bonuses as high as $7 million. However, the record-setting signing bonus for a Latin American prospect was just set this year by 16-year-old Jasson Dominguez at just north of $5 million.[6] In addition to that, players in the US can be drafted in several ways such as out of high school, junior college, after their junior year of college, and after their senior year of college. But a Latino prospect will only be offered the highly desired million-dollar signing bonus as a 16-year-old kid because it contains an implicit subsequent cost. Players in the US can only be drafted as early as 16. However, that age sharply reduces a Latino player's signing bonus opportunities. When a player who is older than 16 signs a deal with an MLB club, they are considered not as developed and therefore more of a risk for the team. The older the player is, the less time he has to catch up with the rest of his peers.

Although it is called international "free" agency, free does not describe the situation accurately. Every year on July 2nd, teams can sign all of the eligible international prospects. Their spending is capped so that each team begins with the same signing pool. As the 2017 collective bargaining agreement states "…shall be reallocated equally among the international Signing Bonus Pools among clubs that did not forfeit Pool space…" Some teams trade away their spending money throughout the year and others do not. The end result is that there is a range of spending from team to team. This means that a player really cannot sign with whatever team he wants because some teams do not have the spending money to do so. The MLB draft has caps on certain slots, but the limitations are nowhere near as great as they are in international free agency.

The majority of players who wish to pursue a life in baseball from Latin American countries usually drop out of school by the age of 13 at the latest or they risk falling behind their peers. Despite this, the estimated success rate for these kids getting signed by an MLB team is

[6] Wagner, James. 2019. "This One Was Worth It: Yankees Set Record With Jasson Dominguez. The New York Times. https://www.nytimes.com/2019/07/02/sports/jasson-dominguez-yankees.html

just around two percent,[7] leaving many kids without a job or education by their mid-teens. The tole of MLB is to raise and protect their own players, but little is done to target the players who fail in their attempt to reach the big leagues. This alone shows the risk that these kids must take to find their way out, and possibly deliver their family life of prosperity.

On top of that, once Latinos reach the minor leagues, it is not as if they get the best treatment. Minor league wages are extremely low. In the lowest levels (Single-A) players make between $6,050 to $8400 per year and in the highest level (Triple-A) players make between $11,825 to $14,850 per year (Williams 2019). Minor leaguers are not treated as U.S. workers because they are exempt from the United States minimum wage laws. Major League Baseball is the only sports league in the United States exempt from antitrust laws (Williams 2019). This has led to a labor dynamic unlike anywhere else in sports or the United States. The major leaguers have extremely high wages relative to the majority of U.S. workers, but the minor leaguers face the complete opposite. A minor leaguer's salary averages just over $1000 a month for between five to six months. Their time commitment to their job is just as severe as it is in the majors. To rise in the minors and receive promotions, training is a full-time job in the off-season. There is not much time for part-time work to increase income. For players who come from well-off backgrounds, this is not a problem because they have the necessary resources to live this life without sacrificing too much in terms of their standard of living, but it is harder for players from poorer backgrounds, like in the case of many Latinos and black players.

In the Dominican Republic, remittances are a vital source of foreign exchange and a major reason why Dominican players continue their life in the minor leagues, even if they do not truly believe in their chances of making the major leagues (Koble 2008). These players send the majority of their salaries home during the season. This means they have little money to support themselves and their families during the season, and they have minimal spending money for food and other resources during the season. Minor league players constantly travel. Their bus rides last hours upon hours. With these harsh conditions and little money, it makes surviving in the minor leagues very difficult. But that does not matter to Latino players because they are making an income to send back to their families. Perhaps, this is the reason for the demographic imbalance in the minor leagues. Half of the players are Latino, but this number does not carry

[7] Lagesse, David. 2016. *Baseball Is A Field of Dreams And Dashed Hopes For Dominicans.* National Public Radio. https://www.npr.org/sections/goatsandsoda/2016/04/03/472699693/baseball-is-a-field-of-dreams-and-dashed-hopes-for-dominicans

over to the major leagues. Despite the MLB being well aware of this situation, the commissioner still proposes ideas that will harshly hurt these players.

In this past winter, Commissioner Manfred introduced a proposal that would eliminate over 40 minor league teams. There are 25 players per team, so that would potentially affect 1000 players and since half of them are Latino, that means roughly 500 Latino players would lose their jobs. Manfred stated the potential benefits for this would be that top prospects have an easier time getting through the minors because it would streamline the process. The top prospects would face better competition from an earlier stage and of course, the league would save a lot of money. However, the MLB receives billions of dollars in revenue every year. This cost-saving idea is like saving pennies. The effect it would have on their finances is minimal, but the effect it would have on those who lost their jobs, especially Latinos, would be devastating.

In Bernie Sanders first elected political office as the mayor of Burlington, Vermont he helped bring a minor league team of the Cincinnati Reds to the town. Sanders argued that to rural neighborhoods, these minor league stadiums are important for culture, and even local revenues. That, combined with the negative effect the proposal would have on hundreds of players, should be enough for Manfred to reconsider his proposal. In fact, Senator Sanders guaranteed that it would not happen after he sat down with the commissioner and discussed why it was wrong. While Sanders fights for minor league teams, President Donald Trump seems to perpetuate an issue that has always hurt minor leaguers. One of his first residential actions was to renew the Save America's Pastime Act. This act exempts minor league baseball players from the minimum wage laws of the United States. It is just another step into Trump fully supporting the nation's corporations into eating up higher profits. The dynamics of the minor leagues are very unique in comparison to any other sector in the United States.

This chapter was meant to provide perspective on the current environment facing Latino players in their path to the major leagues and once they reach it. The road to free agency is very long and sometimes even the most talented Latino players do not make it. Although this chapter provided information on Latino players before they reach the major leagues, the rest will focus on those who have obtained six years of service time in MLB. The information of this chapter was meant to describe how Latinos face discrimination outside of the labor market.

CHAPTER IV: INVESTIGATING MARKET DISCRIMINATION IN BASEBALL

Baseball players have a wide variance of talent and performance. At the top, we have the most elite players who consistently perform well above the average player. As in many other areas, distribution of skill in sports is right-skewed because of the scarcity of elite talent. The sport has accurate ways to define performance which makes it easy for fans and teams to know how well a player has performed. This makes it a nice tool or proxy for skill in labor market studies. We are most concerned with how well a player performed and if their salaries reflect their performance, especially across different races.

I can create a model which uses performance statistics, and other descriptive variables to explain whether a player is paid as much per his performance, in comparison to other races/ethnicities. Holmes (2011) and Palmer and King (2006) found little evidence of discrimination according to a player's performance in MLB. However, Holmes (2011) found evidence of discrimination against minorities in the lowest quantile of earners. An expansion of this research applying a different set of performance statistics as explanatory variables may provide more evidence of discrimination. Holmes (2011) found evidence of discrimination in the bottom-tier of earners. It is possible that using more accurate predictors of performance could find different results. In a similar method to Holmes (2011), I will investigate if the dynamic changed, or stayed the same. With a growing number of elite Latino players signing extensions before they reach free agency, I want to investigate if there is evidence of discrimination in the upper quantiles.

This chapter begins by explaining how other types of models, specifically log-linear least squares regression, can show the effects of performance and race on salary. Following that, I explain how quantile regression can tell a more detailed story about different types of players. Last, I introduce the variables of the models, and why an update from Holmes (2011) is necessary.

Section I: The Models

In a similar procedure to Holmes (2011), I start by analyzing the performance data using Ordinary Least Squares (OLS). This will tell us the average effects of the performance statistics and other variables on a player's salary. Formally, it takes on a log-linear form:

$$\ln(salary) = a_0 + a_1 performance + a_2 player\ characteristics + b * race + e$$

Where performance is defined as a vector of performance statistics (Figure 1). For more info, refer to the Appendix. The statistics for either a pitcher or hitter are available in Figure 1. Player characteristics describes how long the player has been in the league and the position the player plays. However, in models including WAR, position is not included because it is accounted for in WARs. Race describes the player's race/ethnicity. I am most interested in how these variables estimate performance between Hispanic and white players. The race variable can take the form of a white, Hispanic or black player. Since Holmes (2011), there has been a growth of Hispanic players in the major leagues. Within the same time frame, there has been a fall of black players, suggesting that Hispanic players have replaced black players in the league.

Figure 1. Variables

Variable	Explanation
Race	Race of the Player
Service Time	Years Spent in MLB
Position	Position Played by Hitter
WAR	Wins Above Replacement
wOBA	Weighted on-Base Average
OBP	On Base Percentage
SLG	Slugging Percentage
RBI	Runs Batted In
Stolen Bases (SB)	Stolen Bases in a Season
Base Runs (BsR)	Runs Created in Baserunning
Defense Runs (DEF)	Runs Saved on Defense
wRC+[8]	Weighted Runs Created Plus
Wins	Pitcher Wins
Losses	Pitcher Losses
Saves	Pitcher Saves
ERA	Earned Run Average
HR9	Home Runs Allowed per 9 Innings
K9	Strikeouts Per 9 Innings
Innings Pitched	Total Inning Pitched
FIP	Fielding Independent Pitching
LOB%	Left on Base Percentage
xFIP	Expected Fielding Independent Pitching

8 Weighted Runs Created Plus. Percentage above league average hitter.

Using these controls, I can make a few distinctions. First, I can establish whether Holmes (2011) performance statistics are more accurate estimators in the context of my sample or if the modern sabermetric statistics are. Next, I can estimate whether the trends found by Holmes (2011) persist in today's game. He found that Latinos in the lowest quantile of the distribution of earners (10th quantile) earn about 5% fewer dollars than their equally performing white counterparts. The specification of the model will allow us to control for performance in a variety of ways so that there is consistency in the formation of salary across races and models. We assume that each race is subject to the same salary formation. In baseball, the main source of salary formation is performance. Players at similar stages in their career with equal performance ought to be paid similar salaries. Any differences in salary among players with equal performance could reflect discrimination.

There are several variations of the log-linear model which are used. This is to grasp how each set of performance statistics estimates variables both on their own and together. If results of discrimination are consistent across different estimations, then there is more convincing evidence of its existence. By starting the analysis with OLS regression, we can grasp the average effects of the independent variables on salary. For example, the average effects of performance could tell us if players from different races are affected similarly or with gaps. Quantile regression provides further detection of discrimination because of its ability to examine the effects of the independent variables on different areas of the distribution.

Formally, quantile regression will allow us to observe how each quantile of the conditional distribution responds to the observed covariates (Koenker and Hallock 2001). If the effect of a particular covariate, like performance, is not constant over the range of the dependent variable, then quantile regression will reveal this. Following Holmes (2011), I use a quantile regression approach to explore the presence of discrimination through the entire distribution of earners in Major League Baseball. Holmes (2011) investigates the question of whether owners and general managers discriminate against players in different areas of the earning distribution.

When using identical salary formation, we can detect discrimination if one race is treated differently than another within a given quantile of earners. I use quantile regression to estimate salary quantiles using the same specification as in the OLS models. When comparing players from different races in the same quantile, we can assume they have similar performance and characteristics.

The model is estimated as follows:

$$Q(\ln(wage)\, X) = a_0(q) + a_1(q) * performance + a_2(q) player\ characteristics + b * race + e$$

In OLS regression analysis, we solve the problem of minimizing the sum of the squared residuals. Similarly, we can define the median as a solution to minimizing a sum of absolute residuals (Koenker 2001). If instead, we were to minimize a sum of asymmetrically weighted absolute value residuals, then we could yield different quantiles. The weighted residuals are based on the dependent variable, the natural logarithm of salary. Discrimination could be between Hispanics and non-Hispanics, at the median or at different ends of the distribution of earners. We are most interested in seeing if there is any gap in earnings between races among players with similar performance in the upper and lower quantiles.

In order to solve the optimization problem on a specific quantile, we must equate the number of positive and negative residuals to assure there is symmetry in the observations above and below the median. In other words, our sample may be asymmetrically weighted on one end of the distribution. To adjust the distribution so we can yield specific quantiles, we can give varying weights to all positive and negative residuals (Koenker and Hallock 2001), we need to solve this optimization problem below

$$\beta(\tau) = \min_{\beta(\tau)} \sum \rho_\tau(y - \xi\left(x, \beta(\tau)\right)$$

Where $\rho_\tau()$ represents a tilted absolute value function that yields a given quantile $\xi\left(x, \beta(\tau)\right)$ and is a linear function of the given parameters. A tilted absolute value function is how we can asymmetrically weight the distribution by giving different weights to the residuals (Koenker 2001). By varying , we can vary the quantiles. Different specifications may yield different coefficients in the same quantile. Quantile regression will tell us the both the varying degree of coefficients and significance. For example, Holmes (2011) found significant evidence of discrimination in the bottom quantiles. However, none of his specifications found significant results in the upper quantiles. In the context of his sample, there was not enough evidence to draw a conclusion in any direction, suggesting that there is room for more research.

Quantile regression is an ideal analytical tool for sports labor market research, especially baseball. This analysis focuses on using traditional statistics and modern statistics to estimate salaries in Major League Baseball. While it is true statistics like WAR allow us to accurately define similar players on a comprehensive basis, this does not take away from the explanatory power that traditional statistics have had on wages in prior research. Multiple specifications will serve as robustness checks for the models to assure that any results are accurate.

The goal of the quantile approach is to carry out a regression model controlling for performance and other characteristics. However, unlike an OLS model, it estimates the model explaining players' salaries at a given quantile. In this paper, it can be used to tell us the story of whether teams have adapted from discriminating against those players who are replaceable on teams (Holmes 2011). If teams have changed their behavior to discriminating against the upper quantile of players, then the story changes as well. When discriminating against the replaceable players, teams save minimal dollars. These players earn close to major league minimum salaries. On the other hand, if they discriminate against the most expensive players, there is an opportunity to save tens of millions of dollars, and sometimes hundreds of millions. In free agency, an elite player would cost at least $100 million. Several players earn over $200 million every year. Teams that sign these players at cheaper rates could get a significant bargain for the player's performance.

Section II: The Variables

The data comes from a variety of publicly available sources. All information on players' contracts and service time come from Baseball Prospectus' archive called Cot's Baseball Contracts.[9] The data at hand comes from the 2017 season in MLB through the 2019 season. These are the seasons that have been played since the last collective bargaining agreement. All performance measures come from the baseball analytics website, FanGraphs.[10] Only pitchers who qualified for statistical titles were included in the analysis. Only hitters who with at least 300 Plate Appearances were included in the analysis. Similar to Holmes (2011), this is to ensure that rate statistics are fairly descriptive of ability. Information on race/ethnicity was inferred from the players' hometown, name and bio picture on MLB's team websites. The dependent variable is the natural logarithm of the player's salary. This will take into account any heteroscedasticity which arises from the large range of salaries (Holmes 2011). Independent

[9] Baseball Prospectus. *Cot's Baseball Contracts.* Available at:
https://legacy.baseballprospectus.com/compensation/cots/
[10] Fangraphs, *Leaderboards.* Available at:
https://www.fangraphs.com/leaders.aspx?pos=all&stats=bat&lg=all&qual=y&type=8&season=2019&month=0&season1=2019&ind=0,

variables include traditional performance statistics used from prior literature like RBIs, OBP, and SLG. Modern statistics such as WAR are also introduced into the analysis.

A player's salary for a season is what he earned in the given season of performance. Players who enter free agency and are bid on for any number of years. In each year they are paid a salary which is about the average of the total contract but will sometimes slightly vary. If the player has over six years of service time in the league and signed an extension before he reached free agency, he is included in the sample. Average salary of the entire contract is not used because those players who do not reach free agency have their salaries increase once they reach over six years of service time. This is to reflect the larger salary a player makes in what would have been his post-free agency years. In addition, the salary at the beginning of each season is used so that incentives are not included. In the sample of pitchers with at least six years of service time, the minimum earnings level is $536,250 and the maximum is $38,300,00. In the sample of hitters with at least six years of service time, the minimum level of earnings is $1,115,000 and the maximum level is $34,100,000.

By recreating Holmes (2011) analysis, I can observe if trends have changed since his analysis of players from 1998-2006. Then by adding a new set of performance variables, I will see if the models make any difference in whether discrimination exists or not, and if the magnitude has changed. Lastly, a model using both set of variables is conducted. Teams may use a combination of both when considering offering contracts, so this is controlled for. The idea is that the current events and anecdotal evidence suggest that this research should be updated. The collective bargaining agreements between Major League Baseball Player's Union and its owners have changed since Holmes' analysis. With the growing representation of different demographic groups, it is necessary to reinvestigate this issue through Holmes (2011) methods.

Traditional Statistics

Using a different set of performance statistics, we can further grasp how much value a given player brings to his team. In Holmes (2011), he provides three different statistics which measure offensive output, one for baserunning and one for defense. He also includes how many Gold Gloves a player has won in his career.[11] Holmes (2011) placed the heaviest weight on

[11] A Gold Glove is an award given to a player at every position at the end of the year. It is supposed to be an award for the best defensive player, but it often fails to reward the best statistical performers.

offensive performance in his modeling. Slugging Percentage (SLG) wants to capture "how well a batter has hit for power." Simply put, SLG is the average totals bases per at bat. SLG is the only traditional statistic that gives different weights to different offensive events. A home run is four times more valuable than a single, a triple is three times more valuable, and a double is four times more valuable.[12]

Despite the importance of using OBP as an analytical measure and the rise in returns to OBP since the famous book *Moneyball* (Brown, Charles and Rubin 2015), OBP does not assign different weights to different events. OBP is calculated so that each time on base is worth the same number of runs, but this is not the case. Tango (2006) explains that each time on base is associated with a different run value depending on the specific situation in the game. Unlike SLG, it is calculated using (PA).[13] The evidence from Brown et al. (2015) suggests that OBP may have some analytical significance in terms of its relationship with free agent wages. We will be able to see if this still holds true when recreating Holmes' model. Holmes' final choice of offensive production was Runs Batted In. An RBI is the number of times a run scores as a result of a batter's PA, excluding situations where there is an error or the batter hits into a double play. While the consensus among the baseball analytics community is that RBIs are not the best measure of offensive production, it could still have explanatory power in estimating salaries for hitters. For a more in-depth explanation of RBI, OBP and SLG, refer to the Appendix.

Modern/Advanced Statistics

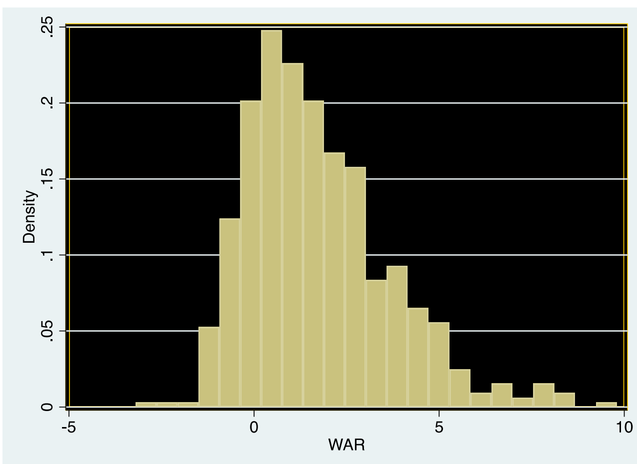
The way that I will add to the analysis is through the use of the comprehensive statistic Wins Above Replacement (WAR) as my main indicator of performance. WAR is an all-encompassing assessment of a baseball player's contributions to his team measured in terms of wins added to the hypothetical replacement level player (Baumer and Matthews 2014). In other words, how many wins is this player responsible for in comparison to a hypothetical replacement-level player that signed for the minimum level contract, easily acquired in a trade, or a minor leaguer with ability high enough that he is on the fringe of the major leagues

[12] Details on variables can be found in the Appendix.
[13] Plate Appearances include every time the batter has come to bat. While at-bats do not include a handful of offensive events. By not including every type of offensive event, positive or negative outcomes are not taken into account.

(Baumer and Matthews 2014). To define a player's total contributions, WAR will have several inputs that may be chosen within the analytical framework (Tango 2006).

Figure 2 displays the distribution of WAR across the entire league from 2017-2019. Just like any distribution of performance, there is a widespread of performers. The biggest benefit of WAR versus traditional statistics is that on their own, traditional statistics cannot show a full distribution of performers in an all-encompassing way. One traditional statistic can only describe one aspect of performance. WAR can tell us about every player's entire measurable contribution to the game. The expectation is that across different races, players with similar WAR totals should make about the same in terms of dollars. If this is not the case, then there could be evidence of discrimination.

Figure 2. WAR From 2017-2019



Source: FanGraphs Leaderboards

WAR can be calculated in different ways, depending on the inputs chosen. It provides a framework in which people can implement their own pathways. There are several different publicly available versions of WAR because each pathway calculates WAR using different inputs. Baumer and Matthews (2014) create their version called OpenWAR because of the lack of references implementation of WAR.[14] Although OpenWAR is not used in this analysis, it is highly correlated with the version, fWAR. fWAR is the FanGraphs implementation of the WAR framework. The choice of WAR is not pivotal in this argument since all of the implementations are highly correlated with one another.

[14] Baseball analytics does not find itself in academic journals because the community is heavily reliant on blog type websites. Since the readership is mainly fans and analysts, they do not have issues reaching their targeted audience.

Baumer and Matthews (2014) acknowledge that the most significant issue with WAR is that there are several options. Since WAR does not have a single definition as other traditional statistics do, it has made it tough to interpret at times. However, as Baumer and Matthews (2014) point out, there is no avoiding WAR. No other statistic is a cumulative assessment of runs created by an individual player on his own on both offense and defense. The conversion into a win total makes it easier to understand why one player is more valuable than another. There are certain immeasurable factors that come into play that WAR cannot grasp, but neither do other statistics. The the idea is that WAR is significantly better than any other option available right now or in the past because of its use of run values and leverage indexes (Tango 2006). WAR is a measure that considers a multidimensional set of variables to explain a player's performance. It is a single measure that is easy to use and interpret. It is a summary index to explain a player's entire performance on the field.

WAR has eight different inputs; batting runs, base running runs, fielding runs, positional adjustment, league adjustment, replacement runs and runs per game. WAR is a standardized index where a 0 WAR represents a replacement player. As was said before a replacement level player is a minimum wage level free agent. As the WAR scale increases, the quality of player increases. In any given season, a player worth about 2 wins is an average player. Anything above that signals an above average player, a great player, then an elite player. Players WAR totals are calculated for each season. However, WAR totals can also be added up over a player's entire career. This allows us to compare players within and between seasons and careers. It is a measurement that attempts to include all the measurable information in the game, assigning a player's value in terms of team wins.

To understand how WAR is an upgrade of Holmes' (2011) choice of statistics, the inputs of WAR must be defined. The offensive input which serves as an upgrade to Holmes' choice of OBP and SLG is wOBA. wOBA is used to obtain how many batting runs a player contributed to his team's offensive output on an objective level. wOBA is on the same scale as OBP. However, it assigns different run values to each offensive event. Although SLG does this as well, the run values for wOBA are calculated to coincide with actual total runs scored by a player's team. SLG's weights are given an arbitrary value.

wOBA combines the essential information from SLG, OBP and RBI for a few reasons. The main purpose of SLG is to estimate a hitter's ability to hit for power. wOBA does this with more accuracy because it assigns the real run values of each batted ball event. OBP measures the frequency in which a player reaches base. wOBA does this as well by assigning the calculated run values of every offensive event, weighing all events differently to coincide with the value of each situation. RBIs are meant to estimate a hitter's ability to drive in runs. wOBA does this with more accuracy because of its use of a leverage index to coincide with situations that are pivotal during a game. For an in-depth description of the rest of the inputs of offensive WAR, refer to the appendix. The other main component of WAR is baserunning.

There is not much of a variance in baserunning in baseball today as there was in the past. Slowly, it has become less of a factor in the game of baseball. Because of this, it may have a larger explanatory power in Holmes' analysis than it does in this one. However, BsR still provides an improvement in comparison to Holmes' (2011) use of the traditional statistic, Speed Rating. While Spd rating is mainly focused on a runner's contributions to stolen bases, BsR is comprehensive in the sense that it covers all types of baserunning attempts. This includes stolen bases, advancing on batted balls, and the ability to avoid hitting into double plays. It does so by weighting each of these three components so that we can know how much above or below a runner is in comparison to the league average. For in-depth calculations and description of BsR, refer to the Appendix. Finally, the last main component of WAR is defensive runs.

In the baseball community, defensive rating statistics are considered volatile depending on the decade. The growth of tracking in baseball has improved the accuracy in which all statistics are calculated. In the past, defensive statistics were difficult to calculate because there was no way to track how difficult it was to reach a particular batted ball. Despite this, the statistic Zone Rating was developed. This was Holmes' (2011) choice of defensive output. Zone Rating assesses a player's fielding ability by measuring the percentage of batted balls that are fielded cleanly within a fielder's defined zone. If a player reaches a ball out of his zone, it is recorded as a chance and a play made, but if the player does not make the play, he is penalized. Players with more range[15] are unfairly penalized. In reality, a player that can get to balls out of his zone is more valuable than one that cannot. Even if the play is not made, the fact that the player stopped the ball could have prevented another runner from advancing or scoring. This is why I will use the stat, Ultimate Zone Rating (UZR). UZR has the same purpose as Zone Rating. However, it is an upgrade because it gives greater value to fielders that make plays

[15] Ability to reach batted balls at a greater distance than the average fielder.

outside of their own zone. This player has more opportunity to save runs for his team. In addition to range, it accounts for errors made by a fielder, outfielder arm strength[16], and double play ability. In the sample of seasons from 2017-2019, UZR uses tracking data, implying it has a high level of accuracy. UZR has a positional adjustment to account for some positions being more difficult than others. This yields the input into WAR, Defensive Runs (DEF).

Now, we must see how all this adds up to WAR, but that is simple because we have almost all the inputs.

$$WAR = \frac{BsR + DEF + Batting\ Runs + Positional\ Adj + Lg\ Adj + Replacement\ Runs}{Runs\ Per\ Win}$$

Where we adjust for position, league and add in the number of runs a replacement level (0 WAR) player delivers and divide by the runs per win throughout the league. Also, a quick note on the catcher position because their defense is measured differently. Catchers define their defense through framing and throwing out runners attempting to steal. Framing is calculated through pitch tracking data. The data reveals if a catcher was able to properly manipulate the ball that is on the inside or outside edge of the strike zone. If the catcher manipulates the pitch correctly, it appears to the umpire as a strike call. Catchers could be good or bad at framing. It has a significant impact on the outcome of the game, which is why their defensive prowess must be calculated differently. Depending on how many strikes above or below average a catcher is, that number is converted to a run total. This run total tells us a catcher's DEF score so we can compare them to other fielders.

Now that we have WAR for position players, we can move onto WAR for pitchers. Again, remember that Holmes did not analyze pitchers, so this is a completely new addition to the investigation.

Pitcher Performance Statistics

For the purpose of this analysis, there are two types of pitchers, the starting pitcher and the relief pitcher. They will be analyzed separately because the differences in how much they can influence a game. Pitcher WAR focuses on how that pitcher controlled the game and ignores the quality of fielders because fielders can largely help or hurt pitchers' outcomes. It does so by using its key input, Fielding Independent Pitching (FIP). FIP, like wOBA and the WAR

[16] The faster velocity in which an outfielder makes a throw, the better the chance the fielder can make an out. If a fielder gains more outs with the strength of his arm than the average fielder, then his arm strength positively contributes to his UZR.

framework, was created by the statistician Tango (2006) so that we could evaluate pitcher's without reliance on how their fielders influence their outcomes. In addition to WAR, other advanced statistics like Expected Fielding Independent Pitching (xFIP) and Left on-Base Percentage (LOB%). WAR for pitchers is calculated as such:

$$Pitcher\ WAR = \left(\frac{Lg\ FIP - FIP}{Pitcher\ Runs\ Per\ Win} + Replacement\ Level\right) \times \frac{IP}{9} + Lg\ Correction$$

For a reliever we multiply this equation by the leverage index. For further description of the inputs, refer to the Appendix.

Models using traditional statistics to evaluate pitcher's performance are also included. Wins (W), Losses (L), Saves (SV), Earned Run Average (ERA), Home Runs per nine innings (HR/9), Strikeouts per nine innings (K/9) and innings pitched. All the statistics evaluate performance for starting pitchers and relief pitchers. W, L and SV are counting statistics and only signal the outcome of a game. A pitcher can win or lose a game for his team, and a reliever can obtain a save by closing the final innings of a win. ERA is the rate at which a pitcher gives up earned runs per nine innings, and HR/9 is the rate at which a pitcher gives up home runs. For rate stats, the goal is to have the lowest number possible. Like Holmes (2011), pitchers who qualified for the statistical title ERA are the only ones included in the analysis.

There is no prior literature indicating which source of statistics has more explanatory power in estimating pitcher's salaries because the literature has mostly focused on hitters. Due to this, I will calculate them separately and together, so that going forward, there is a basis in the literature for how to carry out this type of analysis.

The race/ethnicity for each player was inferred manually. In order to do so, I used name, place of birth and roster photo to indicate whether a player was white, black, Latino, Asian or other. This data is all publicly available. Asian and other races are excluded from the results because of a tiny sample. I do not indicate between black Latino or white Latino players. In addition, it should be specified that I chose Latino instead of Hispanic. In much US labor market research, they are used interchangeably, but there is a key difference. Latino refers to somebody with origins from Latin America and Hispanic refers to somebody who is from a Spanish-speaking country. They do often intersect, but for the purpose of this analyses we will stick to Latino. After inferring each players race, I created two different variables. The first was a dummy variable for each race. Other qualitative variables included in the analysis are service time in the major leagues and the season in which the statistics coincide with.

WAR is not perfect, no statistic is. However, it undeniably the best statistic that is available for comparing players in a wholesome way. If we are looking to analyze whether discrimination exists in free agency through a "equal pay for equal pay" framework, then WAR must be used so we can capture every aspect of performance. Combining quantile regression and WAR, I will attempt to find if any discrimination exists in Major League Baseball.

CHAPTER V: DATA AND RESULTS

The goal of this chapter will be to explore if there is any statistical evidence of discrimination among Latino players and their white counterparts in MLB. However, in addition to Latino players, there were enough people in the sample to carry out an investigation which I also extended the analysis to include black players. The inclusion of these players in the sample comes with a caveat because they are extremely underrepresented in MLB compared to the U.S. population. Before discussing the results of the analysis, I will describe the data.

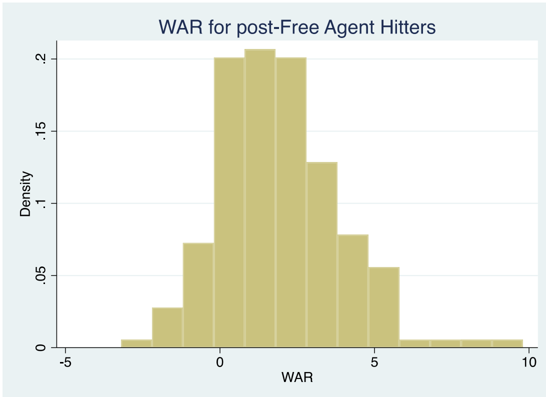
For this analysis, I investigate the three seasons of performance which have followed the most recent collective bargaining agreement that was agreed upon in the Winter of 2016 and will last until December 1st, 2021.[17] My goal is to extend this analysis once the 2020 and 2021 seasons are complete so that we can see the results across the entirety of the collective bargaining agreement. The chapter starts with an analysis of hitters with more than six years of service time in MLB. After that, I conduct the same analysis for pitchers in MLB, breaking it up between starting pitchers and relief pitchers.

Section I. Analysis of Post-Free Agent Hitters

Figure 3 shows the distribution of hitter WAR in the sample of post-free agent hitters. In terms of WAR, there are many types of players who have more than six years of service. In other words, of those players with over six years of service time, there is a wide distribution of overall performance.

Figure 3. Wins Above Replacement for Post-Free Agency Hitters

[17] Major League Baseball Players Association. 2016. 2017-2021 Basic Agreement. Major League Baseball. https://www.mlbplayers.com/cba



Source: FanGraphs Leaderboards

Before conducting a regression analysis, I will summarize what the salary information for hitters according to their race (Figure 4). In addition, I include the average earnings in specific ranges of WAR across the distribution of hitters (Figure 5). The expectation is that discrimination does not exist between different races/ethnicities. This is because players with over six years of service can enter the free agency market. They should be able to negotiate salaries, so they are paid according to their performance. They can also sign extensions before these six years. However, those players will only be included if they have reached over six years of service time. No discrimination across the means would go in-line with prior research.

Figure 4. Salaries (in millions) and WARS of Post-Free Agent Hitters

Race	Mean Salary	Median Salary	Q10	Q90	Mean WAR	Median WAR	n
White	$13.8	$13.0	$5.0	$22.2	2.05	1.9	95
Latino	$13.4	$12.0	$3.5	$26.0	1.72	1.6	41
Black	$12.8	$14.1	$3.1	$21.7	1.99	1.65	30
All	$13.4	$13.2	$4.0	$22.2	1.98	1.8	166

Source: Author's Calculations. Salaries correspond to base player at the beginning of the season without inflation adjustment. Salaries in Millions of Dollars per year.

Figure 5. Salaries (in millions) Across Distribution of WAR for Post-Free Agent Hitters

Race	< 2.0 WAR	2.0-4.0 WAR	> 4.0 WAR	n
White	$12.8	$13.4	$16.9	95
Latino	$13.5	$12.9	$14.4	41
Black	$10.6	$15.0	$18.4	30
All	$12.5	$13.5	$16.8	166

Source: Author's Calculations. Salaries correspond to base player at the beginning of the season without inflation adjustment. Salaries in Millions of Dollars per year on average.

In the context of this sample, all races of players earn around the same salaries on average. According the median salaries, Latino players earn around $1 million less than white player, while black players earn around $1 million more than white players. In the 10^{th} percentile of earners, both Latino and black players earn over $1 million fewer dollars than white players. However, in the 90^{th} percentile of earners this relationship flips for Latino players. They earn over $3 million more than white players, while black players earn about $.5 million fewer than white players. According to both the mean and median WAR, Latino players are slightly worse performers than white players.

Figure 5 separates players based on ranges of WAR. The idea is that quality of player increases as we go from less than 2.0 WAR to greater 4.0 WAR. For players with less than 2.0 WAR, Latinos earned the most on average from 2017-2019. In the range if 2.0-4.0 WAR and greater than 4.0 WAR, Latinos earn the least of all races. Black players earn the most in both ranges. This tells us that for above average performers (2.0-4.0 WAR) and good to great performers (> 4.0 WAR) Latino players earn the least according to WAR. In the regression analysis, several models are estimated. One of them includes WAR as the only performance statistics and the others include a range of performance statistics.

Regression Results for Post-Free Agent Hitters

To start the analysis, first I use Holmes' mode specification using Ordinary Least Squares to assess the presence of discrimination. Then, I use my preferred measures of performance, to revisit the question using what I consider a superior set of performance measurements. For this paper, the goal is not to prove that one group of statistics is better at evaluating salary than the other, but to test the sensitivity of the conclusions using different measures of performance. Even though advanced stats like WAR have proven to be better evaluators of overall performance, this does not necessarily mean that the others do not play a role at all in explaining salaries. The inclusion of advanced statistics is to reflect that teams may have begun to base their decision making more on these statistics than in the past. To reflect this, models have been estimated which include each set of statistics on their own, then together. I assume that the effect of any performance statistic on salary is linear and independent of race for all models across both hitters and pitchers.

Figure 6. Regression Results Baseline White Hitter

Variable	Holmes Stats Model 1	Advanced Stats Model 2	WAR Model 3	RBI+WAR+wOBA Model 4	WAR*Race Model 5
Intercept	14.646 (.603)	14.083 (.804)	15.378 (.218)	16.063 (.863)	15.007 (.671)

Variable	Holmes Stats Model 1	Advanced Stats Model 2	WAR Model 3	RBI+WAR+wOBA Model 4	WAR*Race Model 5
Black	-.190 (.155)	-.185 (.156)	-.163 (.142)	-.121 (.142)	-.491** (.698)
Latino	-.258* (.134)	-.172 (.139)	-.181 (.129)	-.243* (.130)	.046* (.689)
Black*WAR					.184** (.081)
Latino*WAR					-.072 (.075)
Service Time	.078*** (.023)	.083*** (.025)	.087*** (.023)	.082*** (.023)	
WAR			.0997*** (.028)	.106** (.051)	.061* (.032)
wOBA		4.52** (1.95)		-3.372 (2.92)	
OBP	1.112 (2.16)				
SLG	-.178 (1.50)				
RBI	.009*** (.004)			.007** (.003)	
Position	-.002 (.016)	.005 (.017)			
Stolen Bases	.018** (.008)				
Base Runs		.015 (.018)			
Defense Runs	.004 (.007)	.003 (.007)			
wRC+		(.000) (.001)			
R_2	.1821	.1192	.1391	.1700	.1040
Adj R_2	.1279	.067	.112	.1323	.0704
n	162	162	162	162	162

Note: * = Significant at 10% level | ** = Significant at 5% level | ******* = Significant at 1% level
Source: Author's Calculations. Coefficients are log point differences which is approximately equal to percentage change.

Figure 6 presents the OLS models for hitters in Major League Baseball with at least 350 plate appearances in any season from 2017 to 2019. This number is chosen so that all players in the sample have played the majority of his teams' games. In Holmes' (2011) investigation, he dealt with a slightly larger sample size because his population ranged across more years (1998-2006). However, my analysis includes the seasons from the most current CBA. Holmes (2011) found significant evidence in his OLS model of white and Hispanic/Latino make between 14%-16% more than their black counterparts with equal skill.

When using Holmes' (2011) performance statistics (Model 1), the two variables which were statistically significant in estimating players' salaries were RBIs (p-value = .004) and SBs (p-value = .028). The significance of these variables could mean that that teams still value them when offering the contracts. In the context of this sample, we would expect both variables to have significant, but small, positive effects on a player's salary. The key finding of this model is that it estimates significant evidence of possible discrimination against the average Latino hitter with over six years of service in MLB.

The model estimates with statistical significance (p-value = .056) that the average Latino player makes about 25% fewer dollars than their white counterpart with equal ability. Note that the coefficients are log points, which are about the same as percentage points. In this case, the null hypothesis can be rejected that discrimination does not exist between Latino and white players. The next OLS model uses the inputs from which WAR is calculated. The results do not find enough evidence to reject the null hypothesis. However, the model does esimate that wOBA is a significant predictor of salary (p-value = .022). Model 3, with only WAR as the performance statistic, reports about an 18% penalty for the average Latino player in the MLB, but with no statistical significance (p-value = .163). It also estimates that if a player has a WAR 1 point higher, his salary would increase by about 10%.

The combination of significance and magnitude of the coefficient make WAR an ideal starting point for estimating salaries for hitters. Although model 1 reports the highest Adjusted R_2, model 3 reports one slightly below. This indicates that on its own, WAR comes close to explaining salaries as much as a combination of traditional statistics. Model 5 provides an alternative approach to analyzing the effect of performance on salary. By interacting the race variable with WAR, we can see the average effect that an additional WAR would have on each group of hitters according to their race. In other words, this tells us what the returns to WAR is for each of the races. In the model, I do not include any other independent variables so that we can observe the relationship between WAR and salaries on its own. Although it is not statistically significant, the model reports that an additional WAR for a Latino hitter is worth about 7% less than that for a white hitter on average. However, for black hitters an additional WAR is worth about 18% more than one for a white player (p-value = .025).

Models 1-3 suggest that traditional statistics and WAR likely play a role in predicting a player's salary. For that reason, Model 4 estimates salaries using WAR, wOBA and RBI as performance statistics. In the model, Latinos experience a statistically significant penalty of about 24% relative to white hitters (p-value = .064). We would expect an increase in a player's WAR by 1 to increase salary by about 10% (p-value = .040). RBI also has a significant positive effect of just under 1% for each additional RBI (p-value = .020). Across these OLS models, we would expect a penalty of between 17%-26% for Latino hitters on average. While only two of the models report statistically significant penalties, consistency in the signs of the coefficients provides robustness to the results. This opposes prior results found by Holmes (2011) and Palmer and King's (2006). However, these studies did find evidence of discrimination against black and Latino hitters in the lower quantiles of earners. To find if this still persists, I estimate a quantile regression.

Using the same specifications as in Figure 6, I estimate quantile regressions for the 10th, 25th, 50th, 75th and 90th quantiles, to analyze if the wage gap that Latino hitters experience differs across the distribution. I will carry out the same analysis for types of pitchers in the next section. Figure 7 provides the coefficients, standard errors and significance of the Latino hitters. No models show the relationship that was found in Holmes (2011) investigation; there is not enough evidence to conclude discrimination against the 10th and 25th percentile of earners among Latino hitters. While all the models estimate differences in the 25th percentile or lower, the evidence is not statistically significant. Controlling for performance and other characteristics, there appears to be a glass ceiling for Latino players relative to similar white players.

Holmes (2011) suggests that teams are unable to maximize winning if they do not pay top-dollar for the highest earning players. However, the evidence from the quantile regression results (Figure 7) points to differentials among the highest earning Latino and white players. In addition, the quantile results from Model 5 estimate significant differences in the value of one WAR for Latino players above the 75th quantile of earners. This result shows that in the highest quantiles of earners, Latino hitters do not get the same returns for their WAR totals. Model 5 differs from Models 1-4 because it estimates a return to WAR framework, while the others control for performance.

Figure 7. Quantiles for Each Model: Baseline White Hitter

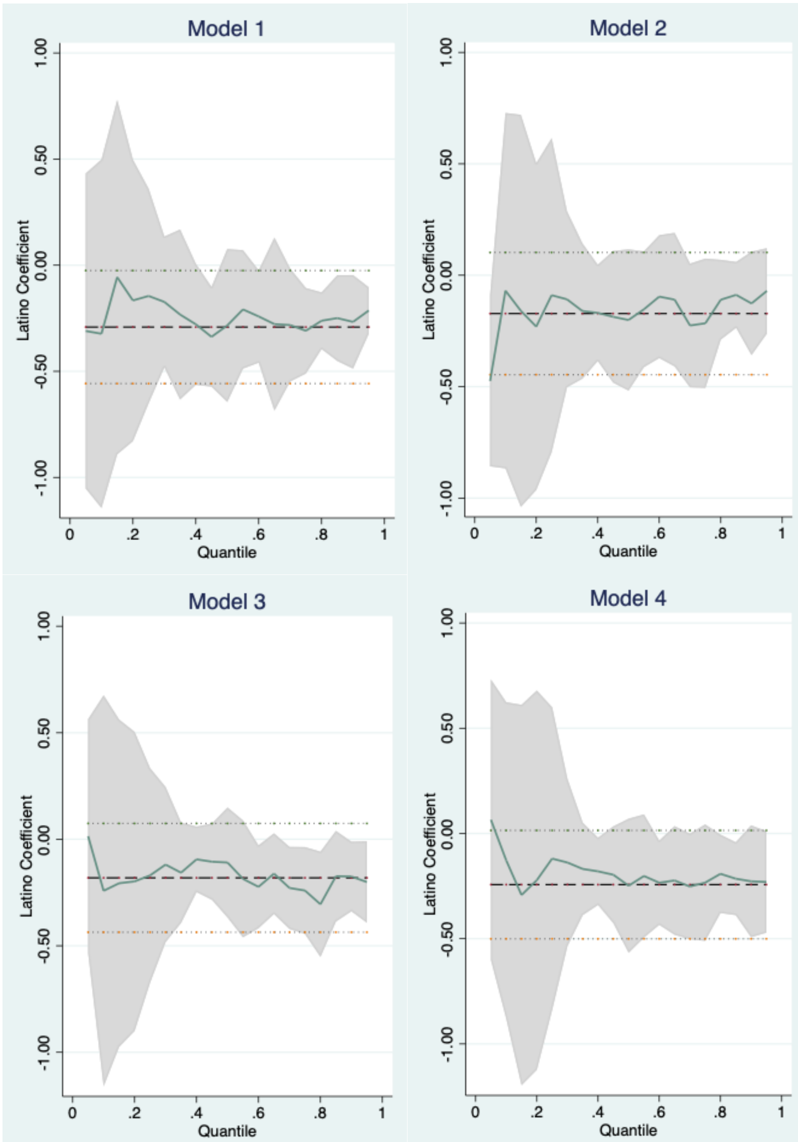
Quantile	Latino Holmes	Latino Advanced	Latino WAR	Latino RBI+WAR+wOBA	Race*WAR
10%	-.399 (.380)	-.070 (.433)	-.242 (.394)	.128 (.287)	.163 (.198)
25%	-.170 (.291)	-.089 (.270)	-.170 (.302)	-.119 (.328)	-.054 (.145)
50%	-.342** (.136)	-.201 (.130)	-.109 (.124)	-.248* (.127)	-.092 (.097)
75%	-.190* (.113)	-.215* (.116)	-.241** (.112)	-.234** (.109)	-.145** (.059)
90%	-.289** (.115)	-.126 (.075)	-.174* (.098)	-.227** (.093)	-.133** (.055)

Note: * = Significant at 10% level | ** = Significant at 5% level | *** = Significant at 1% level
Source: Author's Calculations. Coefficients are log point differences which is approximately equal to percentage change.

When controlling for several performing statistics, evidence suggests that there are statistically significant wage differentials between Latino and white player in the 50th quantile and above. Models 1-4 all estimate significant differences in salaries at some point in the distribution. Model 4 combines the most significant predictors of salaries from each of the first three models. It estimates penalties of at least 20% for all Latino players above the 50th quantile of earners. Statistical significance is found every model in the 75th quantile. In the 90th quantile, coefficients remain negative across all models, but I find statistical significance models 1, 3 and 4. In the 75th quantile, Latinos experience between a 19%-24% penalty relative to white hitters, conditional on similar performance and characteristics. In the 90th quantile, we would expect a penalty for Latinos between 17%-29%. Overall, the evidence shows a change in the dynamic observed in prior research. After adjusting for performance, the top Latino earners experience a significant penalty in comparison to the top white earners. Figure 8 displays the coefficients across all quantiles for Models 1-4. While the estimates show slight variation across the distribution, the effects seem fairly constant. The consistency across models adds robustness to the result of wage differentials above the 50th quantile of earners.

Model 5 estimates significant differences in the value of an additional WAR. Relative to white hitters, Latino hitters in the 75th quantile and above experience a penalty between 13%-14% in their returns to additional WAR. In other words, Latino players in top of the distribution of earners do not receive the same returns for their overall performance. This goes in line with the results from Models 1-4.

Figure 8. Coefficients Across All Quantiles



Source: Author's Calculations

Section II. Analysis of Post-Free Agent Pitchers

When summarizing what the data looks like for pitchers, I will break up the analysis between starting pitchers (SP) and relief pitchers (RP). This is because starters tend to have higher WARs and salaries on average due to the volume of innings they contribute to games. Figure 9 shows the distributions of both starters and relievers in terms of WAR. While starters maintain the right-skewed distribution that we see from hitters, the reliever distribution is more concentrated towards the center of the distribution. Similar to hitters, any pitcher who enters free agency is included in the sub-sample. If the pitcher signed an extension and has reached over six years of service, he is included. The same expectation will be maintained; players earn what they perform to and wages are not dependent on race.

Figure 9. Wins Above Replacement for Post-Free Agent Pitchers

Source: FanGraphs

Figure 10 shows information on salaries and average WARs. Without controlling for performance, the median salary is the highest for white players and lowest for Latino players among starting pitchers. The same trend is seen for Latinos in the median WAR. The median white pitcher's salary is in-line with the median of the overall group, while the Latino pitcher is below the median and the same relationship persists in WAR. For relief pitchers, Latinos have the highest median WAR. However, black relievers have the highest salaries for relievers in terms of both median and mean. In the 10th percentile of earners, salaries are fairly consistent across all races and pitchers. In the 90th quantile, Latino starters average salaries significantly lower than white starters. For relievers, Latinos and whites earn similar salaries.

Figure 10. Salaries (in millions) and WARS of Post-Free Agent Pitchers

Race	Median Salary	Mean Salary	Q10	Q90	Median WAR	Mean WAR	n
White SP	$6.5	$10.0	$.55	$27.5	2.9	3.52	111
White RP	$1.5	$2.8	$.55	$7.4	0.6	.72	252
Latino SP	$5.6	$5.5	$.55	$12.0	2.65	2.9	36
Latino RP	$1.5	$2.9	$.54	$7.3	0.65	.81	88
Black SP	$6.2	$11.4	$3.4	$30.0	3.65	3.55	4
Black RP	$1.9	$4.3	$.54	$11.3	0.6	.89	14

All SP	$6.5	$9.3	$.55	$23.5	2.9	3.37	151
All RP	$1.5	$2.9	$.55	$7.5	0.6	.74	354

Source: Author's Calculations. Salaries correspond to base player at the beginning of the season without inflation adjustment. Salaries in Millions of Dollars per year.

Figure 11 shows the average salaries in certain ranges of WAR. The table splits up ranges for starters and relievers. For starters, the ranges of WAR are the same as for hitters. Relievers have decreased ranges because they do not generally achieve the totals of WAR that starters and hitters do. Among starting pitchers, white players make the highest salaries in both the lowest and highest WAR ranges. For Latino starters, wages decrease when increasing the range, which is unexpected. This could be the result of low-paid Latino pitchers performing incredibly above expectations in the context of this sample. For relievers, Latinos earn similar salaries to white players across all ranges. However, they have higher salaries in the middle range of WAR. In general, the highest average salary is in the middle range of WAR. This is a result of elite seasons, in terms of WAR, from players not in the top-tier of earners among relievers. The volatility of this position

Figure 11. Salaries (in millions) Across Distribution of WAR for Post-Free Agent Pitchers

Race	SP < 2.0 WAR	SP 2.0-4.0 WAR	SP > 4.0 WAR	RP < 1.0 WAR	RP 1.0-2.0 WAR	RP > 2.0 WAR	n
White SP	$11.3	$8.0	$12.5				95
White RP				$2.4	$3.6	$2.8	252
Latino SP	$8.0	$5.5	$3.0				41
Latino RP				$2.1	$4.0	$2.8	88
Black SP	Na	$15.0	$4.9				30
Black RP				$2.7	$5.5	$3.8	14
All SP	$10.2	$13.5	$10.8				166
All RP				$2.4	$3.7	$2.8	354

Source: Author's Calculations. Salaries correspond to base player at the beginning of the season without inflation adjustment. Salaries in Millions of Dollars per year on average.

Regression Results for Post-Free Agent Starting Pitchers

The expansion of pitchers into the analysis of wage differentials in Major League Baseball is an addition from Holmes (2011) and Palmer and King (2006). Before the analytics revolution, teams were simply bad at evaluating pitching. In other words, contracts that did not reflect the actual value of the player were frequent. With the growth of information came improvements in evaluating the assets. Even so, the summary from Figure 11 suggests that salary was slightly random across ranges of WAR. The regression analysis will make the relationship between Latino and white pitchers clearer.

Before highlighting the results of the OLS models (Figure 12), it should be noted that the sample size in the analysis of free agent starting pitchers is small (n = 155). This is because we are only dealing with three seasons of performance and only analyzing those pitchers who qualified for statistical titles in the given season. In order to qualify for statistical titles, pitchers need to pitch a large number of innings. For example, the minimum number of innings in the sample of starting pitchers is 161 and the maximum is 223. A player may have pitched 100 innings and not be included in the sample. Nevertheless, the results are consistent across specifications, providing added robustness to the analysis.

Figure 12: Ordinary Least Squares: Baseline White Starter

Variable	Traditional Stats Model 6	Advanced Stats Model 7	WAR+FIP Model 8	WAR Model 9	Race*WAR Model 10
Intercept	11.832 (1.510)	9.884 (1.472)	10.153 (1.881)	13.633 (.276)	14.871 (.286)
Black	.082 (.534)	.024 (.546)	.154 (.554)	.100 (.174)	3.817 (4.53)
Latino	-.338 (.205)	-.371* (.208)	-.391* (.212)	-.343* (.178)	1.112* (.627)
Black*WAR					-.932 (.887)
Latino*WAR					-.501*** (.190)
WAR			.336** (.150)	.075 (.054)	.139* (.072)
Wins	.019 (.036)				
Losses	-.004 (.047)				
ERA[18]	-.423** (.196)				
HR9[19]	1.39*** (.399)				
K9[20]	-.051 (.057)				

[18] ERA = Earned Run Average
[19] HR/9 = Home Runs per 9 Innings
[20] K/9 = Strikeout per 9 Innings

Variable	Traditional Stats Model 6	Advanced Stats Model 7	WAR+FIP Model 8	WAR Model 9	Race*WAR Model 10
Innings Pitched	.013* (.007)	.017*** (.006)			
Service Time	.015*** (.002)	.016*** (.002)	.016*** (.012)	.016*** (.002)	
FIP[21]		.280 (.264)	.672* (.360)		
LOB%[22]		.003 (.208)			
xFIP[23]		-.131 (.266)			
R^2	.4873	.4516	.428	.4121	.1017
Adj R^2	.4517	.4216	.4024	.3924	.0589
n	155	155	155	155	155

Note: * = Significant at 10% level | ** = Significant at 5% level | ******* = Significant at 1% level

Source: Author's Calculations. Coefficients are log point differences which is approximately equal to percentage change.

Figure 8 displays the results of the OLS models for starting pitchers. Model 6 uses traditional statistics to estimate starting pitchers' salaries in Major League Baseball. As we would have expected, ERA, HR/9 and IP are all significant predictors of salary. The signs are in line with our expectations as well. If a pitcher's ERA increases, then we would expect their salary to decrease. In terms of race, the null hypothesis cannot be rejected because there is no statistical significance in wage differentials between the average Latino and white starting pitcher. However, it is quite close. The model estimates that the average Latino starting pitcher experiences a 33.8% penalty relative to white starting pitchers, with a p-value of .102, just slightly above the threshold for the 10% level.

Model 7 uses advanced statistics plus the main input for WAR, FIP. Now, the null hypothesis of no discrimination can be rejected at the 10% level (p-value = .077). It is estimated that the average Latino starting pitcher experiences about a 37% penalty relative to white starting pitchers. Across all models, we see a 34%-39% penalty against Latino starting pitchers whether we use traditional or advanced statistics. Model 8 estimates salaries with only WAR and FIP as the performance statistics. It estimates a statistically significant penalty of about 39% for Latino starters on average. Both WAR and FIP are statistically significant predictors of

[21] FIP = Fielding Independent Pitching
[22] LOB% = Runners Left on Base Percentage
[23] xFIP = Expected Fielding Independent Pitching

salaries. However, the sign of FIP is counterintuitive. We would expect the coefficient to be negative because better pitchers have lower FIPs. The relatively large standard error shows that this result comes with a caveat.

In model 9, WAR is the only performance statistic. We can reject the null hypothesis at the 10% level that wage differentials do not exist between white and Latino starting pitchers (p-value = .100). Like the other models, there is a large penalty for Latino starters of about 34%. The quantile models will show if the differentials come about in similar areas as they did for hitters. Model 10 estimates an interaction between race and WAR to calculate if different races have different returns to WAR. On average, the model suggests a statistically significant gap in returns to WAR for Latino starting pitchers in comparison to white starting pitchers. We expect Latinos to make about a half as much for an additional WAR relative to whites. Next, I present the results for the Quantile regression for Models 6-10.

Figure 13: Quantiles for Each Model: Baseline White Starters

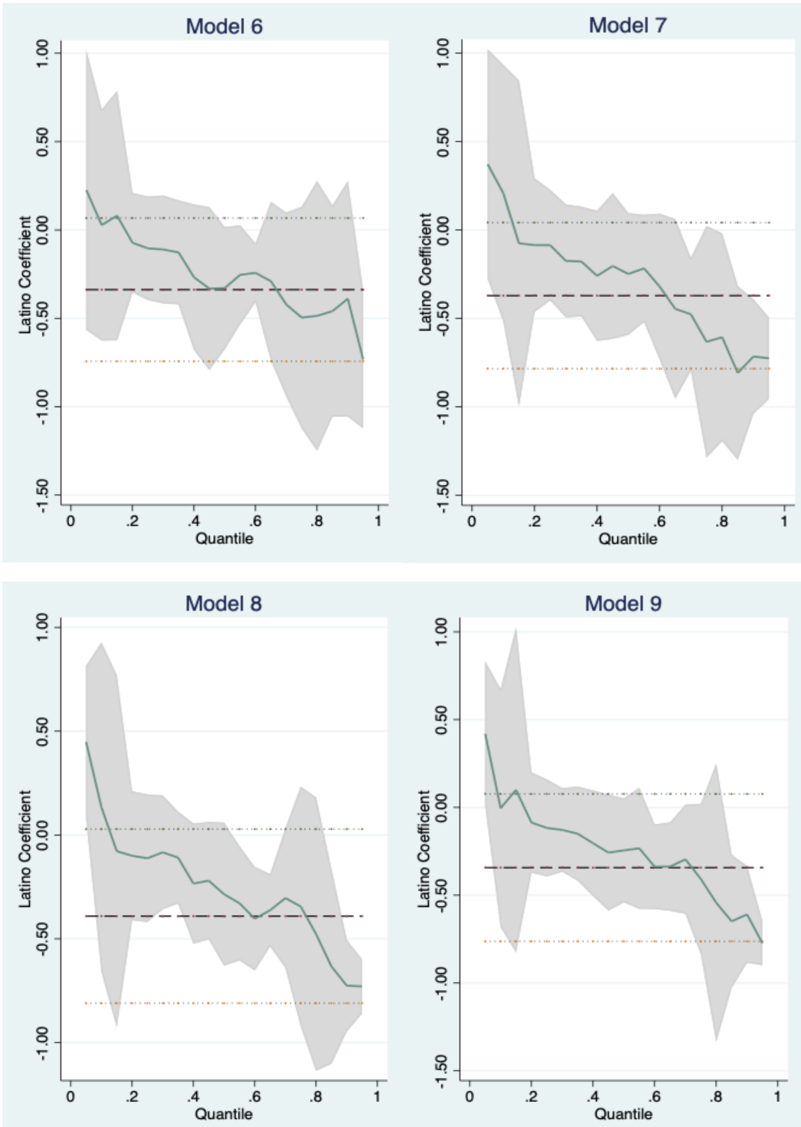
Quantile	Latino Traditional	Latino Advanced	Latino WAR+FIP	Latino WAR	Latino*WAR
10%	.027	.208	.134	-.004	-.077
	(.375)	(.408)	(.382)	(.454)	(.011)
25%	-.104	-.086	-.112	-.117	-.785*
	(.220)	(.203)	(.186)	(.198)	(.424)
50%	-.329*	-.249	-.285*	-.244	-.806***
	(.193)	(.179)	(.151)	(.155)	(.254)
75%	-.496	-.632	-.354	-.407	-.129
	(.423)	(.448)	(.509)	(.506)	(.215)
90%	-.389	-.715***	-.726***	-.609***	-.232**
	(.268)	(.235)	(.264)	(.226)	(.102)

Note: * = Significant at 10% level | ** = Significant at 5% level | ******* = Significant at 1% level
Source: Author's Calculations. Coefficients are log point differences which is approximately equal to percentage change.

Figure 13 shows a similar result that was seen in the analysis of hitters. From the median quantile to the 90th, we would expect a wage penalty for Latino starting pitchers in comparison to their equal performing white counterparts. In the median quantile, there is evidence in two of the four models which allow us the reject the null hypothesis of no wage differentials amongst white and Latino players. In three of the four models, there is evidence which allows us to reject the null hypothesis at the 1% level in the 90th quantile. This means that some of the top earning white starting pitchers have great premiums relative to Latino starting pitchers. The estimated penalty is somewhere between 60% to 70% for Latinos. In the median quantile, we would expect a penalty for Latinos between 29%-33%. In the three seasons of the sample top-paid Latinos earned very low wages relative to top-paid white starting pitchers. The results from Model 10 are in line with what we see in Models 6-9. There is significant evidence that suggests Latino players experience a 23% penalty in returns to WAR relative to white starting pitchers. That penalty significantly increases in the 50th quantile to about 81%, making it clear the average earning Latino pitcher also experiences smaller returns to WAR than the white pitcher.

Figure 14 shows the coefficients across the entire distribution for Models 6-9. Each model begins with positive coefficients in the bottom quantiles, then continues with a downward trend as the quantiles increase. What starts as a premium for the bottom earners eventually turns into a penalty from about the 20th quantile onwards. In each of the four models, the peak in penalties comes between the 80th and 100th quantiles. For hitters (refer to Figure 8), the line fluctuated around 24% for the majority of the distribution. For pitchers, the line steadily decreases, meaning that differentials increase as we increase the quantiles.

Figure 14. Coefficients Across All Quantiles For Starting Pitcher Models



Source: Author's Calculations

Regression Results for Post-Free Agent Relief Pitchers

In the baseball industry, the relief pitcher position is considered to be volatile. Relievers are more likely to get injured, not perform to expectations, or perform inconsistently. While this does not mean volatility in terms of wages, it means that low wage earners could show elite performance and high wage earners could show average or below average performance in any given year. Figure 11 shows that the highest earners are in the middle WAR range. This position has the smallest contracts in terms of years and dollars. There are a handful of players that earn significantly more than the rest of relievers and there is not guarantee they are always the best performer. This is why it is rare that teams invest heavily in this position.

Figure 15: Ordinary Least Squares: Baseline White Reliever

Variable	Traditional Stats Model 11	Advanced Stats Model 12	WAR+FIP Model 13	WAR Model 14	Race*WAR Model 15
Intercept	13.256 (.419)	12.435 (.370)	12.998 (.547)	12.919 (.085)	14.19637 (.088)
Black	.133 (.186)	.232 (.193)	.216 (.196)	.213 (.195)	-.009 (.393)
Latino	.136 (.086)	.240*** (.089)	.184** (.089)	.182** (.089)	-.130 (.179)
Black*WAR					.200 (.301)
Latino*WAR					.097 (.160)
WAR			.098 (.131)	.116** (.048)	.157* (.084)
Saves	.019*** (.004)				
ERA	.038 (.045)				
HR9	-.085 (.102)				
K9	.027 (.019)				
IP	-.010** (.004)	-.011*** (.004)			
Service Time	.014*** (.001)	.014*** (.001)	.015*** (.001)	.015*** (.001)	
FIP		-.016 (.066)			
LOB%		-.0003 (.001)			
xFIP		-.166** (.074)			
R_2	.5888	.551	.5372	.5372	.0297
Adj R_2	.5769	.5393	.5280	.5293	.0074
N	357	357	357	357	357

Note: * = Significant at 10% level | ** = Significant at 5% level | ******* = Significant at 1% level
Source: Author's Calculations. Coefficients are log point differences which is approximately equal to percentage change.

The results for the OLS analysis of relievers are presented in Figure 15. Model 11 finds no evidence of wage differentials between races. However, it concludes that Saves (SV) is a significant predictor of salary. We expect an additional *Save* to increase salary by about 2%. In the baseball world, this result is expected. Over time the team's best reliever almost always earns the role of closer, the position of relivers in which the *Save* can be earned.

For Model 12, we can reject the null hypothesis at the 1% level (p-value = .008) of no evidence of wage differentials. However, the result is in the opposite direction of prior models. In this model, the average Latino is expected to experience a 24% premium relative to his white counterpart with equal skill. This result is surprising because prior models including hitters and starting pitchers estimated significant penalties for Latinos. Model 13 confirms the premium for Latino relievers. On average, we would expect a statistically significant premium for Latino relievers of about 18% (p-value = .041). In the WAR-only model (Model 14), WAR is a statistically significant predictor of salary (p-value = .016). We would expect an increase in WAR to increase salary by about 12% (Figure 15). We can reject the null hypothesis of no evidence of wage differentials at the 5% level (p-value =.041). In this case, the average Latino reliever is expected to experience a premium of around 18% relative their white counterpart with similar skill. Model 15 shows the results for interacting WAR and race, but there are no significant differences.

Figure 16: Quantiles for Each Model: Baseline White Reliever

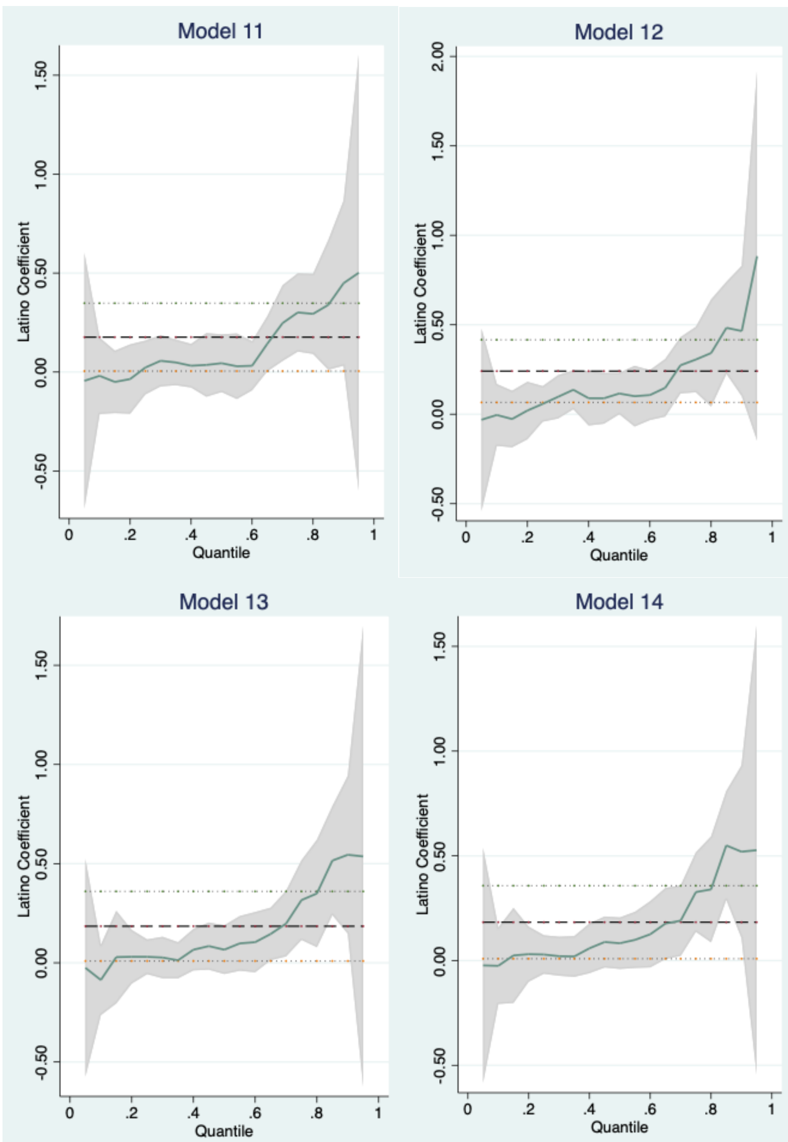
Quantiles	Latino Traditional	Latino Advanced	Latino WAR+FIP	Latino WAR	Latino*WAR
10%	-.063	-.004	.086	-.025	.004
	(.104)	(.196)	(.101)	(.099)	(.010)
25%	.028	.068	.030	.029	.003
	(.119)	(.121)	(.112)	(.108)	(.099)
50%	.129	.117	.083	.083	.487
	(.096)	(.108)	(.107)	(.104)	(.304)
75%	.193**	.255**	.305***	.306***	.284
	(.093)	(.117)	(.117)	(.117)	(.306)
90%	.346	.465**	.545**	.520**	.265
	(.111)	(.208)	(.228)	(.214)	(.249)

Note: * = Significant at 10% level | ** = Significant at 5% level | ******* = Significant at 1% level
Source: Author's Calculations. Coefficients are log point differences which is approximately equal to percentage change.

Using the same specifications as in Figure 15, I conduct quantile regressions for the same relivers. The relationship between the top quantiles of earners will be of particular interest. The expectation is that among players with similar performance and characteristics, the 75th highest earning Latino should make the same as the 75th highest earning white player. While prior models for starters and hitters suggest a wage penalty, the OLS reliever results show an alternative outcome.

There is significant evidence across all the models that the top-tier earners Latino relievers experience wage premiums. Figure 16 shows that premium exists mainly within the 75th and 90th quantiles, similar to what we saw for white starting pitchers. In the 75th quantile, we would expect a premium between 19% and 30%, but that increases in the 90th quantile to 47%-55%. In the context of the three seasons of the sample, top-earning Latino relievers made significant premiums in comparison to elite white relievers.

Figure 17. Coefficients Across All Quantiles for Relief Pitcher Models



Source: Author's Calculations

This is the only area where Latinos experience a statistically significant premium relative to white players. Relievers are the lowest paid position in baseball. Figures 4 and 10 show the significant disparities between relievers and other positions. Figure 17 shows that the premium exists in all models mostly at or above the 60th quantile of earners. This premium comes about unexpectedly, but with statistical significance at the 5% level at the very least. The results of this analysis show that in any future work, the quantitative analysis of starting pitchers and relief pitchers must be calculated separately.

Reflections and Comparison to Prior Research

My estimations suggest that Latino starting pitchers in upper-most quantile of earners experience a wage penalty of at least 60%. On the other hand, Latino relievers above the 75th quantile of earners experience a significant premium compared to white relievers. Latino hitters in the upper-quantiles of earners with over six years of service time in MLB experience a penalty of about 20% in comparison to similar white players. The results of this analysis are significantly different from prior research. For example, Holmes (2011) and Palmer and King (2006) find discrimination in the bottom-tier of earners among hitters in MLB. However, this dynamic no longer comes about. My results suggest that there is high competition between Latinos and whites in the bottom-tier of earners, resulting in no statistically significant differences in earnings. Bodvarsson and Pettman (2010) found discrimination in the form of lower incremental rewards for non-white players. The models which interact WAR and race estimate a similar conclusion for hitters and starting pitcher; Latinos have fewer returns to WAR than white players. The results also oppose the conclusions from Singh, Sack and Dick (2003), who found no evidence of discrimination among players who reached arbitration or free agency.

Among discrimination, there are other reasons why differentials may come about. Link and Yosifov (2011) find that players are more likely to forego monetary returns for a longer guaranteed contract. This provides the player more years of financial security. They do not point out the differences among races, so this remains a possibility to explain the differences observed in this paper. If teams think larger wage dispersion will have a positive effect on winning percentage (Breunig et al. 2012), then they will do their best job to secure elite players for a discounted price. This is so they can still have financial flexibility elsewhere.

For example, the Atlanta Braves managed to sign their best young superstar players Ronald Acuña and Ozzie Albies before they reached their peak-cost in free agency. While these players have yet to reach over six years of service, there are several players in my sample that have. By saving on these players as soon as they got the chance, the Braves created long-term flexibility to sign other players in free agency and create a more balanced team, as Breunig et al. (2012) predict is necessary to win. By targeting players who have more of a sense of urgency to secure financial security, like Latino players, teams like the Braves can ensure a better chance of winning in the short and long-term. Other reasons why Latino players are more likely to sign extensions early in their careers include underprivileged backgrounds, low minor league wages, and smaller signing bonuses as amateurs. These are all motivating factors to take the first big contract offer they receive. While I do not present information on contract extensions, recent trends suggest this may be a reason why we see differentials for Latino hitters and starting pitchers with more than six years of service time.

While WAR proved to be a statistically significant indicator of salary across all positions, the wage differentials were not sensitive to any choice of performance statistics. Going forward, WAR must be included in any baseball labor market research. It takes away the bias of evaluating players based on only one aspect of performance. In addition, it can show us whether players receive the same returns to performance across races and positions.

This analysis proves that all labor market research in Major League Baseball must be updated. There is convincing evidence that shows the highest earning Latino hitters and starting pitchers have been discriminated against since the beginning of the most recent collective bargaining agreement prior to the 2017 season. U.S. labor market research is constantly updated because economists know institutions change, cultures change, and people's behavior change. The results of this analysis show greater levels of differentials between Latinos and whites than in the U.S. labor market. Latino players must be protected in MLB.

Future Research

In future work, there will be an incentive to analyze the likelihood a player will sign an extension before free agency. If there is statistical evidence that shows international Latino players are more likely to sign extensions early in their careers, then there is more evidence that shows owners are actively exploiting players for their backgrounds. In this analysis, a regionality component could be introduced. For example, I would be interested in analyzing if players from areas with certain characteristics are more likely to sign early extensions or make the major leagues in general.

On top of that, a re-opening of the literature on pre-market discrimination using modern statistics is necessary. In Bellemore (2001), the analysis was limited to a certain set of statistics and share of Latinos in the major leagues was much smaller than it is today. The current share of Latinos in the minor leagues suggests that the magnitude of pre-market discrimination has increased. In addition, the current analysis could be expanded to include seasons that spanned the length of prior collective bargaining agreement to see if this relationship also comes about in years prior to 2017. Lastly, an application of The Levy Institute Measure of Time and Income Poverty could be applied to minor leaguers when the necessary data becomes available. The difference in time spent on certain activities between minor leaguers and major leaguers is vast. This could give the league another tool to target reform in the minor leagues.

CONCLUSION AND FINAL REMARKS

The goal of this paper was to investigate the analysis of wage differentials. By observing the qualitative aspects of discrimination faced by Latinos in MLB and the lead-up to the major leagues, we obtained an understanding of the disadvantages faced by Latinos throughout their professional baseball careers. These players are at a serious disadvantage in comparison to white-American players. Language and financial limitations are prevalent. They are less likely to receive promotions to the major leagues than equally performing white players (Bellemore 2001). These conditions all contribute to the growing trend of Latino players signing extensions before they reach free agency. In prior research, the focus was on analyzing wage differentials amongst hitters in MLB. This paper added pitchers into the analysis. The results have shown that to accurately estimate salaries in today's league, a combination of traditional and modern statistics is needed. The creation of Wins Above Replacement has overtaken the analysis of performance in baseball. Using statistics which define every aspect of performance, ensure that all the contributions that players make to the game are accounted for.

There is convincing statistical evidence of wage differentials between Latinos and whites, but in two different ways. Latino hitters and starting pitchers at the top of the distribution experienced a wage penalty in the seasons from 2017-2019. In comparison to the equally performing white players, Latinos face a penalty of at least 20% for hitters and starting pitchers. Relievers experienced a significant premium of at least 20% relative to white relievers. Future research can further investigate why these differentials come about and if discrimination is a significant reason.

If MLB is to ensure equal opportunity in wages for their Latino superstars, then they protect them in the next collective bargaining agreement. Mainly, I would urge the league to decrease the number of years that a player is controlled by his team upon entry into the league. This would take pressure off Latinos to accept deals early in their careers. In addition, minor league pay must be addressed. These players need livable wages during their season. If teams expect them to deliver the best product, then they must be put in the position to do so.

Bibliography

Altonji, J. and C. Pierret. 2001 "Employer Learning And Statistical Discrimination," Quarterly *Journal of Economics, Vol. 116, Issue 3, 313-350.*

Arrow, K. 1973. "The Theory of Discrimination," in Orley Ashenfelter and Albert Rees, eds., Discrimination in Labor Markets. Princeton University Press, pp 3-33.

Baseball Prospectus. Cot's Baseball Contracts. Available at: https://legacy.baseballprospectus.com/compensation/cots/

Baseball Reference. 2020. Minimum Salary. Available at: https://www.baseball-reference.com/bullpen/Minimum_salary

Baumer, B. and G. Matthews. 2014. "A Statistician Reads the Sports Page: There is No Avoiding WAR". *Chance 27(3): 41-44,* DOI: 10.1080/09332480.2014.965630

Becker, G. 1971. "The Economics of Discrimination." *The Journal of Political Economy. 2nd ed,* 813-846.

Bellemore, Fred. 2001 November. "Racial and Ethnic Employment Discrimination: Promotion in Major League Baseball." *Journal of Sports Economics, Vol. 2 No. 4. 358-368.*

Bertrand, M. and S. Mullainathan. 2004. "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *The American Economic Review, Vol. 94, Issue 4, 991-1013.*

Bichsel, J. and J. McChesney. March 2017. Pay and Representation of Racial/Ethnic Minorities in Higher Education Administrative Positions: The Century So Far. Research report. Available At: www.cupahr.org/surveys/briefs.aspx.

Birnbaum, Phil. *A Guide to Sabermetric Research.* Society for American Baseball Research. Available at: https://sabr.org/sabermetrics

Bodvarsson, O. and S. Pettman. 2010 October. Racial Wage Discrimination in Major League Baseball: Do Free Agency and League Size Matter? *Applied Economics Letters 9:12, 791-796*.

Borjas, G. 1982 April. "The Earnings of Male Hispanic Immigrants in the United States." *Industrial and Labor Relations Review. Vol 35, No. 3.*

Breunig, R., M. Jardin, B. Garret-Rumba and Y. Rocaboy. 2012 March. "Wage Dispersion and Team Performance: A Theoretical Model and Evidence From Baseball." *Applied Economics, Vol 46, Issue 3, 271-281.*

Brown, D., C. Link and S. Rubin. 2015. "Moneyball After 10 Years: How Have Major League Baseball Salaries Adjusted?" *Journal of Sports Economics, Vol. 18, Issue 8, 771-786.*

Carneiro, P. 2005 April. “Labor Market Discrimination and Racial Differences in Premarket Factors.” *The Journal of Law and Economics, Vol. 48, Issue 1, 1-39.*

Chang, Alvin, 2013. This is Why Baseball is so White. Vox. Available at: https://www.vox.com/2016/10/27/13416798/cubs-dodgers-baseball-white-diverse

Charles, K. K. and J. Guryan, “Prejudice and Wages: An Empirical Assessment of Becker’s The Economics of Discrimination.” *Journal of Political Economy, Vol. 116, Issue 5, 773–809.*

Darity Jr, W. and P. Mason. 1998 Spring. “Evidence on Discrimination in Employment: Codes of Color, Codes of Gender.” *The Journal of Economic Perspectives, Vol. 12, No. 2, 63-90.*

FanGraphs. 2019. Leaderboards. Available at: https://www.fangraphs.com/leaders.aspx?pos=all&stats=bat&lg=all&qual=y&type=8&season=2019&month=0&season1=2019&ind=0

Findling, M. 2017. “Discrimination in the United States: Experience of Latinos.” Harvard University Department of Health Policy and Management.

Gilette, L. 2017. *Status and Trends in the Education of Racial and Ethnic Groups 2017*. The Department of Education.

Gonzalez, A. 2013. *The Evolution of Hispanic Literacy in the Twenty-First Century.* Th Economic Status of the Hispanic Population: Selected Essays, edited by M.T. Moran and A.Dávila, 23-32.

Heckman, J, E. Lazear, K. Murphy. 2018. “Gary Becker Remembered.” *Journal of Political Economy. Vol. 126, 1-6.*

Holmes, P. 2011. “New Evidence of Salary Discrimination in Major League Baseball. The Economics of Sports Labor Markets.” *Labour Economics, Vol 18(3), 320-331.*

Jayal, A., A. McRobert, G. Oatley and P. O’Donoghue. 2018. *Sports Analytics: Analysis, Visualization and Decision Making in Sports Performance*. Rutledge.

Jewell, T., R.Brown and S. Miles. 2002 January. “Measuring Discrimination in Major League Baseball: Evidence from the Baseball Hall of Fame.” *Applied Economics, Vol 34. Issue 2. 167-177.*

Koble, M. 2008. “Not My Father’s Game: Immigration, Major League Baseball and the Dominican Republic.” The University of Texas at Arlington.

Kochhar, R. 2019 March. *Latinos’ Incomes Higher Than Before Great Recession, but US-Born Latinos Yet to Recover.* Pew Research Center.

Koenker, R., and K. Hallock. 2001. "Quantile Regression." *Journal of Economic Perspectives-Volume 15, Number 4, 143-156.*

Lagesse, D. 2016. *Baseball is a Field of Dreams and Dashed Hopes for Dominicans*. National Public Radio. Available at: https://www.npr.org/sections/goatsandsoda/2016/04/03/472699693/baseball-is-a-field-of-dreams-and-dashed-hopes-for-dominicans.

Lang, K. and K. Lehmann. 2005. "Racial Discrimination in the Labor Market with Posted Wage Offers." *American Economic Review, Volume 94, Issue 4, 1327-1340.*

Lang, K. and M. Manove. 2011 "Education and Labor Market Discrimination," *American Economic Review, Volume 101, Issue 4, 1467–96.*

Lang, K. and J.-Y. K. Lehmann. 2012 "Racial Discrimination in the Labor Market: Theory and Empirics," *Journal of Economic Literature, Vol 50, No. 4, 959-1006.*

Lang, K. and A. Spitzer. 2020 Spring. "Race Discrimination: An Economic Perspective." *Journal of Economic Perspectives, Volume 34, Number 2. 69-69.*

Lichtman, M. 2010 May. The FanGraphs UZR Primer. FanGraphs. https://blogs.fangraphs.com/the-fangraphs-uzr-primer/#2

Major League Baseball. 2016 November. 2017-2021 Basic Agreement. Major League Baseball Players Association. Available at: https://d39ba378-ae47-4003-86d3-147e4fa6e51b.filesusr.com/ugd/b0a4c2_95883690627349e0a5203f61b93715b5.pdf

Meléndez, E. and J. Hinojosa. 2017 October. *Estimates of Post-Hurricane Maria Exodus of Puerto Rico*. Center for Puerto Rican Studies. Hunter College.

Neal, D. and W. Johnson. 1996. "The Role of Premarket Factors in Black-White Wage Differences." *Journal of political Economy, Vol .104, Issue 5. 869-95.*

Oaxaca, R. and M. Ransom. 1994 March. "On Discrimination and the Decomposition of Wage Differentials." *Journal of Econometrics, Volume 61, Issue 1. 5-21.*

O'Neill, J. 1990. "The Role of Human Capital in Earnings Differentials between Black and White Men." *The Journal of Economic Perspectives, Vol. 4, Issue 4, 25-45.*

Oreopoulos, Phillip. 2013 May. "Making College Worth it: A Review of Research on Returns to Education." *National Bureau of Economic Research.*

Palmer, M. and R. King. 2006 March. "Has Salary Discrimination Really Disappeared from Major League Baseball." *Eastern Economic Journal Vol. 32, Issue 2, 285-295.*

Reimers, C. 1983 November. "Labor Market Discrimination Against Hispanic and Black Men." *The Review of Economics and Statistics, Vol. 65, Issue 4, 570-578.*

Richardson, D., G. Hamra, R. MacLehose, S. Cole and L. Chu. 2015 July. "Hierarchical Regression for Analyses of Multiple Outcomes." *American Journal of Epidemiology, Volume 182, Issue 5.*

Rivera-Batiz, F. 1999 February. "Undocumented Workers in the Labor Market: An Analysis of the Earnings of Legal and Illegal Mexican Immigrants in the United States." *Journal of Population Economics, Vol. 12, 92-114.*

Sanchez-Soto, G., A. Bautista-León and Joachim Singelmann. 2018 May. "The Return-on-Education Gap Between Hispanics and non-Hispanic Whites." *Papeles de Pobalcoin, Vol. 24, no. 98.*

Singh, P., A. Sack and R. Dick. 2003. "Free Agency and Pay Discrimination in Major League Baseball." *Sociology of Sport Journal, Vol. 20, Issue 3, 275-285.*

Swazrtz, M. 2014. "Searching for Racial Earnings Differentials in Major League Baseball". FanGraphs. The Hardball Times. Available at: https://tht.fangraphs.com/searching-for-racial-earnings-differentials-in-major-league-baseball/

Swartz, M. 2017. "The Linearity of Cost per Win." Fangraphs. Available at: https://blogs.fangraphs.com/the-linearity-of-cost-per-win/

Szymborski, Dan. 2020. ZiPS. FanGraphs. Available at: https://www.fangraphs.com/projections.aspx?pos=all&stats=bat&type=zips

Tainsky, S. and J. Winfree. 2010 March. "Discrimination and Demand: The Effect of Internatinoal Player on Attendance in Major League Baseball." *Social Science Quarterly, Vol. 16, Issue 4, 353-374.*

Tango, T., M. Lichtman and A. Dolhin. *The Book: Playing the Percentages in Baseball*. Potomac Books.

Tango, Tom. "How to Calculate WAR." http://www.insidethebook.com/ee/index.php/site/comments/how_to_calculate_war/

Tells, E. and E. Murouia. 1998 March. "Phenotypic Discrimination and Income Differences Among Mexican Americans." *Social Science Quarterly, Vol 83, Issue 2, 612-623.*

Tienda, M and Faith Mitchell. 2006. "Hispanics and the Future of America." *National Academy of Sciences.*

Vanderhoeven, N. 2018 May. *Cultural Expression in Baseball: The Exploitation of Latin American Players Justifies Flair*. Medium. Available at: https://medium.com/@vanderhoeven.noah/cultural-expression-in-baseball-the-exploitation-of-latin-american-players-justifies-flair-479096b6af82

Vargas, A. 2000. "The Globalization of Baseball: A Latin American Perspective." *Indiana Journal of Global Studies. Volume 8 Issue 1.*

Volz, B. 2012 November. "Race and the Likelihood of Managing in Major League Baseball." *Journal of Labor Research, Vol. 34, Issue 1, Article 3.*

Wagner, James. 2019. "This One Was Worth It: Yankees Set Record With Jasson Dominguez. The New York Times. https://www.nytimes.com/2019/07/02/sports/jasson-dominguez-yankees.html

Weinber, N. 2014. "Stats to Avoid: Runs Batted In." Fangraphs. Available at: https://library.fangraphs.com/stats-to-avoid-runs-batted-in-rbi/

Williams, D. 2019. "Major League Baseball's Indentured Class: Why the Major League Baseball Players Association Should Include Minor League Players." *The University of San Francisco Law Review. Vol 53, Issue 3, 515-550.*

APPENDIX

Equation A1. Slugging Percentage

$$SLG = \frac{Singles + (Doubles \times 2) + (Triples \times 3) + (Home\ Runs \times 4)}{Total\ At\ Bats}$$

The measure of SLG does not include walks which are not counted in the statistics as at-bats (AB), but as plate appearances (PA). Technically, a walk does not count as an at-bat, but it does count for a plate appearance, so SLG only focuses on hits and is calculated as such:

Equation A2. On Base Percentage (OBP) OBP is the rate at which a batter safely reaches base

$$OBP = \frac{Hits + Base\ on\ Balls + Hit\ By\ Pitches}{Plate\ Appearances}$$

OBP is the rate at which a batter safely reaches base. OBP served as a jump start to the analytical revolution, as Brown et al. (2015) explain. However, it is not considered a sabermetric statistic because it does not reflect accurate run values. Each time on base is associated with a different run value depending on the precise state of the game. Runs Batted In (RBI) counts how many runners scored safely while a hitter was at the plate. The key issue is that not all players come to the plate with equal runners on each base during the course of a game or season. This means a player could have several more opportunities to obtain RBIs. For example, certain hitters in better lineup spots and on better teams will have more opportunity for RBIs than others. In the baseball community, the narrative around players with high RBI totals is that they are "clutch". Clutch refers to a hitter getting timely hits with runners on base. wOBA accurately measures this using game environment-specific run values in describing offensive events. For RBIs, each one counts the same.

Equation A3. Weighted on-Base Average (wOBA)

$$RAA = \left(\frac{wOBA - lgwOBA}{wOBAScale}\right) * PA$$

$$Batting\ Runs = wRAA + \left(\frac{lgR}{PA} - (\frac{PF \times lgR}{PA}\right) \times PA + (\left(\frac{lgR}{PA}\right) - \left(\frac{league\ nonpitcher\ wRC}{PA}\right)) \times PA$$

Just like WAR, the goal is to observe how many runs better, or worse, a player was than the replacement level player. By doing so, we calculate the stat called Weighted Runs Above Average (wRAA). The equation for wRAA appears as

$$wOBA = \left(\frac{.69 \times uBB + \ .72\ HBP + \ .89 \times 1B + 1.27 \times 2B + 1.62 \times 3B + 2.10 \times HR}{AB + BB - IBB + SF + HBP}\right)$$

Where lgwOBA is the league (lg) average wOBA for any given year and wOBA scale is a scale factor that allows us to compare players from different years and generations. The idea here is that a player with more plate appearances should be able to create more runs. A player who is good for a long time may be more valuable than a player who is great for short time, or only in certain situations. The manipulation of wRAA to batting runs controls for the players league and home park so we can have an easier comparison. It has been thoroughly argued that wOBA is a better indicator of player's hitting than the traditional statistics.

Baserunning

Equations A4, A5, A6, A7

$$BsR = UBR + wSB + wGDP$$

$$wSB = SB \times runSB + CS \times runCS - lgwSB \times (1B + BB + HBP - IBB)$$

$$wlgSB = \frac{SB \times runSB + CS \times runCS}{1B + BB + HBP - IBB}$$

$$runCS = 2 \ \times RunsPerOut + .075$$

WAR assigns a run value to what a player does while running the bases through the use of the stat BsR. BsR can be calculated with a combination of three things: Ultimate Base Running (UBR) which measures non-stolen base type running, Weighted Stolen Bases (wSB) which measures the number of runs above or below the average player contribute to his team by stealing bases or being thrown attempting to steal, and Weighted Ground into Double Play Runs (wGDP) which takes all the extra outs a player costs or saves his team by hitting into more or fewer double plays than average given his opportunities. The process is similar in that the goal is to investigate how many more, or fewer, runs a runner contributes compared to the average. Equations 2A-5A shows the calculations. Where league stolen base runs (lgwSB) is equal to the league average and where runCS is the run value of being caught attempting to steal. Runs per out is the simply the run value of each out. Each year, the value of a stolen base changes, just like the run value of a certain batted ball event.

Finally, UBR and wGDP calculations are not publicly available because they take the run expectancy of advancement and credit the base runner depending on the frequency of with which the average runner advances, or reaches first, in the same situation. These calculations are reliant on tracking data. The tracking data would tell us both the quality of a jump, or a runner's reaction, and the speed of the runner. For a long period of time Speed Score (Spd), the baserunning statistic chosen by Holmes, was the best way to measure a player's baserunning output. However, Spd does not assess running in a runs above average framework. The best runners are given a Spd rating of 7.0 and the worst are given 2.0. Spd does not involve tracking, because there was not any available. It does not compare the player to the average runner. For these reasons, BsR is a clear analytical improvement.

Defense

UZR factors in several aspects of a defender's play. It attempts to do the same as previous statistics, quantify the run value, saved or lost, of a player's defensive prowess. There are four components to UZR. They include Outfield Arm Runs (ARM) which is the amount of runs above average an outfielder saves with their arm by preventing runners from advancing, Double-Play Runs (DPR) which the amount of runs above average an infielder is by turning double-plays, Range Runs (RngR) which shows if the player gets to more balls than the average and Error Runs (ErrR) which says if a player commits more or fewer runs compared with the league average at their position. While the calculations for UZR are not available, a description from its creator, Mitchell Lichtman, is provided.[24] Here, the methodology for UZR is described in detail.

The main caveat of the statistics is that there is a positional adjustment because some positions are harder than others according to UZR. Adding the positional adjustment, we get to the statistic Fielding Runs (DEF), which is the official input into WAR. The goal is to compare players across different positions. In addition, there are park adjustments because in baseball, no park is created equal. But in its natural form, the calculations for UZR are simple because all four components are simply added up to get a number above, at or below 0.

Pitcher Statistics

Equations A8 and A9. Calculating FIP

$$ifFIP = \frac{(13 \times HR) + \big(3(BB + HBP)\big) - (2(K + IFFB))}{Innings\ Pitched} + ifFIP\ constant$$

$$ifFIP\ constant = lgERA - \frac{((13 \times lgHR) + \big(3(lgBB + lgHBP)\big) - (2(lgK + lgIFFB)))}{lgIP}$$

24 https://blogs.fangraphs.com/the-fangraphs-uzr-primer/#2

Where IFFB is infield fly balls because in WAR, they are treated as the same as strikeouts, hence ifFIP. lgERA is the Earned Run Average of the league's pitchers. In other words, the runs the average pitchers gives up per nine innings and the added constant so that we can compare from year to year and era to era.

Equations A10 and A22. Other Inputs for WAR

$$dRPW = (\frac{\left(\left(18 - \frac{IP}{G}\right) \times (league\ FIPR9)\right) + \left(\left(\frac{IP}{G}\right) \times pFIPR9\right)}{18} + 2) \times 1.5$$

$$Wins\ Per\ Game\ Above\ Average\ (WPGAA) = \frac{RAA9}{dRPW}$$

After FIP, we need to calculate the Pitcher Specific Runs per Win. This is done so using a dynamic equation that controls for the pitcher's innings pitched per game which allows us to specific between starters with varying innings totals and relievers. The equation is complicated but not so hard to interpret. What this equation does is get the pitcher specific run dynamic of their own games by taking the run average of when they are on the mound and when they are not and then rescales it to get it to a runs per nine inning scale. It is then added to two then multiplied by 1.5 in order to get it to a Runs Per Win framework which is desirable for WAR. It is then converted to WPGAA by dividing by Runs Above Average per 9 which is calculated by subtracting the pitcher specific FIPR9 from their leagues average.

Equations A12, A13 and A14. Replacement Level Pitcher and Leverage Multiplier for Wins Above Replacement

$$Replacement\ Level = .03\left(1 - \frac{GS}{G}\right) + .12(\frac{GS}{G})$$

$$Wins\ Per\ Game\ Above\ Replacement\ (WPGAR) = WPGAA + Replacement\ Level$$

$$Leverage\ Multiplier = \frac{1 + gmLI}{2}$$

If the pitcher was a full-time reliever, the replacement level is .03 times the share of your games that took place as a reliver and as full-time starter it is .12 times the share of your games that took place as a starter. The entire equation is needed for pitchers who have had both roles. The following equation, WPGAR, gives us the wins scale. The Leverage Multiplier is to give relievers a benefit for pitching in high-pressure situations versus those who pitch in low-pressure. It is simply multiplied by the reliever WAR at the end of the calculation. Also, remember that every pitcher's WAR has park and league adjustment to ease the comparison between players.

Table A1. Estimates of Model 1 Quantile Regressions (10th, 25th, 50th, 75th, 90th)

	(1)	(2)	(3)	(4)	(5)
	10th	25th	50th	75th	90th
Service	0.040	0.075	0.104***	0.076***	0.066***
OBP	3.159	1.542	1.066	2.239	2.130
SLG	0.398	-0.012	0.204	-0.244	-0.811
RBI	0.017	0.011	0.004	0.001	0.004
Def	-0.006	-0.005	-0.008	-0.003	-0.002
SB	0.051**	0.026	0.014	0.003	-0.006
Pos	0.027	-0.007	-0.000	0.000	-0.004
A	0.187	-0.787	-1.466**	-1.754***	-1.975***
B	-0.238	-0.535	-0.077	-0.072	-0.165
L	-0.399	-0.170	-0.342**	-0.190*	-0.289**
Constant	12.265***	14.058***	14.798***	15.359***	15.843***
Observations	162	162	162	162	162

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A2. Estimates of Model 2 Quantile Regressions (10th, 25th, 50th, 75th, 90th)

	(1)	(2)	(3)	(4)	(5)
	10th	25th	50th	75th	90th
Service	0.014	0.105**	0.111***	0.080***	0.050***
wOBA	11.213*	6.560*	4.160**	2.502	2.049*
BsR	0.021	0.018	-0.001	0.009	-0.008
Def	0.037	0.010	0.002	-0.001	0.000
wRC.	-0.003	-0.001	0.001	0.001	0.000
Pos	0.018	0.022	0.011	-0.000	-0.012
A	-0.221	-1.214	-1.474**	-1.789***	-1.925***
B	0.166	-0.194	0.077	-0.092	-0.018
L	-0.069	-0.089	-0.201	-0.215*	-0.126
Constant	11.738***	13.072***	13.941***	15.181***	15.822***
Observations	162	162	162	162	162

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A3. Estimates of Model 3 Quantile Regressions (10th, 25th, 50th, 75th, 90th)

	(1)	(2)	(3)	(4)	(5)
	10th	25th	50th	75th	90th
Service	0.046	0.105*	0.106***	0.080***	0.048***
WAR	0.214**	0.137**	0.066**	0.036	0.020
A	-0.130	-1.003	-1.433**	-1.776***	-1.962***
B	-0.104	-0.228	0.075	-0.107	-0.091
L	-0.242	-0.170	-0.109	-0.241**	-0.175*
Constant	14.520***	14.877***	15.359***	16.001***	16.525***
Observations	162	162	162	162	162

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A4. Estimates of Model 4 Quantile Regressions (10th, 25th, 50th, 75th, 90th)

	(1)	(2)	(3)	(4)	(5)
	10th	25th	50th	75th	90th
Service	0.095*	0.086	0.101***	0.075***	0.060***
WAR	0.372***	0.070	0.027	0.006	0.021
RBI	0.004	0.007	0.005*	0.002	0.003
wOBA	-9.746	0.631	0.140	0.980	-0.357
A	0.172	-0.936	-1.436**	-1.772***	-1.925***
B	-0.009	-0.016	-0.009	-0.057	-0.146
L	-0.128	-0.119	-0.248*	-0.234**	-0.227**
Constant	16.608***	14.529***	15.162***	15.660***	16.331***
Observations	162	162	162	162	162

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A5. Estimates of Model 5 Quantile Regressions (10th, 25th, 50th, 75th, 90th)

	(1)	(2)	(3)	(4)	(5)
	10th	25th	50th	75th	90th
WAR	0.152*	0.092	0.023	0.030	0.017
A	-0.225	-0.862	-1.291	-1.671***	-1.816***
B	-0.393	-0.887**	-0.577**	-0.325*	-0.114
L	-0.672	-0.011	0.180	0.240*	0.234*
W	0.000	0.000	0.000	0.000	0.000
A # WAR	0.000	0.000	0.000	0.000	0.000
B # WAR	0.148	0.271*	0.188*	0.063	0.013
L # WAR	0.163	-0.054	-0.092	-0.145**	-0.133**
Constant	15.141***	15.838***	16.337***	16.709***	16.867***
Observations	167	167	167	167	167

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A6. Estimates of Model 6 Quantile Regressions (10th, 25th, 50th, 75th, 90th)

	(1)	(2)	(3)	(4)	(5)
	10th	25th	50th	75th	90th
Service	0.019***	0.022***	0.021***	0.013***	0.002
W	0.023	0.008	-0.001	-0.003	0.009
L	0.122	0.054	-0.028	-0.057	-0.085
IP	0.013	0.006	0.007	0.016	0.015*
ERA	-0.465	-0.141	-0.171	-0.635	-0.117
HR/9	0.532	0.117	0.648*	2.114**	1.126**
K/9	0.044	-0.002	-0.037	-0.179	-0.062
A	1.340	0.871	1.073**	0.236	-0.042
B	0.792	0.430	0.133	-0.234	-0.010
L	0.028	-0.104	-0.329*	-0.496	-0.389
Constant	9.241***	11.430***	12.631***	13.990***	14.122***
Observations	155	155	155	155	155

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A7. Estimates of Model 7 Quantile Regressions (10th, 25th, 50th, 75th, 90th)

	(1)	(2)	(3)	(4)	(5)
	10th	25th	50th	75th	90th
Service	0.017***	0.021***	0.021***	0.013***	0.001
FIP	-0.060	-0.068	0.189	0.374	0.091
IP	0.018	0.008	0.008	0.011	0.004
LOB%	0.002	-0.001	0.002	0.007	0.001
xFIP	0.186	0.022	-0.141	-0.010	-0.180
A	1.803	1.053*	1.301***	0.828	-0.103
B	0.951	0.411	0.072	-0.383	0.051
L	0.208	-0.086	-0.249	-0.632	-0.715***
Constant	8.436***	11.536***	11.470***	10.601***	16.525***
Observations	155	155	155	155	155

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A8. Estimates of Model 8 Quantile Regressions (10th, 25th, 50th, 75th, 90th)

	(1)	(2)	(3)	(4)	(5)
	10th	25th	50th	75th	90th
Service	0.018***	0.021***	0.022***	0.014***	-0.001
FIP	0.578	0.198	0.448*	0.380	-0.277
WAR	0.388	0.110	0.211**	0.152	-0.048
A	1.726*	0.909*	1.053***	0.637	-0.093
B	1.195	0.503	0.116	-0.316	0.162
L	0.134	-0.112	-0.285*	-0.354	-0.726***
Constant	8.817**	11.723***	10.706***	12.611***	18.449***
Observations	155	155	155	155	155

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A9. Estimates of Model 9 Quantile Regressions (10th, 25th, 50th, 75th, 90th)

	(1)	(2)	(3)	(4)	(5)
	10th	25th	50th	75th	90th
Service	0.018***	0.021***	0.022***	0.013***	-0.001
WAR	0.255**	0.038	0.054	-0.004	0.091
A	1.736	0.951*	1.241***	0.683	-0.086
B	1.070	0.523	0.157	-0.378	0.289
L	-0.004	-0.117	-0.244	-0.407	-0.609***
Constant	11.466***	12.679***	13.027***	14.755***	16.799***
Observations	155	155	155	155	155

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A10. Estimates of Model 10 Quantile Regressions (10th, 25th, 50th, 75th, 90th)

	(1)	(2)	(3)	(4)	(5)
	10th	25th	50th	75th	90th
WAR	0.002	0.291*	0.152	0.028	0.111***
A	5.002***	5.027	3.529	0.346	0.191
B	0.677***	0.702	4.441	2.723	2.568
L	0.014	2.314	1.865**	-0.327	-0.148
A # WAR	-0.546***	-0.835	-0.782	-0.028	-0.111
B # WAR	0.333***	0.044	-1.085	-0.889	-0.972**
L # WAR	-0.007	-0.785*	-0.806***	-0.129	-0.232**
Constant	13.222***	13.198***	15.016***	16.560***	16.716***
Observations	155	155	155	155	155

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A11. Estimates of Model 11 Quantile Regressions (10th, 25th, 50th, 75th, 90th)

	(1)	(2)	(3)	(4)	(5)
	10th	25th	50th	75th	90th
Service	0.008***	0.011***	0.015***	0.018***	0.018***
SV	0.005	0.011**	0.025***	0.022***	0.024**
IP	-0.003	-0.003	-0.007	-0.007	-0.009
ERA	0.049	0.035	0.017	0.075	0.089
HR/9	0.008	-0.049	-0.058	-0.124	-0.171
K/9	0.019	0.014	0.013	0.020	0.018
A	1.474**	1.049	1.714***	1.473***	1.203
B	0.146	0.028	0.221	0.039	-0.198
L	-0.063	0.028	0.129	0.193**	0.346
P	0.412	-0.040	-0.699	-1.002	-1.338
Constant	12.594***	12.867***	13.123***	13.034***	13.431***
Observations	357	357	357	357	357

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A12. Estimates of Model 12 Quantile Regressions (10th, 25th, 50th, 75th, 90th)

	(1)	(2)	(3)	(4)	(5)
	10th	25th	50th	75th	90th
Service	0.008***	0.011***	0.015***	0.019***	0.018***
FIP	-0.035	-0.002	0.002	-0.039	0.057
IP	-0.006	-0.003	-0.004	-0.007	-0.013
LOB%	0.001	-0.000	-0.001	-0.001	-0.001
xFIP	0.077	-0.013	-0.108	-0.165	-0.329*
A	1.519***	1.010	0.470	1.517**	0.927
B	0.149	0.089	0.136	0.070	0.086
L	-0.004	0.068	0.117	0.255**	0.465**
P	0.530	0.186	-0.239	-0.731	-1.235
Constant	12.969***	13.186***	13.619***	14.308***	15.563***
Observations	357	357	357	357	357

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A13. Estimates of Model 13 Quantile Regressions (10th, 25th, 50th, 75th, 90th)

	(1)	(2)	(3)	(4)	(5)
	10th	25th	50th	75th	90th
Service	0.007***	0.011***	0.015***	0.019***	0.019***
FIP	0.112	-0.012	-0.002	-0.081	-0.277
WAR	0.110	-0.015	0.064	0.052	-0.090
A	1.575***	1.056*	1.731***	1.616**	1.276
B	0.140	0.030	0.152	0.084	0.022
L	-0.086	0.030	0.083	0.306***	0.545**
P	0.601	0.242	-0.178	-0.605	-1.031
Constant	12.336***	12.969***	12.869***	13.234***	14.435***
Observations	357	357	357	357	357

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A14. Estimates of Model 14 Quantile Regressions (10th, 25th, 50th, 75th, 90th)

	(1)	(2)	(3)	(4)	(5)
	10th	25th	50th	75th	90th
Service	0.008***	0.011***	0.015***	0.019***	0.020***
WAR	-0.001	-0.006	0.066	0.168***	0.181
A	1.595***	1.047*	0.476	1.597**	1.321
B	0.249	0.026	0.212	0.095	0.074
L	-0.025	0.029	0.083	0.306***	0.520**
P	0.619	0.242	-0.176	-0.617	-0.957
Constant	12.761***	12.919***	12.860***	12.876***	13.130***
Observations	357	357	357	357	357

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A15. Estimates of Model 15 Quantile Regressions (10th, 25th, 50th, 75th, 90th)

	(1)	(2)	(3)	(4)	(5)
	10th	25th	50th	75th	90th
WAR	-0.003	0.004	0.191	0.267*	0.043
A	2.009***	1.985***	1.196	0.293	-0.537
B	-0.018	0.001	-0.244	-0.093	0.295
L	-0.011	-0.019	-0.361	-0.192	-0.344
P	0.779***	0.744	-0.307	-1.317	-1.833
A # WAR	-0.635***	-0.642	-0.830	-0.906	-0.682
B # WAR	0.007	-0.005	0.530	0.143	0.550
L # WAR	0.004	0.003	0.487	0.284	0.265
P # WAR	0.000	0.000	0.000	0.000	0.000
Constant	13.224***	13.248***	14.037***	14.941***	15.770***
Observations	357	357	357	357	357

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$