Free Agent Salaries in Baseball: Performance and Non-Performance Factors

BA Economics

20/03/20

Executive Summary

Over the past ten years the baseball industry has undergone an analytical revolution, whereby the methods and approaches of valuing certain skills and abilities have evolved considerably. Moreover, prior literature has shed light on a number of non-performance factors that have influenced the salaries of baseball players in the past. Yet there has been little recent work investigating the wage determination process, namely on a sample of free agent from the last 7 years. Therefore this dissertation investigates the effect on modern salaries of performance factors, as well as non-performance factors such as the signing team, player agent and race.

This dissertation finds that modern, untested variables denoting certain fielding and pitching abilities mark an improvement over traditional variables in explaining salaries. This is deemed to indicate that these particular skills are valued by teams in their salary offers to players. However, there is evidence to suggest that variables utilised in past literature still have a role in determining modern salaries. Further, both the market size and income tax rate of a team's location are found to have no effect on salaries, suggesting that teams do not have labour market advantages in terms of salary offers. The addition of agents into the salary model further reveals that four particular agents were able to increase their clients salaries beyond the level their performance would suggest.

Modern econometric methodology is then used to establish whether salary discrimination against African Americans is evident, and whether this applies to certain levels of the salary distribution. However, no evidence of discriminatory wage differentials is found, implying in the baseball industry that either the cost of discrimination has risen, or rather that the preference for discrimination amongst fans has declined.

Acknowledgements

I would like to express my thanks to my supervisor, Angel Hernando-Veciana, for his advice and guidance throughout the process of completing this dissertation.

Dissertation Declaration

This dissertation is the result of my own work. Material mentioned in the dissertation from the published or unpublished work of others is credited to the author in question. The dissertation is approximately 11,979 words in length. Note that tables 1 and 2 are included in this word count.

Contents

List of figures:

List of Tables:

1 Introduction

1.1 Introduction

The publication of Moneyball in 2003 triggered major shifts in skill compensation patterns in Major League Baseball, capturing the attention of both labour economists and sports fans. Lewis' novel famously told how the lowly Oakland Athletics identified certain skills that were severely undervalued in the baseball labour market. Consequently, they acquired undervalued players to punch above their financial weight and become the best team in Major League Baseball by 2002. As Hakes and Sauer (2006, p.173) highlight, this labour market inefficiency was all the more surprising given that "more is known about pay and quantified performance in this market than in any other labour market in the American economy."

Since the early 2000s, statisticians inspired by Moneyball have taken advantage of the rich informational availability to develop new metrics and variables that seek to provide an even better picture of productivity and account for a greater variety of skills. The way teams value labour inputs has therefore changed even further over the last decade (Congdon-Hohman and Lanning, 2018), but no work has been done to examine potential new determinants of salary in this time. Whilst in regards to Moneyball, various literature has highlighted that salary determination quickly adjusted to reflect newly valued skills. This therefore poses the question of whether these even newer metrics are reflected in modern free agent salaries, and thus mark an improvement over older variables in explaining salaries.

Furthermore, given that the baseball industry has the informational availability to capture performance in such a rich manner, Kahn (2000, p.76) contends that "sports labour markets can be seen as a laboratory." Indeed, once performance has been accounted for, a variety of economic issues can be examined, potentially providing insights for labour markets in general. One such non-performance issue in particular is that of discrimination, where studies in conventional labour markets are often limited by the inability to accurately measure performance (Rosen and Sanderson, 2001). The baseball industry has too had a long, fraught history with racial discrimination, with Jackie Robinson breaking the colour barrier in the 1940s. Since then, empirical studies have generally found little evidence of racial wage differentials, until the use of new econometric techniques found evidence to reignite the debate.

Another such area of interest may be the effect of the signing team on players' salaries. Past literature has suggested a team's characteristics, in particular its market size and state tax rates, may have implications for the baseball labour market. Whether these non-performance effects still exist again remains to be seen. Furthermore, the influence of agents has been rising in the professional sports industry, with notorious super-agents such as Jorge Mendes in football and Scott Boras in baseball attracting interest from fans and media outlets alike. Forbes (2019) estimates four of the top ten agents in sports to be baseball agents, with Scott Boras having recently negotiating $1 billion worth of contracts in only three weeks (Anderson, 2019). However this begs the question of when performance is held constant, do these agents actually generate extra contract value for their clients, and if so how much.

1.2 Outline and Overview

The overall purpose of this study is therefore two pronged. Through modelling wages more may be unearthed about salary determination in a $10.7 bil industry (Forbes, 2019) that has seen further progression in attitudes towards skill valuation in recent years. Meanwhile, given performance can be accounted for in such great detail, certain non-performance conclusions may be generalised and applied to other labour markets with similar structures.

This dissertation will try to accomplish these aims through building up a series of salary models first using performance variables. The focus will then move to non-performance factors such as the identity of the signing team, as well as player specific characteristics such as their race and agent. All models will be estimated using a dataset from 2012 to 2019, on which wage determination has not been previously studied. This will be primarily through standard OLS regression, with wage discrimination investigated using quantile regression.

Following this section, chapter 2 provides some brief explanations of certain intricacies of baseball and its labour market to inform proceeding discussion. The third chapter reviews the prior literature on wage determination in baseball, through discussion of study into both performance and non-performance salary factors. Chapter 4 will then describe the data and methodology employed in detail, whilst outlining the various models to be estimated.

Subsequently, the model estimations will be presented in chapter 5, discussing potential reasons for the results as well as emphasising any implications. Finally, the key conclusions will be outlined in chapter 6, as well as some limiting factors and avenues for future research.

2 Baseball and its Labour Market

2.1 Performance as Productivity

In order to gain an understanding of the way productivity is captured in the baseball industry, it is necessary to give some brief explanation in regards to the game of baseball itself. The game centres around a dual between pitcher and a hitter. The pitcher is attempting to get the hitter out, either by getting three strikes (called a strikeout) or getting him caught out or thrown out by a fielder. Meanwhile, the hitter is attempting to hit the ball and round the bases, or get onto first base through a walk. A walk refers to where the pitcher misses the strike zone with four pitches. Rounding all four bases is rewarded a run, with the team with the most runs after nine 'innings' awarded the win.

Considering this, productivity can be thought of differently for pitchers and hitters. For pitchers it is thought of in terms of preventing runs, whilst for hitters it is primarily in terms of batting ability. In particular, the ability to hit the ball further and more often, thus getting round more bases to score more runs (Hakes and Sauer, 2006). These distinctions are highlighted in figure 1 below.

Figure 1: The Components of Player Productivity

Hitters			**Pitchers**
Batting	Baserunning	Fielding	Run prevention

As shown by figure 1, hitter ability also includes baserunning, the ability to run more effectively around bases. Further, fielding is important given that hitters have to field whilst the other team takes turn hitting. Therefore their skills here will have an impact on overall productivity.

Traditionally, performance variables in the literature have focused on the batting and run prevention aspects of figure 1. However, since the Moneyball revolution, new variables have been developed by statisticians which try to account for baserunning and fielding abilities. Meanwhile, further attempts have been made to more accurately capture the traditional aspects of batting and run prevention ability.

2.2 The baseball labour market

This subsection details some unique features of the baseball labour market, seeking to clarify some key terminology and inform discussion in proceeding chapters. Unlike a typical labour market, baseball players were once legally bound to teams for the duration of their career by contract clause known as the reserve clause. This clause essentially allowed teams to renew a player's contract at the end of the season, meaning that players could not negotiate contracts with any other team (Hill and Spellman, 1983). Although this kind of labour clause was deemed unconstitutional in almost every other US industry, monopsony exploitation persisted in baseball up until a free labour market for players was established in 1976 (Rosen and Sanderson, 2001).

Despite the abolishment of the reserve clause, the contemporary baseball labour market still does reflect some monopsony power, as only those players with over 6 years of experience are eligible for this free agency. Figure 2 below illustrates the current structure of the baseball labour market.

Figure 2: Years of Experience and Contractual Eligibility

1	2	3	4	5	6	6+
Rookie minimum			Arbitration Eligible			Free Agency

Source: (MLB, n.d.)

As shown above, teams do still exhibit a significant amount of control over players with less than 6 years of experience. Player's with 1-3 years of experience see their contracts renewed at the rookie minimum of $563,500, whist those with 3-6 years of experience see their contracts renewed at a salary determined by independent arbiters unless they agree on a contract with their current team (MLB, n.d.). Only once players surpass 6 years of experience are they allowed to negotiate a contract on the free market and crucially only here will they receive the returns to performance conferred by their marginal revenue product

(Krautmann,1999). The negotiation process for free agents involves team executives who use past performance data to form expectations about future performance. Salary bids are then submitted to the agent of a given player of interest. Further, unlike other American sports, Major League Baseball (MLB) does not have a hard salary cap.

3. Literature Review

3.1 Introduction

The Literature regarding the determinants of salary in the baseball labour market is young, yet relatively wide, and has developed considerably since the first publications in the late 50s. This chapter will start with a discussion of its origins, in particular, the debate on marginal revenue product and the introduction of free agency in 1976. The review will then investigate approaches to modelling the effect of performance on player's salaries. Following this, focus will turn to developments in the literature regarding the importance of team specific characteristics and player agents, as well as the wide literature on wage discrimination in baseball that is driven by the ability to adequately control for job performance. Finally, the review will conclude by outlining the contributions the dissertation proposes to make to the literature.

3.2 Origins of the Baseball Salary Determination Literature

Although the neoclassical school of thinking espoused a marginal revenue product theory long before, the earliest such literature applying this line of thinking to baseball was work by Rottenburg (1956). He highlighted that the existence of the reserve clause gave baseball owners potential monopsony power over players, thus suggesting they were receiving wages far below their marginal revenue product (MRP). Contrastingly, in a free labour market, neoclassical wage theory suggests wages should be consistent with a players MRP. Therefore Rottenberg questioned the extent of this exploitation, as well as the most appropriate method of testing for it.

These questions in turn inspired the first wave of econometric analysis of the baseball labour market. Most notably, Scully's (1974) piece in the American Economic Review approached the problem through comparing estimates of players' actual MRP with the returns to performance reflected by their wages. He first regressed team revenue on team win totals to find the effect of team performance on revenue, then regressing win totals on individual player performance to estimate the actual MRP of players. This was then compared to actual returns to performance, found through modelling the determinants of salary. In doing so, he found that salaries from 1968-69 were at least five times lower than their marginal revenue products, thus substantiating Rottenburg's predictions.

Scully's findings generally held until the introduction of a free market for baseball players in 1976, constituting a large structural change for salary determination. Most importantly, a free market for baseball players could theoretically end this monopsony exploitation, by causing a convergence between MRP and wages for those eligible for free agency. Further, such a change in process could potentially lead to a change in salary distributions as well as the individual determinants of wages.

Thus two broad approaches emerged to investigate the effects of structural change on the baseball labour market. The first such approach, championed by Zimbalist (1992), sought to again compare players' wages to estimates of their MRPs. He argued that Scully's original method overestimated actual MRPs as he didn't take into account the possibility that another player would have still produced some performance output in their place. Thus Zimbalist compared the effect of individual performance with that of an average replacement, then following Scully's procedure. In doing so he found salaries to be virtually in line with MRP estimates and thus concluded that monopsony exploitation had almost entirely subsided a decade after the advent of free agency.

The other main approach to investigate structural change abstracted away from MRP estimation and instead focussed on modelling players' salaries on data before and after the change. For instance, papers such as Gius and Hylan (1996) corroborated Zimbalist's findings, with positive and significant structural change dummies reflecting that salaries had undergone a large increase on the free market. Fort (1992) established this salary rise was through greater returns to performance, with such determinants seeing large increases in magnitude in his post 1976 model. In terms of distributional effects, Fort further recognised that due to only certain players being eligible for free agency (see figure 2), the structural change actually lead to greater salary inequality as only more experienced players saw significant increases in their performance compensation.

Ultimately, the investigation of structural change gave impetus for the modelling of salaries in the style of Scully (1974). This wave of literature developed a fundamental modelling approach, whereby free agent salaries are determined primarily by performance factors such as hitting and pitching ability, but also secondarily by non-performance factors which are not directly related to measures of players' output (Krautmann and Oppenheimer, 2002). These may consist of signing team characteristics such as the location of a team, or player specific characteristics such as race and agent representation. The review will continue by examining developments in the literature in these areas.

3.3 Developments in Capturing Productivity

Given that the introduction of free agency had brought salaries more or less in line with players' marginal products, questions then arose regarding how to best capture performance when modelling salary. The process in baseball is very different to a mainstream labour context, where productivity is usually proxied by education and experience for instance. Here, the large variety of performance metrics and statistics presents a number of different options in the modelling process. Thus this section will detail how the approach to modelling performance factors has evolved considerably since its origins.

Scully's (1974) salary model has provided the general basis for the selection of performance variables in baseball salary studies. Scully first established that hitters and pitchers need separate salary models, given their separate on-field roles and thus separate production functions. The next question is how to best capture this productivity or output. He first proposed that playing time may have an important effect on salary, as those who feature more throughout the season contribute more to team performance and thus receive higher salaries. This is embodied in his models by variables for plate appearances (*PA*) for hitters and innings pitched (*IP*) for pitchers. Next, there is a need to establish how well a player actually performed during these playing time opportunities. This is where there is the most debate surrounding the best way to capture ability.

In terms of measuring hitter ability, Scully emphasised the use of the slugging statistic (*SLG*). This is favoured over the batting average (*AVG*) variable used in Pascal and Rapping (1972), as *SLG* accounts for players that hit the ball further and thus round more bases. For pitchers, ability is chiefly represented by strikeouts to walks (*SW*), the ratio of times a pitcher strikes out a batter to the times he walks them over a season. Yet, Scully's pitching model initially received criticism from Zimbalist (1992), who argued the *SW* statistic presented a poor representation of a pitchers productivity. Instead, he suggested the use of earned run average (*ERA*), which signifies runs conceded by a pitcher per game. Zimbalist found this measure better explained team winning percentage as it further accounts for pitchers who get hitters out through forcing them into weak hits, rather than striking them out.

Further, Scully failed to account for a variable denoting a players fielding ability, which as figure 1 outlines, is still a key component of a hitter's productivity. Both Kahn (1993) and Marburger (1994) have sought to include a measure of fielding ability into their salary models (Brown and Jepsen, 2009). However, each study found the fielding variable to be of contrasting statistical significance, which may have been testament to the fact that both papers used subjective measures of fielding ability that were based on judges opinions of players' fielding ability. As Law (2017) outlines, such measures are poor indicators of actual ability at best, and thus there is a need to use an objective variable when capturing fielding skill.

Inspired by observations in Lewis' (2003) *Moneyball*, the focus of the literature moved to the study of inefficiencies in the pricing of certain skills of baseball players. As Lewis illustrated, the contribution of walks to offensive productivity was largely ignored, meaning that players with high on-base percentages (*OBP*) had salaries below their relative contributions to team performance. This therefore created questions in regards to whether other teams had followed the Oakland Athletics and started to value the skill of getting walks. If so, the *OBP* variable could be used in future salary modelling. Hakes and Sauer (2006) and Brown et al. (2017) therefore tested this Moneyball hypothesis, finding wage determination adjusted rapidly to take account of *OBP* following the book's publication. As early as the 2004 season *OBP* became a significant explanatory variable, therefore reflecting that the determinants of baseball salaries can adjust quickly to changes in valuation of certain skills.

Whilst Moneyball was concerned primarily with hitting variables, Bradbury (2007) questioned the use of pitching variables in the pre-existing literature. In particular, he questioned the *ERA* variable given that its use is inconsistent with the assumption in salary equations that there are no externalities to performance. Although as Krautmann (1999) outlines, players’ marginal products are mostly separable, a certain portion of the runs given up by a pitcher is dependent on the quality of the fielders around them. This creates further issues as previous variables may not have provided an accurate representation of pitcher’s true abilities. Thus Bradbury suggested the use of variables such as walks, strikeouts and homeruns given up, that do not involve fielders’ input. He further showed that these variables have better determined pitcher salaries than variables such as *ERA* from 1986 to 2004.

The most recent advancement in approach to capturing performance in salary equations is the use of wins above replacement (*WAR*). The variable attempts to sum up all aspects of a player’s performance into one single figure representing their marginal product. Similar to Zimbalist (1992), this is then compared to the marginal product of a replacement level player that can be freely acquired by teams from reserves (Law, 2017). Further, this crucially addresses the problem of separate wage equations for hitters and pitchers, as the *WAR* variable uses the same units for both the groups, wins. Yet, despite the promise of such an all-encompassing measure, only Wasserman and Paul (2015) and Krautmann (2018) have made use of it in modelling the determinants of salaries.

3.4 Studies on Team Factors

A significant non-performance issue identified in the literature is the team a player signs for, and whether this has an effect on the player’s salary. In particular, the financial strength of a given team and its location may have a theoretical impact on free agent salaries. Scully (1974, p.925) first granted a role to the market or metropolitan area (*SMSA*) a team locates in. Interestingly, he conceded that inclusion in his analysis was the result of a “common notion” that players in larger markets were higher paid, with said phenomenon then unexplained by economic theory.

Ultimately, Scully did not find market size to be a statistically significant determinant of salaries, but his inclusion of the variable did lead to further analysis by Cymrot (1983). He posited that teams located in large and growing metropolitan areas had comparatively higher marginal revenues, and thus the revenue generating effects of increased performance were greater for these teams. Therefore, players signing for teams with higher marginal revenues would receive higher salaries. Using a dataset comprised exclusively of free agents, Cymrot did find *SMSA* to be a statistically significant determinant of salaries in the late 70s. Further, studies by Burger and Walters (2003) and Brown and Link (2008) have verified that large discrepancies in marginal revenue exist between teams. However, Burger and Walters reasoned that large market teams may instead be more likely to bid for a given player, rather than offering them higher salaries than their performance level would suggest.

Despite this, work by Hakes and Sauers (2006) and Healy (2008) suggested that certain teams valued various abilities differently. If valuation methods were linked to financial resources or revenue for instance, then the identify of a given team might have an effect on the skills they compensate in salary offers. Thus these papers provided renewed promise for the use of a market size variable in salary regression, with Brown and Jepsen (2009) investigating this specifically. Unlike previous literature, they adopted hierarchical linear modelling so that determinants of salary could vary between teams in their analysis. Yet they found the valuation of certain performance variables did not vary by team, and further that salaries did not vary in accordance to team revenue. Therefore, they reasoned that large market teams may sign a greater number of players in free agency, rather than paying more for individual performance.

Alm et al. (2012) investigate another factor, the marginal income tax rate (*MTR*) in the state a team resides, in order to establish whether this too has an effect on salary offers. Bearing in mind that baseball free agents are extremely mobile, they theoretically should bear little of the burden of an income tax. Therefore, Alm et al. proposed that teams in low tax states have a labour market advantage through not having to compensate signings for higher income tax rates. They found the inclusion of the team specific *MTR* variable to be positive and significant in explaining free agent salary, with a 1% increase in state marginal tax rate leading to a $21,000 dollar rise in annual salary. Thus they consider the location of the signing team to an important determinant of a player's salary.

3.5 Wage discrimination in Baseball

The issue of racial wage differentials has been of particular interest in the literature on baseball salaries. Since the process of desegregation in baseball began in 1947 with Jackie Robinson breaking the colour barrier, there have been questions over whether wage discrimination still remains in league, and if so, to what extent. This, coupled with the ability to account for performance in much greater clarity than other industries has therefore motivated a wide range of studies.

Much of the research on wage discrimination in baseball is founded on Becker's (1957) seminal work on the economics of discrimination, which outlines that discrimination may originate through employer, employee and customer led channels. Indeed discriminatory wage differentials may exist in the short run through any of these channels. However, in the long run, employer and employee led discrimination is likely to wane as the costs of such discrimination in terms of lost revenue will likely negatively harm the team (Kahn, 2000). For instance, Gwartney and Haworth (1974) provided empirical evidence that segregated teams in 1950s were worse off in terms of on field performance than those that started to integrate African American players. Conversely, customer led wage discrimination could theoretically continue into the long run as discriminating fans continue to keep profits afloat by supporting their discriminatory team (Holmes, 2011). Therefore, most studies in baseball assume away short run employer and employee led discrimination, and the focus is very much on customers as the source of potential wage differentials.

Indeed Scully (1974) found early evidence to show that teams with more African American players received lower ticket revenues. Moreover, perhaps the most notable investigation of customer discrimination is Nardinelli and Simon's (1990), which used the case study of the baseball memorabilia market. They found baseball cards for African American pitchers and hitters sold for 13% and 10% less than their White counterparts respectively in 1970, implying that baseball fans preferred White players. If such an attitude continued to permeate through to ticket demand, the costs of employing a given African American player would be higher and thus teams could suffer financially if their playing roster consisted of more non-white players.

Despite this, through adding dummies representing a players race to a standard performance based salary equation, studies investigating wage discrimination found little evidence of any racial differentials in the 80s and 90s. For instance, both Kahn (1991) and Medoff (1995) found African American race dummies to be wholly insignificant in a number of specifications. Meanwhile, Christiano (1988) even found evidence to suggest that white players were under-compensated in relation to minorities. Together, this led Kahn (1992) to declare that salary discrimination had for the most part faded within the baseball labour market.

However, such conclusions were based on standard OLS regressions that ignore the possibility of wage differentials outside the mean of the salary distribution. As Palmer and King (2006) illustrate, this may be troublesome if salary discrimination is deemed to differ by salary group. They suggested that wage discrimination may be more prevalent in lower echelons of the salary distribution, as the costs of discrimination in terms of on field performance are lower. Further, this is consistent with a customer discrimination approach if a fixed view is adopted, whereby the benefits of discrimination to fans do not differ with performance level. Palmer and King do find evidence to substantiate this fixed view, with African American and Hispanic wages lower than those of White players at low levels of the salary distribution.

On the other hand, Hamilton (1997) instead proposed that the benefits of discrimination to customers actually rises with performance levels, as high performing star players will be more visible to fans. Thus the psychic benefits of discrimination to fans here would be greater and would instead reflect that discrimination is more likely for highly paid, top performing players.

More recently, Holmes (2011) has utilised modern quantile regression techniques to address the issue of discrimination varying with salary level. His findings mirror those of Palmer and King, consistent with a fixed view of customer discrimination. Indeed dummies for both White and Hispanic hitters were found to be positive and significant at lower salary quantiles. Further, the magnitude of this discrimination was estimated to be large, with a differential of more than 20% in wages found between minorities and white hitters in the bottom 20% of the salary distribution.

3.6 The Marginal Value of Agents

The literature surrounding the marginal value of agents is currently very narrow, with only one such paper by Wasserman and Paul (2015) attempting to estimate their value in the sports industry. Akin to Krautmann (2018), they used the *WAR* variable to control for performance, then adding dummies for players belonging to certain agents to their salary model. In all, they found 5 separate agents to have a statistically significant effect on player salaries, including super-agent Scott Boras. These effects are primarily chalked down to differences in negotiating ability amongst baseball agents.

3.6 Conclusion

The literature has come a long way since the initial debate surrounding monopolistic exploitation, taking Scully's salary estimation techniques and applying them to a number of different issues. The capture of performance has evolved considerably since the beginnings of the literature, aided by Lewis (2003) popularising new ways of looking at performance. Further, the investigation of discrimination in particular saw a revival due to the use of inventive methodology to explore the possibility of prejudice on the lower margins.

Despite the advances in approach to modelling salary, literature investigating salary determination in baseball using data from the last 7 years has been sparse. Therefore it has failed to exploit the even greater wealth of data generated in that period in regards to performance especially. Consequently, this study seeks to incorporate new, untested performance variables into the modelling of baseball salaries.

Once a suitable performance foundation is established in the salary model, the pertinence of team factors used in studies such as Cymrot (1987) and Alm et al. (2012) will be explored. Following this will be investigation of whether player agents hold statistically significant marginal values and thus should be considered a factor in baseball salary determination. Both such issues have not been investigated using a recent sample of data. Finally, the issue of discrimination will be reinvestigated using Holmes' (2011) quantile approach, this time including pitchers in the sample, to shed light on whether discrimination still exists for lower salary groups.

4 Data and Methodology

4.1 Introduction

This section will start by outlining the form and collection of the data used in the study, followed by discussion of the salary models and econometric techniques to be employed.

4.2 Data Format

Whilst many studies of wages take a panel data approach, a cross-sectional dataset may actually be more appropriate to examine the determinants of baseball salary. Unlike other wage studies where the researcher may be interested in the effect of certain variables on wages over time, this is not the case when modelling free agency salaries. Here, only performance in prior seasons can be taken into account when a contract is offered and thus actual performance in subsequent years to signing is not of interest.

Despite this, if players were to sign multiple contracts over the sample period (2012 to 2019), then the nature of panel data could perhaps be exploited. Yet due to the fixed nature of longer term contracts, as well as the possibility of retirement for instance, this is relatively uncommon with 64% of players in the dataset signing only one contract in the sample period. Therefore, this extremely unbalanced panel would be econometrically problematic, with time-demeaning rendering the majority of the observations useless in a potential fixed effect model (Wooldridge, 2013). Furthermore, a fixed effect model would not allow investigation of time invariant factors that are of key interest in this study, such as agent and race. Accordingly, many papers modelling the wages of free agents such as Brown et al. (2017) and Krautmann (2018) employ a cross-sectional dataset.

However, a potential drawback of a cross sectional dataset is the existence of players signing multiple contracts during the period. Although performance and contract size differ for these repeat player observations, their inclusion could bias estimates if an unobserved characteristic was causing the relationship between their pay and performance to differ. Therefore a procedure is needed for picking a single contract of a given repeated player from the sample. This will be done randomly in order to prevent any selection bias from simply keeping the first or last season observation for instance. Holmes (2011) tests this procedure, noting that that the results were identical to the weighted least squared approach he implemented.

4.3 Data Collection

Due to the lack of a single centralised source for baseball data, the data used in this study has been taken from a variety of different sources. Observations of the dependent variable, annual average salary, were obtained from Baseball Prospectus' contract tracking hub, whilst all performance data was obtained from Baseball Info Solutions. Data for the rest of the non-performance explanatory variables were obtained from a number of different sources, listed in table 2.

4.4 Variable Specification

This Section will outline the variables used in this study, with all definitions and sources summarized in tables 1 and 2. Data for free agent contracts is in the form of annual average pre-tax salary in dollars. The sample encompasses 8 seasons, ranging from 2012 to 2019 and contains 488 contracts, excluding both repeat observations and instances where the player in question had no performance data in the prior year.

The observations outline the contract size and length (in years) for those filing for free agency with Major League Baseball and subsequently signing a contract. The sample consists purely of free agents as their contracts are negotiated on the free market through a fierce bidding process by agents, thus reflecting valuations by teams on their expected performance in the contract period (Krautmann, 1999).

In line with the literature, contracts are expressed in terms of dollars per year, due to the fact that they may vary significantly in length. Furthermore, it is commonplace in salary models to transform the salary variable into log form, in order to control for heteroskedasticity given that the salaries of high performing players tend to be more variable in nature (Holmes, 2011). Therefore, each model is log-linear, meaning that interpretations refer to the percentage change in salary given a 1 unit increase in an independent variable. Further, given this is in strong accordance with the baseball literature, use of this functional form will allow comparisons to be made.

All performance variables used are specified in table 1 on the next page. These are recorded for the season prior to a contract being signed, as offers are based on expected future performance given performance in the previous season (Hakes and Sauer, 2006).

Table 1: Performance Variable Specification

Group	Variable	Definition	Units
Dependent	log(Salary)	The logarithm of total salary per season	Dollars
-	EXP	Experience: The number of seasons of playing time a player has accumulated.	Years
Hitters	PA	Plate appearances: The number of times a player came up to bat during the season.	Plate appearances, in 100s
	AVG	Batting average: The proportion of opportunities a player successfully got on base through hitting the ball.	Percentage points
	SLG	Slugging: The total bases hit per the amount of times a player came up to bat.	Bases per at bat
	OBP	On-base percentage: The proportion of plate appearances that end with the hitter at first base. Importantly includes walks .	Percentage points
	wOBA	Weighted on base average: A weighted combination of SLG and OBP to provide a more accurate picture of batting ability.	Percentage points
	Bsr	Baserunning rating: Rating denoting how effectively a player is able to run round the bases.	Runs
	Def	Defensive rating: Objectively captures the fielding ability of a player. Uses past data to determine how difficult each fielding opportunity was, then comparing this to the outcome of their fielding actions.	Runs
Pitchers	IP	Innings pitched: Number of innings pitched, with an inning consisting of 3 outs.	Innings pitched
	ERA	Earned run average: Number of runs given up by a pitcher per 9 innings.	Runs
	SW	Strikeout to Walk Ratio: Ratio of strikeouts (3 strikes) to walks (4 balls) given up by pitcher.	Ratio
	Closer	Control dummy denoting a pitcher who's role is to pitch the last inning of the game.	-
	xFIP	Expected fielding independent pitching: Number of runs given up per 9 innings, accounting for luck and fielding externalities.	Runs
	WPA	Win probability added: Accumulation of how much a player's actions changes their team's odds of winning over a season.	Probabilistic wins
Pooled	WAR	Pooled variable which expresses a players marginal product in terms of team's wins they contributed to. For hitters: (Batting Runs + Base Running Runs + Fielding Runs + Positional Adjustment + League Adjustment +Replacement Runs) / (Runs Per Win) For pitchers: (Pitching Runs Prevented + Positional Adjustment)/ (Runs per win)	Wins

Source: Fangraphs (n.d.)

The description of team and player specific variables can further be found below in table 2. Team characteristics refer to demographic factors such as market size as well as state tax rate. The *AGENT* variable refers to series of dummy variables each representing an individual agency with over 8 clients in dataset. Following both Palmer and King (2006) and Holmes (2011), race data was collected primarily from the Lahman Database as well as player pages from both Baseball Reference and MLB. *RACE* is a series of dummy variables referring to whether a player is White, Hispanic or Asian. In the sample there are 289 White players, 20 Asian, 140 Hispanic and 39 African Americans. Summary statistics for all variables can be found in table 3 on the following page.

Table 2: Non-Performance Variable Specification

Group	Variable	Definition	Source
Team Factors	MTR	Top bracket marginal income tax rate for state in which signing team is located	Tax Foundation
	SMSA	Size of population in metropolitan area where the signing team is based in 2018. A proxy for market size.	US Census data
	GRO	Growth of SMSA from 2010 to 2018	
Player Factors	AGENT	Dummy variables, = 1 if a player's contract was negotiated by a certain agent. An agent must have negotiated over 8 contracts to qualify for a dummy.	MLBTR
	RACE	Series of dummy variables, = 1 if race is White, Hispanic or Asian	Lahman database Baseball Reference MLB
Time trend	TIME	Time trend variable. If contract signed in 2012, Trend = 0 If contract signed in 2013, = 1 . . If contract signed in 2019, = 7	Baseball Prospectus

Table 3: Summary Statistics

Variable	Observations	Mean	Standard Deviation	Minimum	Maximum
Salary	488	7,195,036	6,318,829	500,000	34,416,666
Log(Salary)	488	15.38	0.965	13.122	17.354
EXP	488	7.98	2.646	1.05	19.042
Trend	488	3.35	2.378	0	7
WAR	488	1.33	1.309	-1	6.46
PA	232	422.73	170.935	6	712
SLG	232	0.420	0.071	0.174	0.690
AVG	232	0.260	0.033	0.116	0.346
OBP	232	0.330	0.035	0.197	0.423
wOBA	232	0.325	0.0377	0.178	0.430
Bsr	232	-0.438	3.368	-8.4	10.7
Def	232	-1.781	8.034	-20.5	26.5
IP	256	95.324	61.001	2.1	227
SW	256	3.033	1.424	0	11.448
ERA	256	3.674	1.389	0	10
xFIP	256	3.932	0.751	1.92	9.04
WPA	256	0.503	1.458	-3.18	6.76
Closer	256	0.101	0.302	0	1
PIT	488	0.525	0.500	0	1
MTR	488	6.661	4.184	0	13.3
SMSA	488	6,894,519	5,183,784	1,576,113	19,979,477
GRO	488	6.527	5.535	-1.339	18.190
Scott Boras	488	0.078	0.268	0	1
Aces	488	0.078	0.2682	0	1
Jet Sports	488	0.018	0.134	0	1
Mike Moye	488	0.016	0.127	0	1
White	488	0.592	0.492	0	1
Asian	488	0.041	0.198	0	1
Hispanic	488	0.287	0.452	0	1

4.5 Salary Modelling

4.5.1 Overarching Salary Model

To investigate the determinants of modern free agent salaries, a series of wage equations will be estimated following the broad salary model represented by figure 3 below (Adapted from Krautmann, et al., 2002).

Figure 3: The Broad Baseball Salary Model

$$LOG(SALARY_i) = f\left(E\left(PERF_i\right), TEAM_i, PLAYER_i\right)$$

This study will first focus on the relationship between expected performance and salary, $E\left(PERF_i\right)$, through estimating a series of models using OLS regression for both hitters and pitchers. The significance of a series of team characteristics ($TEAM_i$) will too be tested, followed by player characteristics ($PLAYER_i$), such as race and agent representation.

Before presenting the salary models, it is necessary to detail a few assumptions preceding the analysis. It is assumed for the moment that hitters and pitchers cannot be aggregated into the same sample, as their roles and thus performance variables are wholly different. Moreover, players are assumed not to exhibit productivity externalities. Whilst these may exist in reality, traits such as leadership are intangible in nature. In addition, the modelling crucially assumes teams are trying to maximise their win totals and are thus profit maximising.

4.5.2 Performance Models

The first salary model (1) is shown below, with HIT_PERF_i a vector of performance variables denoting the productivity of hitters. This vector includes both traditional and modern variables.

$$LOG(SALARY_i) = HIT\_PERF_i + EXP_i + EXP_i^2 + Trend_i + e_i \qquad (1)$$

The traditional variables used for hitters will be *PA, SLG, AVG* and *OBP*, all of which have featured heavily in prior literature. The first untested hitting variable is *wOBA*, which attempts to weight the *SLG* and *OBP* statistics into a more accurate measure of batting ability (see figure 1). Further, a baserunning rating variable (*Bsr*) will be tested, to account for this second aspect of hitters' productivity.

In addition, the Fielding rating variable (*Def*) will be included. The variable refers to the amount of a runs a player either saved or conceded through their fielding ability, and is computed through comparing the outcome of a fielders actions to historical data on how well other players did in similar situations, meaning that it is an objective metric. Finally, *WAR* will also be included, to establish whether the use of such an all-encompassing measure is detrimental to overall fit. It is also worth noting that due to potential multicollinearity, traditional and modern variables will be tested in separate regressions for hitters.

Equation (1) also includes *EXP*, EXP^2 and *Trend*, to account for the effects of both experience and time on salary. As Krautmann and Oppenheimer (2002) outline, the justification for including experience here is that less experience means a smaller data sample for team executives to base signing decisions on, making such decisions more risky. Further, there is also a negative quadratic term as high experience may indicate a player is older and nearing the end of their career. Therefore they may be more injury prone, leading to smaller salary offers reflecting this greater risk. Similar to Krautmann (2018), a time trend denoting the season the contract was negotiated is also included to account for the gradual rise in revenue over the sample period. However, trend effects are not strictly of interest in this study, so the variable will not be of main interest in the results and discussion section.

The salary equation for the pitcher salary model (2) can again be seen below, with PIT_PERF_i referring to a vector of pitching performance variables.

$$LOG(SALARY_i) = PIT\_PERF_i + EXP_i + EXP_i^2 + Trend_i + e_i \qquad (2)$$

The traditional variables tested will be *IP, SW* and *ERA*, whilst the first untested variable included will be *xFIP*, which builds on work by Bradbury (2007). Such a variable attempts to give an indication of the runs a pitcher would give up over nine innings if they had average fielders behind them. Therefore it is similar to the *ERA* variable, but corrects for potential externalities that good or bad fielders may confer. Further, an 'expected' version of the *FIP* variable is used here. This essentially means the variable tries to strip out luck in regards to the outcomes of balls hit in the air against pitchers. Indeed removing aspects of performance related to luck may give a better indication of an underlying level of performance that can be expected (Slowinski, 2010).

The second untested variable is *WPA* or win probability added, which looks at a team's win probability before and after each plate appearance. *WAR* will again be included with the motivation the same as it was in the hitting model. Yet in line with Krautmann et al. (2003), a control dummy will be added here to account for different types of pitchers that either start or finish the game. This effect is normally captured by the playing time variable, *IP*, as those that start the game usually play most of it. However, when including *WAR*, the *IP* variable becomes redundant due to the fact that the *WAR* variable also captures playing time and is thus cumulative.

The hypothesises are the same for both hitting and pitching performance models, being that the modern, untested variables detailed will mark an improvement over traditional variables in explaining salaries. The criteria to test this will be the significance and robustness of these variables, as well as the R^2 of the individual models. An accepted hypothesis may indicate that these newer variables represent skills that are valued in the baseball labour market.

4.5.3 Team Characteristics Model

Once a suitable performance base ($PERF_i$) has been chosen, team specific variables can be added to the salary model, giving equation (3).

$$LOG(SALARY_i) = PERF_i + SMSA_i + GRO_i + MTR_i + e_i \qquad (3)$$

The individual significance of the $SMSA_i$ and GRO_i variables will be be tested to shed light on whether teams with large and growing markets, and thus higher marginal revenues, do indeed pay higher salaries. It is also worth noting that some studies use a revenue variable instead of a marginal revenue proxy such as market size. In reality, these variables are similar, yet $SMSA_i$ is chosen here as its inclusion has a stronger theoretical basis. Further, the significance of the MTR_i variable will shed light on whether teams in low tax states still have the advantage of being able to offer lower salaries, in accordance with Alm et al. (2012). The hypothesis tested here is that team characteristics are still a factor in salary determination, accepted if at least one of the team characteristic variables proves to be significant and robust.

4.5.4 Player Factors: Measuring the marginal value of agents

The marginal value of an agent refers to their ability to better negotiate contracts for their clients when both performance and team factors have been accounted for. In order to investigate agent value, dummy variables representing the agent who negotiated each contract will be added to the existing salary model. This will only be done for players with agents who have more than eight clients in the sample.

However, as Wasserman and Paul (2015) emphasise, in order to estimate the marginal value of agents, there is a need to verify that certain agents are not just representing the best performing players. Thus they suggest regressing performance on a series of agent variables to establish whether performance means are more or less consistent across agents. This is where the *WAR* statistic is very useful, as a single statistic for performance enables its use an a dependent variable in equation (4a) below.

$$WAR_i = AGENT_i + e_i \tag{4a}$$

The significance of a given $AGENT_i$ dummy would suggest that the average performance of their clients is statistically different from the reference group. If the majority of these individual variables are deemed insignificant, we can then go ahead with the analysis of agents in the salary model as shown in equation (4b).

$$LOG(SALARY_i) = PERF_i + TEAM_i + AGENT_i + e_i \tag{4b}$$

Indeed a positive, significant coefficient would reflect that the agent in question can extract a greater return to client performance. Thus the null hypothesis here is that at least 1 $AGENT_i$ dummy is considered positive in (4b), thus indicating that agents hold some marginal value and thus have a statistically significant effect on salary.

4.5.5 Player Factors: The Use of Quantile Regression to Investigate Discrimination

As shown by equation (5) below, dummy variables representing White, Hispanic and Asian players will be included in the salary equation, in order to investigate potential discrimination against African American players. Given that this is the final salary model, the results of relevant diagnostic tests for this model will be presented and discussed in section 5.5.

$$LOG\left(SALARY_i\right) = PERF_i + TEAM_i + AGENT_i + WHITE_i + HISPANIC_i + ASIAN_i + e_i \tag{5}$$

The above equation will be first estimated using OLS, giving an indication of racial wage differences around the mean. A dummy coefficient will have to be significant and robust in nature if discrimination is evident here. However, in line with Holmes (2011), quantile regression can be used in order to better investigate the possibility of discrimination. Whereas OLS regression is focused on the effect of independent variables on a dependent variable at its mean, quantile regression allows this effect to be investigated at different quantiles of the dependent variable.

Figure 4: Comparison of Coefficient Estimation in OLS and Quantile Regression

<u>OLS:</u>

$$min \sum_{i=1}^{n} (y_i - x_i\beta)^2$$

<u>Quantile Regression:</u>

$$min \sum_{i=1}^{n} Q\left|(y_i - x_i\beta)\right| + (1 - Q)\left|(y_i - x_i\beta)\right|$$

As shown in by figure 4, OLS estimates coefficient values through minimising squared residuals. On the other hand, quantile regression obtains coefficient values through minimising the sum of absolute errors below and above a given quantile (Q), which is equivalent to the quantity of positive and negative residuals (Koenker and Hallock, 2001). By solving the second equation, the values of coefficient estimates at (Q) can be found.

Like OLS, the significance of a race dummy will simply reflect a significant effect on salary at a given quantile, but quantile regression can also reveal whether this effect is evident at high and low salary quantiles. In accordance with the literature, this analysis will assume away short run employer and employee led discrimination channels. Holmes (2011) finds a pattern of discrimination consistent with a customer source with a fixed benefit, where wage discrimination falls as salary rises. However, evidence from Hamilton (1997) suggests that discrimination may instead rise with salary level due to greater visibility to fans at higher levels. Therefore the null hypothesis here will be that the coefficients on race dummy variables differ with the quantile estimated. If the dummies do appear significant then a spearman's rank correlation coefficient can be estimated to establish the direction of this correlation. Equation (5) will be estimated at quantiles of Q = 0.1, 0.25, 0.5, 0.75 and 0.9, following the suggestion by Hamilton (1997) that such quantiles are less likely to pick up effects of outliers at each end of the salary distribution.

5 Results and Discussion

5.1 Performance Models

5.1.1 Hitter Performance Model

First and foremost, it is worth noting that all models failed the Breusch Pagan test for heteroskedasticity (see table 11) and thus robust standard errors have been used in the estimation of each equation. Table 4 presents the results for the hitter models, with the first column detailing the results for the first salary model (1a) based on traditional performance variables used in the literature.

The variable accounting for playing time, *PA*, is positive and significant as expected. Further, the *SLG* variable, the most widely used hitting variable in the literature, is a positive and significant determinant of salary in a modern sample. Interestingly, the *AVG* variable, largely cast aside as a determinant in the literature, is found to be significant at the 5% level. Although this particular result may reflect that it is being used to value players again in the present day, as shown by column 1b, this is likely because it was picking some of the effect of the *OBP* variable. Indeed, with the inclusion of the *OBP* variable, batting average now becomes insignificant. Thus although *AVG* may appear to be significant in explaining salary, this does not necessarily imply it is a skill valued by team executives, as statistical significance is likely through correlation to the *OBP* variables. Considering this, the *OBP* variable appears to mark an improvement over *BA* in determining salary, confirming that the findings of Hakes and Sauer (2006) still hold for a contemporary dataset of salaries.

The third column illustrates estimates of a salary model based on variables previously untested in the literature; *wOBA*, *Bsr* and *Def*. Due to fears of multicollinearity, such variables have not been included with those previously discussed as the mean variance inflation factor (VIF) of a regression including every hitting performance variable would have been very large, at 33.60.

Table 4: Results for the Hitters' Performance Model (1)

	Model			
	1a	1b	1c	1d
Constant	10.81*** (0.470)	10.0*** (0.557)	9.73*** (0.547)	13.89*** (0.3267)
Exp	0.119 (0.073)	0.132 (0.0816)	0.116 (0.083)	0.126* (0.069)
Exp^2	-0.0068* (0.0035)	-0.00765* (0.0039)	-0.00663 (0.00409)	-0.00755** (0.00355)
Trend	0.066*** (0.018)	0.06528*** (0.0181)	0.0622*** (0.0174)	0.0632*** (0.0169)
PA	0.288*** (0.028)	0.286*** (0.0279)	0.295*** (0.0287)	
SLG	0.0393*** (0.0069)	0.0294*** (0.00697)		
AVG	0.0388** (0.016)	0.00460 (0.0194)		
OBP		0.0630*** (0.0175)		
wOBA			0.115*** (0.012)	
Bsr			0.00561 (0.0110)	
Def			0.0184*** (0.00437)	
WAR				0.513*** (0.0273)
R^2	0.61	0.63	0.65	0.65
Observations	232	232	232	232

Standard errors are in parentheses.
* is significant at 10% level, ** at 5% level, *** at 1% level

A key finding can be found in the results surrounding the *Def* variable, which accounts for the effect of fielding or defensive ability on salary. As outlined in the literature review, other explanatory variables attempting to take into account fielding ability have often been found insignificant. However, using this objective measure of fielding ability appears to have a significant effect, with each extra run saved through fielding ability signifying a 1.8% increase in salary. Further, this result is robust to inclusion with traditional performance variables, whilst the overall implication here may be twofold. It is possible traditional fielding variables did not provide an accurate representation of actual fielding ability in the past, or rather teams have only started to value this ability in salary offers as of the last 7 years.

The *wOBA* variable, providing a more accurate representation of batting productivity, is also significant. Given that it essentially weights SLG and OBP into a single variable, its significance is unsurprising. In contrast, the *Bsr* variable, accounting for the ability to more effectively run round the bases, is insignificant in model 1c. Thus the evidence here may potentially imply that such abilities aren't considered as important components of expected performance by teams.

Column 4 further details the results of replacing performance variables with the *WAR* variable. As expected it is positive and significant, with a one standard deviation increase in marginal product leading to an 80% increase in salary. This result is very similar to Kruatmann's (2018) result of 75% for the 2012 to 2014 seasons, reflecting that returns to performance may have increased only marginally in the 5 years since. Further, with an R^2 of 0.65 identical to model 1c, the explanatory power of the model is not reduced by amalgamating productivity variables into a single variable.

5.1.2 Pitcher Performance Model

Table 5 details the results from the estimation of pitching models. Starting with model 2a, the traditional performance variables are again seen to be statistically significant determinants of modern salaries. The significance of the *IP* variable again reflects that salaries rise with playing time, whilst the significance of the *ERA* and *SW* variables seemingly support both Scully (1974) and Zimbalist's (1992) arguments for inclusion. The negative sign on *ERA* is to be expected, as a pitcher giving up more runs is less productive and thus should receive a lower salary.

Table 5: Results for the Pitchers' Performance Model (2)

Variable:	Model 2a	2b	2c
Constant	14.35*** (0.507)	15.1*** (0.333)	14.6*** (0.112)
Trend	0.0851*** (0.0183)	0.08350*** (0.0175)	0.0864*** (0.0164)
EXP	-0.0313 (0.0937)		
EXP^2	0.00179 (0.00488)		
IP	0.0926*** (0.00566)	0.0910*** (0.00538)	
SW	0.164*** (0.0359)	0.0623* (0.0344)	
ERA	-0.140*** (0.0374)	-0.0312 (0.0412)	
xFIP		-0.265*** (0.0569)	
WPA		0.129*** (0.0322)	
Closer			0.444*** (0.109)
WAR			0.516*** (0.0372)
R^2	0.56	0.61	0.62
Observations	256	256	256

Standard errors are in parentheses.
* is significant at 10% level, ** at 5% level, *** at 1% level

Furthermore, the insignificance of *EXP* and EXP^2 in models (1) and (2) deserves some additional discussion. This insignificance is not necessarily uncommon in the baseball salary literature, especially when using a dataset comprised of free agents. Krautmann et al. (2002) suggest this may be due to the fact that only experienced players are eligible for free agency (see figure 2), and thus any risks associated with lack of experience are much less of an issue for these players. Therefore, going forward, the experience variables are excluded from the modelling process.

Column two illustrates the results of including my contributions to the literature, *xFIP* and *WPA*, to the previous salary equation. In contrast to the set of batting models, these were added in along with the more traditional variables, due to reduced fears surrounding multicollinearity. Yet the inclusion of the two new variables does noticeably change the coefficient estimates and standard errors of *SW* and *ERA*. This newfound insignificance could in one way be interpreted as the variables being of little importance in modern salary determination, but given their significance in the first column, this type of interpretation could potentially be misleading.

Despite this, of main interest in table 5 are the untested *xFIP* and *WPA* variables, which are both significant in model 2b. Given that *xFIP* is on the same scale to *ERA*, in terms of runs conceded per 9 innings, it can be compared to *ERA* in 2a. The coefficient on *xFIP* indicates that an extra expected run conceded per 9 innings reduces salary by 27% here, compared to 14% for ERA. Therefore the salary return to this measure of pitching ability is clearly much larger, reflecting that it may indeed better explain modern salaries as expected. This result is one of the main findings in regards to performance variables and builds on work by Bradbury (2007) who advocated fielding independent variables in salary regressions. Further, given that the variable is an expected statistic that attempts to strip out luck in regards to on field outcomes, its significance may suggest that teams are strongly valuing variables that specifically try to build a better picture of what a players performance should be in the future.

The significance of the *WPA* variable perhaps indicates that teams also value the ability to pitch at times where the game is on the line and actions have a greater impact on win probability. However, it is worth noting that unlike *xFIP*, this measure is not robust to inclusion in a regression with additional variables. Thus this modern variable cannot be considered a reliable salary determinant.

Model 2c is estimated in the same vein as model 1d, with the results largely similar. Indeed the coefficient of 0.516 is almost identical to 0.513 for hitters in table 4. Meanwhile, it is clear again that the individual use of the WAR variable has similar explanatory power to a series of individual variables, whilst further suppressing any fears regarding multicollinearity.

Ultimately, in regards to the performance models described in equations (1) and (2), the hypothesis that newer and untested variables mark an improvement over traditional ones can be accepted. This is due to the findings in regards to the *Def* and *xFIP* variables in particular, as well as the favourable characteristics of *WAR*. Unlike the traditional variables, the significance of these is robust, whilst where comparison is appropriate, salaries seem to greater compensate these measures. This may indeed suggest that salaries now reflect compensation of skills reflected by these variables, such as fielding ability for hitters and fielding independent ability for pitchers.

However, as demonstrated by the results of estimating equation 1a and 2a, traditional variables may still constitute an appropriate model of modern salaries. Furthermore, the overall gain in fit from using modern variables is marginal, with R^2 only improving by 4% and 6% for hitters' and pitchers' models respectively. Thus the improvement can only be considered marginal from a modelling perspective, whilst multicollinearity potential creates issues when concluding whether such variables are actively used by team executives in valuing certain skills. Further, moving forward in the analysis of non-performance factors, *WAR* will be used with a control dummy denoting whether the player is a pitcher (*PIT)*. This is primarily due to the fact that *WAR* enables pooling of all players into one sample, with both models (1d) and (2c) confirming similar returns to this measure of marginal product for both hitters and pitchers.

5.2 Team Factor Model

Table 6 presents results of the pooled regression with both performance and team factors included as explanatory variables. Again we can see that *WAR* is a significant determinant of salary for both hitters and pitchers, with a positive and significant *PIT* dummy indicating that pitchers on average receive more than hitters when controlling for marginal product. In contrast, the three team factor variables prove to be insignificant in model 3, with this result robust to alternative specifications as well as splitting up hitters and pitchers.

Table 6: Results for the Team factor Model (3)

Model	3
Constant	14.3*** (0.0930)
WAR	0.542*** (0.0210)
PIT	0.247*** (0.0560)
MTR	0.00342 (0.00717)
SMSA	0.000323 (0.000708)
GRO	-0.00316 (0.00546)
Trend	0.0746*** (0.012)
R^2	0.60
Observations	488

Standard errors are in parentheses. * is significant at 10% level, ** at 5% level, *** at 1% level

Firstly, The insignificant *MTR* variable suggests that teams in low marginal tax states no longer experience the labour market advantage that Alm et al (2012) found for a dataset ranging from 1995 to 2001. This may be because such state marginal tax differentials do not actually influence player decisions regarding contract offers. For instance, players are only taxed on income in the state in which they earn it, so are only exposed to their own teams state tax rate when playing home games. Further, state tax rates are normally dwarfed in size by federal marginal tax rates, which are 37% for top bracket earners such as baseball players. This is in comparison to state tax rates which average only 6.66%. In all, these results would question the advantages of being able to offer comparatively lower wages in these areas.

The insignificance of the *SMSA* and *GRO* variables again appear in contrast to Cymrot's (1983) findings for free agents in the 1970s. This may perhaps reflect that metropolitan area statistics aren't necessarily the best proxy for team market size. For instance, there are multiple teams in cities such as Chicago and New York, with it hard to differentiate between boundaries of potential markets for tickets and televised games. An example of where this may be acute is the Chicago White Sox. Whilst the Chicago metropolitan area is the third largest in the US, the White Sox share it with the more popular Chicago Cubs. Consequently, their average attendance was only 21,442 in 2019, ranked 23rd out of 30 teams (ESPN, 2020).

However, the lack of significance here is more likely a reflection that large market teams tend to sign more players during free agency, rather than pay over the market rate for a given player (Brown and Jepsen, 2009). Meanwhile, the marginal revenue status of a team may mean they are more likely to attempt to sign a particular high performing player, but again this is not to say that they pay a premium for a given level of performance. Therefore, with all team factors insignificant in table 6, the hypothesis that team factors constitute a non-performance determinant of salaries is rejected. Consequently, team factors are excluded from further non-performance models in this study.

5.3 Player Agent Model

Table 7 presents the results for regressions investigating the marginal effects of players agents in baseball. The first column illustrates results of the equation verifying that the performance means of certain agents are not significantly different to the reference group (those agents with less than eight clients). Indeed only two agent dummies are found significant at the 5% level when regressed on *WAR*, indicating for the most part that specific agents don't simply represent the best performing players. Therefore the analysis can continue in line with Wasserman and Paul (2015).

Column 4b presents the results when salary is regressed on both *WAR* and agent dummies. Dummies for Scott Boras, ACES (levinson brothers), JET sports and Mike Moye are all positive and significant at the 1% level, suggesting that they negotiate their clients higher salaries than their level of expected performance would indicate. As outlined by Halvorsen and Palmquist (1980), the interpretation of dummy variables are slightly different in a log-linear context. For Scott Boras, despite the 0.40 coefficient in 4b, his marginal value over agents with less than 8 clients is actually 49%.

Table 7: Results for the Agent Model (4)

	4a	4b
Dependent:	WAR	Log(Salary)
Constant	1.27***	14.2***
	(0.100)	0.0756
WAR		0.539***
		(0.0210)
PIT		0.257***
		0.0563
Trend		0.0674***
		(0.0123)
Scott Boras	0.469**	0.404***
	(0.233)	(0.124)
ACES	-0.217	0.313***
	(0.233)	(0.0990)
BH Sports	-0.481	0.111
	(0.350)	(0.156)
CAA	-0.0515	0.197
	(0.269)	(0.122)
Excel	0.708***	0.156
	(0.250)	(0.117)
GSE	-0.0478	0.176
	(0.265)	(0.114)
ISE	-0.0422	-0.0912
	(0.244)	(0.117)
MVP	0.442	0.186
	(0.322)	(0.160)
Wasserman	-0.0406	0.167
	(0.265)	(0.132)
Octagon	0.467	0.182
	(0.307)	(0.134)
Apex	-0.139	-0.0280
	(0.470)	(0.208)
Jet Sports	0.216	0.436***
	(0.444)	(0.140)
Magnus Sports	-0.186	0.0632
	(0.422)	(0.184)
Meister Sports	-0.716	0.162
	(0.469)	(0.274)
Mike Moye	0.156	0.446***
	(0.469)	(0.150)
R^2	0.049	0.62
Observations	488	488

For comparison, Wasserman and Paul also find a Scott Boras dummy to be statistically significant from 2002 to 2012, but with a slightly more conservative estimate of his individual marginal value at 34%.

Further, if a player belonging to an agent in the reference group had their contract negotiated by any of the four agents previously mentioned, the results imply their salaries would have been on average 48% higher. Although these marginal values do seem surprisingly large, it may be the case that these agents charge a higher fee for their services. Unfortunately, such data on agent fees is not publicly available.

This large salary premia may be attributed to differing negotiation abilities amongst agents. If agents can influence team judgements over expected performance, then larger salaries may be offered. Further, different agents tend to have different market strategies, with Scott Boras notorious for delaying all salary negotiations until the end of the offseason period where teams may become more desperate to fill gaps in their playing rosters and thus become more impatient regarding player acquisition.

In all, due to the significance of at least one agent dummy, the hypothesis that certain agents can tangibly effect the salary of their clients is accepted. This finding may too have further implications for the entertainment industry for instance, where agents are prevalent and performance may be difficult to capture and thus hold constant.

5.4 Investigation of Salary Discrimination

The results of the OLS regression testing for discriminatory wage differentials are seen in table 8 below. Although some variables are excluded from the table for conciseness, the three race dummies were regressed with the same performance variables used in models (3) and (4), alongside the four significant agent dummies.

From table 8, it can be seen that none of the race dummies have a significant effect on salary at the 5% level. This indicates that salary discrimination is not evident at the mean of the salary distribution.

Table 8: OLS Race Dummy Estimates for Model (5)

Variable:	5
White	0.108 (0.109)
Asian	0.0184 (0.160)
Hispanic	0.111 (0.110)
R^2	0.62
Observations	488

Standard errors are in parentheses. * is significant at 10% level, ** at 5% level, *** at 1% level

Table 9 illustrates the results of estimating race dummies using quantile regression, with bootstrapped standard errors used in order to account for heteroskedasticity. It is clear again that none of the race variables are significant at any quantile.

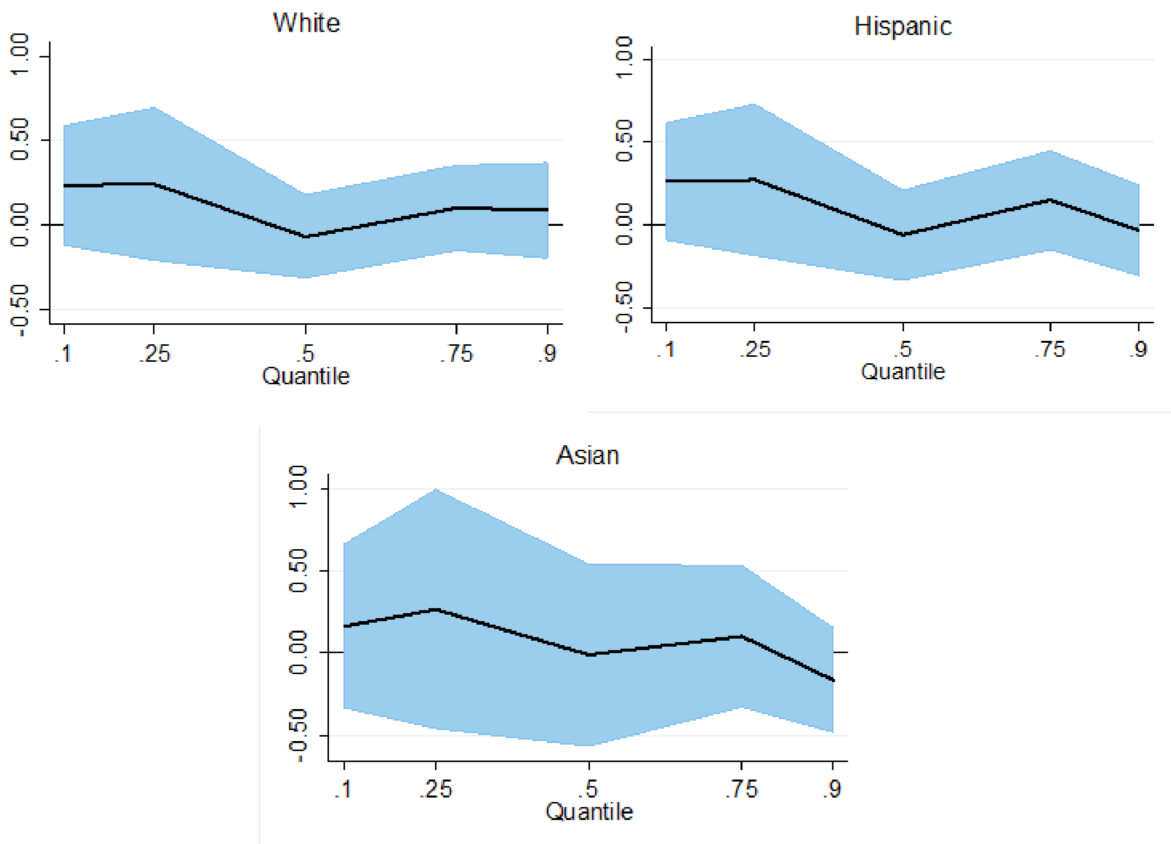
Table 9: Quantile Regression Estimates for model (5)

	Dummy		
Quantile	White	Asian	Hispanic
0.1	0.237 (0.179)	0.164 (0.254)	0.264 (0.180)
0.25	0.244 (0.230)	0.267 (0.370)	0.273 (0.233)
0.5	-0.0654 (0.124)	-0.0122 (0.282)	-0.0588 (0.138)
0.75	0.104 (0.128)	0.102 (0.219)	0.151 (0.152)
0.9	-0.164 (0.162)	0.0889 (0.142)	-0.0319 (0.138)
Observations of Race in the sample	289	20	140

Standard errors are in parentheses.
* is significant at 10% level, ** at 5% level, *** at 1% level

Figure 5 graphically illustrates the coefficient estimates at each quantile. Although the coefficient estimates do exhibit some downward trend over the period, the blue confidence interval overlaps zero at every quantile, meaning that none of the variables can be considered statistically significant. Therefore the hypothesis that discrimination differs with salary quantile is rejected, as each dummy estimate cannot be considered significantly different from zero at any quantile.

Figure 5: Graph illustrating Race Dummy Estimates by Quantile



This result differs considerably from recent studies surrounding racial wage discrimination in baseball by Holmes (2011) and Palmer and King (2006), who both find econometric evidence of racial discrimination at lower levels of the wage distribution. The contrasting results here may reflect a difference in both sample and model used. The sample in this study was much broader than those previously used, including both hitters and pitchers, as well as Asian players. Yet robustness tests revealed no change in results both when the sample was split and when Asian players were excluded.

Further, from a modelling perspective, *WAR* was used instead of individual hitting performance variables by Holmes. This may suggest these regressions were unable to pick up certain individual abilities, such as baserunning, which Holmes highlights may differ with race. Therefore, equation (5) is again estimated via quantile regression, using variables in 2c as $PERF_i$. However, the coefficients of race dummies are once again found to be statistically insignificant at each quantile, reflecting this may not be the source of difference.

Further, given that they used much older samples (up to 2006), it is possible that these racial wage differentials may have dissipated since. The source of this may depend on whether a fixed or variable view of the benefits of customer discrimination is adopted. If benefits are assumed to be fixed, a fall in discrimination at lower levels of the salary distribution may reflect that the marginal costs curve has shifted outwards. Namely, the costs of discrimination may have increased for players of all performance levels. This could be due to a rise in revenues for instance, which can be seen to effect salaries through the positive and significant time trend in each model. Thus if costs even at the lowest levels of the salary distribution now exceed the fixed benefit, discrimination would no longer theoretically be seen.

Alternatively, if a variable benefit approach is adopted in similar vein to Hamilton (1997), the result here may reflect that the benefits to discrimination may have actually fallen since Holmes' sample ended in 2006. Indeed a fall in the benefits of discrimination would reflect that baseball fans' preference for discrimination has waned. Despite this, evidence from the political science literature suggests that prejudice towards African Americans has not actually declined since Obama took office. Indeed Yadon and Piston (2018), through looking at voting patterns of the representative White American, find that opposition to African American public figures and pro African American policies has not changed since 2008. Yet to apply these findings here would assume the average baseball fan was equivalent to the representative White American, which may be far from the case.

Thus overall, the hypothesis that wage discrimination differs at different levels of the salary distribution is rejected, as no evidence of salary discrimination in baseball is found. The source of this difference may not be distinguishable without making assumptions in regards to the variability of benefits in a customer discrimination framework. In regards to generalisations, Holmes (2011) suggests that findings regarding discrimination in baseball may be applicable to other industries where employees have more or less same function, with large discrepancies in ability between them.

5.5 Diagnostic Testing

Although diagnostic tests are seldom discussed in the literature, a few are needed to check that assumptions of OLS are verified in the salary model. Table 10 outlines the relevant diagnostic test results for OLS estimation of the final salary model (5). Starting with heteroskedasticity, the salary model failed the Breusch Pagan test indicating that residuals cannot be considered to be of constant variance. As previously mentioned, this has been remedied though the use of robust standard errors in the estimation of every model. Thus OLS estimates can still be considered efficient. Further the Jarque Bera test p value is greater than 0.05, indicating the normality of residuals can be accepted. This is again the same for each model estimated, meaning that hypothesis tests can be considered reliable. This is very important, as the significance of explanatory variables is central to the conclusions of this dissertation.

Table 10: Results of Select Diagnostic Tests for the Salary Model

	Test		
	Breusch-Pagan	Jarque-Bera	VIF
Model (5)	0.001	0.117	1.61

Note: Statistics for Breusch-Pagan and Jarque-Bera test are p-values

The final column of table 10 outlines the mean variance inflation factor (VIF) for the final model. Although no formal hypothesis test for multicollinearity exists, the VIF result indicates that the final model is less likely to be subject to multicollinearity. Caution has been taken to keep VIFs low in the performance models especially, but instances where it may have been an issue have been discussed accordingly. Overall, given that robust standard errors have been used, the assumptions of OLS can be verified for the salary model.

6 Conclusion

6.1 Summary of Main Findings

This study aimed to establish the effect of various performance and non-performance factors on modern baseball salaries. The main contribution to the literature surrounding performance variables is the use of the *xFIP* and *Def*. Both were found to be significant and robust in their respective models, deemed to mark an improvement over more traditional variables employed by Scully (1974) and Zimbalist (1992). This improvement sheds light on the appeal of using both expected pitching variables and objective fielding measures in future salary studies for pitchers or hitters. Furthermore, this may suggest that the skills captured by these variables are valued by team executives submitting salary offers for players. Yet this finding is made with the appreciation that traditional variables are still appropriate determinants of salaries, and thus in reality the modelling gains to be made from using modern variables as determinants of salary are somewhat slim.

In regards to non-performance factors, additional regressions beyond the base salary models revealed some of these factors to be of little importance in the salary determination process. The identity of the signing team seemingly has no statistical effect on the overall salaries of free agents. This is in congruence with Brown and Jepsen (2009), reflecting that large market teams' labour market advantage is embodied in their ability to sign a greater quantity of players. Meanwhile, low income taxes are not found to confer a labour market advantage for teams situated in certain states.

Further, the investigation of player agents finds evidence to corroborate Wasserman and Paul's (2015) finding that agents have a statistically significant effect on client earnings through negotiating ability. This further suggests the need to account for player representation as a non-performance factor when modelling salary determination in baseball. In addition, this may reflect that choice of agent is important in industries where performance may be harder to capture, such as the entertainment industry.

Meanwhile, a quantile regression approach was utilised to investigate potential racial discrimination beyond the mean of the salary distribution. Using a wider dataset than previous studies, all race dummies were found to be insignificant in the salary model, pointing to an absence of wage discrimination in the baseball industry. These results appear in contrast to more recent evidence from Holmes (2011), but provide renewed support for Kahn (1992). This is attributed to a decline in customer discrimination, but the source of this decline is unidentifiable without making further assumptions about the nature of customer discrimination. Ultimately, these results indicate that the salaries in this modern sample are almost entirely determined by performance, with only the identity of a player's agent identified as a statistically significant non-performance wage factor. Moreover, such non-performance findings may be applicable to labour markets where labour is very mobile and employee function is very similar within firms.

6.2 Limitations and Future Research

As eluded to in the results section, perhaps the largest limitation of this study surrounds the multicollinear nature of certain variables in models 1 and 2. Given that some performance variables are trying to capture certain abilities in different ways, there can inevitably be some correlation between them. Consequently, OLS may struggle to separate out the effects of these correlated variables and so may fail to hold them constant before varying others. This has been mitigated by using WAR in non-performance models as well as running separate performance regressions. Yet at times this was at the expense of being able to compare the significance of determinants directly in the same regression. Thus it can be hard to come to conclusions on whether certain skills are explicitly valued by teams in salary bids, and at times statistical significance may be through correlation to another variable. Despite this, the main performance findings described are robust to any combination of performance variables used, unlike traditional variables.

Furthermore, the analysis in regards to agents was most vulnerable to the exclusion of contracts for players who appear more than once in the sample (see section 4.2). Indeed the exclusion of a given well negotiated contract does have the potential to tangibly effect conclusions in regards to which particular agents might hold marginal value. Despite this, random resampling has revealed the significance of individual agents to be relatively consistent amongst different random samples, whilst the conclusion that player agents constitute a significant non-performance factor remains unchanged. Future research in this area may look to verify these findings in other sports. Although it may be slightly harder to capture performance in football for instance, the modelling of salaries here could further provide insights into the marginal value of sports agents.

In Addition, the performance models developed in this study could be used to gain a further understanding of another non-performance factor, players' attitude towards risk. Prior literature from Krautmann and Oppenheimer (2002) and Link and Yosifov (2012) has already verified that there exists a trade-off between length of contract and annual salary for hitters. However, there exists no empirical evidence in regards to these compensating differentials for pitchers and therefore models 2b or 2c could be used to investigate this. Yet econometric issues surrounding endogeneity mean that such a study would require an instrument variable, likely in the form of a proxy for time spent out injured. At this time no reliable injury database exists, so such a study could be undertaken when one becomes available in the future.

Reference List:

Alm, J., Kaempfer, W. & Sennoga, E. (2012) ‘Baseball salaries and income taxes: The ‘home-field advantage’ of income taxes on free-agent salaries’, Journal of Sports Economics, vol. 13, pp. 619–634.

Anderson, R. (2019) Scott Boras clients give agent billion dollar offseason. [online] Available from: https://www.cbssports.com/mlb/news/scott-boras-clients-give-agent-billion-dollar-offseason-after-hyun-jin-ryus-blue-jays-deal/ [accessed 16th March 2020]

Becker, G. (1957) *The Economics of Discrimination*, 1st ed., Chicago: University of Chicago Press.

Bradbury, J. C. (2007) ‘Does the baseball labor market properly value pitchers?’, Journal of Sports Economics, vol. 8, pp. 616-632.

Brown, D. & Link, C. (2008) ‘Population and bandwagon effects on local team revenues in Major League Baseball’, Journal of Sports Economics, vol. 9, pp. 470–487.

Brown, D., Link, C. & Rubin, S. (2017) ‘Moneyball After 10 Years: How Have Major League Baseball Salaries Adjusted?’, Journal of Sports Economics, vol. 18, pp. 771–786.

Brown, K., & Jepsen, L. (2009) ‘The impact of team revenues on MLB salaries.’, Journal of Sports Economics, vol. 10, pp. 192-203.

Brown, M. (2019) MLB Sees Record $10.7 Billion In Revenues For 2019. [online] Available from: https://www.forbes.com/sites/maurybrown/2019/12/21/mlb-sees-record-107-billion-in-revenues-for-2019/ [accessed 16th March 2020]

Burger, J. & Walters, S. (2003) ‘Market size, pay, and performance: A general model and application to Major League Baseball’, Journal of Sports Economics, vol. 4, pp. 108–125.

Christiano, K. (1988) ‘Salaries and Race in Professional Baseball: Discrimination 10 Years Later’, Sociology of Sport Journal, Vol. 5, pp. 136-149.

Congdon-Hohman, J. M. & Lanning, J. A. (2018) ‘Beyond Moneyball: Changing Compensation in MLB’, Journal of Sports Economics, vol. 19, pp. 1046–1061.

Cymrot, D. (1983) ‘Migration trends and earnings of free agents in major league baseball’, Economic Inquiry, vol. 4, pp. 545–556.

ESPN (2019) MLB Attendance Report – 2019. [online] Available from: http://www.espn.com/mlb/attendance/_/year/2019 [accessed 16th March 2020]

Fangraphs (2020) Fangraphs Library. [online] Available From: https://library.fangraphs.com/ [accessed 16th March 2020]

Forbes (2019) 2019 top sports agents. [online] Available from: https://www.forbes.com/sports-agents/list/#tab:overall [accessed 16th March 2020]

Fort, R. (1992) ‘Pay and performance: Is the field of dreams barren?’ In P. M. Sommers (Ed.), *Diamonds are forever: The business of baseball*, pp. 134-160, Washington, DC: Brookings Institution.

Gius, M., & Hylan, T. (1996) ‘An interperiod analysis of the salary impact of structural changes in Major League Baseball: Evidence from panel data’, In J. Fizel, E. Gustafson, & L. Hadley (Eds.), *Baseball economics: Current issues*, pp. 77-84.

Gwartney, J. & Haworth, C. (1974) ‘Employer costs and discrimination: The case of baseball. Journal of Political Economy’, vol. 82, pp. 873-881.

Hakes, J. K. & Sauer, R. D. (2006) ‘An economic evaluation of the moneyball hypothesis’, Journal of Economic Perspectives, vol. 20, pp. 173-185.

Halvorsen, R. & Palmquist, R. (1980) ‘The Interpretation of Dummy Variables in Semilogarithmic Equations’, The American Economic Review, vol. 70, pp. 474-475.

Hamilton, B. (1997) ‘Racial discrimination and professional basketball salaries in the 1990s.’ Applied Economics, vol. 3, pp. 287–296.

Healy, A. (2008) ‘Do firms have short memories? Evidence from major league baseball’, Journal of Sports Economics, vol. 9, pp. 407-424.

Hill, J. & Spellman, W. (1983) ‘Professional baseball: The reserve clause and salary structure’, Industrial Relations: A Journal of Economy and Society, vol. 22, pp. 1-19.

Holmes, P. (2011) ‘New evidence of salary discrimination in major league baseball’, Labour Econ. vol. 218, pp. 320– 31.

Kahn, L. (1991) ‘Discrimination in professional sports: a survey of the literature’, Industrial and Labor Relations Review, vol. 3, pp. 395–418.

Kahn, L. (1992) ‘Discrimination in baseball.’ In P. M. Sommers (Ed.), *Diamonds are forever: The business of baseball,* pp. 109-133, Washington, DC: Brookings Institution.

Kahn, L. (1993)’ Free agency, long-term contracts and compensation in Major League Baseball: Estimates from panel data’, Review of Economics and Statistics, vol. 75, pp. 157-164.

Kahn, L. (2000) ‘The sports business as a labor market laboratory’ Journal of Economic Perspectives, vol. 3, pp. 75–94.

Koenker, R. & Hallock, K. (2001) ‘Quantile regression’, Journal of economic perspectives, vol. 15, pp. 143-156.

Krautmann, A. & Oppenheimer, M. (2002) ‘Contract length and the return to performance in major league baseball’, Journal of Sports Economics, vol. 1, 6–17.

Krautmann, A. (1999) ‘What’s wrong with Scully’s estimates of a player’s marginal revenue product?’, Economic Inquiry, vol. 37, pp. 369-381.

Krautmann, A. (2018) ‘Contract Extensions: The Case of Major League Baseball’, Journal of Sports Economics, vol. 19, pp. 299–314.

Krautmann, A., Gustafson, E. & Hadley, L (2003) ‘A note on the structural stability of salary equations: Major league pitchers’, Journal of Sports Economics, vol. 4, pp. 56-63.

Law, K. (2017) *Smart Baseball: The Story Behind the Old STATS That Are Ruining the Game, the New Ones That Are Running It, and the Right Way to Think about Baseball*, New York, NY: HarperCollins

Lewis, M. (2003) *Moneyball: The art of winning an unfair game*, New York: W.W. Norton and Company, Inc.

Link, C. & Yosifov, M. (2012) ‘Contract length and salaries compensating wage differentials in Major League Baseball’, Journal of Sports Economics, vol. 13, pp. 3-19.

Major League Baseball (n.d) Service Time. [online] Available from: http://m.mlb.com/glossary/transactions/service-time [accessed 16th March 2020]

Marburger, D. (1996) ‘A Comparison of Salary Determination In the Free Agent and Salary Arbitration Markets’, In J. Fizel, E. Gustafson, & L. Hadley (Eds.), *Baseball economics: Current issues*, pp. 67- 76.

Marburger, D. R. (1994) ‘Bargaining power and the structure of salaries in Major League Baseball. Managerial and Decision Economics’, vol. 5, pp. 433-441.

Medoff, M. (1975) ‘A reappraisal of racial discrimination against blacks in professional baseball’, The Review of Black Political Economy, vol. 5, pp. 259-268.

Nardinelli, C. & Simon, C. (1990) ’Customer racial discrimination in the market for memorabilia: the case of baseball’, Quarterly Journal of Economics, vol. 3, pp. 575–595.

Palmer, M. & King, R. (2006) ‘Has salary discrimination really disappeared from major league baseball?’ Eastern Economic Journal, vol. 2, pp. 285–297.

Pascal, A. & Rapping, L. (1972) ‘The Economics of Racial Discrimination in Organized Baseball,’ in A. H. Pascal (Ed.), *Racial Discrimination in Economic Life*, pp. 119-156, Lexington:

Rosen, S. & Sanderson, A. (2001) ‘Labour Markets in Professional Sports’, The Economic Journal, vol. 111, pp. F47-F68.

Rottenberg, S. (1956) ‘The Baseball Players’ Labor Market’, Journal of Political Economy, vol. 64, pp 242-258.

Scully, G. W. (1974) ‘Pay and performance in major league baseball’, American Economic Review, vol. 64, pp. 915-930.

Slowinski, S. (2010) xFIP. [online] Available from: https://library.fangraphs.com/pitching/xfip/ [accessed 16th March 2020]

Wasserman, T. & Paul, R. (2015) ‘Estimating the Marginal Value of Agents in Major League Baseball’, New York Economic Review, Vol. 46, pp. 3-21.

Wooldridge, Jeffrey M. (2003) *Introductory Econometrics: A Modern Approach*, Mason, OH: Thomson South-Western.

Yadon, N. & Piston, S. (2019) ‘Examining whites’ anti-black attitudes after Obama’s presidency’, Politics, Groups, and Identities, vol. 7, pp. 794-814.

Zimbalist, A. (1992) ‘Salaries and performance: Beyond the Scully model’, In P. M. Sommers (Ed.), *Diamonds are forever: The business of baseball,* pp. 109-133, Washington, DC: Brookings Institution.