2004

Racial Discrimination in Major League Baseball: Can it Exist When Productivity is Crystal Clear?

Will Irwin '04
Illinois Wesleyan University

Racial discrimination in Major League Baseball, can it exist when productivity is crystal clear?

Will Irwin
Department of Economics, Illinois Wesleyan University
Honors Thesis, Spring 2004

Minority workers in many industries have dealt with wage and hiring discrimination. In Major League Baseball, it is typically assumed that fair hiring practices are used and that fair wages are offered due to the intense negotiation process involved with each player contract. However, past research shows that both wage and hiring discrimination has been present during some recent time periods. With most industries, racial discrimination is difficult to measure due to the implicit bias in the productivity variables. However, in baseball, productivity can be easily measured with offensive and defensive performance statistics, which are determined completely independent of race. This paper attempts to use a human capital theory based regression and chi-squared analysis to determine whether or not racial wage and hiring discrimination is present among outfielders in Major League Baseball. The key findings are that offensive production and games played are the determinants of salary, while race is not and that racial hiring discrimination is not present for the data used.

I. Introduction

Major League Baseball (MLB) players make millions of dollars a year to play a game. The players come from all over the world to the United States' largest cities to perform in front of millions of fans each year. From children in the Dominican Republic playing on gravel fields with broomsticks for bats, socks for balls, and milk jugs for gloves to children in the wealthy suburbs of southern California, little boys grow up dreaming of playing for their favorite team. Typically these dreams revolve around hitting a home run to win the World Series or making a diving catch to save a game, and for extremely underprivileged children the promise of large contracts to bring their entire

family out of poverty, provide extra inspiration to make it to the big leagues. But are their dreams tainted by the specter of discrimination?

American labor markets have a long history of racial discrimination. In labor markets, this discrimination often takes the form of pay discrimination and/or hiring discrimination. Affirmative action programs and equal employment opportunity laws have been instituted to ensure that minority races are not discriminated against during the hiring process. However, affirmative action programs and equal employment litigation are rare to nonexistent in professional sports, and little has been done by the government to guarantee that minority players receive pay equal to that of their white teammates.

Human capital theory demonstrates that, on average, workers should be paid according to their likely productivity. However, in many industries, productivity is hard to measure. Rather than measure productivity directly, researchers use proxies such as education, previous work experience, and on-the-job training. Since minorities have often been discriminated against their entire lives in education and training, it becomes very difficult to accurately measure discrimination in the work force. All of the proxies for productivity have implicit racial bias built into them, which blurs the data and makes any finding on discrimination biased. With baseball, the proxies are a more direct measure of productivity and do not fall victim to the problems found with using traditional proxies which are more indirect measures of productivity.

With Major League Baseball, productivity factors that would be included in a human capital theory based equation might include things such as home runs, runs batted in, batting average, years of major league experience, and defensive ability. When a white player hits 50 home runs it should theoretically be treated the same as when a

Latino player hits 50 home runs, which makes for variables that have no implicit bias. Another benefit of using MLB players to assess the state of racial wage discrimination is that many players, including some of the top players in the league, come from the poorest areas in Latin America, which shows that professional baseball is an industry with little class resistance. In many industries, where advanced levels of education are necessary, it is virtually impossible for a child from a poverty stricken country to attain employment in that industry. Examples might include doctors, lawyers, accountants, and business executives. Since MLB, as an industry, does not depend on education, but on athletic ability, a trait that can develop in even the poorest of circumstances, it does not fall victim to built in class biases.

By studying racial discrimination in MLB, I hope to show that it does not exist. Since baseball teams that discriminate would only end up hurting their chances of winning, it seems that there is a strong incentive to use fair hiring practices and to provide fair wages. Also, the strong influence of agents on contract negotiations should insure that all contracts are awarded solely on the basis of performance. I hypothesize that racial discrimination does not exist in MLB and that race is not a significant predictor of players' salaries after controlling for productivity. I also hypothesize that firms do not discriminate against minority players when making hiring decisions.

The following section, Section II, discusses some of the past research about discrimination in professional sports. Section III presents my theoretical model, and then Section IV develops my empirical model. Section V shows my variables and the regression model. Section VI describes my data sources and gives some critique of their usefulness. Section VII presents my results from the regression equations, and Section

VIII outlines the results of the chi-square analysis. Finally, Section IX discusses the conclusions reached from the results and any possible policy implications that may be needed.

II. Literature Review

In 1957, Gary Becker produced the seminal work for the economic study of discrimination, The Economics of Discrimination. The theory from his book will be discussed later in the Theoretical Model section. Many economists have used Becker's theories to perform empirical studies on various issues of discrimination.

Gwartney and Haworth (1974) examined hiring discrimination for the integration period of MLB between 1947 and 1956. They compiled a statistic called "black player years" for each team and then compared it to games won for the 10 year time span. The results show that the teams with the five highest number of black player years held five of the six highest spots in the rankings for games won. The New York Yankees was the only team able to rank in the top six while taking a discriminatory position during integration.

Medoff (1975) followed the research of Gwartney and Haworth by trying to discover what level of wage discrimination existed in MLB for the 1968 season. His model found that racial wage discrimination was not present for his data, but Medoff recognized that wage discrimination was not the only type of discrimination facing minority players.

Prior to 1947, an absolute barrier to entry existed for all nonwhite players. An important research question for the years that followed was whether or not barriers to

entry into MLB persisted after the color barrier was broken. Medoff recognized that even though owners might have paid the nonwhite players they have hired fair wages does not mean that they were hiring as many nonwhite players as deserved to be hired. In order to test this possibility, he examined whether or not nonwhite career statistics variables were significantly better than white career statistics. If nonwhites did perform better than whites, the fact that there were many more white players in the league than nonwhite would suggest that nonwhite players were not being hired as much as they should have been. The data were divided between the two leagues since the American League (AL) had integrated more slowly than the National League (NL). He postulated that the league with the faster entry rate should have statistically greater productivity for nonwhite players because the faster entry rate suggests that they have greater access to the better nonwhite players. The slower entry league would thus have a productivity for nonwhite players which would be closer to the true mean value, and if this value is not statistically different from that of the white players in that league, then it can be said that no barriers to entry exist (Medoff, 1975).

The results showed that the rates of entry were indeed significantly different showing that the NL did have better access to nonwhite players. The nonwhite players in the NL were significantly more productive than both the white NL players and the nonwhite AL players. Furthermore, the AL nonwhites were not significantly more productive than AL whites. These results show that no barriers to entry existed for nonwhite players in the AL for this data.

In an article by Bodvarsson and Pettman (2002), entitled "Racial wage discrimination in major league baseball: do free agency and league size matter?" pitchers

are used to determine whether or not racial wage discrimination exists. This study uses a logarithmic regression model with variables such as race (white or non-white), population of the player's metropolitan area, productivity variables, and racial make-up of the player's metropolitan area. Since the study dealt with pitchers, the performance variables are very different. Examples include: wins, losses, saves, strikeouts, and earned run average. The study attempts to find out whether or not the addition of more teams to MLB would eliminate any racial discrimination that existed. The theory was that more teams would increase the level of competition between teams for good players, thereby causing teams to offer fair wages to get the best players. The results of this study show that expansion in 1993 did indeed eliminate discrimination for pitchers.

A study by Fizel, Krautman, and Hadley (2002), "Equity and Arbitration in Major League Baseball," looks at a theory called "Equity Theory", which states that workers compare their efforts to their co-workers that are in comparable situations and then adjust their behavior based on that comparison. They hypothesize that MLB players look to players of similar abilities to decide how much they should earn when they file for arbitration and free agency. The study uses both regression and bivariate techniques to determine whether or not a player will decide to file for arbitration. The results indicate that players do look at their co-workers performance and salary when they apply for arbitration. They do not include a measure of race in their study. If they had, perhaps they might have found whether minority players make different decisions than white players on this topic, but by demonstrating that MLB players look at their peers' wages when negotiating, they suggest a mechanism for how market forces can drive the system to long run equilibrium where minority players receive equal wages to whites.

In the context of the current literature, it is extremely important to test for the presence of wage discrimination amongst MLB outfielders. Since no such research is currently available, this issue must be analyzed to ensure that minority players are being treated fairly. Additionally, in order to complete the study, the presence of hiring discrimination must also be tested for.

III. Theoretical Model

The theoretical basis for this research comes from Gary Becker's work on discrimination and from Human Capital Theory.

Becker's theory was based on the premise that discrimination was not about conspiracies against minorities or irrational business practices, but rather that discrimination arose as a result of the "tastes" of individual employers. In the case of baseball, if a team owner has a taste for discrimination against Latino players, he will experience a decrease in utility each time he signs a Latino player. This decrease in utility will be represented in two possible ways: a lower wage for Latino's or fewer Latino players on the team. The team owner with the taste for discrimination will then end up paying white players, of equal skill to the Latino, higher wages (Becker, 1971).

Becker's theory indicates that the owners with a taste for discrimination will be less successful than nondiscriminatory employers. The theory behind this is that in a free market league, a minority player will receive offers from both discriminatory and nondiscriminatory owners, and since the discriminatory owners have a negative utility for signing minority players, the wage they offer will be lower than the nondiscriminatory owner. Therefore, the nondiscriminatory owner will be able to sign minority players

more often than the discriminatory owner, giving him a competitive advantage in building his roster. Also, since Becker's theory suggests that minority wages will be lower than white wages, nondiscriminatory firms will have lower labor costs. This cost advantage also gives the nondiscriminatory firm a competitive advantage (Becker, 1971).

In the long run, lower labor costs put nondiscriminatory firms at an advantage over discriminatory firms. As discriminatory firms decrease in number and nondiscriminatory firms increase in number, the demand for minority players increases which causes their wages to increase. This process continues until wages are equal between white and minority players. Also, in the long run, nondiscriminatory employers will integrate their firms since white and black wages become equal for each skill level (Becker, 1971).

According to Becker, employers may also respond to employee and consumer prejudice in their hiring decisions. If a team owner in MLB is a profit maximizer, as he should be assumed to be, he will put players on the team that his fans want to watch and that his employees want to work with. We would expect that minorities would be underrepresented on those teams where these prejudices exist. If employers respond to consumer and employee prejudice, there will be lower demand for minority players and their wages will be lower. Thus, consumer and employee prejudice has the same short run implication as employer discrimination. In the long run however, Becker argues that competitive forces should still eliminate the effect of these prejudices just as they eliminate the effects of employer prejudices. Later in the paper, we look for evidence of consumer prejudice by exploring whether the distribution of players on teams does not vary across cities, even though those cities have vastly different racial compositions.

Becker's theory of employer discrimination indicates that hiring discrimination results from firms who have a negative utility for hiring minority workers. This negative utility will cause them to hire minority workers at a lower rate than nondiscriminatory firms. Therefore, on average, the discriminatory firms should have fewer minority workers.

Thus, in the short run, Becker's theory predicts the possibility of wage differentials and racial segregation. The work by Haworth and Gwartney cited above suggests that this is exactly what was happening in MLB during the 1950's.

However, Becker's theory of discrimination suggests very different outcomes in the long run. The theory states that both wage differentials and racial segregation should be eliminated by market forces (Becker, 1971). It has been 57 years since Jackie Robinson broke the color barrier in 1947, which should be enough time to reach a long run equilibrium. Thus, we would hypothesize that MLB has reached a long run equilibrium and that after controlling for differences in human capital between players there should be no remaining wage differentials or hiring discrimination.

Since it is essential to control for human capital differences between major league players in order to determine whether racial wage differentials exist, it is important to have a grasp of human capital theory. Human Capital Theory is used to explain how productivity factors of a worker's history influence how much they are paid in their jobs. "Human capital" is seen as the result of investment. People can invest in schooling in order to increase their level of human capital. Similarly, firms can invest in on-the-job training for their workers. Since education increases a worker's knowledge and skills that can be applied to a job, theory suggests that this increase in knowledge will allow the

worker to be more productive, which will give them the ability to bargain for higher wages (Rima, 1981). Firms will pay higher wages since the marginal productivity of their business will be increased with the addition of more human capital to their work force. Also, since investing in human capital can have costs, such as tuition for schooling and the opportunity cost of lost wages during the time period of schooling, employers can be forced to pay higher wages to people with more human capital (Mincer, 1958).

In a traditional work setting, the factors that go into human capital would include educational attainment level, years of work experience, on-the-job training, and age. All of these variables theoretically should give an indication of how productive a worker will be. Once an employer knows how productive the worker will be, he/she can pay them a wage that is fair for that level of productivity.

For baseball players, the factors that employers look at to determine salary are very different from the traditional job market. Since players are simply born with innate abilities, they are compensated for their natural ability to produce in an athletic industry. However, just as job experience is a form of human capital investment for traditional industries, the amount of time training and playing high-level professional baseball should be an important human capital determinant for baseball players. Athletic talent should be considered more of an amplifier of this training and experience than a form of human capital for baseball players.

Therefore, a human capital model for baseball will have some different factors, but the factors in the model will represent productivity in the same way that education, experience, training, etc. do in a traditional human capital model.

In conclusion, given Gary Becker's model of discrimination and the human capital theory of wage determination, I hypothesize:

1) According to Becker's theory of discrimination in the long run, there will be no statistically significant wage differences between white and minority players after controlling for productivity differences.
2) According to Becker's theory of discrimination in the long run, there will be no statistically significant differences in racial distribution among teams.

IV. Empirical Model

Hiring Discrimination

In order to determine whether or not baseball teams discriminate in their hiring practices, chi-squared analysis will be used to compare the distributions of minority and majority players across teams.

Wage Discrimination

In order to determine whether there is racial wage discrimination, dummy variables for race are included in a regression model that also controls for productivity differences across players.

Equation 1.

Wage = f (productivity, race)

Productivity depends on innate ability and a set of investments in human capital. In MLB, unlike other industries, productivity can be measured directly by performance statistics such as runs batted in, home runs, and batting average. Also important in determining productivity are experience related variables such as games played and age.

More experienced players may generate more revenue for the club independent of their productivity because they have loyal fans who are more likely to attend games and purchase other team sponsored goods.

When studying race as a determinant of wage compensation, it is important to account for individual differences in productivity due to the factors mentioned above in order to determine whether or not race has an effect. Since, theoretically, race alone should not affect a person's productivity, it would be expected that different races would receive equal pay, *ceteris paribus*.

In order to determine whether or not MLB players are receiving fair pay, race is included as one of the determinant variables. Other determinant variables account for traditional productivity factors. These include measures for offensive and defensive ability, experience, and reliability of future production.

It is important to note that MLB teams are located in major cities across the country which range widely in size from New York City, with a metropolitan area of over 20 million people, to Milwaukee, with a metropolitan area just over one million people (www.Census.gov). Since teams in smaller cities would have to pay just as much as teams in large cities to attract players, some may say that it is doubtful that a team in a small market can pay a player equally well as a large market team due to the large differences in total revenues. However, for the purpose of this study, it is assumed that a player would not accept a lower salary to play in a small market, and, therefore, that all salary figures are true values awarded for the level of productivity that is expected of the player. Given this assumption, no measure of market size is included in the model.

Conceptually, a players contract value could be thought of depending on several categories of variables:

Equation 2.

Player Contract Value = f (offensive ability, defensive ability, experience, race)

From this equation, I build a regression equation that I use to test my hypothesis that racial discrimination does not exist in MLB. This hypothesis combines implications of Gary Becker's theory of discrimination and ideas from Human Capital Theory. Recall that Becker's theory suggests that, in the long run, racial wage differences should disappear as a result of competitive forces in the labor market. Human Capital Theory suggests that wage differences between players should reflect differences in actual productivity. Thus, the first research hypothesis is that after controlling for differences in productivity, race will not be a significant determinant of player contract value. Also, Becker's theory of discrimination suggests that in the long-run teams should not be segregated. Thus, the second research hypothesis is that the racial distribution should not differ significantly across clubs.

V. Variables and Regression Model

The regression model for wage discrimination will take the pieces described in the theoretical model and apply them specifically to the research problem by using the available data. The data, which are described in detail in Section VI, consists of 52 outfielders who were eligible for free agency status and signed a new contract between 1997 and 2003.

The dependent variable is salary. The measure for this variable consists of a one-year contract value, the first year of a new contract. The number of years of the contract is not included. This is unfortunate because length of contract could be another variable in which discrimination could play a role. However, data for length of contract are not readily available.

The regression equation attempts to account for different aspects of a player's abilities. For the purpose of limiting the data and eliminating positional biases, only outfielders are used in the study. Typically, a player is judged by scouts and general managers on the basis of "five tools": (1) Hitting for average (2) Hitting for power (3) Ability to steal (4) Ability to catch (5) Ability to throw.

1. Offensive Statistics and Variables

Offensive productivity is typically measured by traditional statistics like batting average, runs batted in, runs scored, home runs, and steals. The problem with using such statistics is that they do not fully capture the ability of all players to contribute to their teams. For example, Player A plays for a team in which the players in front of him and behind him in the line-up all hit for high average and power, while Player B plays for a team in which the players immediately surrounding him in the lineup all hit for poor average and power. If Player A and Player B have exactly the same inherent offensive capability, but are playing with these different teammates, Player A will have higher runs scored and runs batted in numbers because there will be more men on base ahead of him to drive in and better hitters behind him in order to drive him in. Some may say that this effect will only bias the runs scored and runs batted in measures of performance, but it is

generally assumed and widely accepted that home runs and batting average can also be positively and negatively influenced depending on the players surrounding a player in the line-up. In order to insure that statistical studies of baseball hitters are not flawed due to this "team influence" a statistic has been created by baseball scholar Bill James called "Runs Created" (www.baseballstuff.com).

The formula for Runs Created has many different variations. It has been changed many times by its creator and also by other researchers that have used it. For this study, the most basic formula is used due to the availability of data from www.baseball-reference.com, which has already calculated the Runs Created numbers. The formula used is as follows:

Equation 3.

*Runs Created = (Hits + Walks) * Total Bases / Plate Appearances*

"Total bases" is defined as the number of bases reached by a player due to hits, so for a single: total bases = 1, a double = 2, a triple = 3, and a home run = 4. "Plate appearances" is the total number of times a player comes to the plate. It differs from "At bats" because "At bats" subtracts out walks, sacrifices, and hits-by-pitch. "At bats" is the statistic used to calculate batting average (www.sports-wired.com).

Runs Created is an attempt to estimate a player's total contribution to his team's total output. It tries to eliminate the "double-counting effect" that occurs when a run is scored. To explain, each time a run is scored, the player who actually crosses the plate gets credit for a "run scored", while the player that got the hit to drive the other player in

gets credit for a "run batted in." Statistically, the run is counted twice. Therefore, the players on a team that scores 700 runs in a season might have a sum of 1400 runs scored and runs batted in (minus special cases where runs score without a run batted in being awarded). In order to try to measure which players on each team are actually most responsible for the total offensive output, the Runs Created variable gives one player more credit than the other when each run is scored (www.wfu.edu). This principle is further illustrated with this hypothetical situation:

> *Player A hits a triple to lead off an inning. Player B sacrifices Player A home with a long fly ball to centerfield. In this situation, the statistical breakdown would be: Player A – 1 hit in 1 at bat for a 1.000 batting average and 1 run scored, Player B – 0 hits in 0 at bats for no change to his batting average and 1 run batted in (sacrifice fly outs do not count as an official at bat). If these players' statistics are compared, Player A would only receive a slight advantage over Player B in the batting average category. Their run production would be considered equal.*

This example of traditional stat-keeping shows how the player that "did the most" in the scoring of the run does not earn the credit that he might technically deserve. Runs Created solves this problem by giving "extra credit" to Player A by giving him 3 total bases and giving Player B 0 total bases.

For regression analysis, Runs Created appears to be a good alternative to traditional statistics as a measure of offensive productivity because it eliminates some of

the team bias that is implicit in the data. Also, by reducing the number of independent variables in the regression, using Runs Created saves valuable degrees of freedom.

In order to further normalize the variable, each players' total Runs Created, for the years before they sign their contract, is divided by the number of games they have played and then multiplied by 162 (the number of games in a full season). This should show their average ability to produce offensively.

The ability to steal is measured, obviously, with a steals statistic. The same procedure used with Runs Created is also used with steals. Each player's steals value shows the number of steals they have had on average per 162 game season. On a side note, this statistic could be extremely interesting due to the fact that some players' main purpose for being in a line-up is to steal bases. It will be interesting to see if these stealing specialists are rewarded for their rare talent.

2. Defensive Variables

The defensive productivity measures, such as the ability to catch and to throw, are a little more difficult to quantify. For the ability to catch, the most commonly used statistic is fielding percentage, but the problem with this statistic is that some slower players do not have the opportunity to attempt to make difficult plays and therefore end up having higher fielding percentages than the faster, flashier fielders. This problem is virtually unavoidable with the data, so fielding percentage will not be included in the regression.

The ability to throw can be easily quantified with outfield assists. An assist is awarded to a player when they throw the ball to another player and this action results in

an out, force or tag. For outfielders, this statistic is extremely important to their overall defensive ability. For the model, assists are included as a 162-game average, calculated in the same way as the offensive statistics described above.

3. Experience Variables

In order to account for the traditional Human Capital Theory variable of experience, the number of games played in the MLB league will be included in the regression equation. Since many players play a few games in a year before playing their first full year, it would be impossible to include a variable such as years of experience. Number of games played will clearly differentiate between young players and veterans, while also allowing all available statistics prior to contract signing to be used in calculating potential productivity.

An element of productivity that has nothing to do with the game of baseball but will still have an impact on the human-capital thought processes that general managers have when deciding how much money to give to a player is age. Often times, a younger player will be awarded a large amount of money over a long period of time, while an older player may receive a one or two year contract worth an extremely large amount of money. Since no research has been found on how age affects the awarding of salaries in MLB, it is uncertain what affect age will have in the regression equation. On one hand, older workers have more experience, even after controlling for games played. Thus, wages would be directly related to age. On the other hand, reflexes slow and the probability of injury increases with age. Thus, wages would be expected to be inversely related to age. Given these opposite effects, the sign on age is indeterminant.

4. Racial Variables

For this category of variables, dummy variables must be used. For both black and Latino players, a one or a zero is used to designate a player in either of those racial categories. Being white is the omitted variable. The placement of players into these categories was completed by judgments based on country of origin, physical appearance, and surname.

Variables used in the analysis are presented in Table 1. The following equation will be estimated using OLS:

$$Player\ Salary = \alpha_1 + \beta_1(RC) + \beta_2(ASSIST) + \beta_3(STEALS) + \beta_4(GP) + \beta_5(AGE) + \beta_6(BLACK) + \beta_7(LATINO)$$

Table 1: Variables and Descriptions

Dependent Variable	Description	Mean	Max	Min	Std. Dev.
Salary	Value of Contract	2,440,064.1	7,550,000	250,000	1,985,099.9
Independent Variables (Expected Sign)					
RC (+)	(Runs Created / GP) * 162	78.1	125.6	39.9	20.9
ASSIST (+)	(Assists / GP) * 162	9.1	16.7	2.8	3.8
STEALS (+)	(Steals / GP) * 162	17.3	61.6	1.0	13.7
GP (+)	Games played before contract was awarded	523.5	982	168	208.7
AGE (+ or -)	Age on July 1st of the year contract was awarded	27.9	34	23	2.3
BLACK (0)	1 = Black, 0 = Other	.37	1	0	.486
LATINO (0)	1 = Latino, 0 = Other	.33	1	0	.474

VI. Data

The sample consists of 52 outfielders who have achieved free agency status and signed a new contract between 1997 and 2003. Free agency status occurs after a player has been under contract to a major league team for six years. After those six years, as a free agent, the player has the option to negotiate with any team he wishes. Thus, free agent wages are determined purely according to market forces.

The wonderful thing about working with MLB data is that people are fanatical about keeping track of statistics about every aspect of the game. From offensive statistics to defensive statistics, to player salaries and free agent filings, all of the data found in this paper is readily accessible on reputable and reliable websites. It is important to note the ease of Internet research when working with baseball statistics. Efficiency is drastically increased due to the extreme quickness by which player data can be found on these websites. Additionally, definitions of statistics and explanations of how to interpret them are found easily on the Internet.

The data for offensive and defensive statistics can be found on www.Baseball-Reference.com. This website is basically an encyclopedia of baseball players and has records of almost every player that has ever played in the major leagues.

Data for player salaries can be found at a website run by the chairman of the Business of Baseball Committee of the Society for American Baseball Research (www.roadsidephotos.com). The site is particularly useful in determining when the players had received their salary payments as a result of a free agent signing because it has a list of all free agent filings made by players. Using this list, the year in which the

player entered the free agent market can be found and their performance data from the previous years can be used to determine on what basis their new contract was awarded.

The data for the race variable is less concrete. In many cases the player's country of origin (www.baseball-reference.com) will indicate into which category they should be placed. For example, all players from the Dominican Republic are considered to be Latino. The same can be said for players from any other country in Central and South America. For minority players born in the United States, the distinction is more difficult to make. Typically, a determination can be made based on the origin of the player's last name. A name of Hispanic origin would tend to indicate that the player is Latino. Conversely, a recognizable "American name" would categorize the player as black. Two cases where no determination could be made with some certainty are excluded.

The data for the hiring discrimination portion of the study was also found using www.Baseball-Reference.com. This website lists complete team rosters. From the rosters, the number of players of each race that a team has employed for a given year can be determined. The same method as was described before was used to determine the race of each individual player. For this test, player rosters from the 2000 season were used.

VII. Regression Results

For independent variables, the unstandardized coefficients, t-statistics, and significance values are presented below in Table 2.

Table 2. Coefficients and Significance

	B	Std. Error	*t*	*Sig.*
(Constant)	-1,319,631	3,248,449.0	-.406	.687
RC	43,127.6	13,601.4	3.171	.003
ASSIST	-21,341.3	67,442.0	-.316	.753
STEALS	-520.7	17,834.0	-.029	.977
GP	5,447.6	1,324.3	4.114	.000
AGE	-90,453.0	108,941.0	-.830	.411
BLACK	555,729.0	525,330.0	1.058	.296
LATINO	202,386.1	506,230.1	.400	.691

$R^2 = .545$, n = 52

The two significant variables are Runs Created and Games Played. The rest are not significant with extremely low significance values.

Runs Created has a significance value of .003 and, as expected, a positive coefficient of 43,128, indicating that an additional run created results in 43,128 more dollars in the value of a player's contract. To put this in perspective, adding a Runs Created value of 10 for a season would add $430,127 to a player's salary. The average Runs Created value for the sample is 78.2. Traditionally, a season of 100 runs created is considered a very good season.

The other significant variable, Games Played, has a positive coefficient of 5,448 and a significance value of .000. This variable indicates that an additional year of experience (162 more games played) should translate, on average, into 882,511 more dollars. In order to test for possible nonlinearities, a regression was run which included a

squared transformation of Games Played. Since the results from this regression were very poor, the squared term was dropped from the regression reported here. Since the group of players that is used in this group is relatively young their games played numbers are also relatively low, it is to be expected that more experience will benefit these younger workers. The highest number of games played in this study is 982, so these results can only indicate a rate of return on games played up to that point. After that the return to games played may become negative, reflecting the older age of the player, which may make a player less desirable.

Steals, Assists, and Age are all not statistically significant variables in this regression. All of these variables were expected to have positive coefficients, but their low significance values suggest that the signs on the coefficients are relatively meaningless. Therefore, the regression results failed to show that steals, defense, and age are important determinants of baseball players' salaries.

Most important for the purpose of this study are the race dummy variables, both of which are not statistically significant. Since the variables are not significant, the signs and magnitudes of the coefficients are not important.

Table 3. Coefficients and Significance

	B	Std. Error	*t*	*Sig.*
(Constant)	-2,056,674	3,055,410.1	-.673	.504
RC	42,380.2	13,482.2	3.143	.003
ASSIST	-15,254.1	66,497.9	-.229	.820
STEALS	1357.7	17,530.4	.077	.939
GP	5,410.3	1,315.7	4.112	.000
AGE	-63,860.5	101,514.7	-.629	.532
MINORITY	365,410.1	446,860.1	.818	.296

$R^2 = .550$, n = 52

In order to attempt to further show that race is not a factor in salary determination, a separate regression was run using a variable, MINORITY, instead of the broken up racial groups, black and Latino. The results are amazingly similar to the first regression. In both regressions, the only significant variables are Runs Created and Games Played. All other variables were not significant in both regressions.

Race, no matter how it is defined, simply does not have an effect on salary for these data.

VIII. Chi-Squared Results

A chi-squared test was used to test for hiring discrimination by comparing the proportions of races on the 30 teams in MLB. By comparing the proportions, whether or not certain teams hire proportionately fewer minorities can be discerned.

The chi-squared analysis comparing the number of players of each race on the 30 MLB teams was not statistically significant. Since the employment distribution across teams is not statistically different, no hiring discrimination was apparent in the data.

IX. Conclusion

The purpose of this study is to determine whether or not racial discrimination exists in MLB for outfielders. After running the regressions and finding race variables that were not significant, it can be concluded that under this set of circumstances, racial wage discrimination does not exist for outfielders in MLB. Also, the chi-squared analysis did not detect racial hiring discrimination. It appears that MLB has reached the long run equilibrium that Becker hypothesized where players will be hired and paid fairly. The positive implications of this are obvious. It is an indication that if productivity is clear and unbiased in nature, employers will provide fair wages to all employees and will hire based solely on that productivity. Obviously, very few industries exist with such clear cut and easily measured productivity, but just the fact that minority workers are paid fairly when it is, is a positive sign for the state of the American social consciousness.

It should be noted that we hesitate to generalize this research beyond baseball and potentially other major professional sports leagues. We fear that by generalizing without justification, more harm would be done than good by potentially down playing the importance of testing each industry individually for racial discrimination. Therefore, in those industries where productivity is easily measured (e.g. NFL, NBA, NHL, handicrafts industries, and some manufacturing industries) future research should be done to see if racial discrimination exists after controlling for productivity. However, while productivity is very easily measured in some industries, most do not exhibit such a large degree of transparency. It would be a mistake to generalize the results of this study to those industries.

Additional future research could examine total contract value, which would capture the length of contracts to see if baseball owners prefer one race to another for long term contracts. Additionally, researchers could expand the regression to include all positional players, to try to get a better representation of the entire league's salary structure.

Works Cited

Becker, Gary S. The Economics of Discrimination. 2nd edition. Chicago: University of Chicago Press, 1971.

Bodvarsson, Orn B. and Shawn P. Pettman. "Racial wage discrimination in major league baseball: do free agency and league size matter?" Applied Economics Letters 9 (2002): 791-796

Fizel, John, Anthony C. Krautmann, and Lawrence Hadley. "Equity and Arbitration in Major League Baseball." Managerial and Decision Economics 23 (2002): 427-435.

Gwartney, James and Charles Haworth. "Employer Costs and Discrimination: The Case of Baseball." Journal of Political Economy 82.4(1974): 873-81.

Medoff, Marshall H. "Racial Discrimination in Professional Baseball." Atlantic Economic Journal 3.1(1975): 37-44.

Mincer, Jacob. "Investment in Human Capital and Personal Income Distribution." Journal of Political Economy LXVI (4) (1958): 281-302.

Rima, Ingrid H. Labor Markets, Wages, and Employment. New York: W.W. Norton & Company, 1981.

www.Census.gov. 2003. U.S. Census Bureau 20 November 2003
<http://eire.census.gov/popest/archives/metro/ma99-02.txt>.

www.sports-wired.com. 2003. 20 November 2003
<http://www.sports-wired.com/content/glossary.asp>.

www.baseballstuff.com. 2003. 20 November 2003
<http://www.baseballstuff.com/fraser/gloss.html>

www.baseball-reference.com. 2003. 20 November 2003
<http://www.baseball-reference.com>

www.wfu.edu, 2003. 20 November 2003
<http://www.wfu.edu/users/dunninrb/stuff/baseball/runs_created.html>

www.roadsidephotos.com, 2003. 20 November 2003
<http://roadsidephotos.com/baseball/index.htm>