

8-2013

EXPLAINING THE WAGE DIFFERENTIAL BETWEEN AMERICAN AND LATINO BASEBALL PLAYERS IN MAJOR LEAGUE BASEBALL

Kristopher Kusterman
Clemson University, kkuster@clemson.edu

Follow this and additional works at: https://tigerprints.clemson.edu/all_theses

 Part of the [Economics Commons](#)

Recommended Citation

Kusterman, Kristopher, "EXPLAINING THE WAGE DIFFERENTIAL BETWEEN AMERICAN AND LATINO BASEBALL PLAYERS IN MAJOR LEAGUE BASEBALL" (2013). *All Theses*. 1719.
https://tigerprints.clemson.edu/all_theses/1719

This Thesis is brought to you for free and open access by the Theses at TigerPrints. It has been accepted for inclusion in All Theses by an authorized administrator of TigerPrints. For more information, please contact kokeefe@clemson.edu.

EXPLAINING THE WAGE DIFFERENTIAL BETWEEN AMERICAN AND
LATINO BASEBALL PLAYERS IN MAJOR LEAGUE BASEBALL

A Thesis
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Master of Arts
Economics

by
Kristopher Kusterman
August 2013

Accepted by:
Dr. F. Andrew Hanssen, Committee Chair
Dr. Raymond Sauer
Dr. Robert Tollison

ABSTRACT

This paper examines the wage differential between Latino and American baseball players in the Major Leagues. First, a literature review is prepared to look at how, historically, economists have gone about comparing different races' and ethnicities' compensation in professional sports. Historically, these comparisons are made between white and African American players. Next, a detailed look at the dataset that is used in the paper, which is from the Sean Lahman 2012 Baseball Database which consists of data going back to 1871, although this paper's scope is only for the years between 1985 and 2009. Regressions are then prepared in the fashion that follows Lawrence Kahn's seminal work in the field, and we find that there does seem to quite a significant differential in American and Latino baseball players. I find that there does seem to be a significant wage gap between Latinos and American players, that Latino players do seem to earn, on average, higher salaries. Following, I discuss possible explanations for this that are obviously present in the regression. Concluding, I discuss possible future work and what could be done to improve.

TABLE OF CONTENTS

| | Page |
|---------------------------------------|------|
| TITLE PAGE | i. |
| ABSTRACT | ii. |
| LIST OF FIGURES | iv. |
| CHAPTER | |
| INTRODUCTION | 1 |
| LITERATURE REVIEW | 2 |
| ECONOMIC MODEL | 4 |
| DISCUSSION OF DATA | 6 |
| INTRODUCTORY REGRESSIONS | 11 |
| FORMAL REGRESSION AND RESULTS..... | 13 |
| POSSIBLE EXPLANATIONS | 18 |
| TENURE AND IT'S POSSIBLE EFFECT | 22 |
| ROBUSTNESS TESTS | 25 |
| CONCLUSIONS AND FUTURE WORK..... | 27 |
| REFERENCES..... | 29 |

LIST OF FIGURES

| Figure | Page |
|---|------|
| 1.1: Definitions of a couple rate statistics | 8 |
| 1.2: Makeup of Race and Position..... | 9 |
| 1.3: Summary Statistics..... | 10 |
| 1.4: Comparing Latinos and Americans across various offensive statistics..... | 11 |
| 1.5: Initial Regression | 12 |
| 1.6: Final Regression..... | 15 |
| 1.7: Graph depicting expected salaries using the model | 17 |
| 1.8: Changes among makeup of at-bats among role players and starters..... | 20 |
| 1.9: Average Number of At-Bats among different nationalities..... | 21 |
| 1.10: Comparing salaries across nationalities and ages..... | 23 |
| 1.11: Comparing Average Age, Tenure, Salary, and Debut Age between Latinos & Americans | 25 |
| 1.12: Graph depicting differences in salary for Middle Infielders only | 26 |
| 1.13: Oaxaca Decomposition | 27 |

Introduction

Wage discrimination is undoubtedly a topic that has been researched often; especially as interest in sports economics increases. Sports offer a look into a market that has clear-cut beginnings and endings as well as a copious amount of information with which to examine consumer, employer, and player behavior. This paper will examine differences in Major League Baseball salaries among players born in the United States, as compared to players born outside of the U.S., most notably, countries in Central and South America.

Lawrence Kahn is perhaps one of the best-known economists in examining discrimination in the market of professional sports. He has consistently found that there is little to no significant salary discrimination against black players in Major League Baseball (Kahn, 1991). He has, however, found that there is significant wage discrimination against blacks in the NBA. Tollison and McCormick (2001) have also found similar results, as they have found that NBA owners pay black players less even though there does not seem to be any consumer discrimination against teams with more black players. Perhaps more startling in the two papers found, is that this discrimination has lasted long after free agency was implemented in the NBA. Typically, we assume that free markets would eliminate any sort of inefficiency, and the fact that there remains a wage gap between black and white NBA players, holding player production constant, this would eliminate both employee and firm discrimination. This idea of wage discrimination in the NBA has seemingly dissolved more recently, however, as Kahn has found that the wage differential has nearly diminished.

Literature Review

Kahn notes in his *Discrimination in Sports: A Survey of the Literature*, that typically the log of salary is regressed “on a list of productivity indicators and a dummy variable for race” (Kahn, 1991). When this regression is done, it returns a measure of the market discrimination coefficient, which does make a large assumption regarding the returns to higher performance levels, in that whites and nonwhites (in my case, Americans and Latinos) receive similar returns to their increasing productivity levels. Kahn also notes that there will be biases in the results should the productivity measures be correlated with race and if the measures that are being used in the regression do not accurately reflect a player's overall productivity. Omitted variable bias and sample size are two more issues that Kahn raises in his methodology summary of how discrimination is typically measured. Sample size is less likely to be an issue, as the sample spans eighteen years of baseball across 30 teams and eight everyday players per team, all with differing salaries.

One issue that may arise with regards to omitted variable bias is the inability to quantify a player's defensive ability. Errors are certainly not an accurate measurement, as they are a counting statistic and good defensive players who have long careers will indubitably have more errors than a bad defensive player who lasts for only a couple of seasons. Within a season span, this is also the case, as bad defensive players will typically be moved to corner infield or outfield positions, where errors are less likely to occur. Fielding Percentage may be an accurate measurement, although this once again skews against those players playing middle infield positions, who are typically regarded as the best defenders, but who typically are faced with defending much more difficult balls in play off the batter's bat, and have many more opportunities to make fielding mistakes than a lumbering left fielder

who simply has to field a grounder or a fly ball. More recently, baseball has seen an uptick in the ability to quantify defensive acumen. Fangraphs uses a statistic known as UZR¹, or Ultimate Zone Rating, which divides the baseball diamond in sections, and measures the ability of the defender based upon his ability to field his own section as well as the ability to perhaps assist his teammate with balls in play in their section. Unfortunately, this technology sometimes fickle and still has a good amount of kinks to work out, especially with regards to seasons prior along with the consistency season to season. Working in position dummy variables will allow for the most accurate available measurement of defense, as it is assumed that the best fielders play “up the middle,” that is, in the middle infield positions of Shortstop and Second Base.

There are various forms with which discrimination may show up in salary regressions. They include the simply unequal pay for equal work, or just flat-out racism. This form has, as previously noted, little evidence to back up the notion and has seemingly dissipated over time. There is also the idea raised that black players face a higher barrier to entry than white players, or that black players must show a higher level of productivity to gain a spot on a team than white players. Positional discrimination is also another form, although Kahn (1991) notes this is more with respect to football, as many were hesitant to put a black player at Quarterback, for example. One last example is endorsement income, and while this is immeasurable using salary data, it could be the fact that teams find white players more marketable to their fans, which is also a form of consumer discrimination.

¹ <http://www.fangraphs.com/blogs/the-fangraphs-uzr-primer/>

With respect to baseball, there have not been many findings with respect to wage discrimination among white and black players. Hanssen and Anderson's paper (1999) regarding the number of All-Star players, which is still voted on by fans, found statistical evidence that discrimination been completely erased by the 1980s and that black players garner as many, or more, votes as their white counterparts. This perhaps provides some statistical evidence that consumer-driven discrimination has seemingly disappeared in United States professional athletics. Previous economic work on discrimination of sport's league wages seem to show that there is very little to no existing discrimination, among black and white players in the MLB (Kahn, 2009). Since I am examining United States born players and players born outside of the United States, perhaps this will provide different results.

In the early results regarding baseball discrimination, there were many issues that Kahn points out. Whether biases resulting from a lack of actual black players in the MLB to properly measure the bias coefficient, or the fact that some of the results were found following the institution of free agency, which would skew some of the data; although, Kahn does find that the amount of discrimination did not change following free agency (Kahn, 2009). Overall, in the various papers Kahn reviews, almost none of the research found a significant discrimination coefficient among black and white players, which suggests that there is little to no wage discrimination, at least among black and white players.

Economic Model

There are various reasons as to why we might expect there to be a significant wage gap between International players and American born players. Players born in America reach the major leagues via only one way, the First-Player draft, where players can either be

drafted out of High School, or, should they choose to go to college, they can enter the draft following their Junior year of college. Upon being drafted by a team, there are “slots” for each pick in the draft based upon previous drafts, where the contract given to the player changes after each pick. The player has little to no bargaining power, and the only bargaining chip in favor of the player is if the player is in High School, to which he can threaten to enter college, thus the team that drafted the player cannot sign the player. This is known as “signability” among those in baseball, and players who may not be as good as others may be drafted higher because the team has more confidence that they will be able to sign the player. There is a soft budget in place for the teams that can be used in the first 10 rounds, and each pick has an assigned² value, decreasing in value after the first overall pick. For example, the assigned value for the first overall pick in the 2013 Draft is \$7.7 million and decreases to \$6.7 million for the second overall pick.

The rules for signing international players has drastically changed in the last year³, although the data I am using has the old international signing rules applied during the years of the sample, which was much more favorable to international players than the new rules. Previously, teams could treat international players like unrestricted free agents where the teams competed against each other in courting the players with no cap on the offers the

² This is also referred to as the “slot.” Each pick in the draft is given a dollar value by MLB for which the team can sign the player. Some teams may draft players because they know that will sign “under slot,” and thus the teams can use the money saved for higher value, higher dollar contracts in the rest of the draft.

³ Pages 268-274 of the 2012 Major League Baseball Collective Bargaining Agreement offers further explanation on the new rules.

team could make to the player. This led to lucrative contracts being given to players the same age as those players getting drafted out of high school and college, while receiving free agent type contracts rather than rookie level contracts. A recent example is Yasiel Puig, a 22-year-old outfielder from Cuba who was able to defect and thus started receiving lucrative offers. He signed a 7-year, \$42 million contract with the Los Angeles Dodgers as a 22-year-old. For comparison's sake, Mike Trout, who was drafted in 2009 as the 25th overall pick, and just recently finished one of the greatest seasons of all time and finished second in the MVP voting as a 21-year-old, and was recently granted a \$510,000 salary for 2013, a year after the amazing season. So, while Puig had never played a professional baseball game in the United States receives \$6 million per year, Mike Trout makes \$510,000 due to the nature of the international signing and MLB First Year Player Draft rules. Should Trout's potential have been more recognizable and been drafted number 1 overall, rather than 25th he still would only make around \$1.5 million rather than \$500,000. Major League Baseball has only recently changed the international signing rules to assign each team an international pool of money to be spent on international players based upon franchise value and recent success. This should certainly limit the amount of money being handed out to international players and keep salaries mostly in line with domestic and international players. This rule only applies to those players who are younger than 23 and who have less than two years of experience at the professional levels; those players older than 23 and with three years of experience still face no limit on the contract they may be offered.

Age is one of the other factors that would have a large effect on salary, as we would expect players who are younger to earn more money. Players who enter the Major Leagues begin their "professional clock" which means that they earn a base salary for the first three

years of their career and then after those first three years, players can enter arbitration for the next three years, where players and team representatives settle on a 'fair' salary based upon age, position, and their career performance. After these three years, the player can then enter free agency where any team can bid for their services, which is similar to how the international player market had previously operated, before the rules were changed for the 2013-2014 season. Players can also sign contracts with their teams at any time, and most do to avoid free agency, and these contracts are typically quite lucrative. In the data for this paper, the average Latino player is 28.69 years old while the average American baseball player is 30.01 years of age. This is quite a large difference; especially considering the average career length of a player is about 5.6 years, according to a study done by the University of Colorado in 2005 (Rogers, 2005).

One other possible explanation for the wage gap between International players, specifically those players from Latin and Central America, and American born players is that a larger percentage of Latino players play the Middle Infield positions. Those include Shortstop and Second Base. These positions are the most difficult positions to field when playing defense, and thus only the most talented of defenders play there. The less athletic players typically play the corner outfield positions, First Base, etc. The fact the players from Latin and Central America typically do play these middle infield positions at a higher percentage than American born players would also lead to higher salaries, considering equal offensive output. My main model will consist of:

$\text{Log}(\text{salary}) = F(\text{offensive statistics, position dummy variables, age, birth country})$

Data

Sean Lahman's Baseball Database (Lahman, 2012) will be the main data source I will be using, as it contains salary data as well as performance indicators for those players, from years 1985 to 2009. The database also has the birth country of each player, so I will be able to determine whether the player is American or Latino. The database has only counting baseball statistics, such as runs, Home Runs, Runs Batted In (RBI), etc., so various rate statistics were created from the counting statistics.⁴ A quick explanation of two important rate statistics follows:

| | |
|---------------------|--|
| Batting Average | The number of total hits over the total number of plate appearances |
| On-Base Percentage | The total number of times the player reaches a bases via either a hit or a walk over the total number of At-Bats |
| Slugging Percentage | Gives a measure of the player's power. More weight is given to Home Runs, Triples, and Doubles rather than simply referring to them as hits. |

Figure 1.1: Definitions of a couple rate statistics

Along with those counting statistics, Slugging Percentage, Batting Average, and On-Base Percentage are also used in the performance factors. Rate statistics allow for a closer examination of the player's true value, as stats like runs and RBIs are quite team-dependent.

⁴ Counting statistics include those statistics that add up over time. Rate statistics are statistics that provide an average or rate of success at getting on base, getting a hit, etc. over the course of a period of time, in this case, a season.

A dummy variable for birth country is also used⁵. Salary in 2000 dollars⁶ is the dependent variable, and I will be using the log of this variable, as the data has such a wide range, ranging from around \$300,000 to \$30,000,000 and the log will help to smooth some possible skewness.

Makeup of Dataset

Looking at the overall makeup of the MLB, we can see that Latino players do play Middle Infield positions at a higher clip than their American counterparts:

| Position | Percentage (Latino) | Percentage (American) |
|----------------|---------------------|-----------------------|
| Catcher | 0.13 | 0.15 |
| Middle Infield | 0.31 | 0.16 |
| Corner Infield | 0.27 | 0.34 |
| Outfield | 0.29 | 0.35 |

Figure 1.2: Makeup of Race and Position

If we do find that position does have a relationship with the player's salary, this would almost certainly explain some of the variability in salary, especially with regards to Latinos

⁵ The birth country variables were not separated by race, and thus the model may be picking up some underlying discrimination between Black and White American baseball players; although this is unlikely and, if present, miniscule.

⁶ The salaries are also taken from Sean Lahman's data set, and is typically taken from the total salary amount of the contract divided by the number of years, this is known as the Average Annual Value (AAV). There may be some possible bias with using AAV as some contracts are back-or-front loaded, depending on the age of the player; although this is expected to be minimal.

and American players. While there are a smaller number of Latinos in the sample than American players, Latinos make up about 20% of the sample while Americans make up 77%, with the rest of the world coming in at 3%. The fact that 30% of Latinos play Middle Infield positions while 16% of Americans play Middle Infield positions is certainly noteworthy. Below is a table of the summary statistics that will be used in the regression, split between the overall sample as well as looking at only American or Latino players:

| Variable | Mean (Total Sample, 7753 Obs) | Mean (American Players, 5988 Obs) | Mean (Latino Players, 1590 Obs) |
|-------------------------|-------------------------------|-----------------------------------|---------------------------------|
| At-Bats | 392.31 | 389.48 | 403.33 |
| Hits | 106.85 | 105.79 | 110.62 |
| Home Runs | 12.18 | 12.17 | 12.12 |
| RBI's | 52.85 | 52.50 | 53.98 |
| SB's | 8.26 | 8.16 | 8.68 |
| Batting Average | 0.267 | 0.267 | 0.269 |
| On-Base Percentage | 0.272 | 0.273 | 0.266 |
| Slugging % | 0.418 | 0.419 | 0.416 |
| Catcher Position | 14.75% | 15.31% | 13.08% |
| Middle Infield Position | 18.58% | 15.59% | 30.50% |
| Corner Infield Position | 32.98% | 34.25% | 28.86% |
| Outfield Position | 33.67% | 34.83% | 28.86% |
| Salary | \$2,144,924.00 | \$2,011,952.00 | \$2,572,535.00 |

Figure 1.3: Summary Statistics

One explanation is that perhaps MLB teams are paying more for a certain skill set that Latinos have, and that perhaps USA-born players may not. Using the dummy variable as a dependent variable would allow us to see if Latinos are outperforming their American counterparts in some areas:

| Skill | Regression (R2) | Number of Observations | t Ratio on race dummy |
|---|-----------------|------------------------|-----------------------|
| Skills with a negative race coefficient | | | |
| At Bats | .0031 | 6724 | -2.96 |
| Runs | .0001 | 6724 | -0.99 |
| Hits | .0017 | 6724 | -3.43 |
| RBI's | .0002 | 6724 | -1.21 |
| Stolen Bases | .0003 | 6724 | -1.37 |
| On Base percentage | .001 | 6724 | -2.64 |
| Skills with a positive race coefficient | | | |
| Slugging Percentage | .0003 | 6724 | 1.44 |

Figure 1.4: Comparing Latinos and Americans across various offensive statistics

Excluded all player samples with <200 AB per season.

These findings suggest that Latinos are perhaps better in almost all offensive facets of the game, which would lead to Latinos receiving higher salaries.

Introductory Regression

A preliminary regression was prepared to look at the sample overall:

| L(Salary) | Coef. (Total) | Std. Error | P-Value |
|--------------|---------------|------------|---------|
| Runs | .0085 | .0015 | 0.000 |
| Hits | .0003562 | .0008525 | .676 |
| Doubles | .0081833 | .0024777 | .001 |
| Triples | -0.296335 | .0061597 | .000 |
| HRs | .0138343 | .0042749 | .001 |
| RBIs | .0012679 | .0013162 | .335 |
| SBs | .0012972 | .001431 | .365 |
| BBs | .0041878 | .0014991 | .005 |
| SOs | -.00069 | .0005572 | .214 |
| OBP | -.87387 | 1.011 | .388 |
| SLG | 1.1924 | .44811 | .008 |
| Intl Dummy | .1632117 | .0722066 | .024 |
| Latino Dummy | .4369881 | .0276 | .000 |
| Catcher | -.073082 | .03503 | .037 |
| MI | .029656 | .0343113 | .387 |
| OF | .00279 | .028525 | .921 |
| Age | 1.096435 | .030815 | .000 |
| Agesq | -.0149763 | .00049 | .0000 |

Figure 1.5: Initial Regression

The offensive statistics all seem to make sense, although On-Base percentage has a negative effect on salary, though this may be due to heteroscedasticity issues among the other variables, as slugging % has quite a large effect on salary. However, this coefficient is quite statistically insignificant and can be ignored for the paper. Most notably, is that both international players and Latino players make, on average and holding offensive statistics the same, 16.32% and 43.69% more, respectively. This is obviously a huge discrepancy and

requires more investigation. Also notable is that the Middle Infield position (Second Base and Shortstop) has the largest positive effect on salary, and we would expect middle infielders to make about 3% more than their corner infield counterparts. The age at which we can expect salary to be at its highest is the age 36 season. This does seem to make sense, as it is likely the end, or nearing the end, of a player's career, where the free agent contract is coming to an end. These findings seem to fall in line with what we expected, as Latinos typically make significantly more in their younger years, while also playing premium positions.

A separate regression was done, where Latinos and American players were separated, and I found a significant increase in the middle infield position for Latinos while the position is insignificant for American players. That is, I found that when looking at Latinos only, we can expect middle infielders to make 9.58% more than corner infielders, and when looking at American players only the difference is only .2% and is statistically insignificant. The age 36 season is once again the expected age for when salary is at its maximum for both race groups.

Model & Regression Results

Below is the log of salary regressed on productivity measures along with interaction terms between the Birth Country dummy variable as well as age and age-squared. This allows us to compare how age affects salary when looking at players from different countries. I also interacted the Latino players with position dummy variables to see if position had a differing effect on salary for Latinos than American born players. Some

further production variables were added to the regression to get an even more accurate depiction of the player's production level.

There is certainly some correlation between these variables, although these specific rate statistics are the most often used statistics when referencing a player's production level over the course of a season. It is necessary to use both rate and counting statistics to assess the value of a player, as small sample sizes (with small counting statistics) may lead to poor or great rate statistics. These statistics are also often used in arbitration hearings when players are arguing for their appropriate salary after their three years of professional tenure has passed.

| Log (Salary) | Coefficient | Std. Error | P-value |
|----------------------|-------------|------------|---------|
| At-Bats | -.0011461 | .0006375 | .072 |
| Hits | .0114114 | .0023927 | .000 |
| Home Runs | .0108826 | .0038097 | .004 |
| RBI's | .0016568 | .001293 | .200 |
| SBs | .0015362 | .0012676 | .226 |
| Batting Average | -7.309891 | 1.169055 | .000 |
| On-Base Percentage | 3.801326 | .494731 | .000 |
| Slugging Percentage | 2.568026 | .4471833 | .000 |
| Catcher Dummy | -.0830287 | .0390894 | .034 |
| Middle Infield Dummy | .0558769 | .0382275 | .144 |
| Corner Infield Dummy | .0148174 | .0306787 | .629 |
| AL Dummy | -.0619403 | .0218959 | .005 |
| Latino*Age | 1.073031 | .0569224 | .000 |
| Latino*Age2 | -.0149379 | .0009414 | .000 |
| Intl*Age | .9677375 | .2381644 | .000 |
| Intl*Age2 | -.0126605 | .0038508 | .001 |
| USA*Age | 1.151591 | .0373002 | .000 |
| USA*Age2 | -.0157506 | .0006013 | .000 |
| Latino Dummy | -17.047 | .8930618 | .000 |
| Intl Dummy | -16.24322 | 3.576376 | .000 |
| USA Dummy | -19.08428 | .6390896 | .000 |
| Latino*MI | -.0215939 | .0725854 | .766 |
| Latino*CI | -.0207059 | .0706898 | .770 |
| Latino*Catcher | .0092505 | .0882311 | .917 |

Figure 1.6: Final Regression

Batting Average is statistically insignificant and involves a large error, and thus can be ignored when interpreting. The Middle Infield dummy variable continues to have a large positive effect on salary, as this implies a 5.5% higher salary for middle infielders than outfielders with similar offensive output. This is up from 2.9% in the previous regression and this may be due to the increase in dependent variables explaining variations in salary. The interaction terms between the position variables and the Latino dummy variable are all statistically insignificant. This does seem to make sense, as even though we have seen Latinos play middle infield positions at a higher clip than American players, there wouldn't necessarily be a correlation between the position and birth country.

The interaction terms between the birth country dummy variables along with age and age-squared are perhaps most notable. Below is a graph depicting the expected salaries for players from different countries using these interaction terms as an equation:

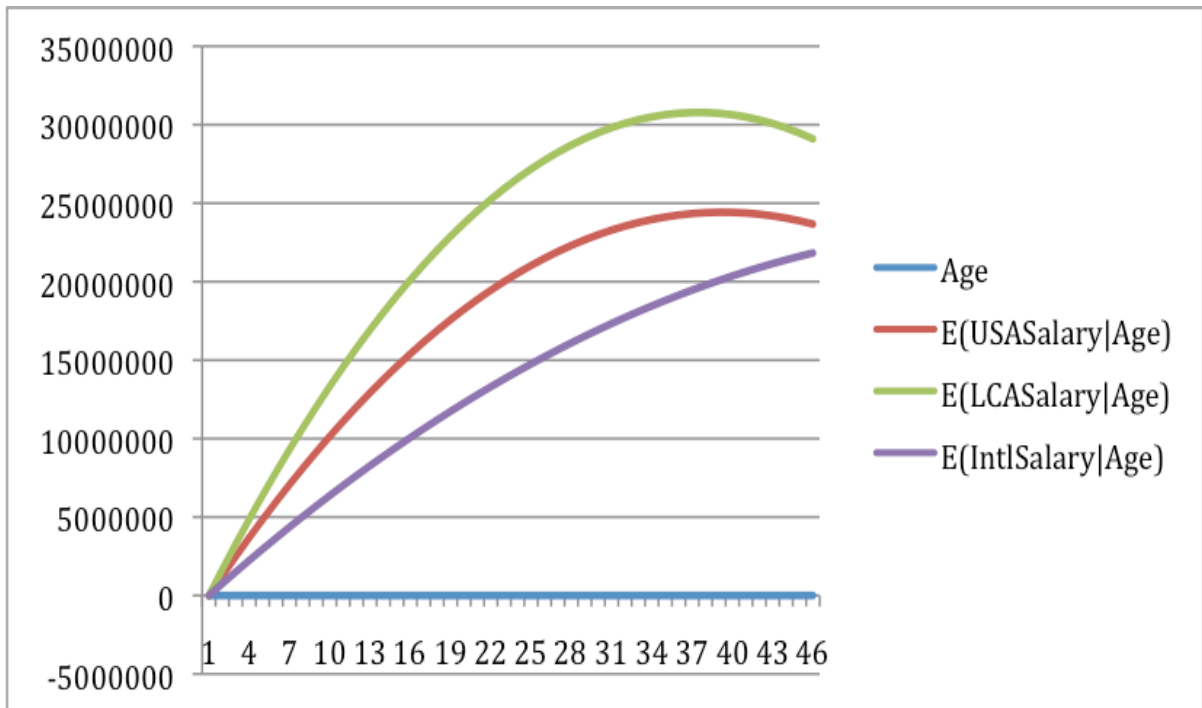


Figure 1.7: Graph depicting expected salaries using the model

The graph implies that the salary for Latin American players begins higher than American born players. This is due to the international free agency rule, which was discussed previously, and the salaries do seem to continue on a similar pace throughout the careers. 36 years of age also does seem to be the max age for each player, and after that the salaries take a steep decline, likely as this is the age when most free agency contracts expire, and the veteran player either retires or signs a modest, one or two year contract with a team. The two main issues that were discussed previously do seem to be the biggest factors when looking at differences between Latino and American players' salaries. Middle Infielders do earn, on average and holding offensive production constant, around 5% more than outfielders, which is also higher than both the Catcher and Corner Infield positions. Latin American players also play the Middle Infield position at a much higher clip, as around 31%

of Latin American players play these positions while only 16% of American players play the Middle Infield positions. The other main factor is age and the salary that the players earn at the young ages. Since international players were previously under different rules than American players, these players, specifically Latin American players, earn free agency-type contracts at ages where American born players are earning around \$500,000. This leads to quite a large gap in the salaries, as is depicted in the above graph. There are some other issues that may explain these gaps in salary, which cannot be explained in the above regressions. Perhaps only the best players from the Latin American countries are being signed since they are being more thoroughly vetted due to the large nature of their contracts. Another factor is that Latino players are simply reaching free agency faster, since they are often signed at younger ages.

A Couple of Possible Explanations

There are up to 40 rounds in a MLB draft, and a minimum of 32 draft picks per round. There are various competitive balance and compensatory picks that are also doled out to teams prior to the draft. This gives a minimum of 40 players drafted per year by a team, of which about 5% per year will actually make it to the Major Leagues. The economic risk is also quite low, per the rookie contracts explained earlier, so while teams obviously don't want to "miss" on a draft pick, especially in the first round, it is not so damaging that the team cannot recover for years. The upside is also obviously huge; if you can get a player on your team that is highly valuable to your team's success while only making \$500,000 the team can divert their salary towards other more expensive veterans to fill out their roster. Mike Trout is, again, a good example for this, as the Los Angeles Angels are paying \$550,000

this year and were able to go out into free agency and gave Josh Hamilton, a slugging outfielder previously of the Texas Rangers, a 5 year/\$125 million contract this year. This low risk/high reward is not so much when teams are going after Latin American and other International players. Since there is no draft and no way to regulate how much a player can get paid, good players are often pursued like American players who have entered free agency.

Yu Darvish, a talented pitcher from Japan, was making 500 million Yen per year in Japan before signing a 6 year/\$60 million contract with the Texas Rangers despite ever pitching on American soil. This contract has since worked out, as Darvish has been marvelous for the Rangers. However, there were likely quite a bit more resources diverted towards the scouting of Darvish before the Rangers were willing to offer such a lucrative contract. Should Darvish not have passed the Rangers scouting department's qualifications, he would likely have never ended up in a Rangers uniform. If International players are older than 23 and have more than 3 years of professional experience, they are not subject to the same international rules as those players younger than 23; and thus can be signed as if they were an unrestricted free agent. There is such a higher risk with signing expensive International players that the scouting that goes into the players weeds out players who might have been drafted in the MLB First Year Player Draft should they have been born in America. This means that those players who come over from Latin America or Japan are highly talented, highly scouted and are thought to make immediate impact at the Major League level and thus are going to be spending little to no time at the Minor League level. These international players are not getting signed by Major League teams to spend time in the minor leagues, whereas teams that draft players from college or high school may be

much more willing to divert training resources to these players in the minors. So the overall theme regarding the possible bias in the data is that the players who are from Latin America are typically more thoroughly vetted, reach the Major League level at a much faster pace, and make higher salaries immediately rather than waiting 4-6 years to receive a contract offer.

One other explanation that requires further analysis is with regards to the “professional clock” that begins once a player reaches the Major Leagues. Below is a table depicting the difference in percentage of players when looking at role players and bench players (less than 250 At-Bats in a season) compared to everyday players (more than 250 At-Bats):

| Less than 250 At-Bats | |
|-----------------------|-------|
| American | 78.6% |
| Latin American | 19.3% |
| International | 2.1% |
| More than 250 At-Bats | |
| American | 76.8% |
| Latin American | 20.8% |
| International | 2.4% |

Figure 1.8: Changes among makeup of at-bats among role players and starters

Clearly there is a marked decrease in the percentage of American players when moving from role players to everyday starters, as we see a 2.2% decrease. This coincides with a 1.5% increase in Latino born players as well as a 0.3% increase in International players. Of course, this difference is not really that big, but the finding is somewhat explanatory. While the changes in percentages aren't so extreme, the difference is significant and displays the fact that players who are signed from Latin America are signed to play everyday and not to simply be a role player.

| Average Number of At-Bats Per Season | |
|--------------------------------------|--------|
| American | 389.48 |
| Latin American | 403.33 |
| International | 388.89 |

Figure 1.9: Average Number of At-Bats among different nationalities

Again, we see that the Latino players' average more playing time per season, which bolsters the idea that players signed from Latin American countries are signed to be starting players and are the best of the best that these countries have to offer.

Tenure and its Possible Effect

Contract negotiations are quite convoluted⁷; as it takes six years of service time before players become an unrestricted free agent. For the first three years, the player simply receives his basic rookie contract, which is typically between \$400,000 and \$500,000, and the player receives a bonus of around \$20,000 per year. After those three years are up, and before the player reaches six years of service time, a player can apply for arbitration; this is when the team and the player's agent settle on a reasonable salary as compared to players with similar production at similar positions on the baseball diamond. Teams and players have reached a settlement about 90% of the time before entering arbitration, and teams are increasingly signing players to long-term deals, and thus buying out their arbitration years. This layout almost certainly favors international players, which are not eligible for the MLB First-player draft, and can only be signed from their country by the MLB team. International players, who don't go to college like many USA born players, can begin their "professional clock" much earlier in their career and thus reach the lucrative free agent market much sooner and at a much more long-term deal friendly age. One example to explain this idea is to compare two fantastic MLB players who recently signed long-term deals. Elvis Andrus is a shortstop from Venezuela who was signed by the Atlanta Braves at age 17; he was traded to the Rangers and reached the MLB at age 20. He recently signed a long-term deal at age 24 (four years of service time). Justin Verlander is a pitcher from the United States who attended college and was drafted in the MLB First Player draft at the age of 21. He reached

⁷ For further explanation, please refer to Article VI of the 2012 Major League Baseball Collective Bargaining Agreement.

the MLB at the age of 23 and signed his first long-term contract at the age of 27 (again, four years of service time). This leads to the younger player possibly getting two long-term deals in their career, while the USA born player who attends college may only receive one long-term deal to go along with a short contract as their career winds down. Below is a table depicting the average salary for players below their age-25 season compared to the average salary for players above their age-25 season:

| Average Salary, below age 25 | |
|-------------------------------------|----------------|
| American | \$289,061.60 |
| Latin American | \$433,229.40 |
| International | \$214,772.70 |
| Average Salary, above age 25 | |
| American | \$2,257,413.00 |
| Latin American | \$3,202,946.00 |
| International | \$3,121,080.00 |

Figure 1.10: Comparing salaries across nationalities and ages

The difference in salaries before age 25 is quite significant between players born in the United States and those players born in Latin America. This lends credence to the notion expressed above that those players who aren't given large contracts straight from their

country are still able to receive lucrative contracts much earlier in their career than players that are forced to go through the First-Year Player Draft. Players who are good enough to get through approximately their first four years of service time at the Major League level are able to sign contract extensions, and if a player signs this contract at age 24 rather than age 27, the contract will also be more lucrative. This is due to the simple fact that teams will be paying for seasons in which the player is still considered in his prime and ending around age 30 rather than ending around age 35, where the player's skills are in a rapid decline.

Various Robustness Tests

Since gathering all of the data on tenure and experience of each player is simply beyond the scope of the paper, I decided to single out a single season included in the sample and gather the tenure, age, debut age (simply the difference between their current age and tenure to find debut age), nationality, and salary of each of the 8 starting position players (9 for American League teams, since they have a Designated Hitter position) on each team for the 2003 season. While this obviously is a small sample, it could lend some credence to the idea that Latino players simply accrue more tenure at a younger age than American born players. This test of robustness should provide a small snapshot of a single year to provide further insight into this idea:

| Average Age (American Born) | Average Age (Latino) |
|-------------------------------------|-----------------------------------|
| 29.34 years | 28.70 years |
| Average Tenure (American) | Average Tenure (Latino) |
| 7.16 years | 7.45 years |
| Average Salary (American) | Average Salary (Latino) |
| \$3,811,428.00 | \$4,062,576.15 |
| Average Debut Age (American) | Average Debut Age (Latino) |
| 22.18 | 21.25 |

Figure 1.11: Comparing Average Age, Tenure, Salary, and Debut Age between Latinos & Americans

This small test of robustness provides some interesting results, as these findings do provide some data in defense of the possible explanation above. Latinos have an average debut age almost an entire year younger than their American peers and also have a longer average tenure or years experience in the Major Leagues, despite the average Latino being about 0.6 years younger. This higher tenure despite being younger leads to players reaching free agency at a younger age and thus being able to sign more expensive contracts, which leads to their average salary also being higher than their American counterparts. Another test to check the robustness of the results is to single out a single position and compare salaries while holding a single position constant. Looking at only Middle Infielders yielded

similar results as when including position dummy variables to hold those constant. The findings are below:

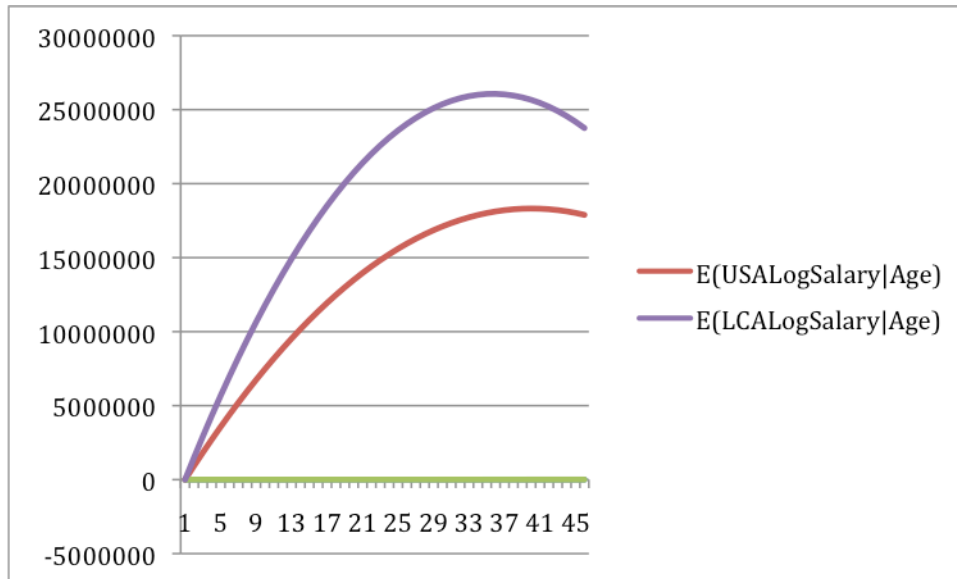


Figure 1.12: Graph depicting differences in salary for Middle Infielders only

The same ideas hold when looking at middle infielders only, as we see a large wage gap between Americans and Latinos once again, holding the same offensive statistics constant.

Oaxaca Decomposition is a procedure developed by Ronald Oaxaca and is widely used when identifying the presence of labor market discrimination between two or more groups. Below is an Oaxaca Decomposition test run for this regression, including only Latino and American baseball players. Obviously, for this paper the two groups are Latino MLB players and American MLB players.

| | Coefficient | Std. Error | P>z |
|----------------------------|-------------|------------|------|
| Group 1 (American players) | 13.6873 | .01719 | 0.00 |
| Group 2 (Latino players) | 13.82678 | .03608 | 0.00 |
| Difference | -.1394841 | .03997 | 0.00 |
| Endowments | .2613849 | .03252 | 0.00 |

Figure1.13: Oaxaca Decomposition

First off, we see that there seems to be a 13.94% wage gap between Latinos and American players, which is obviously consistent with our findings so far. This was found using the same explanatory variables in the previous regressions. Perhaps the more important finding when working with the Oaxaca Decomposition is the mean increase in wages we would see for American players should they have the same characteristics of Latino players; that is age, position, etc. We find this to be around 26.13%, or that, on average, we would expect American players' wages to increase by 26.13% should they have the same characteristics as Latino players. This could be due to some unexplained variables that are omitted, including that Latino players are specializing in baseball at a further age, are able to play outdoors year-round as compared to young American players, et cetera.

Conclusion and Possible Future Work

While there may not be discrimination in the fashion that Kahn defines, there does seem to be a significant salary differential between players born in the United States and those players born in Latin American countries. Using data from the Sean Lahman Baseball

Database, I was able to show that Latin Americans do, on average, receive higher salaries than their American peers. I suggest that this was due to two main reasons. First, players are not subjected to the salary restrictions that stem from being drafted in the First-Year Player draft, and the best International players are treated like free agents where teams are bidding against each other and thus the contracts rival those signed by American players after they have finally reached free agency. Including data on years of experience and tenure would assuredly offer some improvement. Secondly, there is higher risk in spending so much money on a player, so the players signed internationally are exceptional as the scouting has been so thorough. Players signed internationally are typically everyday players who receive a higher number of at-bats than the average American player while also receiving higher salaries in their below age-25 season, although this explanation seems less plausible and harder to provide data on than the tenure and experience discrepancy.

Concluding, there does seem to be some significance, and the changes that are starting next season will certainly curtail the lucrative contracts, so I would expect the average salaries to become more comparable and the size of salary differential to be much smaller. This rule change would provide for some further analysis in the coming years to compare what, if any, changes there are in the salaries among American and Latino baseball players.

WORKS CITED

- "Baseball-Reference.com - MLB Stats, Standings, Scores, History." *Baseball-Reference.com*. USA Today Sports, n.d. Web. 11 July 2013.
- Hanssen, F. Andrew, and Torben Andersen. "Has Discrimination Lessened Over Time? A Test Using Baseball's All-Star Vote." *Economic Inquiry* 37.2 (1999): 326-52. Print.
- Kahn, Lawrence M. "Discrimination in Sports: A Survey of the Literature." *Industrial and Labor Relations Review* 44.3 (1991): 395-418. *Jstor.com*. Web. 15 May 2013.
- Kahn, Lawrence M. "The Economics of Discrimination: Evidence from Basketball." *Institute for the Study of Labor* (2009): n. pag. Web.
- Lahman, Sean. "Lahman's Baseball Database." *Lahman's Baseball Database*. Creative Commons, 9 Jan. 2012. Web.
- McCormick, Robert E., and Robert D. Tollison. "Why Do Black Basketball Players Work More for Less Money?" *Journal of Economic Behavior & Organization* 44 (2001): 201-19. Print.
- "Major League Baseball Players Association: Basic Agreement." *Major League Baseball Players Association: Basic Agreement*. N.p., n.d. Web. 10 July 2013.
- Rogers, Richard G., Jarron Saint Onge, and William D. Witnauer. "Baseball Career Length in the Twentieth-Century: The Effects of Age, Performance, and Era." 2005. MS. University of Colorado, Boulder. 21 Sept. 2005. Web. 15 Mar. 2013.