Japanese Players in Major League Baseball: An Econometric Analysis of Labor Discrimination

The Honors Program
Senior Capstone Project
Kyle Audet
Faculty Advisor: Jongsung Kim
May 2009

Table of Contents

Japanese Players in Major League Baseball: An Econometric Analysis of Labor Discrimination
Senior Capstone Project for Kyle Audet

ABSTRACT

In recent years, Major League Baseball (MLB) teams have increasingly drawn talent from the international market. This study employs econometric techniques to test the various labor discrimination issues facing Japanese players playing in America's MLB. There have been remarkably few studies focusing on Japanese players in MLB. Data from the 1995 to 2007 season has been analyzed. High degrees of discriminatory pay scales are found for Japanese pitchers. The findings suggest that Japanese batters do not face such pay discrimination. Positional discrimination has also been the subject of investigation. Japanese players are found to be disproportionally pitchers. Suggestions for future research topics are presented.

INTRODUCTION

This paper will investigate issues pertinent to the assimilation of Japanese baseball players in the United States' Major League Baseball system. In order to gain a more complete understanding of this issue, broad research has been conducted in the area of baseball's historical place in Japanese history. In Japan, "America's" game thrives in its own right. This study includes passages detailing various aspects of baseball in Japan. The fundamentals of Japanese baseball are introduced first. Following this section, a brief historical context of baseball in Japan and an explanation of the posting system is provided. Koshien, the annual high school tournament is also profiled. Game play differences between Major League Baseball (MLB) and Japan's Nippon Professional Baseball (NPB) league will be interesting to even the most casual baseball spectator. Several of baseball's factual and hypothetical globalization efforts are also provided in the introductory pages of this study.

The main purpose of this study is to investigate the issue of discrimination, using econometric analysis, regarding Japanese baseball players playing in America's Major League Baseball. Of this topic, salary discrimination will be the main focus. Various econometric and statistical devices, including linear regression, the Chow Test, "bootstrapping", random sampling, and Oaxaca decomposition will be utilized in order to best reveal the true nature of labor market discrimination towards Japanese-born players in Major League Baseball. Positional segregation as well as salary discrimination will be investigated. Demographic influences and inclusive team performance have also been analyzed, albeit to a lesser extent. The following is a list of specific questions that this study attempts to addresses:

1. What are the best salary predictors for Japanese-born players playing in the United States and do these differ from non-Japanese players?
2. Does the average salary level of equally talented players[1] differ between Japanese and Non-Japanese?
3. Do returns-to-experience[2] in MLB differ between Japanese and Non-Japanese?
4. Do Japanese-born players playing in the United States face positional discrimination?

[1] For both batters and pitchers
[2] For both batters and pitchers

5. Do Japanese-born players playing in the United States face customer discrimination?
6. Do inclusive teams[3] perform better[4] than their homogeneous counterparts?
7. What are the characteristics of inclusive teams[5] and do they differ from their homogeneous counterparts?

> Please note: "Call-out" boxes like this one have been utilized in an effort to help bring attention to the main conclusions of this study; look for them throughout.

[3] With respect to Japanese born players
[4] In terms of team wins
[5] With respect to Japanese born players

FUNDAMENTALS OF JAPANESE BASEBALL

In order to obtain a solid foundation of background knowledge concerning the history of Japanese-MLB relations, five texts were read in preparing to write this paper. The first section of this paper represents a summary of said works. The content of the books varied from historical overviews of baseball in Japan, to modern Japanese players and their transition to America's Major League Baseball system, to the globalization of the sport of baseball.

The first book I read was Baseball Samurais: Ichiro Suzuki and the Asian Invasion by Ron Rains. This book, although slightly outdated (copyright date 2001), provides solid introductory information concerning the influx of Japanese professional baseball players in the United States. The book commences with a very brief historical overview of baseball in Japan. Following this introduction, Rains discusses the first few Japanese players' transition from Nippon Professional Baseball (NPB) in Japan to the Major Leagues in America. Rains focuses his attention on Masanori Murakami, Hideo Nomo, Kazuhiro Sasaki, Hideki Irabu, and Ichiro Suzuki among some other future prospects. The purpose of this study is not to focus on certain individual player's transitions from Japan's Nippon Professional Baseball to Major League Baseball. However, learning of the different paths each player took to the MLB, each with his own obstacles, proved fundamental to researching the history of relations between NPB and MLB.

Next, I read a book entitled You Gotta Have Wa, written by Robert Whiting. Robert Whiting is an American journalist who resides in Tokyo, having penned several works regarding Japan's Nippon Professional Baseball league. Arguably Whiting's most famous book, You Gotta Have Wa discusses in much detail some of the main differences between professional baseball in Japan and professional baseball in America. There are several differences, both on the field and off. These differences will be highlighted in a subsequent section of this paper.

Another book written by Whiting, entitled The Meaning of Ichiro, was read next. Admittedly there was much overlap between Whiting's two books. The more recent installment included topical information and stories on the latest Japanese and American baseball interactions. The updated storylines focused namely on Ichiro Suzuki, Alfonso Soriano, Hideki Matsui and Bobby Valentine.

The fourth book I read was Baseball without Borders: The International Pastime, edited by George Gmelch. Gmelch organizes sixteen essays detailing baseball's presence in various nations around the globe. The essays are organized by region—Asia, The Americas, Europe, and The Pacific. The section on baseball in Asia is clearly the most relevant for my paper. However, the remaining essays provided good information on baseball's development in the outstanding sections of the globe.

The last book read in my quest to acquire a solid background context was a very interesting book written by Alan M. Klein. In his book, entitled Growing the Game: The Globalization of Major League Baseball, Klein conducts several case studies related to the continuing internationalization of the game of baseball. A typical MLB small-market (in this case the Kansas City Royals) team's efforts to acquire international talent is juxtaposed with that of a typical MLB large-market (here the Los Angeles Dodgers) team's. Continuing, Klein writes of the Major League academies as well as "buscones" (local agents, oftentimes corrupt) so prevalent in the Dominican Republic baseball scene. Chapters focusing on Japanese, European, and South African baseball are also included. Finally, Klein examines Major League Baseball's past, present, and future efforts in internationalizing its game.

BRIEF HISTORY OF BASEBALL IN JAPAN

An American Professor teaching in Tokyo, Horace Wilson, introduced the Japanese people to the game of baseball in 1872 (Solloway, 2007). Albert Bates and Hiroshi Hiraoka are credited for creating the first organized game in 1873 and team in 1883, respectively (Rains, 2001). A national high school tournament started in 1915; this tournament is known as "koshien" and is detailed in a subsequent section of this paper. By 1936, a professional league, called Nippon Professional Baseball, was created (Rains, 2001).

Opposing the United States in the Second World War, the Japanese created new words for things such as "base" "strike" and "pitch" because previously the English words were used (Rains, 2001). After the Second World War, Allied troops and officials encouraged baseball to be played, both for recreation and as a way for the Japanese and Allied troops to interact (Rains, 2001). The growth of baseball in Japan quickened after the war and by the 1950's, there were two leagues in Nippon Professional Baseball, the Central and Pacific.

Both before and after the Second World War, American teams made several tours playing baseball against local Japanese teams throughout the nation. In addition to these touring trips, Major League teams invited Japanese players to play with them during spring training. This was a mutually beneficial situation as the Japanese players could use their time in America to fine-tune their skills, while the American teams could explore the Japanese's talent level. In 1964 Masanori Murakami played in the farm system of the San Francisco Giants. He played so well, that the Giants desired to have Murakami play the 1965 season with the Major League club. This set off a large dispute between the two nation's leagues.

The Japanese club believed they had the rights to Murakami, however, according to the contract they had with the Giants, those rights were forgone if the Giants paid $10,000 (Rains, 2001). In the end, an agreement allowed the Giants to retain Murakami in 1965; after the season Murakami would be free to decide where he wanted to play. Murakami pitched very well for the Giants, but because of intense pressure politically and socially, he decided to return to Japan in 1966. No other Japanese players would come to play in America until Hideo Nomo in 1995.

The "modern" (ie. from 1995 to present) influx of Japanese-born players in Major League Baseball began when Hideo Nomo signed with the Los Angeles Dodgers in 1995. The story of how Nomo came to play in America necessitates some explanation. At the time, Nomo was pitching well in Japan, earning $1.5 million per year. Nomo's competitive nature drove him to bring his pitching to the highest level of play, the United States' Major Leagues. However, Japanese rules stated that a player had to be a ten-year veteran before becoming a free agent. To evade this rule, Nomo's agents advised him to announce his "retirement"—therefore becoming a free agent (Rains, 2001). After announcing his "retirement" in Japan, Nomo was free to negotiate contracts with Major League clubs in America.

The contractual loophole exposed by Don Nomura and Hideo Nomo was subsequently closed. A clever system termed "posting" has been developed to facilitate the movement of players from Japan's Nippon Professional Baseball to Major League Baseball. A delineated description of this system is provided below.

POSTING SYSTEM

Due to Hideo Nomo's clever actions in evading the Japanese free agency rules, the NPB quickly modified them. Under the new system, players in Japan become free agents after nine years of tenure. If a team anticipates a star player wishes to declare free agency as soon as eligible, the team will utilize the "posting" system. The posting system allows a team to receive some compensation for a player that they would otherwise not receive. This system has received much attention in the Northeast with the coming of such players as Hideki Matsui and Daisuke Matsuzaka; however, further explanation will be provided.

In the posting system, Major League organizations are given a time period to provide sealed bids for the opportunity to negotiate with a Japanese player. The Japanese player is someone who would otherwise not be a free agent, but the player's Japanese club realizes the player intends to leave the team soon (as soon as he becomes free agent eligible). This allows a cash-strapped Japanese team to get something in return for the talent they have helped develop from a youth to a talented star.

Sealed bids are submitted by any interested MLB club. Bids must be submitted within four days of a player being posted. The club with the highest bid possesses the sole negotiation rights with the player for a certain length of time. If the Major League club signs the player, they have to pay the Japanese club the bid submitted earlier. If an agreement between the Japanese player and American team cannot be reached, the player continues to play for the Japanese team and the American team pays nothing to the Japanese team. The posting system is very unique and interesting. Some critics accuse the posting system of being in violation of U.S. antitrust law as well as an infringement of basic human rights (Whiting, 2004). Indeed, an entire thesis could be written on the posting system; but that is not the purpose of this paper. However, it is important to understand the basics of the posting system as part of the basic relations between NPB and MLB.

The first positional player to come from Japan and play in the Major Leagues was Ichiro Suzuki. Ichiro went through the posting system in 2000, as his Japanese club, the Orix Blue Wave, knew of his intentions to sign in America as soon as becoming free-agent eligible

(Rains, 2001). The Seattle Mariners won the rights to negotiate with Ichiro with a sealed bid of $13.125 million (Rains, 2001). At the time, Seattle already had a prominent Japanese player on their roster in pitcher Kazuhiro Sasaki. Having a fellow countryman on his team may have been reason enough for Ichiro to sign with the Mariners (the 3-year, $14 million deal may have helped as well). Clearly the Mariners, owned by Japanese videogame company Nintendo, have a high interest in Japanese players.

Why The Posting System Hurts Small-Market Teams

The player market in Japan is well developed and mature. The level of play in Japan is high, to such a degree that exhibition tours of American teams foster stiff competition. Because of the high level of play and wide media coverage received by Japanese teams, players' talents can be thoroughly evaluated by MLB scouts. The posting system inherently hurts teams with less money to spend on player contracts. Allard Baird, former General Manager (GM) of the Kansas City Royals (the archetypical small-market MLB club) described the penny-pinching ways of small market clubs as: "Going shopping without a credit card" (Klein, 2006). The posting system essentially aides teams that can afford to pay large player contracts. The posting system is a silent auction! Therefore, small-market teams cannot afford to sign Japanese talent via the posting system because large-pocketed clubs [i.e. - New York (Yankees and Mets), Boston, Los Angeles (Angels and Dodgers)] will routinely outbid them. Small-market teams cannot afford to gamble on a proven Japanese star—if the expensive player flops in America, most of the teams' salary expenditures are wasted.

KŌSHIEN

In Japan, an annual national baseball tournament of high school teams occurs each August. The tournament, held at Kōshien, is the most-watched high school sporting event in the world (Gmelch, 2006). A stellar performance at the tournament, which is viewed by millions on television and in person, can catapult a player to the status of a national hero. Only the best high school teams make it the coveted tournament; many players in the Nippon Professional Baseball league had the honor of participating during their youth. Dramatic games filled with emotion are the norm at this exciting event. America's nationally televised and dramatic Little League tournament in Williamsport, Pennslyvania pales in comparison. The pressroom contains nearly eighty reporters, from all of the nation's major newspapers (Gmelch, 2006).

Once making it to the tournament, coaches frequently overextend (by American standards) their players, especially the pitching aces. A story that epitomizes this overextension involves current star for the Boston Red Sox, Daisuke Matsuzaka. In 1998, "Dice-K" pitched all seventeen innings (250 pitches) of an extra innings game one day, nine innings the following day, and threw a no-hitter in the tournament finals. While not on the hill (an infrequent occurrence indeed), Matsuzaka was kept busy by patrolling the outfield (Gmelch, 2006).

Baseball's popularity among Japan's youth population has recently seen a decline. Perhaps the creation of a Japanese professional soccer league and the corporal punishment and rigid training requirements of many youth baseball teams can account for the declining popularity of baseball among Japan's youth. A committee, the Looking Forward to the 21st Century Committee, was created during the mid 1990's to examine how to improve the popularity of baseball in Japan. It was suggested that coaches stress the enjoyment of the game, and reduce the rigid and discipline training requirements inherent in Japanese baseball. However, the Japanese system has largely been slow in adopting the suggestions (Gmelch, 2006). However, baseball's popularity is increasing as players like Ichiro Suzuki and Hideki Matsui flourish in the major leagues.

GAME-PLAY AND CULTURAL DIFFERENCES

The similarities between American Major League Baseball and Japanese Nippon Professional Baseball all but end after the fact that there are nine players on the field at a time. Perhaps this is an overstatement, but the cultural and on-the-field differences are striking. Please note: the vast majority of these differences have been revealed by Robert Whiting in his two books on the subject You Gotta Have Wa and The Meaning of Ichiro.

- Japanese professional teams are owned by large corporations (media conglomerates, railroad companies, large retailers etc.) mostly as public relations vehicles. It is not uncommon for Japanese teams to earn negative profits as they are frequently run by a businessman with little baseball knowledge.
- Japanese teams play a 130 game season, as opposed to MLB's 162. Because of the rainy season each summer, many games are postponed until the end of the season.
- Ties are allowed. In the Pacific League, no new innings may be started after four hours or twelve innings of play. In the Central League, a game lasting fifteen innings results in a tie.
- The Japanese farm system (player development) is much less developed than the system in America. There are only twelve minor league teams. Each professional club has one affiliated minor league team that plays only eighty games per season.
- The Japanese baseball is slightly smaller than the American version while the Japanese strike zone is slightly larger.
- On the whole, Japanese field dimensions are smaller than American ballparks'.
- Team aspects, the sacrifice bunt, the squeeze play and the hit-and-run are used much more in Japanese baseball than in American. In 2003, Tokyo Giants infielder Masahiro Kawai set the record for career sacrifice bunts with *514*.
- Practice
- Spring training in Japan begins in mid-*January*, in the bitter cold of winter. Practices are seven hours long, consist of ten miles of running and are followed by evening strategy sessions with coaches in the team dormitory. This stands in marked contrast to American spring training which begins in March, lasts five or six weeks, and consists of three to four hours practicing per day.
- Japanese teams conduct difficult training nearly every day during the regular season, even on afternoon game days and getaway travel days. This is certainly not the case in America, where in the heat of summer, many players do not even take on-field pregame batting practice.

- The “1000 Fungo Drill” – players are forced to field batted balls hit slightly out of their reach to their sides or rear until they collapse from exhaustion.
- In Japan, a quota is in place on how many foreign players are allowed on each team. The quota has varied but currently stands at four foreigners per team (with no more than three position players or three pitchers). One Asian player (from one of the leagues in China, Korea, or Taiwan) may be excluded from the foreigner limit.
- In general, the salary level in Japanese baseball is lower than in Major League Baseball. To attract former Major Leaguers, gaijins (“foreigners”) are paid a higher salary level (two to three times what comparable Japanese players make) and treated to such luxuries as free cars, and living and travel accommodations not offered to local players.
- In Japan, a player is deemed a free agent after nine years of service; in America, six years gets a player to free agency.
- Japanese players are on average much more discipline at the plate than their American counterparts. Take this example—Ichiro Suzuki never struck out swinging *during his entire high school career* (536 at bats)!
- The Japanese player’s union is much less forceful than Major League Baseball’s player’s union. In fact, the players have never struck in Japanese baseball history whereas there have been several work stoppages in the history of MLB.

INTERNATIONALIZATION OF MAJOR LEAGUE BASEBALL

There is no denying the increasing internationalization of Major League Baseball. The international growth of the game is remarkable. Take this striking example—only two of the nine Florida Marlins starters in the final game of the 2003 World Series were American citizens (Whiting, 2004). The growth of the game has faced obstacles, namely underfunding in some nations, namely South Africa and Europe, as well the International Olympic Committee's decision to remove baseball (and softball) from the summer Olympiad's lineup beginning with the London games in 2012. Baseball may be reinstated by the committee for action in the 2012 games, that prospect seems unlikely. Major developments in the internationalization of baseball have taken the form of a FIFA (Fédération Internationale de Football Association) style world-cup, called the "World Baseball Classic", in addition to the continued relations between Major League Baseball and Nippon Professional Baseball—taking the form of tours, demonstrations, exhibition, preseason, and opening day games being played in and around Japan.

World Baseball Classic

In the fall of 2001, Bud Selig, MLB Commissioner, first publicly announced his intentions for a world-cup style international baseball tournament. In stark contrast to FIFA's organizational structure where the tournament is ruled by a board with representatives from competing countries, the WBC is a Major League Baseball International (MLBI) run invitational tournament. This structure caused some uproar from baseball powerhouses Korea and Japan. After some negotiations, the tournament was able to be played with all major baseball-playing nations in March of 2006. The revenue sharing process of the inaugural WBC was as follows: the costs to run the tournament will be skimmed off the top. After paying all costs, the national federations from each country would divide 53% of the remaining revenue along the following lines: MLB 17.5%, MLBPA 17.5%, all other professional leagues 13%, and International Baseball Federation 5%. The other 47% of revenue remaining would be allocated as prize money (Klein, 2006). Klein is quick to point out that on a team basis (30 in MLB while only 12 in NPB) the revenue sharing plan is quite fair.

It should be noted that Japan was given two seats on the WBC steering committee in an effort to improve the tournament and relations between MLBI and NPB. The next WBC will commence on March 5, 2009, with the winning team crowned champion on March 23, 2009. Venues include Tokyo, Mexico City, Toronto, San Juan, San Diego, Miami, and Los Angeles. Again, a 16-team format is being employed with the same nations scheduled to return (although initial pools have been slightly altered from the first WBC). MLBI has shifted the tournament from round-robin pool play to double-elimination pool play. This modification was undertaken in order to avoid awkward tie-breaking procedures in-place during the first tournament.

The WBC is scheduled to occur every four years (although, only three years elapsed between the inaugural tournament and the second occurrence of the WBC). This scheduling appears to mesh nicely with the schedules of other major international events—Summer Olympics will occur one year before the WBC, Winter Olympics will occur one year after the WBC, and the FIFA World Cup (men's tournament) will occur one year after the WBC.

MLB-NPB Merger

Currently a MLB-NPB merger seems quite unrealistic due to cultural, business, and logistical reasons. However, the situation has been occasionally mentioned and pondered. According to Whiting, a Japanese division could be formed and included in the American and or National League. Due to extremely long flight times as well as visa issues, a potential merger between MLB and NPB is highly unlikely (Whiting, 2004).

Asian League

Perhaps an Asian League, consisting of teams from the professional leagues in South Korea, Taiwan and Japan may be formed. The winner of the Asian League could then perhaps be invited to participate in the MLB postseason as a wildcard team (Whiting, 2004). Alan Klein also offers a suggestion of how to realign MLB on a global scale. Klein suggests creating a

"Pacific Rim" division consisting of most current MLB west coast teams and teams from Tokyo, Seoul, Taiwan and Sidney. Continuing on that structure, a "Pan-American" division consisting of teams from Florida, Tampa Bay, Texas, Houston, Arizona, Havana, Monterrey, and Caracas could be developed (Klein, 2006). Currently, this realignment seems far-fetched at best, however, it remains a very intriguing way of creating an actual "world" champion.

INTRODUCTION TO ECONOMETRIC STUDY

Many people believe that the sporting industry represents a true "melting pot" of diverse people and culture. The motivations of winning, both the pride of winning in amateur sports and the financial incentives in professional sports, are thought to rid the sporting world of any labor discrimination. Because of their popular appeal, sports often become the focus of scholarly research. Many research endeavors have focused on evaluating the degree of labor discrimination within various professional sports, towards a variety of ethnic and racial groups. In my review of the literature, I surveyed many of these articles. *I have chosen the issue of labor discrimination surrounding Japanese players playing in America's Major League Baseball as the topic of my research.*

The vast majority of labor discrimination research in professional baseball concerns itself with investigating the integration of African-Americans in Major League Baseball since Jackie Robinson first broke the color barrier in 1947. In addition, a small portion of the literature concerns itself with the issue of labor discrimination of Hispanic-born players in Major League Baseball. The specific issue of labor discrimination of Japanese-born baseball players in the United States' Major League Baseball (MLB) has gone largely unresearched. In some studies, players of Asian descent were even excluded on the basis of their being too small a sample size.

Horace Wilson introduced the Japanese people to the game of baseball in 1872 (Solloway, 2007) and yet, over the 135 years that have since passed, a mere thirty-four Japanese players have come to play in Major League Baseball. With baseball's strong popularity in Japan and a 2007 population base of over 127 million people (*Japan*, 2008), why has the number of Japanese players to come to America been so insignificant? Relations between Japan's Nippon Professional Baseball (NPB) League and America's Major League Baseball (MLB), as well as cultural differences between Japanese and American people may perhaps begin to answer a portion of this question. The issue of labor discrimination also indeed presents itself. Relevant questions that come to mind include:

- Have fewer Japanese-born players come to play in American Major League Baseball because they have faced labor discrimination or have fewer Japanese-born players come to play in American Major League Baseball because of other reasons?

- Is the recent surge of Japanese players coming to America due to the weakening of labor discrimination or has the surge been caused by other reasons?
- What fielding positions have Japanese players in Major League Baseball traditionally played and why has this been the case?

This study will employ econometric regression analysis to best attempt to answer these and other relevant questions concerning Japanese integration in Major League Baseball

LITERATURE REVIEW – ECONOMETRIC STUDY

Gerald Scully's 1974 pioneering article, *Pay and Performance in Major League Baseball,* has been heralded as one of the first significant articles examining salary determinants in professional sports. As the title of his work indicates, Scully focused his attention on Major League Baseball. Using data from the 1968 and 1969 MLB seasons, Scully estimated salary functions and marginal revenue products of the factors of production; he then found the rates of monopolistic exploitation by comparing salary and marginal revenue product over various performance levels and career lengths (Scully, 1974).

The most important part of this study, as it applies to my examination of discrimination towards Japanese players in Major League Baseball, resides in Scully's salary regression equation's specification. Scully attributes four main factors responsible for determining baseball players' salaries in Major League Baseball—hitting/pitching performance; the weight of the players' contributions to team performance; the number of years spent in the majors; and the greater bargaining power of "superstar" players (Scully, 1974). Scully ran two regressions, one for batters and one for pitchers.

For the batters equation, salary was regressed as the dependent variable with lifetime slugging percentage,[6] the weight of a hitter's contribution to team performance, experience (measured as years of Major League experience), and variables accounting for the size of team market, intensity of fan interest and membership in the then-superior National League, as the independent variables.[7]

The pitchers' regression again regressed salary as the dependent variable. The independent variables used were: the lifetime strikeout-to-walk ratio, lifetime average percentage of innings pitched out of total innings, experience (measured as years of Major League experience), size of team market, intensity of fan interest and a dummy variable for

[6] To include the more popular batting average statistic, Scully included a dummy variable if a player had above average batting average and a lesser than average slugging percentage.
[7] The log (base 10) of each variable was used in the regression equation and all performance measures were lagged one year to reflect the contractual nature of baseball contracts.

membership in the then-superior National League. Once again, performance measures were lagged and logs for all variables were used.

Scully's batters' equation accounted for 81% (R-squared=0.81) of the variation in players' salaries; the pitchers' equation accounted for 78% (R-squared=0.78) of the variation in players' salaries. The performance variables were all found to be significant at the 10% level with all but one significant at 1% significance or better. These high R-squared values and significant variables indicate that Scully's regression was well specified econometrically. The variables accounting for team market size, fan intensity and membership in the National League were all interestingly found insignificant. Overall, Scully's work confirmed the concept that ballplayer salaries are related to performance (Scully, 1974).

Using data from the 1976 season, Hill and Spellman investigated labor discrimination in Major League Baseball. Their study was conducted in a similar way to Scully's. Using both linear and logarithmic formats, only small wage differentials between black and white players were found to be attributable to discrimination (Hill & Spellman, 1984). Included as independent variables in the batters' specification were runs scored per year and years of major league experience. For pitchers, career earned run average, career innings pitched per year and years of major league experience were included as independent variables. Rookies were dropped from the regression. Hill and Spellman used both a log-linear and log-log specification and found the log-log specifications to have a higher R-squared values in most cases.

Lawrence M. Kahn's summary of all substantial previous research concerning labor discrimination in professional sports, published in his 1991 article in Industrial and Labor Relations Review, possesses tremendous value. In this survey of the literature, Kahn conducts a "study of studies" presenting the previously discovered findings (from Pascal and Rapping in 1972 to Kahn in 1989) from the scholarly community up through the late 1980's. In general, Kahn disseminates labor discrimination into three broad categorical forms—unequal pay for equal work, unequal hiring standards, and positional segregation. To a large extent, these three forms of labor discrimination, modeled specifically towards the issue of Japanese players in Major League Baseball, sparked my curiosity and prompted this study.

Let me focus on previous studies relating to Kahn's first form of labor discrimination, unequal pay for equal work. Most of the previous scholarly research has been conducted in a generally similar way.[8] This method can be described as regressing workers' salaries (or the natural log of workers' salary) as the dependent variable, and using varying measures of productivity as the independent variables. To examine the issue of discrimination, a dummy variable, or several if examining several races, (e.g. - White, African-American, Hispanic, Asian decent, etc.) is also included as an independent variable in the regression equation. The coefficient of the dummy variable, and its significance (or lack of significance) indicates the degree of discrimination (or labor equality) towards that race or group of people. Before including a dummy variable, separate regressions would be run for each group. Scholars then conduct a Chow test[9] to evaluate the difference in the two (or more) equations' estimated coefficients. If the Chow test provided evidence that the separate regressions' coefficients do not differ significantly, the various race observations were allowed to be pooled together and further analysis would be conducted.

Kahn's 1991 examination of previous "unequal pay for equal work" studies included a total of twenty-four prior studies; of these, twelve specifically inspect discrimination in professional baseball. These studies, and others published since Kahn's summary, will be the focus of my research; they are varied, including research on different groups of players such as free agents and non-free agent, as well as infielders, outfielders, and pitchers. The majority of the studies focus solely on black (or nonwhite) versus white discrimination (eg. Scully, Mogull, Medoff, Moguli, Cymrot, Raimondo, Hill and Spellman, Christiano) while a few others include Hispanic and/or Latino (Pascal and Rapping, Kahn) discrimination. Some of the studies examine returns to performance across races, while others examine returns to experience.

[8] Similar to that of Gerald W. Scully, and Hill and Spellman, discussed in detail above.
[9] The Chow Test is an econometric analytical tool named for its inventor, Gregory Chow. This test is a specialized version of the F test.

The results of these studies cannot be summarized in one or two sentences. To a large degree, the conclusions drawn from the studies contradict one another. A significant amount of the findings indicate a presence of salary discrimination in baseball while at the same time a significant amount of the findings report just the opposite. In an updated article in 2000, Kahn indicates that "…regression analysis of salaries in baseball…have not found much evidence of racial salary discrimination against minorities" (Kahn, 2000).

These differences can possibly be explained by the fact that the scholars employed differing statistical and econometrical analysis tools in their research—some used Chow tests, some two-tailed significance tests, some one-tailed significance tests, some no significance tests, etc. In addition, the use of different era and player datasets perhaps sheds light on why the conclusions between scholars differ to the extent that they do. The largest sample used in any of the studies was only 885 players while the smallest sample used consists of a mere 54 players. Perhaps more concrete conclusions can be drawn by increasing the sample size employed in this type of research.

The decision of which performance measures best indicate a player's performance and thus are most important to include in the regression as independent variables can be a heated debate topic. Many early researchers make the common mistake of including too many independent variables. If utilized, this "everything-but-the-kitchen-sink" mentality will produce an estimated regression equation with little value. I will examine traditional performance measures as well as contemporary research in sabermetrics, in order to determine which variables to utilize in my study.

The study of statistics in professional sports, specifically in baseball, has grown tremendously with the advent of computer technologies, rotisserie fantasy sport leagues, and the Internet (Luttrell, 1999). Statistical guru and lifetime baseball fan Bill James has been recognized as the pioneer of the "sabermetrics" movement. Loosely speaking, sabermetrics can be thought of as the use of extensive statistics in empirically quantifying new and old ideas in baseball thinking. James coined the terminology sabermetrics, after the Society for American Baseball Research or SABR (Ackman, 2007).

Oakland Athletics' front office, led by General Manager Billy Beane, and portrayed in Michael L. Lewis' best-seller, Moneyball, employs several new methodologies that challenge many of baseball's traditional thinkers. In general, this new way of thinking emphasizes generating runs for batters[10] and preventing runs for pitchers and defense, as opposed to many of the traditional statistical measurements that blatantly disregard these objectives[11] (McGrath, 2003).

Some of the most traditionally referenced measures of performance for pitchers are Wins (W), Loses (L), Games (G), Games Started (GS), Strikeouts (K), Saves (SV), Hits (H), Earned Runs (ER), Home Runs (HR), Walks (BB), Intentional Walks (IBB), Wild Pitches (WP), Hit by Pitch (HBP), Balks (BK), Batters Faced by Pitcher (BFP), Games Finished (GF) and Runs (R). In popular culture, these statistics gain much esteem; however, in recent years these statistics' ability to reflect truly on pitcher performance has been questioned by those from the sabermetric school of thought. Most pitching statistic measures are (at least somewhat, and some much more so than others) team dependent, and thus they fail to truly reflect the individual ability of the pitcher. To solve this problem, a new category of pitching statistics attempt to remove any teammate dependency and therefore better reflect the pitcher's ability. These statistics fall under the terminology "defensive independent pitching statistics" or DIPS (McCracken, 2001).

For batters, the traditional statistical categories have been: Games (G), At-Bats (AB), Runs (R), Hits (H), Doubles (2B), Triples (3B), Homeruns (HR), Batting Average (AVG), Runs Batted In (RBI), Runs Scored (RS), Stolen Bases (SB), Caught Stealing (CS), Walks (BB), Strikeouts (K), Intentional Walks (IBB), Hit by Pitch (HBP), Sacrifice Hits (SH), Sacrifice Fly (SF) and Ground into Double Plays (GIDP). New waves of "sabermetric" statistical categories have been developed to better stress the importance of scoring runs. These newer statistical categories include – Slugging Percentage (SLG), On-base Percentage (OBS), On-Base plus Slugging (OPS), and Runs Created (RC).

[10] Examples: On-base percentage and slugging percentage.
[11] Examples: Batting average and stolen bases.

It is important to avoid severe multicollinearity as well as omitted variable bias. For these reasons, the choices of which performance measures to include, as well as which specification to select, are of vast importance. Indeed, much of the work in developing a study such as this one lies in determining the variables to include and the specification to use. Sound economic and sporting theory should always dictate these decisions.

It is essential that the performance measures are measured and reported accurately in the datasets. If there are errors or missing data in the data sets, the estimated regression equation will lose its accuracy and significance in depicting how the system truly operates. As in any estimated regression equation the size of the sample dictates how much faith we can place in the output. All else being equal, a larger sample size is preferred. The earlier salary regression studies concerning labor discrimination in professional sports had smaller sample sizes than the more recently conducted studies (Kahn, 1991).

SALARY REGRESSION ANALYSIS

I have attained my dataset for this study from Sean Lahman's comprehensive collection of baseball statistical data available for free private-use online at http://www.baseball1.com. The database is available for download in both Microsoft Excel (CSV) and Microsoft Access (.mdb) format. This dataset summarizes a widespread amount of statistics of professional baseball—everything from player biographical information, statistical performance, postseason performance, awards to salary information is included. Much of this data is available for dates as far in the past as 1871!

The database was modified in order to properly establish the correct relationships between the many tables of data within the database. By establishing proper relationships within Microsoft Access, I was able to conduct many queries that provided the proper data when requested. Further modification of the database was necessary; for my research purposes I aggregated player statistics for players with multiple "stints"[12] or positions by year. Much credit and appreciation is due to Bryant University Computer Information Systems Professor and Honors Program Coordinator, Dr. Kenneth J. Sousa, for his earnest assistance in this process of database management.

In all instances, player data from 1995 to 2007 was used in the empirical research. Data from years prior to 1994 was eliminated because only one Japanese player, Masanori Murakam in 1964, had come to Major League Baseball prior to Hideo Nomo in 1995. Separate regressions have been run for batters and pitchers. Only batters with greater than or equal to 300 At-Bats (ABs) qualified for use in this study. Continuing, only data from pitchers with greater than or equal to 30 Innings-Pitched (IPs) were used in the pitchers' salary regressions.

Only players deemed to have some bargaining power in the salary process have been included in the regression specifications. In Major League Baseball, the team that initially drafts a player has exclusive rights of him for three years. After this three year period, the player may or may not be contractually bound by this same team for three additional years. The rules of free agency are indeed quite convoluted and oftentimes each player's individual situation will

[12] Trades between teams, call-ups / send-downs, in the same season

vary depending on how much playing time at the major league level he has received in the first three years of his career (Kendrick). The decision to include all players playing in at least their fourth year[13] in the regression specifications was made. It is very unlikely that a player would have 300 or more at-bats (for a batter) or pitch greater than 30 innings (for a pitcher) in a season, without possessing some form of salary negotiating rights (i.e.. arbitration).

A review of the literature does not reveal any theoretical problems with running Chow Tests for small samples; but, there is reason to doubt the combining of data with such discrepancies in observation sizes[14]. To make my findings more robust, I have run thirty additional regressions, ten for each of the three specifications. These additional regressions sampled 10% and 20% of the non-Japanese batters. Chow tests were then conducted on the additional regressions. Only two of the Chow Tests indicated that it is incorrect to assume equal coefficients. Thus, it appears that pooling the data and assuming equal coefficients is correct. For simplicity, this output has not been included in the Appendix; however, it may be furnished upon request[15].

The thirty MLB teams have been broken down into groups based on market size. To this extent, I employed the consolidated statistical metropolitan areas, as depicted by Al Streit in an article published on http://www.baseball-almanac.com, to break the teams into Large, Medium and Small markets (with ten teams per category). Finally, in order to fairly assess increasingly rising salary levels in MLB across several years, each player's salary has been converted to year 2000 dollars using GDP deflator information available from the United States Bureau of Economic Analysis (BEA) at http://www.bea.gov. Summary statistics regarding the variables used throughout the study is presented both on the following page and in the appendix.

---

[13] Note that all Japanese players were included because it is assumed that they have negotiating power when coming to play in Major League Baseball.
[14] For batters, the dataset includes 3,061 Non-Japanese observations and 22 Japanese observations. For pitchers, the dataset includes 2,471 Non-Japanese observations and 44 Japanese observations.
[15] Such inquiries should be sent via email to kaudet22@yahoo.com.

		Average Value - Batters			Average Value - Pitchers		
Variable	**Description**	**Japanese**	**Non-Japanese**	**Combined**	**Japanese**	**Non-Japanese**	**Combined**
Observations		**22**	**3,065**	**3,087**	**44**	**2,471**	**2,515**
Age	A player's age, in years.	31.31818	31.32822	31.32815	31.02273	30.8361	30.83936
Age^2	A player's age, in years, squared.	986.3182	992.7794	992.7334	972.0227	965.3962	965.5121
Japan	A dummy variable equal to 1 if the batter is Japanese.	1	0	0.0071267	1	0	0.017495
JapanxAge	An interaction variable equal to Japan*Age	31.31818	0	0.223194	31.02273	0	0.5427435
ln(ABper)	The natural log of the player's ratio of individual at-bats to team at-bats.	-2.611814	-2.814013	-2.812572			
ln(sal)	The natural log of the player's salary, lagged one year, converted to base year 2000 dollars using the GDP deflator.	14.99823	14.50726	14.51076	14.29548	14.38429	14.38274
ln(SLG)	The natural log of a player's slugging percentage.	-0.8472776	-0.8639729	-0.8638537			
lnIP	The natural log of a player's total innings-pitched.				4.70963	4.617302	4.618917
lnSW	The natural log of a player's strikeout to walk ratio.				0.7335818	0.7393728	0.7392715
RC_Bat	The number of runs generated by a player, computed as (Slugging percentage)*(On-base percentage)*(At-bats).	72.4511	60.93429	61.01637			
Runs	The number of runs generated by a player, computed as (0.41*singles)+(.82*doubles)+(1.06*triples)+(1.42*home runs).	80.70091	67.48931	67.58346			
Small	A dummy variable, equal to one, if the player's team is in the bottom third of the league teams' market sizes.	0.1818182	0.3187602	0.3177846	0.1136364	0.3269931	0.3232604
SV_pitch	A player's total number of saves in a season.				4.681818	4.323756	4.33002
W_pitch	A player's total number of wins in a season.				7.5	7.106839	7.113718

Batter Regressions

I will run three sets of regressions for batter—one set will employ the traditional baseball thinking (ie. Scully), the second will use more modern sabermetric statistical theory while the third set of regressions will be conducted by combining the best aspects of the aforementioned as well as my own thoughts and theories. Each set of regressions will have a Japanese, a non-Japanese and a combination equation estimated—thus a total of nine regressions will be initially run. Chow tests will then be performed to evaluate the degree to which the equations within each set differ. If it is possible to pool the data, this will be done and further regression will be run using Oaxaca decomposition. This process will allow quantitative evaluation of differences in salary level attributable between differences in characteristics (i.e. performance measures) and the differences attributable to other sources (i.e. discrimination).

First I will supply reasoning and justification detailing the merit of each specification's theory. This justification will be followed by detailing the findings of each regression. Tables 1, 2 and 3 detail the results from the three regression specifications. A summary table, combining these tables is also included.

Traditionalist Method Specification - Batters

I will follow Scully's methodology regarding the specification of the traditional baseball batter regression. Following Scully's lead, I will regress the natural log of player's salaries[16] as the dependent variable with slugging percentage, contribution to team (calculated as individual at-bats divided by total team at-bats), and experience[17] (proxied by age) as the independent variables. In addition, I have added age squared to account for the diminishing property of returns to age in Major League Baseball. In laymen's terms, as a player gets older, he receives higher salaries, but only up to a certain point. Eventually, if a player plays past his "prime", he will begin to earn decreasing salary levels. It is very important to lag the players' salaries so that this year's salary is aligned with last year's performance.[18] As mentioned earlier, only players with sufficient bargaining power were included in the regression analysis.

Many previous studies used the log of the independent variables in their specifications. This "log-log" specification provides for evaluating the effect of independent variable *percentage* changes with respect to the dependent variable (salary), in *percentage* terms. I have selected a semi-log specification—taking the log of some variables while using the actual variable for others. The natural logs of players' salary, slugging percentage, and team contribution have been used; the players' age has been measured simply in years.

The specification of the equation is as follows:[19]

$$\ln(salary) = \beta_0 + \beta_1 ln(SLG) + \beta_2 \ln\left(AB_{per}\right) + \beta_3(AGE) + \beta_4 AGE^2 + \varepsilon$$

The expected signs of the estimated coefficients were positive for each of the independent variables, except for the square of age. It makes sense to believe that a baseball player's

[16] (converted to base year 2000 dollars using the GDP deflator calculated from Bureau of Economic Analysis data, http://www.bea.gov)
[17] Age is preferred to MLB experience because of the tendency of Japanese players to develop their skills (for several seasons) in Japan's NPB prior to coming to MLB
[18] The downside of this lagging is a smaller sample size. Eight Japanese players' year's statistics were dropped in order to lag their salaries.
[19] Because Scully found fan intensity, market size and inclusion in the National League to be insignificant, I am excluding variables that would account for such factors.

salary will increase as he has a higher slugging percentage, more at-bats per team at-bats and gains seniority. A series of three regressions were run, one for Japanese players, one for non-Japanese players and one combining all batters. Next a Chow Test was performed to evaluate if pooling the data were possible. The regression output from STATA as well as the Chow Tests can be found in Appendix B.

Table 1 - Traditional Method Specification - Batters

Variable	Japanese	Non-Japanese	Pooled Sample
Constant	-33.059480	12.77249*	12.67581*
	(-1.02)	(13.68)	(13.57)
ln(Slugging Percentage)	2.885037	1.474381*	1.480376*
	(1.69)	(19.17)	(19.28)
ln(ABperteamAB)	0.7449075*	0.8940838*	0.8937928*
	(2.37)	(5.69)	(33.94)
Age	3.355589	0.3323444*	0.3391687*
	(1.68)	(5.69)	(5.81)
Age^2	-0.053374	-0.0049254*	-0.0050371*
	(-1.74)	(-5.44)	(-5.56)
R^2	0.5485	0.4781	0.4776
Adjusted R^2	0.4423	0.4774	0.4769
Number of Observations	22	3061	3083
Chow Test			
Critical F value	2.21		
F Statistic	1.601375374		
Conclusion	It is correct to assume equal coefficients.		

't' values in parentheses
* Significant at the 5% level

In the first regression run, of Japanese batters only, each of the independent variables' signs was in the expected direction. The signs of slugging percentage, at-bats per team at bats and age were positive, and the coefficient of the age squared variable was negative.

The only variable deemed significant at the 5% level was the batters' contribution to the team (AB/team AB). However, slugging percentage, age and age squared were all significant at the 11% level. The specification's theory is strong; this regression confirms this. However, an adjusted R-squared of 0.4423 is fairly low for such a small sample size. In the future, with more observations, more robust specifications will be able to be produced.

The regression of non-Japanese batters was very encouraging. Again, all of the independent variables' signs are in their expected direction. Furthermore, each variable is significant at the 1% level. The adjusted R-squared value for this regression was 0.4774. This regression, with 3,061 observations has a better fit than the Japanese batters' regression.

The third regression, of all batters, was similar to that of non-Japanese batters. All of the independent variables were in the expected direction and significant at the 1% level. The adjusted R-squared value was 0.4769. Scully was able to achieve much stronger levels of overall fit[20] and thus it appears that this specification is not without faults.

Finally, a Chow Test was conducted to determine the degree of difference between the constrained regression's estimated coefficients. An F statistic of approximately 1.61 was calculated. This F-value is less than the critical value of the F distribution (approximately 2.21). We can conclude that it is correct to pool the data. In terms of labor market discrimination, this conclusion supports the notion that the wage structure across Japanese and Non-Japanese batter populations is the same.

Traditional Method Specification - In terms of labor market discrimination, this conclusion supports the notion that the wage structure across Japanese and Non-Japanese batter populations is the **same**.

[20] Scully's R-squared values ranged from 0.78 to 0.81.

Sabermetric Method Specification - Batters

In choosing the specification for the sabermetric regression equation, I focused on the performance measure of "Runs". Runs is a composite statistic developed and created by sabermetrician George Lindsey. Essentially, the Runs statistic aims to capture the offensive production of a player or team, in terms of runs generated (Albert, 1999). Bill James (as well as others) has developed more complicated formulas of the "Runs Created" statistic. However, in this specification, Lindsey's simpler linear weight methodology has been utilized (see Appendix A for formula). Lindsey developed the formula in the 1960s using baseball data and probability theory (Albert, 1999).

A semilog specification will be used in order to determine the *unit* increase effects of runs generated and the *percentage* change effect of age on the *percentage* change on salary, holding all other variables constant. Because the runs statistic is thought to do a good job capturing a player's offensive production, it will be the only performance measure included. Because of the importance of player experience in predicting salary, age and the square of age were also included in the sabermetric specification.

The specification of the equation is as follows:

$$\ln(salary) = \beta_0 + \beta_1(Runs) + \beta_2(AGE) + \beta_3(AGE^2) + \varepsilon$$

Both runs and age are thought to contribute positively to a batter's salary. However, the positive age effect is limited and thus, the variable for the square of age is thought to have a negative coefficient. A series of three regressions were run, one for Japanese players, one for non-Japanese players and one combining all batters. Next a Chow Test was performed to evaluate if pooling the data were possible. The regression output from STATA as well as the Chow Tests can be found in Appendix B.

Table 2 - Sabermetric Method Specification - Batters			
Variable	**Japanese**	**Non-Japanese**	**Pooled Sample**
Constant	-46.795750	6.911381*	6.840096*
	(-1.63)	(7.96)	(7.87)
Runs	0.0151046*	0.0236116*	0.0235852*
	(3.00)	(61.56)	(61.69)
Age	3.857727*	0.351889*	0.3568111*
	(2.16)	(6.51)	(6.60)
Age^2	-0.0610776*	0.0050582*	-0.0051389*
	(2.16)	(-6.04)	(-6.13)
R^2	0.5600	0.5538	0.5532
Adjusted R^2	0.4867	0.5534	0.5527
Number of Observations	22	3065	3087
Chow Test			
Critical F value	2.37		
F Statistic	1.83414466		
Conclusion	It is correct to assume equal coefficients.		

't' values in parentheses
* Significant at the 5% level

The Japanese batter regression is quite promising as each explanatory variable's coefficient is in the expected direction and significant at the 5% level. However, the adjusted R-squared value is slightly low (for such a small sample) at 0.4867. Again, with more observations this specification could be strengthened.

The regression including all non-Japanese batters, with 3,065 observations, was much more robust in its specification. The estimated coefficients of all explanatory variables are in the expected direction and significant at the 1% level. The R-squared of this equation was significantly higher at 0.5534. The sabermetric method of thinking regression including all batters was very similar to that of the non-Japanese players' regression—both variables estimated coefficients positive and significant at the 1% level. The R-squared value of this regression was 0.5527.

Finally, a Chow Test was conducted to determine the degree of difference between the constrained regression's estimated coefficients. An F statistic of approximately 1.83 was calculated. This F-value is less than the critical value of the F distribution (approximately 2.37) at the 5% level. We can conclude that it is correct to pool the data. In terms of labor market discrimination, this conclusion supports the notion that the wage structure across Japanese and Non-Japanese batter populations is the same.

Sabermetric Method Specification - In terms of labor market discrimination, this conclusion supports the notion that the wage structure across Japanese and Non-Japanese batter populations is the **same**.

Hybrid Method Specification - Batters

After running the initial two specifications, the traditional and sabermetric, I sought to develop a superior specification. The traditional and sabermetric batter's regressions for Japanese players have very similar overall fits (adjusted R-squared values of 0.4423 and 0.4867, respectively). The same cannot be said for the traditional and sabermetric non-Japanese batter regressions. The sabermetric regression has a better overall fit, with an adjusted R-squared value of 0.5534 versus 0.4774 for the traditional method. Thus, it appears that a sabermetric statistical category like "runs" has more explanatory value than a traditional metric like slugging percentage.

In this new specification, I included a dummy variable to account for the vast spending differences of small-market MLB clubs. Major League Baseball is different from other major sports' leagues like the NBA, NFL and NHL in that no salary cap exists. The luxury tax system attempts to even the playing field between large and small market clubs' payroll. A tax is levied on teams that have payroll in excess of the statutorily set limits ($136.5 million in 2006) of the luxury tax system. Nonetheless, as any MLB fan can attest, the large market teams consistently outspend the small market teams.

Using Al Streit's breakdown of MLB teams ranked according to the size of the consolidated statistical metropolitan areas, I divided the thirty teams into three categories: Small, Medium, Large.

Large	Medium	Small
Baltimore Orioles	Arizona Diamondbacks	Cincinnati Reds
Chicago Cubs	Atlanta Braves	Cleveland Indians
Chicago White Sox	Boston Red Sox	Colorado Rockies
Los Angeles Angels of Anaheim	Detroit Tigers	Florida Marlins
Los Angeles Dodgers	Houston Astros	Milwaukee Brewers
New York Mets	Kansas City Royals	Minnesota Twins
New York Yankees	Philadelphia Phillies	Pittsburgh Pirates
Oakland Athletics	Seattle Mariners	San Diego Padres
San Francisco Giants	Texas Rangers	St. Louis Cardinals
Washington Nationals	Toronto Blue Jays	Tampa Bay Rays

This is not the perfect breakdown—being located in a large market does not necessarily mean that a club will have an elevated payroll (ie. Oakland Athletics). In general, it can be said that teams in the smallest markets have the lowest payrolls. The average salary for a player in the "Small" category is $2,934,577 compared to an average salary of $3,723,200 for a player in the "Large" category. It is therefore expected that the coefficient of the Small variable will be negative. Presented below is a summary of each category:

Market Size	Observation	Mean	Standard Deviation	Minimum	Maximum
Small	**981**	**2,934,577**	**2,757,838**	**116,139**	**15,500,000**
Medium	**1,046**	**3,566,399**	**3,547,483**	**157,211**	**21,500,000**
Large	**1,060**	**3,723,200**	**3,678,443**	**159,825**	**23,000,000**

This new "hybrid" specification is as follows:[21]

$$\ln(salary) = \beta_0 + \beta_1(Runs\ Created) + \beta_3(Small) + \varepsilon$$

The coefficient of Runs Created is expected to be positive while the coefficient of Small variable is expected to be negative.

[21] "Runs Created" has been used instead of "Runs" because using Runs Created better depicted the relationship between offensive production and salary level in this case. See Appendix A for the Runs Created formula.

Table 3 - Hybrid Method Specification - Batters			
Variable	**Japanese**	**Non-Japanese**	**Pooled Sample**
Constant	14.35573*	13.18182*	13.18573*
	(32.04)	(464.47)	(465.32)
Runs Created	0.0119007*	0.0221799*	0.022163*
	(2.36)	(59.57)	(59.72)
Small	-1.208448*	-0.8940838*	-0.0858057*
	(-2.50)	(-2.79)	(-2.93)
R^2	0.5737	0.5386	0.5383
Adjusted R^2	0.5288	0.5383	0.5380
Number of Observations	22	3065	3087
Chow Test			
Critical F value	2.60		
F Statistic	1.966999723		
Conclusion	It is correct to assume equal coefficients.		

't' values in parentheses
* Significant at the 5% level

The regression results were very promising. In all cases, the estimated coefficients' signs were in the expected direction and all of the estimated coefficients were significant at the 3% level. The adjusted R-squared values were 0.5288 for the Japanese batter regression, 0.5383 for the non-Japanese batter regression and 0.5380 for the combined batter regression. A Chow Test was conducted to determine the degree of difference between the constrained regression's estimated coefficients. An F statistic of approximately 1.967 was calculated. This F-value is less than the critical value of the F distribution (approximately 2.60) at the 5% level. We can conclude that it is correct to pool the data. In terms of labor market discrimination, this conclusion supports the notion that the wage structure across Japanese and Non-Japanese batter populations is the same. A table providing a summary of this section's results is displayed on the following page; another summary is provided in the conclusion to this report.

In terms of labor market discrimination, this conclusion supports the notion that the wage structure across Japanese and Non-Japanese batter populations is the **same**.

	Traditional Method Specification			Sabermetric Method Specification			Hybrid Method Specification		
Variable	**Japanese**	**Non-Japanese**	**Pooled Sampe**	**Japanese**	**Non-Japanese**	**Pooled Sample**	**Japanese**	**Non-Japanese**	**Pooled Sample**
Constant	-33.059480	12.77249*	12.67581*	-46.795750	6.911381*	6.840096*	14.35573*	13.18182*	13.18573*
	(-1.02)	(13.68)	(13.57)	(-1.63)	(7.96)	(7.87)	(32.04)	(464.47)	(465.32)
ln(SLG)	2.885037	1.474381*	1.480376*						
	(1.69)	(19.17)	(19.28)						
ln(ABper)	0.7449075*	0.8940838*	0.8937928*						
	(2.37)	(5.69)	(33.94)						
Age	3.355589	0.3323444*	0.3391687*	3.857727*	0.351889*	0.3568111*			
	(1.68)	(5.69)	(5.81)	(2.16)	(6.51)	(6.60)			
Age^2	-0.053374	-0.0049254*	-0.0050371*	-0.0610776*	0.0050582*	-0.0051389*			
	(-1.74)	(-5.44)	(-5.56)	(2.16)	(-6.04)	(-6.13)			
Runs				0.0151046*	0.0236116*	0.0235852*			
				(3.00)	(61.56)	(61.69)			
Runs Created							0.0119007*	0.0221799*	0.022163*
							(2.36)	(59.57)	(59.72)
Small							-1.208448*	-0.8940838*	-0.0858057*
							(-2.50)	(-2.79)	(-2.93)
R^2	0.5485	0.4781	0.4776	0.5600	0.5538	0.5532	0.5737	0.5386	0.5383
Adjusted R^2	0.4423	0.4774	0.4769	0.4867	0.5534	0.5527	0.5288	0.5383	0.5380
Number of Observations	22	3061	3083	22	3065	3087	22	3065	3087
Critical F value		2.21			2.37			2.60	
F Statistic		1.601375374			1.83414466			1.966999723	
Conclusion	It is correct to assume equal coefficients.			It is correct to assume equal coefficients.			It is correct to assume equal coefficients.		

't' values in parentheses * Significant at the 5% level

Oaxaca Decomposition - Batters

A powerful tool in labor discrimination analysis, called the Oaxaca Decomposition Model, was developed by Ronald Oaxaca in the 1970s. Ronald Oaxaca's techniques have been employed in this study in order to break down the differences in compensation between Japanese and Non-Japanese baseball players into two parts—one that can be attributed to differences in characteristics[22] and one that cannot. The residual component represents discrimination—the difference in compensation not attributable to differences in characteristics (Oaxaca, 1973).

Oaxaca decomposition analysis has been conducted on the sabermetric batter regression specification. As a reminder the sabermetric batter regression is:

$$\ln(salary) = \beta_0 + \beta_1(Runs) + \beta_2(AGE) + \beta_3(AGE^2) + \varepsilon$$

The sabermetric specification was chosen as the researcher deemed it the most robust specification. Only one batter regression was chosen to undergo this analysis because the Chow Tests have revealed that pooling the batter data was appropriate. This supports the notion that Japanese batters face little discrimination. Nonetheless, Oaxaca decomposition was still employed to increase the vigor of this study. The complete results and Stata output of the Oaxaca decomposition can be found in Appendix B.

Before explaining the results of this analysis, a question relevant to the theory of such analysis warrants additional explanation. The main objection to running Oaxaca decomposition in this instance is caused by the small sample size of Japanese batter observations.[23] Stemming from the small sample size, the significance of the explanatory variables in the sabermetric batter regression is questionable. To help mitigate this problem, the statistical technique of random resampling (i.e. bootstrapping)[24] was utilized. Unfortunately, the bootstrapping breakdown failed to result in strongly significant estimated coefficients for the Japanese batter sabermetric specification.

[22] As measured by the regression specification's explanatory variables
[23] The batter dataset includes only 22 Japanese batter observations.
[24] An iterative process in which a series of random samplings are taken and linear regression is subsequently run on the random samplings.

The Oaxaca decomposition Stata output indicates that the difference in the *natural log of salary*, between Japanese and Non-Japanese batters, is equal to 0.4895785. Because the salary variable is expressed in log form, this means that Non-Japanese batters are paid salaries 48.95% higher than Japanese batters. This fact alone does not indicate the presence of discrimination whatsoever. For example, certainly no discrimination would exist if we could prove that 100% of the difference in pay is attributable to differences in skill level—this is where the Oaxaca decomposition comes in.

Some explanation of the interpretation of Oaxaca decomposition is needed. From a theoretical perspective, there are four ways to analyze the difference in salary level between two groups. We could compare group A to group B; conversely, we could compare group B to group A. Thirdly, could determine some half-way point and compare each group to that point. And finally, we could apply weights to the observations and compare the results using the weighted point of view. In this study, largely due to the small sample of Japanese player data, it makes the most sense to analyze the output from the weighted point of view.

The weighted results indicate that of the 0.4896 difference in salary level between the two groups, .2305 (47.09%) can be explained by the difference in batter characteristics[25] and 0.2590 (52.91%) cannot be explained by differences in batter characteristics. In a discrimination-free labor market, 100% of the difference in salary level would be attributable to differences in performance characteristics. Because only 47.09% of the difference in salary level can be attributable to differences in batter characteristics, it appears that Japanese batters face salary discrimination in Major League Baseball.

> Oaxaca Decomposition - It appears that Japanese batters face salary discrimination in Major League Baseball.

[25] As indicated by the included explanatory variables

Pitcher Regressions

A similar framework to the batter regressions was employed as pitcher regressions were developed. Again here three sets of regressions were initially run—a traditional specification,[26] a sabermetric specification, and finally a hybrid specification utilizing the best aspects of each prior regression along with my own theory and ideas. Each specification will include three separate regression components—a Japanese pitcher, a non-Japanese pitcher and a pooled pitcher regression. Chow Tests will be run for each specification set in order to evaluate the equivalency of estimated variable coefficients. If pooling the data is deemed appropriate, subsequent Oaxaca decomposition will be conducted to quantify the degree of discrimination present.

The dataset has been modified to only include pitchers with ample playing time. This will in effect eliminate any outlying observations (i.e. pitchers who accumulated no or little pitching statistics because of injury, and September call up pitchers). The cut-off point of thirty innings-pitched was deemed appropriate by the researcher. In addition, only observations with lagged salaries and at least four years of experience[27] were included. In making these modifications to the dataset, a total of 2,381 observations were dropped. In total, there were 2,515 observations—2,471 Non-Japanese pitchers and 44[28] Japanese pitchers in the dataset.

Each regression specification with detailed justification and results will be presented next. The specifications will be presented in the following order—traditional method, sabermetric method, hybrid method. Summary tables, presenting the results of each specification's regressions will be shown in tables 4, 5 and 6. The detailed STATA output, along with Chow Test calculations may be found in Appendix C.

[26] Largely modeled off previous studies, namely Scully's.
[27] To include only pitchers with sufficient bargaining power. Japanese players were included regardless of MLB experience.
[28] 18 Japanese observations were dropped because they lacked a lagged salary.

Traditional Method Specification – Pitchers

For pitchers, as with batters, I will model the "traditionalist" baseball method of thinking regression specification after Scully's initial model. In Scully's pitcher regression, he regressed the log of players' salary as the dependent variable. The independent variables used to account for pitcher performance were the lifetime strikeout-to-walk ratio and the lifetime average percentage of innings pitched out of total innings.[29] In Scully's regression, variables to account for team market size, fan intensity and inclusion in the National League were included. However, these variables were found to be insignificant and will not be included in my specification.

My regression will include mostly similar variables to Scully's, however, I feel that including lifetime performance measures do not accurately depict the economic reality of the baseball world. In today's baseball world, your lifetime statistics are not as important as your current statistics. Baseball can be said to be a "what-have-you-done-for-me-lately?" business. Thus, these performance variables will be included on a season basis. Included variables are the ratio of strikeouts to walks (SW), innings pitched (IP), age,[30] and age-squared. The log of all independent performance variables is used in order to interpret percentage changes of these variables on salary, in percentage terms. The specification of the equation is as follows:

$$\ln(salary) = \beta_0 + \beta_1 \ln(\mathrm{SW}) + \beta_2 \ln(\mathrm{IP}) + \beta_3(\mathrm{AGE}) + \beta_4(\mathrm{AGE}^2) + \varepsilon$$

As indicated in the equation, the estimated coefficient of each variable, with the exception of age-squared, is expected to be positive. The estimated coefficient of age-squared is expected to be negative to depict the eventual diminishing returns to age in Major League Baseball.

A series of three regressions were run, one for Japanese players, one for non-Japanese players and one combining all pitchers. Next a Chow Test was performed to evaluate if pooling the data were possible. The full regression output from STATA as well as the Chow Test can be found in Appendix C. Presented below is a summarizing table of the results.

[29] Measured as nine x games x years.
[30] Again in this instance, age was used, as opposed to experience, to better reflect the tendency of Japanese players to develop in Japan for many years prior to coming to MLB.

Table 4 -Traditional Method Specification - Pitchers

Variable	Japanese	Non-Japanese	Pooled Sample
Constant	-18.491540	1.322308	1.135362
	(-1.70)	(1.41)	(1.22)
lnSW	0.342087	0.6271473*	0.6301392*
	(0.98)	(15.16)	(15.35)
lnIP	0.430970	0.9138889*	0.9043861*
	(1.72)	(30.25)	(30.36)
Age	1.847313*	0.4754024*	0.3886703*
	(2.57)	(8.28)	(.0571649)
Age^2	-0.0275738*	-0.0065061*	-0.006697*
	(-2.34)	(-7.28)	(-7.52)
R^2	0.3705	0.3751	0.3709
Adjusted R^2	0.3060	0.3741	0.3699
Number of Observations	44	2471	2515
Chow Test			
Critical F value		2.21	
F Statistic		2.769937625	
Conclusion		It is incorrect to assume equal coefficients.	

't' values in parentheses
* Significant at the 5% level

The Japanese pitcher regression was not strikingly strong, however, it was not terribly weak. The estimated coefficients of both age and age-squared were significantly in their respective directions, at the 5% level. The other variables were in their expected direction, albeit not significantly. The adjusted R-squared is quite weak, 0.3060. The non-Japanese pitcher regression output is quite strong. Each estimated coefficient was significantly in its respective direction, at the 5% level. The adjusted R-squared was 0.3741, with 2,471 observations. The pooled data largely mirrors the non-

Traditional Method Specification - In terms of labor market discrimination, this conclusion supports the notion that the wage structure across Japanese and Non-Japanese pitcher populations is **not the same**.

Japanese pitcher regression.

The important result from this regression lies in the Chow Test. The calculated F-statistic is greater than the critical F value. Therefore, it is incorrect to assume equal coefficients. In terms of labor market discrimination, this conclusion supports the notion that the wage structure across Japanese and Non-Japanese pitcher populations is not the same.

It appears that the determinants of a pitcher's salary differ between Japanese and non-Japanese players. This issue will be investigated further via Oaxaca decomposition.

Sabermetric Method Specification – Pitchers

A sabermetric specification, utilizing such well-developed statistics as defensive-independent component earned run average (DICE), and walks and hits per innings pitched (WHIP) was attempted to be sculpted. However, after modeling several different specifications, combining various sabermetric pitching statistics and other independent variables, no conclusive specification could be created. The specifcaitions run seemed to have strong predictive value for *either* the Japanese pitcher dataset or the non-Japanese pitcher dataset, but not both. This tendency was evident in nearly each sabermetric specification run. Perhaps this indicates that the statistics that predict pitcher salary vary quite widely between Japanese and non-Japanese pitchers. This may yield a conclusion of discrimination—unequal pay for equal work in the case of pitchers. The issue of discrimination will be investigated in the Oaxaca decomposition of the two other pitcher regression specifications employed.

Sabermetric Method Specification – No conclusive labor market discrimination conclusions could be reached.

Hybrid Method Specification – Pitchers

The traditional specification includes the strikeout-to-walk ratio and innings-pitched of a pitcher. By including these two variables as the only performance measures, overpowering pitchers are favored. Traditional staff "aces" tend to have overpowering fastballs and strikeout many batters, walking few, and aggregating many innings-pitched throughout the season. In sculpting my "hybrid" specification, I sought to develop a specification that better reflects baseball reality. A baseball team succeeds when it wins games, thus a pitcher's wins was included as an independent variable. The wins statistics favors starters as opposed to relievers, so saves was also included as an independent variable. Although one may counter that a pitcher's wins is affected by his team's defensive ability, and thus may not truly depict his talent, wins are still a largely cited statistic in evaluating a pitcher's ability.

Continuing on the logic built in the hybrid batter specification, a variable to account for team market size was included. This dummy variable was equal to one if a team is in the smallest third of league markets. The specification is as follows:

$$\ln(salary) = \beta_0 + \beta_1(\text{Wins}) + \beta_2(\text{Saves}) + \beta_3(\text{AGE}) + \beta_4(\text{AGE}^2) + \beta_5(Small) + \varepsilon$$

A series of three regressions were run, one for Japanese players, one for non-Japanese players and one combining all batters. Next a Chow Test was performed to evaluate if pooling the data were possible. The full regression output from STATA as well as the Chow Test can be found in Appendix C. Presented below is a summarizing table of the results.

Table 5 - Hybrid Method Specification - Pitchers			
Variable	**Japanese**	**Non-Japanese**	**Pooled Sample**
Constant	-17.507370	5.811774*	5.677091*
	(-1.90)	(6.72)	(6.61)
W_pitch	0.0267386*	0.133005*	0.1322973*
	(2.78)	(40.22)	(40.28)
SV_pitch	0.0267386*	0.0325574*	0.0326926*
	(2.43)	(20.61)	(20.88)
Age	1.923978*	0.4298984*	0.4378198*
	(0.6097785)	(7.92)	(8.11)
Age^2	-0.029257*	-0.157071*	-0.0060367*
	(-2.94)	(-7.00)	(-7.18)
Small Market Team	-0.8500109*	-0.157071*	-0.1565574*
	(-2.45)	(-4.63)	(-4.64)
R^2	0.5625	0.4408	0.4391
Adjusted R^2	0.5050	0.4396	0.4379
Number of Observations	44	2471	2515
Chow Test			
Critical F value		2.10	
F Statistic		2.303232731	
Conclusion	It is incorrect to assume equal coefficients.		

't' values in parentheses
* Significant at the 5% level

The regression results were quite strong in each of the three regressions. *In every instance,* the coefficient of each variable was significant, at the 5% level and in its expected direction. The adjusted R-square values were higher than the traditional specification in each regression. Overall this regression appears to be better than the traditional method specification. Again, the important result from this regression lies in the Chow Test. The calculated F-statistic is greater than the critical F value. Therefore, it is incorrect to assume equal coefficients. In terms of labor market discrimination, this conclusion supports

> Hyrbid Method Specification – In terms of labor market discrimination, this conclusion supports the notion that the wage structure across Japanese and Non-Japanese pitcher populations is **not the same**.

the notion that the wage structure across Japanese and Non-Japanese pitcher populations is not the same. It appears that the determinants of a pitcher's salary differ between Japanese and non-Japanese players. This issue will be investigated further via Oaxaca decomposition. A table providing a summary of this section's results is displayed on the following page.

	Traditional Method Specification			Hybrid Method Specification		
Variable	**Japanese**	**Non-Japanese**	**Pooled Data**	**Japanese**	**Non-Japanese**	**Pooled Data**
Constant	-18.491540	1.322308	1.135362	-17.507370	5.811774*	5.677091*
	(-1.70)	(1.41)	(1.22)	(-1.90)	(6.72)	(6.61)
lnSW	0.342087	0.6271473*	0.6301392*			
	(0.98)	(15.16)	(15.35)			
lnIP	0.430970	0.9138889*	0.9043861*			
	(1.72)	(30.25)	(30.36)			
Age	1.847313*	0.4754024*	0.3886703*	1.923978*	0.4298984*	0.4378198*
	(2.57)	(8.28)	(.0571649)	(0.6097785)	(7.92)	(8.11)
Age^2	-0.0275738*	-0.0065061*	-0.006697*	-0.029257*	-0.157071*	-0.0060367*
	(-2.34)	(-7.28)	(-7.52)	(-2.94)	(-7.00)	(-7.18)
Small				-0.8500109*	-0.157071*	-0.1565574*
				(-2.45)	(-4.63)	(-4.64)
W_pitch				0.0267386*	0.133005*	0.1322973*
				(2.78)	(40.22)	(40.28)
SV_pitch				0.0267386*	0.0325574*	0.0326926*
				(2.43)	(20.61)	(20.88)
R^2	0.3705	0.3751	0.3709	0.5625	0.4408	0.4391
Adjusted R^2	0.3060	0.3741	0.3699	0.5050	0.4396	0.4379
Number of Observations	44	2471	2515	44	2471	2515

Chow Test

	Traditional Method Specification	Hybrid Method Specification
Critical F value	2.21	2.10
F Statistic	2.769937625	2.303232731
Conclusion	It is incorrect to assume equal coefficients.	It is incorrect to assume equal coefficients.

't' values in parentheses * Significant at the 5% level

Oaxaca Decomposition – Pitchers

Following the logic outlined in the Oaxaca decomposition analysis of the batter data, the pitcher data has been analyzed. The same commentary regarding theoretical and interpretation issues apply in this case. Oaxaca decomposition analysis has been conducted on both of the original pitchers regressions. The full Stata output can be found in Appendix C. As a reminder the two pitcher regression specifications are:

Traditional Method: $\ln(salary) = \beta_0 + \beta_1 \ln(\text{SW}) + \beta_2 \ln(\text{IP}) + \beta_3(\text{AGE}) + \beta_4(\text{AGE}^2) + \varepsilon$

Hybrid Method: $\ln(salary) = \beta_0 + \beta_1(\text{Wins}) + \beta_2(\text{Saves}) + \beta_3(\text{AGE}) + \beta_4(\text{AGE}^2) + \beta_5(Small) + \varepsilon$

The results can be found below; complete Stata output is located in Appendix C.

The Oaxaca decomposition Stata output indicates that the difference in the *natural log of salary*, between Japanese and Non-Japanese pitchers, is equal to 0.0888. Because the salary variable is expressed in log form, this means that Non-Japanese pitchers are paid salaries 8.88% higher than Japanese pitchers. This fact alone does not indicate the presence of discrimination whatsoever. For example, certainly no discrimination would exist if we could prove that 100% of the difference in pay is attributable to differences in skill level—this is where the Oaxaca decomposition comes in.

Traditional Method

The Oaxaca output from both the traditional and hybrid pitcher specifications is very interesting. The traditional method's data will be presented first. The weighted results indicate that of the 0.0888 difference in salary level between the two groups, -0.1267 (-143.74%) can be explained by the difference in pitcher characteristics[31] and 0.2154 (244.48%) cannot be explained by

> Oaxaca Decomposition (Traditional Method Specification) – Japanese pitchers face **high levels of salary discrimination**.

[31] As indicated by the included explanatory variables

differences in pitcher characteristics. The interpretation of such results can take some time to get used to. This situation is equivalent to a group of *highly educated* workers being *paid less* than a group of *less educated* workers. This situation goes beyond discrimination. The Japanese pitchers exhibit superior pitching talent and receive less! Over 200% of the difference in salary level cannot be attributed to differences in pitcher performance!

Hybrid Method

The results of the hybrid method specification mirror those of the traditional method specification. The weighted results indicate that of the 0.0888 difference in salary level between the two groups, -0.1388 (-156.32%) can be explained by the difference in pitcher characteristics[32] and 0.2276 (256.32%) cannot be explained by differences in pitcher characteristics. Because over 200% of the difference in salary level cannot be attributable to differences in pitcher characteristics, it appears that Japanese pitchers face a high degree of salary discrimination in Major League Baseball.

Oaxaca Decomposition (Hybrid Method Specification) – Japanese pitchers face **high levels of salary discrimination**.

[32] As indicated by the included explanatory variables

Evaluating Average Difference in Salary Level

These regressions investigate the average difference in salary level between Japanese and non-Japanese (a) batters and (b) pitchers.

Batters

A dummy variable, "Japan", equal to one if a player is Japanese and zero if a player is non-Japanese, was added to each pooled batter regression specification previously run. This dummy variable will be tacked-on to each regression and the results will be interpreted. I hypothesize that the dummy variable's estimated coefficient will be positive. This makes sense as it is likely that a Major League Baseball team would be willing to pay a Japanese player a premium to entice him to play in America. A plausible explanation for this is that perhaps American baseball teams feel that they can capitalize from an affluent Asian fan base if their team includes Japanese players. The same hypotheses regarding the remaining independent variables' estimated coefficients from the each pooled regression still apply. Table 6 displays a summary of the effect of including the "Japan" dummy variable.[33]

Table 6 – Average Differences in Salary Level - Batters

Regression	Estimated Coefficient of "Japan"	Sign	P-value	Effect of adjusted R^2
Traditional	0.2594886	+	0.132	Increase
Sabermetric	0.1500573	+	0.347	No Effect
Hybrid	-0.0316807	–	0.317	Increase

[33] The full STATA outputs may be found in Appendix D.

	Traditional Method Specification			Sabermetric Method Specificaition			Hybrid Method Specification		
Variable	**Japanese**	**Non-Japanese**	**Pooled Sample**	**Japanese**	**Non-Japanese**	**Pooled Sample**	**Japanese**	**Non-Japanese**	**Pooled Sample**
Constant	-33.059480	12.77249*	12.71623*	-46.795750	6.911381*	6.866354*	14.35573*	13.18182*	-1.127553*
	(-1.02)	(13.68)	(13.62)	(-1.63)	(7.96)	(7.90)	(32.04)	(464.47)	(-202.98)
ln(SLG)	2.885037	1.474381*	1.481094*						
	(1.69)	(19.17)	(19.3)						
ln(ABper)	0.7449075*	0.8940838*	0.8927026*						
	(2.37)	(5.69)	(33.90)						
Age	3.355589	0.3323444*	0.3363722*	3.857727*	0.351889*	0.3551582*			
	(1.68)	(5.69)	(5.76)	(2.16)	(6.51)	(6.57)			
Age^2	-0.053374	-0.0049254*	-0.0049939*	-0.0610776*	0.0050582*	-0.0051134*			
	(-1.74)	(-5.44)	(-5.51)	(2.16)	(-6.04)	(-6.10)			
Runs				0.0151046*	0.0236116*	0.0235732*			
				(3.00)	(61.56)	(61.62)			
Runs Created							0.0119007*	0.0221799*	0.0042926*
							(2.36)	(59.57)	(59.06)
Small							-1.208448*	-0.8940838*	0.0052367*
							(-2.50)	(-2.79)	(0.91)
Japan			**0.2594886**			**0.1500573**			**-0.0316807**
			(1.51)			**(0.94)**			**(-1.00)**
R^2	0.5485	0.4781	0.4780	0.5600	0.5538	0.5533	0.5737	0.5386	0.5314
Adjusted R^2	0.4423	0.4774	**0.4771**	0.4867	**0.5534**	0.5527	0.5288	0.5383	**0.5309**
Number of Observations	22	3061	3083	22	3065	3087	22	3065	3087

't' values in parentheses * Significant at the 5% level

The conclusion of these regressions, regarding the average differences in salary level between Japanese and non-Japanese batters, is inconclusive. In some cases, the inclusion of the dummy variable strengthened the overall fit of the specification,[34] while compromising the significance of individual estimated coefficients of some of the variables. In some cases the estimated coefficient of the Japan dummy variable was in the expected direction, however, the significance of this coefficient is weak. We cannot conclude that Japanese batters are paid a premium over equally-skilled, non-Japanese players.

We cannot conclude that Japanese batters are paid a premium over equally-skilled, non-Japanese players.

[34] As measured by the adjusted R-squared.

Pitchers

A dummy variable, "Japan", equal to one if a player is Japanese and zero if a player is non-Japanese, was added to each pooled pitcher regression specification previously run. This dummy variable will be tacked-on to each regression and the results will be interpreted. I hypothesize that the dummy variable's estimated coefficient will be positive. This makes sense as it is likely that a Major League Baseball team will be willing to pay a Japanese player a premium to entice him to play in America. A plausible explanation for this is that perhaps American baseball teams feel that they can capitalize from an affluent Asian fan base if their team includes Japanese players. The same hypotheses regarding the remaining independent variables' estimated coefficients from the each pooled regression still apply. Table 7 displays a summary of the effect of including the "Japan" dummy variable.[35] The more complete data is presented on the subsequent page.

Table 7 – Average Difference in Salary Level, Pitchers

Regression	Estimated Coefficient of "Japan"	Sign	P-value	Effect of adjusted R^2
Traditional	-0.2159403	—	0.088	Increased
Hybrid	-0.2288458	—	0.056	Increased

All else equal, Japanese pitchers are **paid less** than comparable non-Japanese.

In both instances, the coefficient of the "Japan" dummy variable is negative. This suggests that all else being equal (performance, age, and team market), Japanese pitchers are paid *less* than comparable non-Japanese pitchers. These results are very interesting, although they disprove my original hypothesis. Perhaps teams feel that the cultural and language barriers between Japanese and non-Japanese pitchers warrants discounting their pitching ability. Furthermore, perhaps Japanese pitchers are unfairly viewed as having less skill during the hiring (scouting in the case of Major League Baseball), process. Further investigation into the discriminatory pay practices against Japanese pitchers can be found in the Oaxaca decomposition performed in this study.

[35] The full STATA outputs may be found in Appendix D.

	Traditional Method Specification			Hybrid Method Specification		
Variable	**Japanese**	**Non-Japanese**	**Pooled Sample**	**Japanese**	**Non-Japanese**	**Pooled Sample**
Constant	-18.491540	1.322308	1.070845	-17.507370	5.811774*	5.616929*
	(-1.70)	(1.41)	(1.15)	(-1.90)	(6.72)	(6.54)
lnSW	0.342087	0.6271473*	0.6298058*			
	(0.98)	(15.16)	(15.34)			
lnIP	0.430970	0.9138889*	0.9057056*			
	(1.72)	(30.25)	(30.41)			
Age	1.847313*	0.4754024*	0.4925648*	1.923978*	0.4298984*	0.441934*
	(2.57)	(8.28)	(8.61)	(0.6097785)	(7.92)	(8.19)
Age^2	-0.0275738*	-0.0065061*	-0.0067567*	-0.029257*	-0.157071*	-0.0061009*
	(-2.34)	(-7.28)	(-7.58)	(-2.94)	(-7.00)	(-7.25)
Small				-0.8500109*	-0.157071*	-0.1604358*
				(-2.45)	(-4.63)	(-4.75)
W_pitch				0.0267386*	0.133005*	0.1323636*
				(2.78)	(40.22)	(40.32)
SV_pitch				0.0267386*	0.0325574*	0.0327139*
				(2.43)	(20.61)	(20.90)
Japan			**-0.2159403**			**-0.2288458**
			(-1.71)			**(-1.92)**
R^2	0.3705	0.3751	0.3717	0.5625	0.4408	0.4399
Adjusted R^2	0.3060	0.3741	**0.3704**	0.5050	0.4396	**0.4385**
Number of Observations	44	2471	2515	44	2471	2515

't' values in parentheses * Significant at the 5% level

Evaluating Returns to Age

These regressions investigate the difference in returns to age between Japanese and non-Japanese (a) batters and (b) pitchers.

Batters

An interaction variable was added to each pooled batter regression specification previously run.[36] These specifications include an interaction variable to account for the different returns to age for Japanese players. I hypothesize that the estimated coefficient of this variable will be positive. When a team hires a young player (Japanese or otherwise), the team bears a large risk due to the uncertainty associated with predicting that player's future performance (i.e. will success in college/high school translate to the major leagues, will that player adapt to professional baseball life, etc.). International scouting divisions are not as robust as their domestic counterparts; therefore, team's have even less information to use in evaluating foreign prospective players' future performance prospects.

The cultural change from playing baseball in Japan to playing in America (or vice versa) is a large one. Indeed, much of Robert Whiting's You Gotta Have Wa, and The Meaning of Ichiro, is devoted to explaining some of the major cultural differences between the two nations' baseball worlds. Therefore, it is logical to believe that a Major League Baseball team would pay more for proven veteran Japanese players (i.e. older players), who have developed and proven their skill set on American turf, in Major League Baseball, as opposed to a Japanese player coming straight out of Japan. Table 8 displays a summary of the effect of including the "Japan" dummy variable.[37] The more complete data is presented on the subsequent page.

[36] The original specifications, as opposed from the specifications designed to evaluate average differences in salary level between Japanese and non-Japanese players (the immediately previous section).
[37] The full STATA outputs may be found in Appendix E.

Table 8 – Returns to Age, Batters				
Regression	Estimated Coefficient of "Japan*Age"	Sign	P-value	Effect of adjusted R^2
Traditional	0.0076373	+	0.164	Increase
Sabermetric	0.0043989	+	0.386	No Effect
Hybrid	0.0068705	+	0.183	Increase

These results are mixed. The inclusion of the interaction term never hurt the specifications overall fit[38]. In all instances the estimated coefficient of the interaction term was positive, indicating that the returns to age for a Japanese batter are greater than the returns to age for a non-Japanese batter. However, the significance of the variable was slightly weak in two cases and very weak in one case.

> We cannot conclude that Japanese batters receive higher returns to age than non-Japanese.

[38] As measured by the adjusted R-squared value.

	Traditional Method Specification			Sabermetric Method Specification			Hybrid Method Specification		
Variable	**Japanese**	**Non-Japanese**	**Pooled Sample**	**Japanese**	**Non-Japanese**	**Pooled Sample**	**Japanese**	**Non-Japanese**	**Pooled Sample**
Constant	-33.059480	12.77249*	12.71332*	-46.795750	6.911381*	6.864153*	14.35573*	13.18182*	13.18464*
	(-1.02)	(13.68)	(13.61)	(-1.63)	(7.96)	(7.90)	(32.04)	(464.47)	(465.15)
ln(SLG)	2.885037	1.474381*	1.480923*						
	(1.69)	(19.17)	(19.29)						
ln(ABper)	0.7449075*	0.8940838*	0.8928816*						
	(2.37)	(5.69)	(33.90)						
Age	3.355589	0.3323444*	0.33661*	3.857727*	0.351889*	0.3553071*			
	(1.68)	(5.69)	(5.76)	(2.16)	(6.51)	(6.57)			
Age^2	-0.053374	-0.0049254*	-0.004998*	-0.0610776*	0.0050582*	-0.0051159*			
	(-1.74)	(-5.44)	(-5.52)	(2.16)	(-6.04)	(-6.10)			
Runs				0.0151046*	0.0236116*	0.023575*			
				(3.00)	(61.56)	(61.63)			
Runs Created							0.0119007*	0.0221799*	0.0221517*
							(2.36)	(59.57)	(59.68)
SmallMarket							-1.208448*	-0.8940838*	-0.0850534*
							(-2.50)	(-2.79)	(-2.90)
Japan*Age			**0.0076373**			**0.0043989**			**0.0068705**
			(1.39)			**(7.90)**			**(1.33)**
R^2	0.5485	0.4781	0.4779	0.5600	0.5538	0.5533	0.5737	0.5386	0.5386
Adjusted R^2	0.4423	0.4774	**0.4771**	0.4867	0.5534	**0.5527**	0.5288	0.5383	**0.5381**
Number of Observations	22	3061	3083	22	3065	3087	22	3065	3087

't' values in parentheses * Significant at the 5% level

Pitchers

The same process detailed above was conducted on the original two pitcher regression specifications. Using the same logic, the hypothesized estimated coefficient of the interaction variable is positive. The results are summarized and explained below. The more complete data is presented on the subsequent page.

Table 9 – Returns to Age, Pitchers

Regression	Estimated Coefficient of "Japan*Age"	Sign	P-value	Effect of adjusted R^2
Traditional	-0.0057914	—	0.153	Increase
Hybrid	-0.0066338	—	0.084	Increase

The results are intriguing. The inclusion of the interaction term strengthened both specifications.[39] In both cases, the estimated coefficient of the interaction term was *negative*. The hypothesis stating that Japanese pitchers with proven experience (measured using age as a proxy) would be paid premiums, compared to non-Japanese pitchers was proven false. These findings are more robust given the significance of the interaction term; in both cases the estimated coefficient was fairly significant (see p-values).

Japanese pitchers with proven success are **paid less** than non-Japanese pitchers.

[39] As measured by the adjusted R-squared value.

	Traditional Method Specification			**Hybrid Method Specification**		
Variable	**Japanese**	**Non-Japanese**	**Pooled Sample**	**Japanese**	**Non-Japanese**	**Pooled Sample**
Constant	-18.491540	1.322308	1.072015	-17.507370	5.811774*	5.611269*
	(-1.70)	(1.41)	(1.15)	(-1.90)	(6.72)	(6.53)
lnSW	0.342087	0.6271473*	0.6298631*			
	(0.98)	(15.16)	(15.34)			
lnIP	0.430970	0.9138889*	0.9053865*			
	(1.72)	(30.25)	(30.39)			
Age	1.847313*	0.4754024*	0.49249*	1.923978*	0.4298984*	0.4421845*
	(2.57)	(8.28)	(8.61)	(0.6097785)	(7.92)	(8.19)
Age^2	-0.0275738*	-0.0065061*	-0.0067547*	-0.029257*	-0.157071*	-0.0061037*
	(-2.34)	(-7.28)	(-7.58)	(-2.94)	(-7.00)	(-7.25)
				-		
Small				0.8500109*	-0.157071*	-0.1599309*
				(-2.45)	(-4.63)	(-4.74)
W_pitch				0.0267386*	0.133005*	0.1323587*
				(2.78)	(40.22)	(40.31)
SV_pitch				0.0267386*	0.0325574*	0.0327212*
				(2.43)	(20.61)	(20.91)
JapanxAge			**-0.0057914**			**-0.0066338**
			(-1.43)			**(-1.73)**
R^2	0.3705	0.3751	0.3714	0.5625	0.4408	0.4391
Adjusted R^2	0.3060	0.3741	**0.3702**	0.5050	0.4396	**0.4379**
Number of Observations	44	2471	2515	44	2471	2515

't' values in parentheses

* Significant at the 5% level

POSITIONAL DISCRIMINATION

Perhaps Japanese players experience on-field discrimination, in addition to or in place of pay discrimination. On-field discrimination takes shape when certain groups of people are restricted from playing certain positions. For example, research has been conducted in examining the utter lack of African-American quarterbacks in the National Football League[40]. Other examples of positional discrimination research include discrimination against African-American infielders in baseball and French Canadian offense and defensemen on English Canadian hockey teams. The hockey example may be a better example of segregation—where certain teams display discriminatory practices. This topic, as concerned with Japanese players in Major League Baseball will be examined in a subsequent section of this paper.

The data used in this analysis is a combination of the datasets used in the above regression analysis. This dataset includes both pitchers and batters from the 1995 season through the 2007 season. 569 records were deleted because positional information was not available in the dataset. The total dataset includes 10,971 records. When a player's record indicated that he play multiple positions, the first positional record was taken to be his primary position for this analysis.

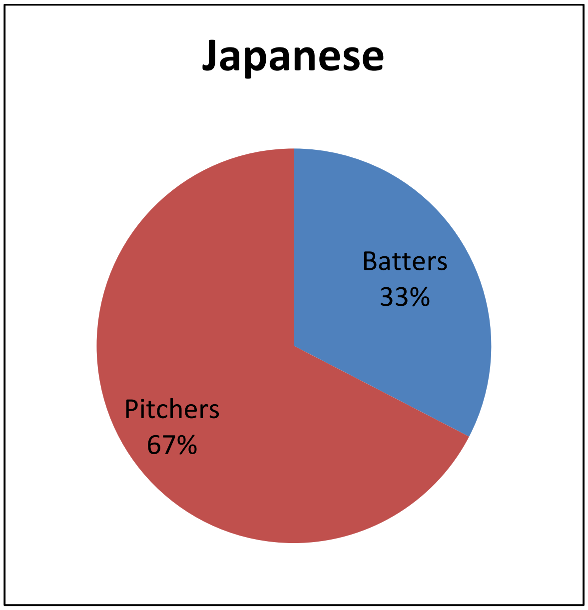
The player records were dichotomized into a Japanese and non-Japanese element. I analyzed positional segregation by examining the proportion of players in each category by position. Positions were aggregated into two (Batters or Pitchers) and four categories (Infielders, Outfielders,[41] Pitchers, or Catchers). Table 10 and Table 11 display the data from these analyses; Figures 1 and 2 display this data visually in the form of pie charts.

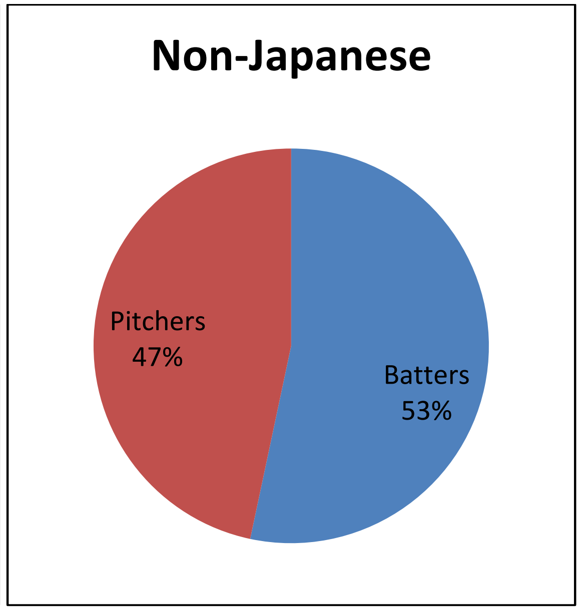
Table 10 - Players by Position, 2 Categories

	Batters	Pitchers
Non-Japanese	5801	5077
Japanese	30	62
Grand Total	5831	5139

[40]A paper by Jason Chung available at the following URL is a good resource: http://ssrn.com/abstract=835204.
[41] Includes designated hitters (DH).

Figure 1 - Players by Position, 2 Categories





The Z-test for two proportions was conducted on the above statistics.[42] This type of statistical test is used to determine whether two proportions are significantly different. The null hypothesis stating that both proportions are equal was tested. For both pitchers and batters, this null hypothesis was not confirmed. At the 90% confidence level, the proportion of Japanese players who are pitchers is significantly different from the proportion of non-Japanese players who are pitchers. The same statement is true for batters.

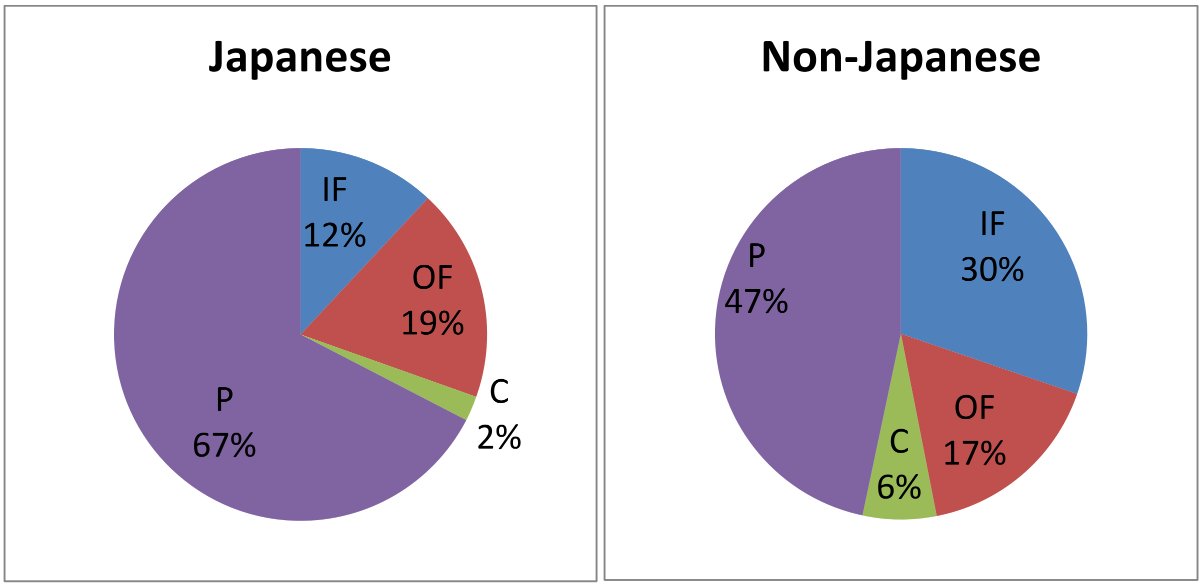
At the 90% confidence level, the proportion of Japanese players who are pitchers is **significantly different** from the proportion of non-Japanese players who are pitchers. The same statement is true for batters.

[42] A handy calculator found at http://www.dimensionresearch.com/resources/calculators/ztest.html was utilized for this analysis.

Table 11 - Players by Position, Four Categories

	IF	OF	C	P
Japanese	11	17	2	62
Non-Japanese	3290	1814	697	5077
Grand Total	3301	1831	699	5139

Figure 2 - Players by Position, 4 Categories



The Z-test for two proportions was conducted on the above statistics.[43] The null hypothesis stating that both proportions are equal was tested. For both pitchers and infielders, this null hypothesis was not confirmed. At the 90% confidence level, the proportion of Japanese players who are pitchers is significantly different from the proportion of non-Japanese players who are pitchers. The same statement is true for infielders. The null hypothesis was rejected on the Z tests performed for catchers and outfielders—the proportion of Japanese players who are outfielders (catchers) is not significantly different from the proportion of Non-Japanese players who are outfielders (catchers).

Between Japanese and Non-Japanese players, the proportion of players that are outfielders and catchers are **not significantly different**.

Between Japanese and Non-Japanese players, the proportion of players that are infielders and pitchers **are significantly different**.

[43] Again, the calculator found at http://www.dimensionresearch.com/resources/calculators/ztest.html was utilized in this analysis.

CUSTOMER DISCRIMIATION: DOES MARKET REPRESENTATION INFLUENCE TEAM REPRESENTATION

For this section of the study, United States' Census data, from the year 2000 census, was analyzed in relation to the number of Japanese players employed by given Major League Baseball teams. The U.S. Census data utilized in this study was downloaded from http://www.census.gov. I hypothesize that teams with large Asian fan bases are more likely to employ larger numbers of Japanese players. This kind of research, examining the relationship between fan base and team representativeness, has been widely conducted, mostly in the area of African-Americans in the National Basketball Association (NBA).[44] The customer-discrimination hypothesis tells us that teams lose revenue and profits when they adopt color-neutral hiring practices (Burdekin, Hossfield, & Smith, 2005). Put more simply, this theory says that teams factor a player's race into the hiring decision due to discrimination by the fan.[45]

For purposes of this study, I will model much of my analysis in the same general methodology demonstrated by Burdekin, Hossfield and Smith in their 2005 article "Are NBA Fans Becoming Indifferent to Race?" For my analysis, I created a spreadsheet file linking: teams, Asian populations,[46] and the number of Japanese players employed (in total). Thus there will initially be thirty-four observations, one for each Major League team.[47] In conducting this analysis, several assumptions and limitations were encountered. The most recent team location and stadiums were used for teams that have previously moved. State populations were used instead of city populations. The Toronto Blue Jays were not included in the study because only 2000 United States Census data was used. Additionally, the Tampa Bay Devil Rays were dropped because they did not have sufficient years of data, as compared to the other organizations.

[44] I most frequently reference a 2005 study by Burdekin, Hossfield and Smith. Also see works by: Brown, Spiro & Keenan (2001); Burdekin & Idson (1991); Hoang & Rascher (1999); and Koch & Vander Hill (1988).
[45] (i.e. white fans prefer watching white players, African-American fans prefer watching African-American players, etc.)
[46] Measured at the state level.
[47] This is not equal to thirty due to teams that moved locations during the sample time period.

The regression equation seeks to shed light on the determinants of a team's racial composition.[48] Therefore, the dependent variable in the regression specification is the team's racial composition. Building on earlier studies[49] the independent variables in this study—those that act to explain a team's racial composition—are: the racial composition of the team's state and the ratio of the team's stadium capacity to the population of its state. This last variable is included because teams with larger markets need not match their teams' roster racial composition to that of their fan bases to the extent of their small market competitors.[50] Ultimately, customer discrimination affects teams in small markets more. In equation format, the regression appears as follows:

$$\text{team composition} = \beta_0 + \beta_1(percentpopwhite) + \beta_2\left(\frac{\text{stadium capacity}}{state\ population}\right) + \varepsilon$$

The regression results provide little value to this study. The full Stata output has not been included because of its lack of significance. In hindsight, this analysis was limited from the beginning. The data pool of observations was limited. The Asian/Non Asian study, as compared to the black/white studies of customer discrimination in the NBA, contains much fewer data observations. As the years continue, more and more Asian players in MLB will increase the validity of such an exercise. Many of the assumptions[51] were made during the data gathering process only compounded the problem. This is an area that should be explored by further researchers.

The customer discrimination analysis provided inconclusive results and should be the basis of a future study.

[48] Race in the context of Asian or Non Asian. The researcher has observed that MLB teams are largely equal in their propensity to employ African-Americans.
[49] Namely Burdekin, Hossfield and Smith's 2005 article "Are NBA Fans Becoming Indifferent to Race?"
[50] Large market teams have a larger pool from which to attract paying fans.
[51] Using state census data instead of standard metropolitan statistical areas; simply using the most recent stadium and locations for teams that have moved; eliminating the one Canadian team (the Toronto Blue Jays); aggregating statistics.

SEGREGATION AND PERFORMANCE ANALYSIS: INCLUSIVE VS. HOMOGENEOUS

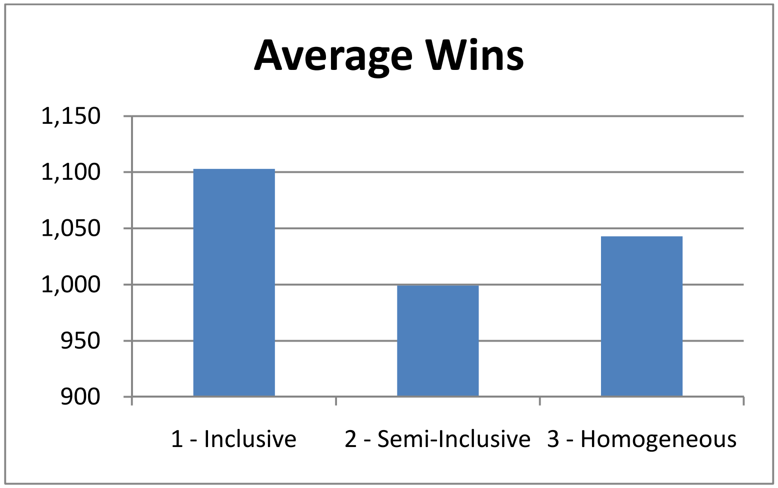
In the workplace, employers have incentives to hire a diverse pool of employees. In many cases, the combination of people with unique histories, talents, and ideas produces superior performance—as compared to homogeneous sets of employees (Davis & Bryant, 2008). A logical question to ask would be—does this affect of performance enhancement via diversity apply to professional sports, specifically Major League Baseball. MLB teams were 100% white until the Brooklyn Dodgers hired Jackie Robinson in 1947. Although some teams were slow to integrate (i.e. the Boston Red Sox through the 1950s), in the recent past all MLB teams stand on roughly equal footing regarding the integration of African Americans. However, teams do differ in their integration with respect to Japanese born players. Teams like the Seattle Mariners, Boston Red Sox, and New York Yankees have integrated their rosters with Japanese players much more rapidly than teams like the Minnesota Twins, Pittsburgh Pirates, and Cleveland Indians. I will investigate which characteristics Japanese inclusive teams share as well as whether their integration has led to increased success—namely in the categories of wins and attendance.

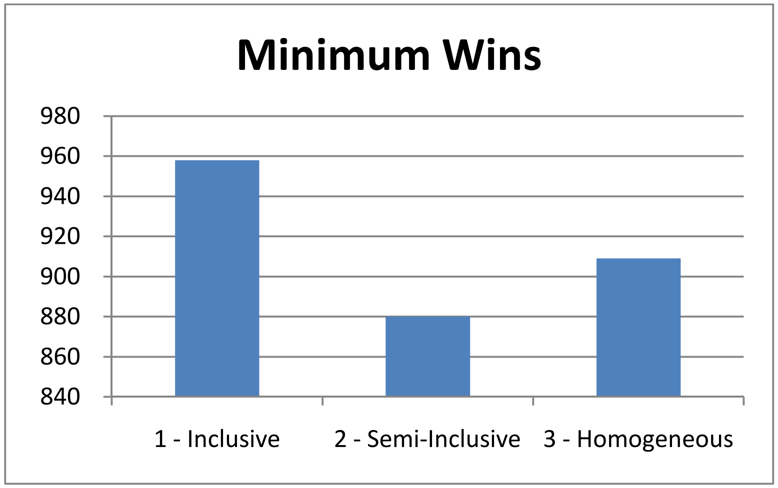
For this aspect of the study, simple statistical tools[52] have been employed. First teams have been categorized according to their nature of inclusion of Japanese players.[53] The team breakdown and accompanying graphs are displayed on the subsequent pages.

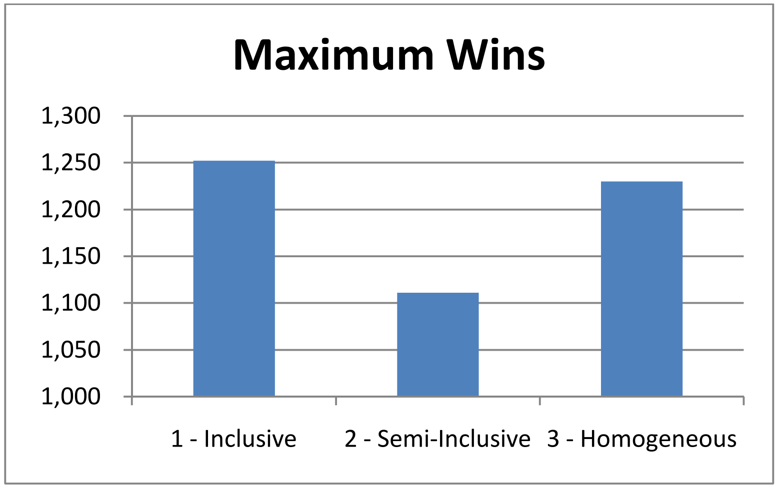
[52] Mean, minimum, maximum and standard deviation

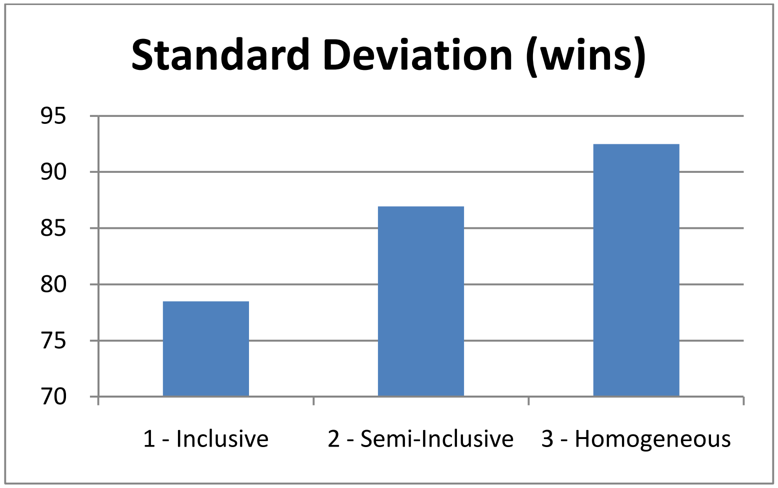
[53] The Toronto Blue Jays were dropped because of their Canadian location. The Arizona Diamondbacks as well as the Tampa Bay Rays were dropped because they are expansion teams for which we do not have complete data.

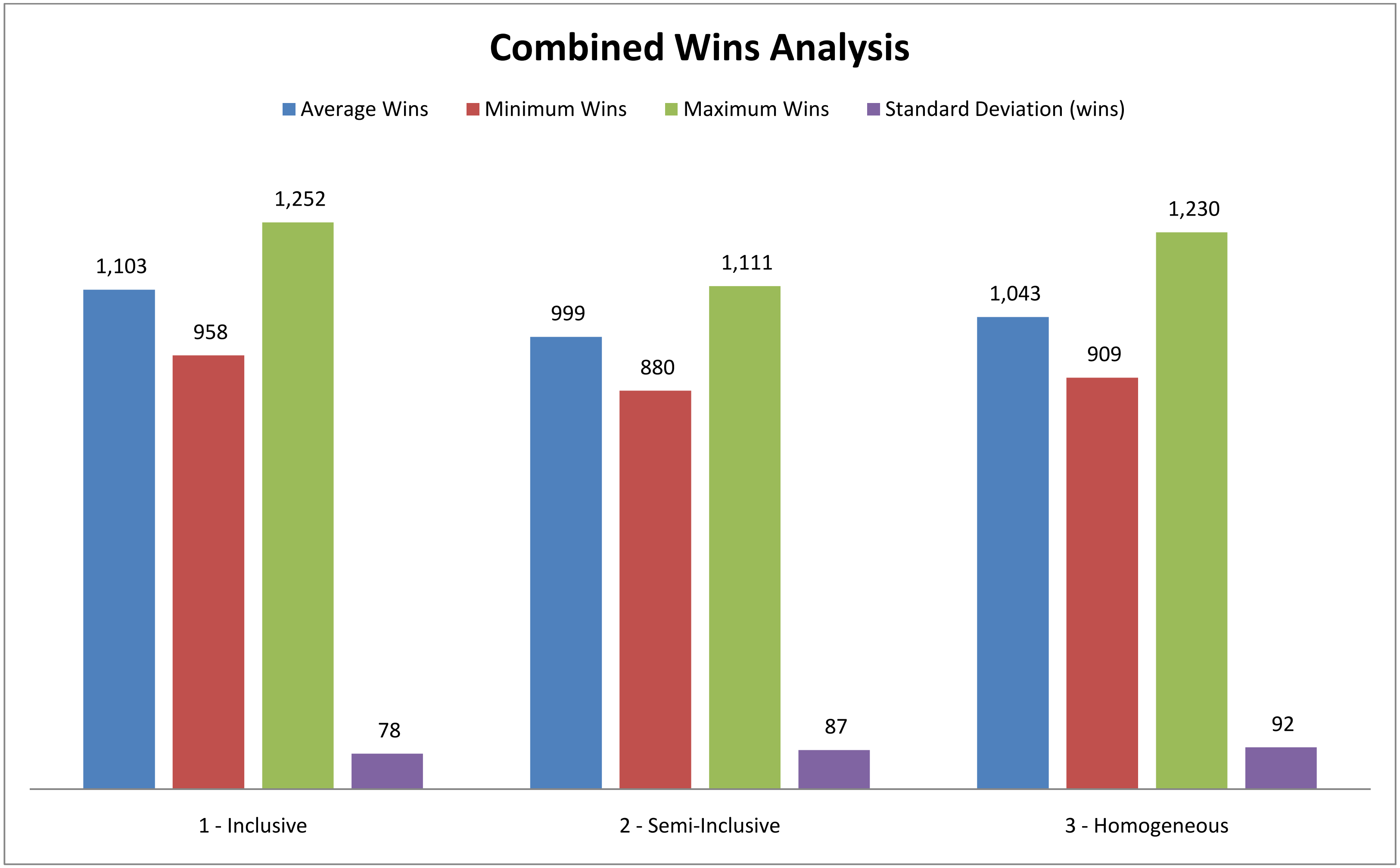
Inclusive		Semi-Inclusive		Homogeneous	
Teams	Japanese Players ('95-'07)	Teams	Japanese Players ('95-'07)	Teams	Japanese Players ('95-'07)
SEA	18	MIL	3	ATL	0
LAD	11	TEX	3	BAL	0
NYM	9	COL	2	CHC	0
NYY	9	DET	2	CIN	0
WSN	7	KCR	2	CLE	0
ANA	5	SDP	2	FLA	0
CHW	5	OAK	1	HOU	0
STL	5	SFG	1	MIN	0
BOS	4			PHI	0
				PIT	0

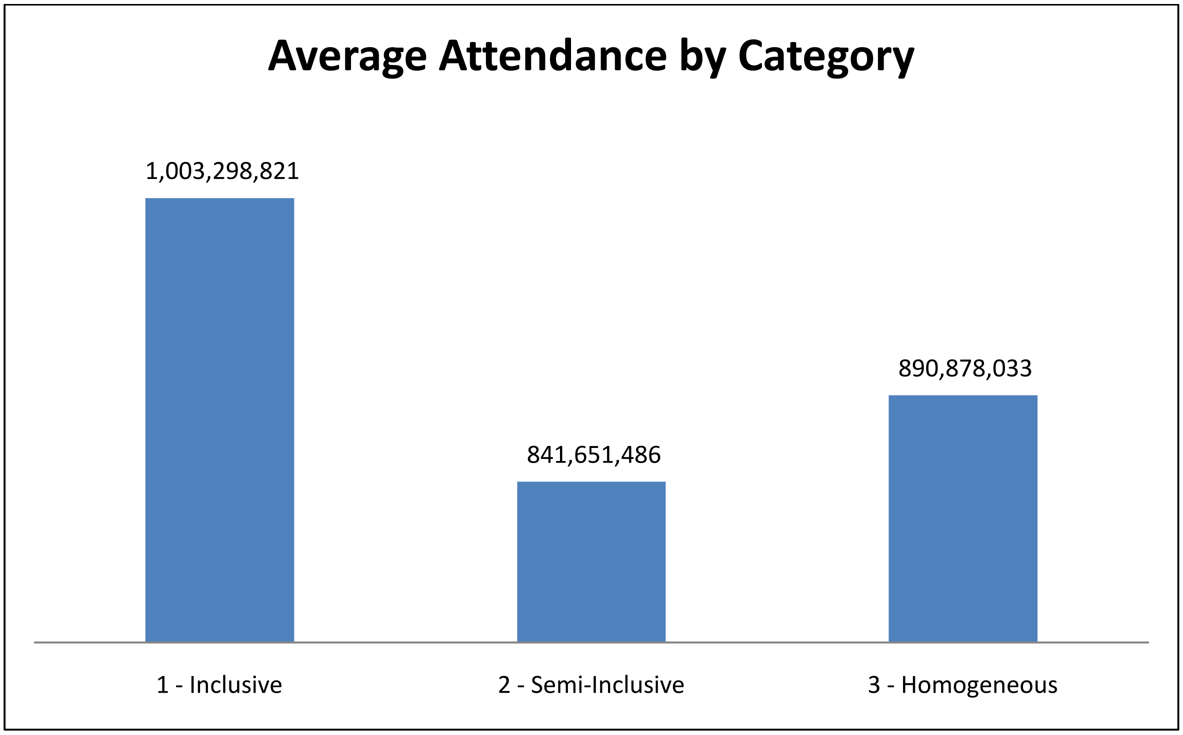

Average Wins
1,150
1,100
1,050
1,000
950
900
1 - Inclusive
2 - Semi-Inclusive
3 - Homogeneous

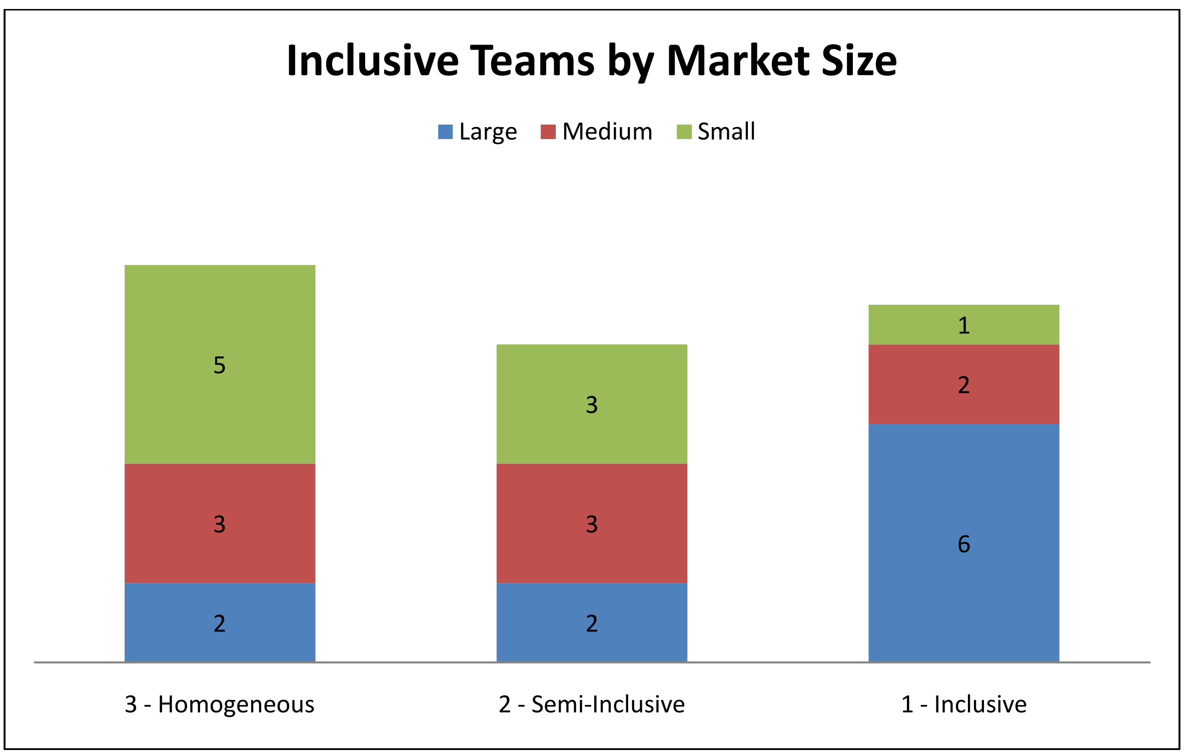

Minimum Wins
980
960
940
920
900
880
860
840
1 - Inclusive
2 - Semi-Inclusive
3 - Homogeneous


Maximum Wins
1,300
1,250
1,200
1,150
1,100
1,050
1,000
1 - Inclusive
2 - Semi-Inclusive
3 - Homogeneous


Standard Deviation (wins)
95
90
85
80
75
70
1 - Inclusive
2 - Semi-Inclusive
3 - Homogeneous


Combined Wins Analysis
Average Wins
Minimum Wins
Maximum Wins
Standard Deviation (wins)
1,103
958
1,252
78
999
880
1,111
87
1,043
909
1,230
92
1 - Inclusive
2 - Semi-Inclusive
3 - Homogeneous


Average Attendance by Category
1,003,298,821
841,651,486
890,878,033
1 - Inclusive
2 - Semi-Inclusive
3 - Homogeneous

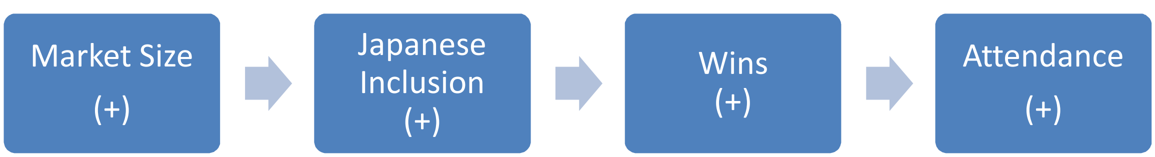

Inclusive Teams by Market Size
Large
Medium
Small
5
3
2
3
3
2
1
2
6
3 - Homogeneous
2 - Semi-Inclusive
1 - Inclusive

This section of the study is very interesting. Inclusive teams tend to be located on the east and west coast and in large metropolitan markets. It appears that "early adopters" of Japanese players[54] have achieved superior performance in the categories of wins and total attendance as compared to those teams that have not integrated Japanese players into their rosters. In addition, inclusive teams' win totals are more tightly centered around its mean, as compared to homogeneous teams. It should be mentioned that this apparent causal relationship (Model A) could indeed be spurious (Model B). It is likely that a second variable[55] drives this relationship. The two models are displayed below. Nonetheless, this is another area that should be further investigated. By controlling for variables like market size, future research will yield more robust results.

More research needs to be conducted prior to reaching concrete conclusions regarding the causality between Japanese player inclusion and superior performance.

Model A:

Model B:



[54] Those classified as Inclusive
[55] Likely market size

CONCLUSION

Using Major League Baseball statistics of the 1995 through 2007 seasons,[56] this study has investigated the issue of labor discrimination, specifically in terms of Japanese players. As outlined in the introductory section, this study intended to answer seven specific questions.

1. What are the best salary predictors for Japanese-born players playing in the United States and do these differ from non-Japanese players?

A detailed summary of the salary regression and Oaxaca decomposition is provided on page 75. Salary discrimination, defined as unequal pay for equal work, was examined first. My results indicate that Japanese batters, compared to non-Japanese batters, do not face salary discrimination. Japanese pitchers, however, do suffer from salary discrimination. In some cases,[57] the level of discrimination is very high.

2. Does the average salary level of equally talented players[58] differ between Japanese and Non-Japanese?

In terms of average salary levels, the distinction between batters and pitchers is also evident. All else equal, Japanese pitchers are paid less than comparable non-Japanese pitchers. However, the results from the batters' regressions are not as concrete. Therefore, we cannot conclude that Japanese batters are paid premiums over equally skilled, non-Japanese players.

3. Do returns-to-experience[59] in MLB differ between Japanese and Non-Japanese?

In answering this question, the results mirror the earlier results—Japanese pitchers face high levels of discrimination. Japanese pitchers with proven success are paid less than non-Japanese pitchers with proven success. However, once again we cannot make concrete conclusions regarding Japanese batters.

4. Do Japanese-born players playing in the United States face positional discrimination?

Japanese players are disproportionally pitchers as compared to batters. However, the gap is closing and more Japanese batters are coming to MLB each year. Between Japanese and

---

[56] The full dataset from Sean Lahman's Baseball Archive can be downloaded at http://www.baseball1.com.
[57] Namely in the Oaxaca decomposition analysis
[58] For both batters and pitchers
[59] For both batters and pitchers

Non-Japanese players, the proportion of players that are outfielders are not significantly different. The same statement is true for catchers. Thus, Japanese players do not face unfair disadvantages in the context of playing the outfield and catcher position. Between Japanese and Non-Japanese players, the proportion of players that are infielders are significantly different. This means that Japanese players do face unfair disadvantages in the context of playing the infield positions.

5. Do Japanese-born players playing in the United States face customer discrimination?

Previous studies have indicated that fans prefer watching players of their own race. This study reached inconclusive results regarding the customer discrimination hypothesis. This area should be the basis of a future study.

6. Do inclusive teams[60] perform better[61] than their homogeneous counterparts?

Teams classified as "Inclusive" have achieved higher levels of wins and total attendance as compared to teams classified as "Homogeneous". These results alone cannot prove causality. Future studies could focus on this topic. In such studies, variables like market size should be held constant.

7. What are the characteristics of inclusive teams[62] and do they differ from their homogeneous counterparts?

Inclusive teams tend to be located on the east and west coast. In general, Homogenous teams tend to be located inland. Inclusive teams tend to be located in large metropolitan markets. Homogeneous teams tend to be located in small metropolitan markets.

[60] With respect to Japanese born players
[61] In terms of team wins
[62] With respect to Japanese born players

Salary Regressions		
Batters		
Traditional Specificaition	It is correct to pool the data	Nondiscriminatory
Sabermetric Specification	It is correct to pool the data	Nondiscriminatory
Hybrid Specification	It is correct to pool the data	Nondiscriminatory
Oaxaca (Sabermetric)	47.09% of the difference in salary level can be attributable to differences in batter characteristics.	Japanese batters face salary discrimination in Major League Baseball.
Pitchers		
Traditional Specificaition	It is incorrect to pool the data	Japanese pitchers face salary discrimination in Major League Baseball.
Sabermetric Specification	N/A	N/A
Hybrid Specification	It is incorrect to pool the data	Japanese pitchers face salary discrimination in Major League Baseball.
Oaxaca (Traditional)	-143.74% can be explained by the difference in pitcher characteristics and 244.48%) cannot be explained by differences in pitcher characteristics.	Japanese pitchers face high degrees of salary discrimination in Major League Baseball.
Oaxaca (Hybrid)	-156.32% can be explained by the difference in pitcher characteristics and 256.32% cannot be explained by differences in pitcher characteristics.	Japanese pitchers face high degrees of salary discrimination in Major League Baseball.

Evaluating Average Differences in Salary Level		
Batters	Mixed results - inconclusive	We cannot conclude that Japanese batters are paid a premium over equally skilled, non-Japanese players.
Pitchers	Dummy variable's coefficient negative, increased R-squared.	All else equal, Japanese pitchers are paid less than comparable non-Japanese.

Evaluating Returns to Age		
Batters	Mixed results - inconclusive	We cannot conclude that Japanese batters receive higher returns to age than non-Japanese.
Pitchers	Interaction variable's coeffcieint negative, increased R-squared.	Japanese pitchers with proven success are paid less than non-Japanese pitchers with proven success.

Limitations and Areas for Future Study

It is important to note the main limitations present in this study. The number of Japanese player observations in Major League Baseball is significantly smaller than the number of non-Japanese player observations. With such a small sample size, it is possible that some of the patterns observed may have occurred simply by chance. Relevant econometric techniques—bootstrapping and repetitive random sampling—have been used to help combat this limitation. None of the econometric literature discounts the relevance of the Chow Test under such discrepancies; nonetheless, the conclusions of this study would be strengthened with the presence of additional Japanese player observations.

Several decisions were made in the preparing the relational database for analysis. The database was modified such that only players with sufficient bargaining power and playing time were included. Also for my research purposes I aggregated player statistics for players with multiple "stints"63 or positions by year. It is believed that any errors in this process would be random and self-cancelling.

This research project has opened the doors for many future studies. Because of the limited number of Japanese players playing in Major League Baseball, there are many areas of this study that could be made more robust, given more observations. All of the salary regressions—the sabermetric, traditional, hybrid, returns to age, average differences in salary level—would benefit from increased Japanese samples. Positional discrimination may prove to be spurious—perhaps an additional variable(s) is determining what positions Japanese ballplayers play in Major League Baseball. Similarly, before making concrete conclusions regarding customer discrimination, future studies need to be conducted to account for other variables. More data observations will help solidify any such theories. The barriers to entry and exit in MLB could be another interesting topic for future researchers.

[63] Trades between teams, call-ups / send-downs, in the same season

APPENDICES

Appendix A – Statistical Category Calculations

Batters' Regressions

$$AB_{per} = \frac{AB_{individual}}{AB_{team}}$$

$$OBP = \frac{H + BB + HBP}{AB + BB + HBP + SH + SF}$$

$$SLG = \frac{(1B) + (2B * 2) + (3B * 3) + (HR * 4)}{AB} = \frac{TB}{AB}$$

$$Runs\ Created = SLG * OBP * AB$$

$$Runs = (.41 * 1B) + (.82 * 2B) + (1.06 * 3B) + (1.42 * HR)$$

Key Abbreviations

1B=Singles
2B=Doubles
3B=Triples
AB=At bats
BB=Base on balls
H=Hits
HBP=Hit by pitches
HR=Home runs
RC=Runs created
SF=Sacrifice flies
SH=Sacrifice hits
SLG=Slugging percentage
TB=Total bases

Pitchers' Regression

$$WHIP = \frac{BB + H}{IP}$$

$$ERA = \frac{ER}{IP} x\ 9$$

$$DICE = 3 + \frac{13HR + 3(BB - HBP) - 2K}{IP}$$

Key Abbreviations

BB=Base-on-Balls (or Walks)
ER=Earned Runs
ERA=Earned Run Average
H=Hits Allowed
HBP=Hit-by-Pitch
HR=Home Runs Allowed
IP=Innings Pitched
K=Strikeouts

Appendix B – Batter Regressions

Variable	Description
Age	A player's age, in years.
Age^2	A player's age, in years, squared.
Japan	A dummy variable equal to 1 if the batter is Japanese.
JapanxAge	An interaction variable equal to Japan*Age
ln(ABper)	The natural log of a player's ratio of a player's at-bats to team at-bats.
ln(salary)	The natural log of a player's salary, lagged one year, converted to base year 2000 dollars using the GDP deflator.
ln(SLG)	The natural log of a player's slugging percentage.
RC_Bat	The number of runs generated by a player, computed as (Slugging percentage)*(On-base percentage)*(At-bats).
Runs	The number of runs generated by a player, computed as (0.41*singles)+(.82*doubles)+(1.06*triples)+(1.42*home runs).
Small	A dummy variable, equal to one, if the player's team's market is in the bottom third of the league's team's markets.

Traditional Method - Batters

Japanese Batters – Traditional Method

```
. reg  lnsal lnslg lnABper age AGE2 if  Japan==1

      Source |       SS       df       MS              Number of obs =      22
-------------+------------------------------           F(  4,    17) =    5.16
       Model |  13.2678572     4  3.3169643            Prob > F      =  0.0066
    Residual |  10.9216381    17  .642449302           R-squared     =  0.5485
-------------+------------------------------           Adj R-squared =  0.4423
       Total |  24.1894953    21  1.15188073           Root MSE      =  .80153

------------------------------------------------------------------------------
       lnsal |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lnslg |   2.885037   1.707777     1.69   0.109    -.7180571    6.488131
     lnABper |   .7449075   .3141022     2.37   0.030     .0822098    1.407605
         age |   3.355589   1.992353     1.68   0.110    -.8479088    7.559087
        AGE2 |  -.0533735    .030708    -1.74   0.100    -.1181616    .0114146
       _cons |  -33.05948   32.56218    -1.02   0.324    -101.7597    35.64072
------------------------------------------------------------------------------
```

$$\ln(\widehat{salary}) = -33.05948 + 2.885037\ln(SLG) + .7449075\ln(AB_{per}) + 3.355589(AGE) - .0533735(AGE^2)$$

100% Non-Japanese Batters – Traditional Method

```
. reg  lnsal lnslg lnABper age AGE2 if  Japan==0

      Source |       SS       df       MS              Number of obs =    3061
-------------+------------------------------           F(  4,  3056) =  699.76
       Model |  1809.46296     4   452.36574           Prob > F      =  0.0000
    Residual |  1975.56559  3056  .646454709           R-squared     =  0.4781
-------------+------------------------------           Adj R-squared =  0.4774
       Total |  3785.02855  3060  1.23693744           Root MSE      =  .80402

------------------------------------------------------------------------------
       lnsal |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lnslg |   1.474381   .0768921    19.17   0.000     1.323616    1.625147
     lnABper |   .8940838   .0264454    33.81   0.000     .8422313    .9459362
         age |   .3323444   .0584219     5.69   0.000     .2177943    .4468945
        AGE2 |  -.0049254   .0009056    -5.44   0.000     -.006701   -.0031498
       _cons |   12.77249   .9339354    13.68   0.000     10.94129     14.6037
------------------------------------------------------------------------------
```

$$\ln(\widehat{salary}) = 12.77249 + 1.474381\ln(SLG) + .8940838\ln(AB_{per}) + .3323444(AGE) - .0049254(AGE^2)$$

Pooled Batters (100% Non-Japanese) – Traditional Method

```
. reg  lnsal lnslg lnABper age AGE2

      Source |       SS       df       MS              Number of obs =    3083
-------------+------------------------------           F(  4,  3078) =  703.49
       Model |  1821.75117     4  455.437793           Prob > F      =  0.0000
    Residual |  1992.70236  3078  .647401678           R-squared     =  0.4776
-------------+------------------------------           Adj R-squared =  0.4769
       Total |  3814.45354  3082  1.23765527           Root MSE      =  .80461

------------------------------------------------------------------------------
       lnsal |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lnslg |   1.480376   .0767654    19.28   0.000      1.32986    1.630893
     lnABper |   .8937928   .0263312    33.94   0.000     .8421644    .9454213
         age |   .3391687   .0584055     5.81   0.000     .2246509    .4536864
        AGE2 |  -.0050371   .0009053    -5.56   0.000    -.0068122    -.003262
       _cons |   12.67581   .9337786    13.57   0.000     10.84492     14.5067
------------------------------------------------------------------------------
```

$$\ln\left(\widehat{salary}\right) = 12.67581 + 1.480376\ln(SLG) + .8937928\ln\left(AB_{per}\right) + .3391687(AGE) - .0050371(AGE^2)$$

Chow Test – Traditional Method Entire Batter Dataset

Step 1: $H_0: \propto_0 = \beta_0, \propto_1 = \beta_1, \propto_2 = \beta_2, \propto_3 = \beta_3, \propto_4 = \beta_4,$
$H_A: Negation\ of\ H_0$

Step 2: $\propto = 5\%$

Step 3:

$$F_{5,3071} = \frac{(1992.70236 - 1986.487228)/(6)}{(1986.487228)/(3071)} = \frac{1.035855333}{.6468535422} = 1.601375374$$

Step 4: Because the F statistic is less than the critical value of the F distribution F_C (approximately 2.21) at the 5% level, we fail to reject the null hypothesis, H_0. It is correct to assume equal coefficients.

<u>Sabermetric Method – Batters</u>

Japanese Batters – Sabermetric Method

```
. reg  lnsal runs age AGE2 if Japan==1

      Source |       SS       df       MS              Number of obs =      22
-------------+------------------------------           F(  3,    18) =    7.64
       Model |  13.5458167     3  4.51527224           Prob > F      =  0.0017
    Residual |  10.6436786    18   .59131548           R-squared     =  0.5600
-------------+------------------------------           Adj R-squared =  0.4867
       Total |  24.1894953    21  1.15188073           Root MSE      =  .76897

------------------------------------------------------------------------------
       lnsal |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        runs |   .0151046   .0050333     3.00   0.008     .0045301    .0256792
         age |   3.857727   1.782648     2.16   0.044      .112522    7.602931
        AGE2 |  -.0610776   .0274847    -2.22   0.039    -.1188208   -.0033345
       _cons |  -46.79575   28.78529    -1.63   0.121    -107.2714    13.67991
------------------------------------------------------------------------------
```

$$\ln(\widehat{salary}) = -46.79575 + .0151046(Runs) + 3.857727(AGE) - .0610776(AGE^2)$$

100% Non-Japanese Batters – Sabermetric Method

```
. reg  lnsal runs age AGE2 if Japan==0

      Source |       SS       df       MS              Number of obs =    3065
-------------+------------------------------           F(  3,  3061) = 1266.60
       Model |  2103.26963     3  701.089877           Prob > F      =  0.0000
    Residual |  1694.32415  3061  .553519814           R-squared     =  0.5538
-------------+------------------------------           Adj R-squared =  0.5534
       Total |  3797.59378  3064  1.23942356           Root MSE      =  .74399

------------------------------------------------------------------------------
       lnsal |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        runs |   .0236116   .0003835    61.56   0.000     .0228596    .0243637
         age |    .351889   .0540567     6.51   0.000     .2458978    .4578802
        AGE2 |  -.0050582   .0008378    -6.04   0.000    -.0067009   -.0034156
       _cons |   6.911381   .8687093     7.96   0.000     5.208068    8.614693
------------------------------------------------------------------------------
```

$$\ln(\widehat{salary}) = 6.911381 + .0236116(Runs) + .351889(AGE) - .351889(AGE^2)$$

Pooled Batters (100% Non-Japanese Batters) – Sabermetric Method

```
. reg  lnsal runs age AGE2

      Source |       SS       df       MS              Number of obs =    3087
-------------+------------------------------           F(  3,  3083) = 1272.22
       Model |   2116.9966     3  705.665534           Prob > F      =  0.0000
    Residual |  1710.05203  3083  .554671434           R-squared     =  0.5532
-------------+------------------------------           Adj R-squared =  0.5527
       Total |  3827.04863  3086  1.24013242           Root MSE      =  .74476

------------------------------------------------------------------------------
       lnsal |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        runs |   .0235852   .0003823    61.69   0.000     .0228355    .0243349
         age |   .3568111   .0540565     6.60   0.000     .2508208    .4628015
        AGE2 |  -.0051389   .0008378    -6.13   0.000    -.0067815   -.0034963
       _cons |   6.840096   .8687111     7.87   0.000     5.136785    8.543407
------------------------------------------------------------------------------
```

$$\ln(\widehat{salary}) = 6.840096 + .0235852(Runs) + .3568111(AGE) - .0051389(AGE^2)$$

Chow Test – Sabermetric Method Entire Batter Dataset

Step 1: $H_0\colon \propto_0 = \beta_0, \propto_1 = \beta_1, \propto_2 = \beta_2, \propto_3 = \beta_3$
$H_A\colon Negation\ of\ H_0$

Step 2: $\propto = 5\%$

Step 3:

$$F_{4,3077} = \frac{(1710.05203 - 1704.970518)/(5)}{(1704.970518)/(3077)} = \frac{1.016302428}{.5541015658} = 1.83414466$$

Step 4: Because the F statistic is less than the critical value of the F distribution F_C (approximately 2.37) at the 5% level, we fail to reject the null hypothesis, H_0. It is correct to assume equal coefficients.

Hyrbid Method – Batters

Japanese Batters – Hybrid Method

```
. reg lnsal RC_bat  Small if Japan==1

      Source |       SS       df       MS              Number of obs =      22
-------------+------------------------------           F(  2,    19) =   12.78
       Model |  13.8774581     2  6.93872906           Prob > F      =  0.0003
    Residual |  10.3120372    19  .542738801           R-squared     =  0.5737
-------------+------------------------------           Adj R-squared =  0.5288
       Total |  24.1894953    21  1.15188073           Root MSE      =  .73671

------------------------------------------------------------------------------
       lnsal |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      RC_bat |   .0119007   .0050486     2.36   0.029     .0013338    .0224676
       Small |  -1.208448   .4832568    -2.50   0.022    -2.219916   -.1969797
       _cons |   14.35573   .4480994    32.04   0.000     13.41785    15.29361
------------------------------------------------------------------------------
```

$$\ln(\widehat{salary}) = 15.412 + .046962(Runs\ Created) - .0324151(Small)$$

100% Non-Japanese Batters – Hybrid Method

```
. reg lnsal RC_bat  Small if Japan==0

      Source |       SS       df       MS              Number of obs =    3065
-------------+------------------------------           F(  2,  3062) = 1787.50
       Model |  2045.56041     2   1022.7802           Prob > F      =  0.0000
    Residual |  1752.03338  3062  .572185949           R-squared     =  0.5386
-------------+------------------------------           Adj R-squared =  0.5383
       Total |  3797.59378  3064  1.23942356           Root MSE      =  .75643

------------------------------------------------------------------------------
       lnsal |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      RC_bat |   .0221799   .0003723    59.57   0.000     .0214499      .02291
       Small |  -.0818273   .0293428    -2.79   0.005    -.1393609   -.0242937
       _cons |   13.18183   .0283802   464.47   0.000     13.12618    13.23747
------------------------------------------------------------------------------
```

$$\ln(\widehat{salary}) = 13.18183 + .0221799(Runs\ Created) - .0818273(Small)$$

Pooled Batters (100% Non-Japanese) – Hybrid Method

```
. reg lnsal RC_bat  Small

      Source |       SS       df       MS              Number of obs =    3087
-------------+------------------------------           F(  2,  3084) = 1798.02
       Model |  2060.19977     2  1030.09988           Prob > F      =  0.0000
    Residual |  1766.84887  3084  .572908193           R-squared     =  0.5383
-------------+------------------------------           Adj R-squared =  0.5380
       Total |  3827.04863  3086  1.24013242           Root MSE      =  .75691

------------------------------------------------------------------------------
       lnsal |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      RC_bat |    .022163   .0003711    59.72   0.000     .0214353    .0228907
       Small |  -.0858057   .0292848    -2.93   0.003    -.1432253   -.0283861
       _cons |   13.18573   .0283367   465.32   0.000     13.13017    13.24129
------------------------------------------------------------------------------
```

$$\ln\left(\widehat{salary}\right) = 13.18573 + .022163(Runs\ Created) - .0858057(Small)$$

Chow Test – Hybrid Method Entire Batter Dataset

Step 1: $H_0: \propto_0 = \beta_0, \propto_1 = \beta_1, \propto_2 = \beta_2$
$H_A: Negation\ of\ H_0$

Step 2: $\propto = 5\%$

Step 3:

$$F_{3,3079} = \frac{(1766.84887 - 1762.345417)/(4)}{(1762.345417)/(3079)} = \frac{1.12586325}{.5723759068} = 1.966999723$$

Step 4: Because the F statistic is less than the critical value of the F distribution F_C (approximately 2.60) at the 5% level, we fail to reject the null hypothesis, H_0. It is correct to assume equal coefficients.

Oaxaca Decomposition – Batters

Sabermetric Specification

$$\ln(salary) = \beta_0 + \beta_1(Runs) + \beta_2(AGE) + \beta_3\left(AGE^2\right) + \varepsilon$$

```
. oaxaca nonJapan Japan;
(high estimates: Japan; low estimates: nonJapan)

                                                 Mean prediction 1 =  14.99823
                                                 Mean prediction 2 =  14.50865
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  difference |   .4895785   .2527598     1.94   0.053    -.0058216    .9849787
------------------------------------------------------------------------------

Linear decomposition
------------------------------------------------------------------------------
       Total |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
three-fold   |
  endowments |   .2298492   .1212733     1.90   0.058     -.007842    .4675405
coefficients |   -.017054   .2557174    -0.07   0.947    -.5182509    .4841428
 interaction |   .2767834   .2278557     1.21   0.224    -.1698056    .7233723
------------------------------------------------------------------------------

. oaxaca nonJapan Japan, weight(1 0.5 0 omega);
(high estimates: Japan; low estimates: nonJapan)

                                                 Mean prediction 1 =  14.99823
                                                 Mean prediction 2 =  14.50865
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  difference |   .4895785   .2527598     1.94   0.053    -.0058216    .9849787
------------------------------------------------------------------------------

Linear decompositions
------------------------------------------------------------------------------
       Total |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
W=1          |
   explained |   .5066326   .2656083     1.91   0.056      -.01395    1.027215
 unexplained |   -.017054   .2557174    -0.07   0.947    -.5182509    .4841428
-------------+----------------------------------------------------------------
W=,5         |
   explained |   .3682409   .1721857     2.14   0.032     .0307631    .7057187
 unexplained |   .1213376   .2058956     0.59   0.556    -.2822103    .5248855
-------------+----------------------------------------------------------------
W=0          |
   explained |   .2298492   .1212733     1.90   0.058     -.007842    .4675405
 unexplained |   .2597293   .2129641     1.22   0.223    -.1576727    .6771313
-------------+----------------------------------------------------------------
W=omega      |
   explained |   .2305448   .1213653     1.90   0.057    -.0073269    .4684164
 unexplained |   .2590338   .2124685     1.22   0.223    -.1573969    .6754644
------------------------------------------------------------------------------
```

```
. oaxaca2 lnsal lnslg lnABper age AGE2, by(Japan) pooled;

Group 1: Japan = 0                                   Number of obs 1   =      3061
Group 2: Japan = 1                                   Number of obs 2   =        22

                                                     Mean prediction 1 =  14.50865
                                                     Mean prediction 2 =  14.99823
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  difference |  -.4895785   .2527598    -1.94   0.053    -.9849787    .0058216
------------------------------------------------------------------------------

Linear decomposition
------------------------------------------------------------------------------
       Total |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
pooled       |
   explained |  -.2305448   .1213658    -1.90   0.057    -.4684173    .0073278
 unexplained |  -.2590338          .        .       .            .           .
------------------------------------------------------------------------------
```

Appendix C – Pitchers Regressions

Variable	**Description**
Age	A player's age, in years.
Age^2	A player's age, in years, squared.
Japan	A dummy variable equal to 1 if the batter is Japanese.
JapanxAge	An interaction variable equal to Japan*Age
ln(salary)	The natural log of a player's salary, lagged one year, converted to base year 2000 dollars using the GDP deflator.
lnIP	The natural log of a player's total innings-pitched.
lnSW	The natural log of a player's strikeout to walk ratio.
Small	A dummy variable, equal to one, if the player's team is in the bottom third of the league teams' market sizes.
SV_pitch	A player's total number of saves in a season.
W_pitch	A player's total number of wins in a season.

Traditional Method – Pitchers

Japanese Pitchers

```
. reg  lnsal lnSW lnIP AGE AGE2 if Japan==1

      Source |       SS       df       MS              Number of obs =      44
-------------+------------------------------           F(  4,    39) =    5.74
       Model |  15.4227371     4  3.85568428           Prob > F      =  0.0010
    Residual |  26.1989891    39  .671768951           R-squared     =  0.3705
-------------+------------------------------           Adj R-squared =  0.3060
       Total |  41.6217262    43  .967947121           Root MSE      =  .81962

------------------------------------------------------------------------------
       lnsal |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        lnSW |   .3420868   .3488879     0.98   0.333    -.3636056    1.047779
        lnIP |     .43097   .2503918     1.72   0.093    -.0754953    .9374353
         AGE |   1.847313   .7180596     2.57   0.014     .3949007    3.299726
        AGE2 |  -.0275738   .0117604    -2.34   0.024    -.0513613   -.0037862
       _cons |  -18.49154    10.8473    -1.70   0.096    -40.43228      3.4492
------------------------------------------------------------------------------
```

$$\ln(salary) = -18.49154 + .3420868\ln(\mathrm{SW}) + .43097\ln(\mathrm{IP}) + 1.847313(\mathrm{AGE}) - .0285738(\mathrm{AGE}^2)$$

Non-Japanese Pitchers

```
. reg  lnsal lnSW lnIP AGE AGE2 if Japan==0

      Source |       SS       df       MS              Number of obs =    2471
-------------+------------------------------           F(  4,  2466) =  370.01
       Model |  1015.53816     4  253.884539           Prob > F      =  0.0000
    Residual |  1692.04845  2466  .686151033           R-squared     =  0.3751
-------------+------------------------------           Adj R-squared =  0.3741
       Total |   2707.5866  2470  1.09618891           Root MSE      =  .82834

------------------------------------------------------------------------------
       lnsal |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        lnSW |   .6271473   .0413619    15.16   0.000     .5460395     .708255
        lnIP |   .9138889   .0300134    30.45   0.000     .8550348     .972743
         AGE |   .4754024   .0574072     8.28   0.000     .3628311    .5879736
        AGE2 |  -.0065061   .0008938    -7.28   0.000    -.0082587   -.0047535
       _cons |   1.322308   .9364749     1.41   0.158    -.5140503    3.158667
------------------------------------------------------------------------------
```

$$\ln(salary) = 1.322308 + .6271473\ln(\mathrm{SW}) + .9138889\ln(\mathrm{IP}) + .4754024(\mathrm{AGE}) - .0065061(\mathrm{AGE}^2)$$

All Pitchers

```
. reg  lnsal lnSW lnIP AGE AGE2

      Source |       SS       df       MS              Number of obs =    2515
-------------+------------------------------           F(  4,  2510) =  370.01
       Model |  1019.89291     4  254.973228           Prob > F      =  0.0000
    Residual |   1729.6564  2510  .689106137           R-squared     =  0.3709
-------------+------------------------------           Adj R-squared =  0.3699
       Total |  2749.54932  2514  1.09369503           Root MSE      =  .83012

------------------------------------------------------------------------------
       lnsal |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        lnSW |   .6301392   .0410637    15.35   0.000      .549617    .7106614
        lnIP |   .9043861   .0297864    30.36   0.000     .8459777    .9627945
         AGE |   .4886703   .0571649     8.55   0.000     .3765752    .6007655
        AGE2 |   -.006697   .0008905    -7.52   0.000    -.0084431   -.0049509
       _cons |   1.135362   .9321406     1.22   0.223    -.6924809    2.963206
------------------------------------------------------------------------------
```

$$\ln(salary) = 1.135362 + .6301392\ln(\mathrm{SW}) + .9043861\ln(\mathrm{IP}) + .4886703(\mathrm{AGE}) - .006697(\mathrm{AGE}^2)$$

Chow Test

Step 1: $H_0: \propto_0 = \beta_0, \propto_1 = \beta_1, \propto_2 = \beta_2, \propto_3 = \beta_3, \propto_4 = \beta_4$
$H_A: Negation\ of\ H_0$

Step 2: $\propto = 5\%$

Step 3:

$$F_{3,627} = \frac{(1729.6564 - 1718.247439)/(6)}{(1718.247439)/(2503)} = \frac{1.9014935}{.6864752054} = 2.769937625$$

Step 4: Because the F statistic is greater than the critical value of the F distribution F_C (approximately 2.21) at the 5% level, we reject the null hypothesis, H_0. It is incorrect to assume equal coefficients.

Hybrid Method – Pitchers

Japanese Pitchers

```
. reg lnsal  W_pitch  SV_pitch AGE AGE2 Small if Japan==1

      Source |       SS       df       MS              Number of obs =      44
-------------+------------------------------           F(  5,    38) =    9.77
       Model |  23.4131683     5  4.68263366           Prob > F      =  0.0000
    Residual |  18.2085579    38  .479172576           R-squared     =  0.5625
-------------+------------------------------           Adj R-squared =  0.5050
       Total |  41.6217262    43  .967947121           Root MSE      =  .69222

------------------------------------------------------------------------------
       lnsal |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     W_pitch |   .0700887   .0251923     2.78   0.008     .0190896    .1210878
    SV_pitch |   .0267386   .0110037     2.43   0.020     .0044628    .0490145
         AGE |   1.923978   .6097785     3.16   0.003     .6895456     3.15841
        AGE2 |   -.029257   .0099627    -2.94   0.006    -.0494254   -.0090885
       Small |  -.8500109   .3476231    -2.45   0.019    -1.553737   -.1462848
       _cons |  -17.50737   9.220435    -1.90   0.065    -36.17316    1.158424
------------------------------------------------------------------------------
```

$$\ln(salary) = -17.50737 + .0700887(\text{Wins}) + .0267386(\text{Saves}) + 1.923978(\text{AGE}) - .029257(\text{AGE}^2) - .8500109(Small)$$

Non-Japanese Pitchers

```
. reg lnsal  W_pitch  SV_pitch AGE AGE2 Small if Japan==0

      Source |       SS       df       MS              Number of obs =    2471
-------------+------------------------------           F(  5,  2465) =  388.54
       Model |  1193.37355     5   238.67471           Prob > F      =  0.0000
    Residual |  1514.21305  2465  .614285214           R-squared     =  0.4408
-------------+------------------------------           Adj R-squared =  0.4396
       Total |   2707.5866  2470  1.09618891           Root MSE      =  .78376

------------------------------------------------------------------------------
       lnsal |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     W_pitch |   .1330055   .0033073    40.22   0.000     .1265201    .1394909
    SV_pitch |   .0325574   .0015798    20.61   0.000     .0294595    .0356553
         AGE |   .4298984   .0542982     7.92   0.000     .3234236    .5363731
        AGE2 |  -.0059235   .0008456    -7.00   0.000    -.0075817   -.0042653
       Small |   -.157071   .0339003    -4.63   0.000    -.2235471   -.0905949
       _cons |   5.811774   .8647519     6.72   0.000     4.116059    7.507489
------------------------------------------------------------------------------
```

$$\ln(salary) = 5.811774 + .1330055(\text{Wins}) + .0325574(\text{Saves}) + .4298984(\text{AGE}) - .0059235(\text{AGE}^2) - .157071(Small)$$

All Pitchers

```
. reg lnsal  W_pitch  SV_pitch AGE AGE2 Small

      Source |       SS       df       MS              Number of obs =    2515
-------------+------------------------------           F(  5,  2509) =  392.76
       Model |  1207.20137     5  241.440274           Prob > F      =  0.0000
    Residual |  1542.34795  2509  .614726165           R-squared     =  0.4391
-------------+------------------------------           Adj R-squared =  0.4379
       Total |  2749.54932  2514  1.09369503           Root MSE      =  .78404

------------------------------------------------------------------------------
       lnsal |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     W_pitch |   .1322973   .0032844    40.28   0.000     .1258569    .1387378
    SV_pitch |   .0326926   .0015657    20.88   0.000     .0296224    .0357627
         AGE |   .4378198   .0539679     8.11   0.000     .3319936    .5436459
        AGE2 |  -.0060367   .0008409    -7.18   0.000    -.0076857   -.0043877
       Small |  -.1565574   .0337087    -4.64   0.000    -.2226571   -.0904577
       _cons |   5.677091   .8590604     6.61   0.000     3.992551    7.361631
------------------------------------------------------------------------------
```

$$\ln(salary) = 5.677091 + .1322973(\text{Wins}) + .0326926(\text{Saves}) + .4378198(\text{AGE}) - .0060367(\text{AGE}^2) - .1565574(Small)$$

Chow Test

Step 1: $H_0: \propto_0 = \beta_0, \propto_1 = \beta_1, \propto_2 = \beta_2, \propto_3 = \beta_3, \propto_4 = \beta_4, \propto_5 = \beta_5$
$H_A: Negation\ of\ H_0$

Step 2: $\propto = 5\%$

Step 3:

$$F_{3,627} = \frac{(1542.34795 - 1532.421608)/(7)}{(1532.421608)/(2489)} = \frac{1.418048857}{.6156776247} = 2.303232731$$

Step 4: Because the F statistic is greater than the critical value of the F distribution F_C (approximately 2.10) at the 5% level, we reject the null hypothesis, H_0. It is incorrect to assume equal coefficients.

Oaxaca Decomposition – Pitchers

Tradiitonal Specification

$$\ln(salary) = \beta_0 + \beta_1 \ln(\text{SW}) + \beta_2 \ln(\text{IP}) + \beta_3(\text{AGE}) + \beta_4(\text{AGE}^2) + \varepsilon$$

```
. oaxaca nonJapan Japan;
(high estimates: nonJapan; low estimates: Japan)

                                                 Mean prediction 1 =  14.38429
                                                 Mean prediction 2 =  14.29548
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  difference |   .0888126   .1590087     0.56   0.576    -.2228388     .400464
------------------------------------------------------------------------------

Linear decomposition
------------------------------------------------------------------------------
       Total |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
three-fold   |
  endowments |  -.1998524   .1158087    -1.73   0.084    -.4268334    .0271285
coefficients |   .2151689   .1485963     1.45   0.148    -.0760745    .5064123
 interaction |   .0734961   .1015974     0.72   0.469    -.1256312    .2726234
------------------------------------------------------------------------------

. oaxaca nonJapan Japan, weight(1 0.5 0 omega);
(high estimates: nonJapan; low estimates: Japan)

                                                 Mean prediction 1 =  14.38429
                                                 Mean prediction 2 =  14.29548
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  difference |   .0888126   .1590087     0.56   0.576    -.2228388     .400464
------------------------------------------------------------------------------

Linear decompositions
------------------------------------------------------------------------------
       Total |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
W=1          |
   explained |  -.1263563   .0748764    -1.69   0.092    -.2731115    .0203988
 unexplained |   .2151689   .1485963     1.45   0.148    -.0760745    .5064123
-------------+----------------------------------------------------------------
W=,5         |
   explained |  -.1631044    .083238    -1.96   0.050    -.3262478    .0000391
 unexplained |   .2519169   .1346316     1.87   0.061    -.0119561      .51579
-------------+----------------------------------------------------------------
W=0          |
   explained |  -.1998524   .1158087    -1.73   0.084    -.4268334    .0271285
 unexplained |    .288665   .1390377     2.08   0.038     .0161561    .5611739
-------------+----------------------------------------------------------------
W=omega      |
   explained |  -.1266728   .0745641    -1.70   0.089    -.2728158    .0194702
```

```
 unexplained |   .2154854    .147636     1.46   0.144    -.0738758    .5048467
------------------------------------------------------------------------------

. oaxaca2 lnsal lnSW lnIP AGE AGE2, by(Japan) pooled;

Group 1: Japan = 0                                 Number of obs 1   =      2471
Group 2: Japan = 1                                 Number of obs 2   =        44

                                                   Mean prediction 1 =  14.38429
                                                   Mean prediction 2 =  14.29548
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  difference |   .0888126   .1590087     0.56   0.576    -.2228388     .400464
------------------------------------------------------------------------------

Linear decomposition
------------------------------------------------------------------------------
       Total |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
pooled       |
   explained |  -.1266728   .0745658    -1.70   0.089    -.2728192    .0194735
 unexplained |   .2154854          .        .       .            .           .
------------------------------------------------------------------------------
```

Hybrid Specification

$$\ln(salary) = \beta_0 + \beta_1(\text{Wins}) + \beta_2(\text{Saves}) + \beta_3(\text{AGE}) + \beta_4(\text{AGE}^2) + \beta_5(Small) + \varepsilon$$

```
oaxaca nonJapan Japan;
(high estimates: nonJapan; low estimates: Japan)

                                                 Mean prediction 1 =  14.38429
                                                 Mean prediction 2 =  14.29548
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  difference |   .0888126   .1580381     0.56   0.574    -.2209365    .3985616
------------------------------------------------------------------------------

Linear decomposition
------------------------------------------------------------------------------
       Total |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
three-fold   |
  endowments |  -.3836826   .1503578    -2.55   0.011    -.6783785   -.0889868
coefficients |   .2272539   .1288806     1.76   0.078    -.0253476    .4798553
 interaction |   .2452413   .1197272     2.05   0.041     .0105803    .4799024
------------------------------------------------------------------------------

. oaxaca nonJapan Japan, weight(1 0.5 0 omega);
(high estimates: nonJapan; low estimates: Japan)

                                                 Mean prediction 1 =  14.38429
                                                 Mean prediction 2 =  14.29548
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  difference |   .0888126   .1580381     0.56   0.574    -.2209365    .3985616
------------------------------------------------------------------------------

Linear decompositions
------------------------------------------------------------------------------
       Total |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
W=1          |
   explained |  -.1384413   .1002681    -1.38   0.167    -.3349631    .0580805
 unexplained |   .2272539   .1288806     1.76   0.078    -.0253476    .4798553
-------------+----------------------------------------------------------------
W=,5         |
   explained |   -.261062   .1129023    -2.31   0.021    -.4823464   -.0397776
 unexplained |   .3498745   .1213365     2.88   0.004     .1120594    .5876896
-------------+----------------------------------------------------------------
W=0          |
   explained |  -.3836826   .1503578    -2.55   0.011    -.6783785   -.0889868
 unexplained |   .4724952    .141429     3.34   0.001     .1952995    .7496909
-------------+----------------------------------------------------------------
W=omega      |
   explained |  -.1388299   .1001779    -1.39   0.166     -.335175    .0575152
 unexplained |   .2276425   .1279548     1.78   0.075    -.0231442    .4784292
------------------------------------------------------------------------------

. oaxaca2 lnsal W_pitch  SV_pitch AGE AGE2 Small, by(Japan) pooled;
```

```
Group 1: Japan = 0                                 Number of obs 1   =      2471
Group 2: Japan = 1                                 Number of obs 2   =        44

                                                   Mean prediction 1 =  14.38429
                                                   Mean prediction 2 =  14.29548
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  difference |   .0888126   .1580381     0.56   0.574    -.2209365    .3985616
------------------------------------------------------------------------------

Linear decomposition
------------------------------------------------------------------------------
       Total |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
pooled       |
   explained |  -.1388299   .1001798    -1.39   0.166    -.3351787    .0575188
 unexplained |   .2276425          .        .       .            .           .
```

Appendix D - Evaluating Average Difference in Salary Level

Batters

Traditional Method Specification – Batters

```
. reg  lnsal lnslg lnABper age AGE2 Japan

      Source |       SS       df       MS              Number of obs =    3083
-------------+------------------------------           F(  5,  3077) =  563.47
       Model |  1823.21938     5  364.643876           Prob > F      =  0.0000
    Residual |  1991.23416  3077  .647134923           R-squared     =  0.4780
-------------+------------------------------           Adj R-squared =  0.4771
       Total |  3814.45354  3082  1.23765527           Root MSE      =  .80445

------------------------------------------------------------------------------
       lnsal |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lnslg |   1.481094   .0767511    19.30   0.000     1.330605    1.631583
     lnABper |   .8927026   .0263357    33.90   0.000     .8410653      .94434
         age |   .3363722    .058423     5.76   0.000     .2218202    .4509242
        AGE2 |  -.0049939   .0009056    -5.51   0.000    -.0067695   -.0032183
       Japan |   .2594886   .1722749     1.51   0.132     -.078297    .5972742
       _cons |   12.71623   .9339718    13.62   0.000     10.88496     14.5475
------------------------------------------------------------------------------
```

$$\ln(salary) = \beta_0 + \beta_1 ln(SLG) + \beta_2 \ln\left(AB_{per}\right) + \beta_3(AGE) - \beta_4 AGE^2 + \beta_5(Japan) + \varepsilon$$

Sabermetic Method Specification - Batters

```
. reg  lnsal  runs  age AGE2 Japan

      Source |       SS       df       MS              Number of obs =    3087
-------------+------------------------------           F(  4,  3082) =  954.35
       Model |  2117.48745     4  529.371862           Prob > F      =  0.0000
    Residual |  1709.56119  3082  .554692143           R-squared     =  0.5533
-------------+------------------------------           Adj R-squared =  0.5527
       Total |  3827.04863  3086  1.24013242           Root MSE      =  .74478

------------------------------------------------------------------------------
       lnsal |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        runs |   .0235732   .0003826    61.62   0.000     .0228231    .0243233
         age |   .3551582    .054086     6.57   0.000     .2491099    .4612066
        AGE2 |  -.0051134   .0008382    -6.10   0.000    -.0067569   -.0034699
       Japan |   .1500573   .1595186     0.94   0.347    -.1627162    .4628309
       _cons |   6.866354   .8691756     7.90   0.000     5.162132    8.570577
------------------------------------------------------------------------------
```

$$\ln(salary) = \beta_0 + \beta_1(Runs) + \beta_2(AGE) - \beta_3(AGE^2) + \beta_4(Japan) + \varepsilon$$

Hybrid Method Specification – Batters

```
. reg lnslg RC Small Japan

      Source |       SS       df       MS              Number of obs =    3083
-------------+------------------------------           F(  3,  3079) = 1163.71
       Model |  76.3732192     3  25.4577397           Prob > F      =  0.0000
    Residual |   67.357521  3079  .021876428           R-squared     =  0.5314
-------------+------------------------------           Adj R-squared =  0.5309
       Total |   143.73074  3082  .046635542           Root MSE      =  .14791

------------------------------------------------------------------------------
       lnslg |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      RC_bat |   .0042926   .0000727    59.06   0.000     .0041501    .0044351
       Small |   .0052367   .0057262     0.91   0.361     -.005991    .0164643
       Japan |  -.0316807   .0316667    -1.00   0.317    -.0937708    .0304094
       _cons |  -1.127553   .0055551  -202.98   0.000    -1.138445   -1.116661
------------------------------------------------------------------------------
```

$$\ln(salary) = \beta_0 + \beta_1(Runs\ Created) - \beta_3(Small) + \beta_4(Japan) + \varepsilon$$

Pitchers

Traditional Method Specification – Batters

```
. reg lnsal lnSW lnIP AGE AGE2 Japan

      Source |       SS       df       MS              Number of obs =    2515
-------------+------------------------------           F(  5,  2509) =  296.82
       Model |   1021.9045     5   204.3809            Prob > F      =  0.0000
    Residual |  1727.64482  2509  .688579042           R-squared     =  0.3717
-------------+------------------------------           Adj R-squared =  0.3704
       Total |  2749.54932  2514  1.09369503           Root MSE      =  .82981

------------------------------------------------------------------------------
       lnsal |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        lnSW |   .6298058   .0410485    15.34   0.000     .5493134    .7102981
        lnIP |   .9057056    .029785    30.41   0.000     .8472999    .9641114
         AGE |   .4925648   .0571884     8.61   0.000     .3804235    .6047062
        AGE2 |  -.0067567   .0008908    -7.58   0.000    -.0085035   -.0050099
       Japan |  -.2159403   .1263401    -1.71   0.088    -.4636818    .0318012
       _cons |   1.070845   .9325483     1.15   0.251    -.7577984    2.899488
------------------------------------------------------------------------------
```

$$\ln(salary) = \beta_0 + \beta_1 \ln(\mathrm{SW}) + \beta_2 \ln(\mathrm{IP}) + \beta_3(\mathrm{AGE}) - \beta_4(\mathrm{AGE}^2) + \beta_5(Japan) + \varepsilon$$

Hybrid Method Specification – Batters

```
. reg lnsal W_pitch SV_pitch AGE AGE2 Small Japan

      Source |       SS       df       MS              Number of obs =    2515
-------------+------------------------------           F(  6,  2508) =  328.26
       Model |  1209.45345     6  201.575575           Prob > F      =  0.0000
    Residual |  1540.09587  2508  .614073313           R-squared     =  0.4399
-------------+------------------------------           Adj R-squared =  0.4385
       Total |  2749.54932  2514  1.09369503           Root MSE      =  .78363

------------------------------------------------------------------------------
       lnsal |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     W_pitch |   .1323636   .0032829    40.32   0.000     .1259262     .138801
    SV_pitch |   .0327139   .0015649    20.90   0.000     .0296453    .0357825
         AGE |    .441934    .053982     8.19   0.000     .3360802    .5477879
        AGE2 |  -.0061009   .0008412    -7.25   0.000    -.0077504   -.0044515
       Small |  -.1604358   .0337516    -4.75   0.000    -.2266197   -.0942519
       Japan |  -.2288458   .1194982    -1.92   0.056     -.463171    .0054794
       _cons |   5.616929   .8591786     6.54   0.000     3.932157    7.301702
------------------------------------------------------------------------------
```

$$\ln(salary) = \beta_0 + \beta_1(\mathrm{Wins}) + \beta_2(\mathrm{Saves}) + \beta_3(\mathrm{AGE}) - \beta_4(\mathrm{AGE}^2) - \beta_5(Small) + \beta_6(Japan) + \varepsilon$$

Appendix E - Evaluating Returns to Age

Batters

Traditional Method Specification

```
. reg  lnsal lnslg lnABper age AGE2 JapanxAGE

      Source |       SS       df       MS              Number of obs =    3083
-------------+------------------------------           F(  5,  3077) =  563.35
       Model |  1823.00584     5  364.601168           Prob > F      =  0.0000
    Residual |   1991.4477  3077  .647204322           R-squared     =  0.4779
-------------+------------------------------           Adj R-squared =  0.4771
       Total |  3814.45354  3082  1.23765527           Root MSE      =  .80449

------------------------------------------------------------------------------
       lnsal |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lnslg |   1.480923   .0767547    19.29   0.000     1.330428    1.631419
     lnABper |   .8928816   .0263353    33.90   0.000     .8412451    .9445182
         age |     .33661   .0584255     5.76   0.000      .222053    .4511669
        AGE2 |   -.004998   .0009056    -5.52   0.000    -.0067736   -.0032223
   JapanxAGE |   .0076373   .0054852     1.39   0.164    -.0031178    .0183923
       _cons |   12.71332   .9340248    13.61   0.000     10.88194    14.54469
------------------------------------------------------------------------------
```

$$\ln(salary) = \beta_0 + \beta_1 ln(SLG) + \beta_2 \ln\left(AB_{per}\right) + \beta_3(AGE) - \beta_4 AGE^2 + \beta_5(JapanxAGE) + \varepsilon$$

Sabermetric Method Specification

```
. reg lnsal  runs age AGE2  JapanxAGE

      Source |       SS       df       MS              Number of obs =    3087
-------------+------------------------------           F(  4,  3082) =  954.28
       Model |  2117.41273     4  529.353183           Prob > F      =  0.0000
    Residual |   1709.6359  3082  .554716386           R-squared     =  0.5533
-------------+------------------------------           Adj R-squared =  0.5527
       Total |  3827.04863  3086  1.24013242           Root MSE      =  .74479

------------------------------------------------------------------------------
       lnsal |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        runs |    .023575   .0003825    61.63   0.000     .0228249    .0243251
         age |   .3553071   .0540866     6.57   0.000     .2492578    .4613565
        AGE2 |  -.0051159   .0008382    -6.10   0.000    -.0067594   -.0034724
   JapanxAGE |   .0043989   .0050788     0.87   0.386    -.0055594    .0143572
       _cons |   6.864153   .8691902     7.90   0.000     5.159902    8.568404
------------------------------------------------------------------------------
```

$$\ln(salary) = \beta_0 + \beta_1(Runs) + \beta_2(AGE) - \beta_3(AGE^2) + \beta_4(JapanxAGE) + \varepsilon$$

Hybrid Method Specification

```
. reg lnsal RC Small JapanxAGE

      Source |       SS       df       MS              Number of obs =    3087
-------------+------------------------------           F(  3,  3083) = 1199.57
       Model |  2061.21583     3  687.071945           Prob > F      =  0.0000
    Residual |   1765.8328  3083   .57276445           R-squared     =  0.5386
-------------+------------------------------           Adj R-squared =  0.5381
       Total |  3827.04863  3086  1.24013242           Root MSE      =  .75681

------------------------------------------------------------------------------
       lnsal |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      RC_bat |   .0221517   .0003712    59.68   0.000     .0214239    .0228795
       Small |  -.0850534   .0292865    -2.90   0.004    -.1424765   -.0276303
   JapanxAGE |   .0068705   .0051584     1.33   0.183    -.0032437    .0169847
       _cons |   13.18464   .0283449   465.15   0.000     13.12906    13.24022
------------------------------------------------------------------------------
```

$$\ln(salary) = \beta_0 + \beta_1(Runs\ Created) - \beta_3(Small) + \beta_4(JapanxAGE) + \varepsilon$$

Pitchers

Traditional Method Specification

```
. reg lnsal lnSW lnIP AGE AGE2 JapanxAGE

      Source |       SS       df       MS              Number of obs =    2515
-------------+------------------------------           F(  5,  2509) =  296.54
       Model |  1021.29867     5  204.259733           Prob > F      =  0.0000
    Residual |  1728.25065  2509  .688820506           R-squared     =  0.3714
-------------+------------------------------           Adj R-squared =  0.3702
       Total |  2749.54932  2514  1.09369503           Root MSE      =  .82995

------------------------------------------------------------------------------
       lnsal |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        lnSW |   .6298631   .0410557    15.34   0.000     .5493566    .7103695
        lnIP |   .9053865   .0297885    30.39   0.000      .846974     .963799
         AGE |     .49249   .0572155     8.61   0.000     .3802955    .6046845
        AGE2 |  -.0067547   .0008912    -7.58   0.000    -.0085023   -.0050072
   JapanxAGE |  -.0057914    .004054    -1.43   0.153    -.0137409    .0021581
       _cons |   1.072015   .9330017     1.15   0.251    -.7575174    2.901547
------------------------------------------------------------------------------
```

$$\ln(salary) = \beta_0 + \beta_1 \ln(\text{SW}) + \beta_2 \ln(\text{IP}) + \beta_3(\text{AGE}) - \beta_4(\text{AGE}^2) + \beta_5(Japan * AGE) + \varepsilon$$

Hybrid Method Specification

```
. reg lnsal W_pitch SV_pitch AGE AGE2 Small JapanxAGE

      Source |       SS       df       MS              Number of obs =    2515
-------------+------------------------------           F(  6,  2508) =  328.06
       Model |   1209.0402     6    201.5067           Prob > F      =  0.0000
    Residual |  1540.50912  2508  .614238084           R-squared     =  0.4397
-------------+------------------------------           Adj R-squared =  0.4384
       Total |  2749.54932  2514  1.09369503           Root MSE      =  .78373

------------------------------------------------------------------------------
       lnsal |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     W_pitch |   .1323587   .0032833    40.31   0.000     .1259204     .138797
    SV_pitch |   .0327212   .0015651    20.91   0.000     .0296521    .0357903
         AGE |   .4421845   .0540054     8.19   0.000     .3362848    .5480843
        AGE2 |  -.0061037   .0008415    -7.25   0.000    -.0077538   -.0044536
       Small |  -.1599309   .0337517    -4.74   0.000    -.2261149   -.0937469
   JapanxAGE |  -.0066338   .0038341    -1.73   0.084    -.0141521    .0008845
       _cons |   5.611269   .8595615     6.53   0.000     3.925746    7.296792
------------------------------------------------------------------------------
```

$$\ln(salary) = \beta_0 + \beta_1(\text{Wins}) + \beta_2(\text{Saves}) + \beta_3(\text{AGE}) - \beta_4(\text{AGE}^2) - \beta_5(Small) + \beta_6(JapanxAGE) + \varepsilon$$

WORKS CITED

Ackman, D. (2007). Sultan of Stats. *The Wall Street Journal*. Retrieved from http://www.opinionjournal.com

Albert, J. (1999). An Introduction to Sabermetrics. Bowling Green State University.

Burdekin, R., Hossfield, R., & Smith, J. (2005). Are NBA Fans Becoming Indifferent to Race?: Evidence from the 1990s. *Journal of Sports Economics*, *6*(2), 16.

Davis, A., & Bryant, P. (2008). Diverse Teams Outperform Homogenous Teams - Everytime. *Universal Advisor*(One), 10-11.

Gmelch, G. (2006). *Baseball without Borders: The International Pastime*. Lincoln, NE and London: University of Nebraska Press.

Hill, J. R., & Spellman, W. (1984). Pay Discrimination in Baseball: Data from the Seventies. *Industrial Relations*, *23*(1), 103-112.

. *Japan* (2008). Retrieved from http://www.cia.gov.

Kahn, L. M. (1991). Discrimination in Professional Sports: a Survey of the Literature. *Industrial and Labor Relations Review*, *44*(3), 395-418.

Kahn, L. M. (2000). The Sports Business as a Labor Market Laboratory. *The Journal of Economic Perspectives*, *14*(3), 75-94.

Klein, A. M. (2006). *Growing the Game: The Globalization of Major League Baseball*. New Haven and London: Yale University Press.

Luttrell, J. (1999). Library/Baseball Statistics: Tracking the Majors, or a League of Your Own. *The New York Times*. Retrieved from http://www.nytimes.com

McCracken, V. (2001). Pitching and Defense: How Much Control Do Hurlers Have? Retrieved April 7, 2008, from http://www.baseballprospectus.com

McGrath, B. (2003). The Professor of Baseball. *The New Yorker*.

Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, *14*(3).

Rains, R. (2001). *Baseball Samurais: Ichiro Suzuki and the Asian Invasion*. New York: St. Martin's Paperbacks.

Scully, G. W. (1974). Pay and Performance in Major League Baseball. *The American Economic Review*, *64*(6), 915-930.

Solloway, S. (2007). Gorham Man's Gift to Japan: a National Pastime. *Portland Press Herald*.

Whiting, R. (2004). *The Meaning of Ichiro: The New Wave from Japan and the Transformation of Our National Pastime*. New York: Warner Books.