Media Bias in Major League Baseball Award Voting:
An Examination of Voting Discrimination in the MVP and Cy Young Awards

Nicholas D. Buonome
College of the Holy Cross
Worcester, MA 01610
(860) 918-0245
ndbuon06@holycross.edu

ABSTRACT:

This paper seeks to investigate potential voting bias of the Baseball Writers Association of America (BBWAA) in the MVP and Cy Young awards. Using Win Shares as a measure of a player's true value, it was possible to analyze the difference between how good someone statistically is and how our national sportswriters perceive them. This study provides evidence of media-based discrimination in both the MVP and Cy Young award voting during the ten year period after Major League Baseball's strike in 1995. Results showed discrimination against Whites in MVP voting, and discrimination against Nonwhites in Cy Young voting. The results add to the empirical literature investigating media-based discrimination in professional sports.

Acknowledgements:

Victor A. Matheson (Advisor)
Department of Economics
College of the Holy Cross
Worcester, MA 01610
(508) 793-2649
vmatheso@holycross.edu

I. INTRODUCTION:

The concept of the MVP and Cy Young awards is quite obvious; the best overall player and pitcher win. While this seems simple, there historically has not been a set metric to objectively measure this. In fact, determining a baseball player's value has and will most likely continue to be one of the most argued topics in sports. For example, how can one compare someone who hits for average versus another with great power? While the former is likely to get on base more often and prolong innings, he may not always induce runs scored or batted in. The latter, while less likely to get on base, can put runs on the board with one swing.

This question of which type of hitting is more beneficial to a team is just one of many uncertainties in baseball. Previously, questions like this never had a right answer; everyone had their own perception of what was important in baseball and used many examples from the past to support their theory. Only until a fully comprehensive statistic could be created allowing one to wholly measure a player's value over another would controversy be solved.

Eventually the idea of a baseball production function came along. It was believed answers to questions of player value can and should be based on the statistical analysis of an algorithm totally inclusive of all the actions in baseball that lead to winning games. Over the past twenty years many studies worked with this new concept, but each fell short in totally encompassing this objective. However recently, statistical advancements in the technology-crazed information age triggered much of baseball's conventional wisdom to be reconsidered. The development of new statistical methodology coupled

with the increasing number of available baseball statistics, has finally led to the creation of a player value statistic to end the debates.

Developed by Bill James in 2002, "Win Shares" is a way to boil down a player's contribution to a single number. This widely accepted statistic has revolutionized baseball mentality and quickly developed into the best unbiased gauge of player quality. As a result, Win Shares can be a powerful tool when examining baseball's most prestigious awards, the MVP and Cy Young. It would seem logical for the order of player Win Shares in a season to closely follow the final Baseball Writers Association of America (BBWAA) voting results. Yet, this is not the case. In fact, rarely does the order of player value as determined by Win Shares match with what is perceived by the media. Therefore, there must be some sort of discrimination, both team and player specific, which has caused sportswriters to vote the way they have over the years.

Discrimination has been a part of human interaction for as long as man has existed. As a result, there have been numerous academic studies of discrimination, especially in the field of sports. In the case of Major League Baseball, several papers investigate customer based discrimination using evidence from the baseball card markets and all-star ballots. However, little research has been conducted questioning any prominent media discrimination. While it is important to focus on studies of the general public, we should also focus on the baseball media and its most famous end of the season awards.

Does the media, in fact, differ from the general public? And if so, in what ways? Debate has risen over the years questioning whether or not the media shows preference in voting for these awards based on a certain race, team location, position, etc. It cannot be

ignored that sportswriters, especially today, have a significant effect on the general public. The fact is, it is impossible for anyone to watch every single Major League Baseball game in a season. Therefore, what we see and believe in baseball is, for the most part, from the eyes of the media. Consequently, how do we know what we read in the Sports pages or watch on *Sportscenter* is not a biased opinion of our national media? The most prestigious awards of our country's national pastime are not being judged by the people who pay for tickets, but rather by select journalists that make up the BBWAA. Is there, in fact a difference of opinion between the public and the media? This paper hopes to find out what discriminatory factors, if any, are part of a national media bias which possibly skewed voting results and robbed worthy winners of the Most Valuable Player and the Cy Young awards for years.

II. HISTORY OF SPORTS DISCRIMINATION STUDIES:

Zech (1981) chooses to look at the case of Major League Baseball and its production function to identify and empirically estimate those factors contributing to team success. This paper uses cross-sectional data from the 26 Major League Baseball teams in the 1977 season to derive the production function for the MLB industry. As an application of those results, the paper focuses on measuring a league's most valuable player by constructing a weighting scheme based solely on a player's performance. To compute each player's marginal product, Zech calculates the team's resulting winning percentage without that player's skills and simply translates this into games won. The difference between that figure and the number of games the team actually won would

serve as that player's MP. Hence, barring any type of media bias, the player with the highest MP should be that league's Most Valuable Player.

A summary of the main results shows MVP's compiled using this application of the baseball production function differed significantly from the official voting list of the Baseball Writers Association of America (BBWAA). This result leads Zech to believe that differences in national publicity received by players and perhaps personality considerations are the reason for the discrepancy.

Woolway (1997) builds upon the analysis of Zech by updating the baseball production function through identifying new factors contributing to team success. In his study, Woolway uses cross-sectional data from the 1993 Major League Baseball season to derive his baseball production function.[1] Although Woolway uses his function to derive players' marginal products like Zech, he does not compare his calculated "Most Valuable Players" with the ones chosen by the BBWAA. Nevertheless, Woolway does address a simple examination of the 1992 and 1993 National League Cy Young Award winner Greg Maddux using his estimated player marginal products. His analysis shows that although Maddux's marginal products are not exceptionally high in those years, they were still higher than that of any other pitcher in the sample. Thus, Maddux was in fact most worthy of the award. This type of evidence contrasts the perceived bias of sportswriters mentioned by Zech. However, it does only reflect two years of the sample and only takes into account the person who won the award. Perhaps the entire voting result was not in agreement with Woolway's marginal product rankings.

Essentially, this conceptual application of a production function provides baseball analysts with a completely objective way to measure value of players. Therefore, it

seems only logical to base these awards on player's relative marginal products. Any deviation from these calculated measures should suggest some type of bias toward other aspects. It must be noted, however, there are serious issues concerning these previous studies, and thus their analysis might not be completely factual. For starters, while the Cobb-Douglas model does seem to be appropriate for this estimation, some of the chosen independent variables are not completely representative of what wins baseball games. Even though Woolway builds upon Zech using more statistically significant variables like on base percentage (OBP) and slugging average (SA), advanced sabermetric statistics have been recently developed that could serve as better measures. Additionally, the method used to determine a player's marginal product is extremely flawed. Comparing predicted wins to predicted wins without a player does not account for the fact a team would undoubtedly get some support from a player they picked up to fill the empty spot. It would make more sense to instead replace the player under examination with the 31st best player at his position and then compare to predicted wins from the production function model.[1] However, this method also has its caveats.

As outlined previously, Zech and Woolway are hitting on important facts about the game. However they are just touching the surface of how this measure of player quality should be obtained. With the increased number of published baseball datasets, use of high-powered computers, and the growing interest of brilliant statisticians, there has finally been a method deduced which has grabbed the attention of baseball analysts and fans alike. The new method of determining what wins in baseball can be explained by the concept of Win Shares.

The arrival of Bill James' "Win Shares" (2002) has given sports analysts and team managers the best measure of a player's value to his team. Win Shares is a relatively new statistical phenomena. It uses an extremely complex methodology factoring in everything from outs made, runs created, park effects, etc. to output a very simple and intuitive value which can be used to compare baseball player's contributions to their teams. James writes,

> "[Win Shares] removes illusions of context, putting a hitter from Yankee Stadium on equal footing with a hitter from Colorado, and putting a hitter from 1968 on equal footing with a hitter from 2000. Second, the Win Shares system attempts to state the contributions of pitchers and of fielders in the same form as those of hitters."

Not only are Win Shares a better measure than the previous marginal product methods, but they are also much more practical when factoring in the time constraint of calculating marginal products for every year of data. As a matter of fact, the algorithm has been used to output Win Shares for all players from the 1870s to the present. All Win Shares data during this time period along with various applications are contained in James' book.

As reported by James in his article "Win Shares and MVP Voting", from 1931 to 2001 only 58 of 142 Most Valuable Players (41%) have had or have shared the highest Win Share totals in their respective leagues. In fact, the 1940s had the lowest percentage (6 out of 20), while the 1970s (11 out of 20) had the highest. An updated list of the past ten years (1995-2004) Most Valuable Players, as determined by Win Shares, and as determined by the BBWAA voters is listed below in Table 1. Notice during this time period only 6 out of 20 actual MVPs correspond with the MVPs as determined by Win Shares.

Table 1. Actual vs. Win Shares MVP (1995-2004)

Year	League	Method	MVP Winner
1995	AL	Win Shares MVP	Edgar Martinez (32)
1995	AL	Actual MVP	Mo Vaughn (24)
1995	NL	Win Shares MVP	Barry Bonds (36)
1995	NL	Actual MVP	Barry Larkin (30)
1996	AL	Win Shares MVP	Alex Rodriguez (34)
1996	AL	Actual MVP	Juan Gonzalez (21)
1996	NL	Win Shares MVP	Jeff Bagwell (41)
1996	NL	Actual MVP	Ken Caminiti (38)
1997	AL	Win Shares MVP	Frank Thomas (39)
1997	AL	Actual MVP	Ken Griffey Jr. (36)
1997	NL	Win Shares MVP	Piazza and Gwynn (39)
1997	NL	Actual MVP	Larry Walker (33)
1998	AL	Win Shares MVP	Albert Belle (37)
1998	AL	Actual MVP	Juan Gonzalez (25)
1998	NL	Win Shares MVP	Mark McGwire (41)
1998	NL	Actual MVP	Sammy Sosa (35)
1999	AL	Win Shares MVP	Alomar, M. Ramirez and Jeter (35)
1999	AL	Actual MVP	Ivan Rodriguez (28)
1999	NL	Win Shares MVP	Jeff Bagwell (37)
1999	NL	Actual MVP	Chipper Jones (32)
2000	AL	Win Shares MVP	Jason Giambi (38)
2000	AL	Actual MVP	Jason Giambi (38)
2000	NL	Win Shares MVP	Jeff Kent (37)
2000	NL	Actual MVP	Jeff Kent (37)
2001	AL	Win Shares MVP	Jason Giambi (38)
2001	AL	Actual MVP	Ichiro Suzuki (36)
2001	NL	Win Shares MVP	Barry Bonds (54)
2001	NL	Actual MVP	Barry Bonds (54)
2002	AL	Win Shares MVP	Alex Rodriguez (35)
2002	AL	Actual MVP	Miguel Tejada (32)
2002	NL	Win Shares MVP	Barry Bonds (49)
2002	NL	Actual MVP	Barry Bonds (49)
2003	AL	Win Shares MVP	A. Rodriguez and C. Delgado (32)
2003	AL	Actual MVP	Alex Rodriguez (32)
2003	NL	Win Shares MVP	Albert Pujols (41)
2003	NL	Actual MVP	Barry Bonds (39)
2004	AL	Win Shares MVP	Gary Sheffield (31)
2004	AL	Actual MVP	Valdimir Guerrero (29)
2004	NL	Win Shares MVP	Barry Bonds (53)
2004	NL	Actual MVP	Barry Bonds (53)

The discrepancy above raises questions about what sort of discrimination could exist in MLB award voting. What helps in answering this question is that fact that a large amount of research has focused on sports discrimination. In the article, “Biases in MVP

Voting", James (2002) examines many possible explanations. He starts by looking at a generalized population of players having seasons of 30 or more Win Shares to figure out how many Most Valuable Player Awards should be expected from that population. From there, James first looks at a possibility of team specific bias. He questions a bias in MVP voting for players who play on New York teams. While many sportswriters stand firm that there is such a bias, James ultimately disproves this common assumption. New York players simply do not have an advantage in MVP voting.

Additionally, James (2002) examines another team specific bias by showing that a player on a pennant-winning team is more likely to win the MVP Award than the same player playing on a non-championship team. In fact, only four of the previous 148 MVPs have played for losing clubs. Verducci (2003) argues that since no rule states the MVP must come from a winning team, it must be rather just a natural preference. It is conventional wisdom that MVPs are typically defined to be players who produce in important spots for playoff teams all season long, especially down the stretch. However, this way of thinking will not always honor the best player in the league and ultimately penalizes someone by holding the rest of his team against him.

Another major team specific factor to consider for both awards is a team's market size. It is no secret that some MLB cities have a greater fan base than others. For example, a team like the Milwaukee Brewers, with an estimated market of 1,691,294 people, cannot compare to a market of 21,404,951 people associated with the New York Yankees and Mets. Data used to determine market size was based on the 2000 Bureau of Economic Analysis "Combined Metropolitan and Micropolitan Statistical Areas, Metropolitan Divisions, and Combined Statistical Areas CA1-3 Population" records.[1] By

using this category of data it was possible to gauge a team's market size not solely on its metropolitan city, but also on the combined metropolitan statistical areas in close proximity. Therefore, a team like the New York Yankees was measured over New York-Newark-Bridgeport, encompassing the states NY-NJ-CT-PA, to give a true sense of its total fan base. While this seems arbitrary at first, this way off grouping provides a balanced separation of teams and fits in with other studies of baseball markets (Streit 2005).

Moving onto the topic of player specific bias, suspicion has been raised for positional bias in MVP awards. The most common argument, and one that has grown since 1986, is that pitchers are not equally considered for MVP awards.[1] The rationale is that pitchers only play every five days, therefore how can they be as important to their team as someone who plays all 162 games? Heyman (1999) argues that after putting together one of the most dominant seasons ever by a baseball player, Pedro Martinez did not receive his fair share of first-place votes in the American League. What is even more disturbing to Hayden is that Martinez was completely left off of some of the 10-player ballots. Today baseball wisdom is increasingly following the notion that to be MVP, you must not only be in the lineup, but in the field everyday to help win games – something pitchers simply cannot do. BBWAA sportswriter La Velle E. Neal III of the *Star Tribune* in Minneapolis has even gone on record saying "I've determined that pitchers should not win the award." Still, some argue there is no such widespread prejudice in MVP voting since it has been won by a pitcher twenty times in its history. Nevertheless, statements like the one by Neal must draw attention to this topic. Additionally, catchers have historically been judged by the media as viable candidates for the MVP award. However,

James (2002) notes that most of the worst MVP selections, in reference to the Win Shares system, are catchers.[1] In fact, the system rarely ranks catchers as the best player in the league. Thus, there appears to be some certain qualities about catchers that drive the media to vote in their favor versus towards another positional player who is more deserving. Perhaps other positional bias could be considered as well.

Furthermore, another area of player specific bias is player reputation. Zillante (2005), in his analysis of reputation effects in Gold Glove award voting, shows that voters may be relying more on the reputation of a player previously receiving the award rather than on actual current defensive performance.[1] Since the Gold Glove award is solely for defensive performance in the current season, Zillante proves reputation does affect Gold Glove voters when casting their ballots. However, James (2002) finds contrary evidence of reputation bias in his examination in MVP voting. He shows that players who have their first outstanding season are more likely to win an MVP Award than players who have been performing at an equal or slightly better level for years. James stresses that a player's chance of winning the MVP increases as he reaches a new level of success. In other words, a player who is performing at the same exceptional level over a long period of time does not catch a voter's eye like someone who has a surprising exceptional season. When looking at the 2000-2001 AL MVP race, Ichiro Suzuki (36), according to win share totals, was not the best player. Other players such as Roberto Alomar (37), Alex Rodriguez (37), and Jason Giambi (38) were slightly better. However Ichiro was new and the others had a history of seasons like this before. Essentially James proves reputation has a reverse effect in MVP voting than it does in Gold Glove voting. With this said, it is critical to know James goes about his analysis of MVP bias

differently than in this study. Therefore it is still important to incorporate measures of reputation effects in this analysis.

The next obvious direction to move is to examine race discrimination, as it is always an issue to come up when looking for evidence of player specific bias. Nardinelli and Simon (1990) set out to find a valid and explanatory measure for consumer discrimination to examine whether race directly affects the entertainment value of a player in the market for baseball cards.[1] Their results indicate that consumer discrimination exists in the market for baseball cards. The regression shows that for hitters, the cards of nonwhites sell for 10 percent less than the cards of white players of comparative ability. Similarly, among nonwhite pitchers the discount is 13 percent (Nardinelli 1990).[1] Although the market for baseball cards is small, the results do suggest something about the problem of consumer discrimination. Analysis of their findings in reference to the lack of personal contact in the market for baseball cards is somewhat shocking. Nardinelli and Simon state, "The absence of personal contact [between card owners and players] should reduce the potential for consumer discrimination. That it does not eliminate [consumer discrimination] may be our most important result". Although this paper looks at the general public side to racial bias in baseball, it does provide a stepping-stone in my ultimate analysis to find if there is any form of racial discrimination by the media in voting for a league's MVP and Cy Young winner. Using the same idea of including dummy variables for race should lead to more explanations of the difference between how good someone statistically is and how sportswriters perceive them.

Similarly, Depken and Ford (2003) also examine customer-based discrimination in Major League Baseball. Yet they act rather through the lens of All-star votes from 1990 through 2000. Contrary to Nardinelli and Simon, their study finds no evidence of customer-based racial discrimination against minority players. In fact, Black and Latinos were actually preferred by MLB all-star voters. This approach uses an alternative measure of player value different from the analysis of baseball card prices. Their model uses a player's vote total as the dependant variable and controls for individual and team characteristics, which could affect discrimination of players. For one, position dummy variables are used as all positions may enjoy different levels of notoriety. Also, name recognition dummy variables were used testing for reputation bias. They include: if a player was a previous All-Star, if a player won MVP in the previous year, player age, and MLB experience. Lastly, team recognition dummy variables were added, controlling for if a team made it to the playoffs in the previous year, whether the team was an expansion team, if the All-star game was held there, attendance of the player's team, and where a team is located.

This type of discrimination analysis is extremely relevant because it measures many different types of potential bias factors. While Depken and Ford concentrate mainly on customer-based racial discrimination in MLB All-star voting, this paper will examine MVP and Cy Young voting to assess any type of media-based bias for these awards. Application of this analysis could be useful to determine if the press has different perceptions than the public when voting for Major League Baseball's best players.

Lastly, Desser, Monks, and Robinson (1999) look at the effects of race on induction into the Baseball Hall of Fame by testing whether there is discrimination against African and Latin Americans. Their results propose strong voting bias against both African Americans and Latin Americans in Hall of Fame balloting while also suggesting a preference in placing white players on the ballot. On the other hand, Jewell, Brown, and Miles (2002) find different evidence by extending research to also test for players who are both Black and born in Latin America. This concept is very important in Major League Baseball as many players we consider "Black" can be from the Dominican Republic, Venezuela, Panama, etc. – all Spanish-speaking countries. Their data showed limited evidence of discrimination against Latin players. However, no evidence supported bias among voters against black players, with the exception of black players born in a Latin American country. Ultimately, their findings show that race did not affect whether the players in their sample received enough votes for induction on the first ballot.

The results of these papers specifically relate to the topic of MVP voting as the sportswriters also vote on the Hall of Fame ballots. These studies use a model to control for lifetime performance statistics, single season achievement, career awards, and race. Although this paper will not use lifetime statistics because they should not matter in an MVP decision for a certain year, other notable variables to consider are: leading the league in batting or homeruns, number of times selected to the All-star team, gold glove awards, MVP awards, number of times the player appeared on the ballot, if they played in a large media market, and relatives in the league.

III. DATA/METHODOLOGY:

In this study, analysis of bias in baseball media awards is based on data ranging from the 1995 to 2004 seasons.[1] This time series not only provides a decade worth of data, but it also marks the beginning of a new era for Major League Baseball as the league attempts to recover from the effects of the 1994 strike causing the first ever cancellation of the World Series. Much of the information used in this analysis is directly from Baseball-Reference.com. This site lists complete voting results along with individual player statistics and biographies for all of MVP and CY Young awards since their inceptions.

Data used for MVP and Cy Young voting results were reported by the Baseball Writers Association of America (BBWAA), the determining body in MLB player awards. The MVP voting process follows the Borda count method, which is based on a 14-9-8-7-6-5-3-2-1 basis. Under this method 1st place votes are worth 14 points, 2nd place votes are worth 9 points, and so on. Cy Young voting, although similar, is more scaled down. It follows a system where 1st place votes are 5 points, 2nd place votes are 3 points and 3rd place votes are 1 point. Thirty-two sportswriters vote for these awards in the National League while twenty-eight sportswriters vote in the American League. Due to the fact the American and National Leagues differ with the number of maximum points in the voting process, it is necessary to take the share of votes received in order to obtain a comparable variable to be used in this analysis. As a result, the maximum number of points a player can receive in his league must divide the total number of points actually received by that player

In this analysis, I have kept the model similar to those used in studies of customer-based discrimination in trading card markets and all-star balloting. I estimate a linear relationship SHARE = f(X, β) + e, where SHARE is a vector of player share of voting points received, X is a matrix of various player-specific and team-specific characteristics, β is a vector of parameters to be estimated, and e is the error term. A list of variables is reported in Table 2. The dependent variable (SHARE) is the calculated player share of MVP and Cy Young voting points received. The independent variables include player and team attributes which could possibly uncover certain types of media bias in MLB award voting.

Table 2. Variable Names and Definitions

Variable Type	Name	Definition
Dependent	**SHARE**	Ratio of MVP votes received divided by the max points
Reputation	**PSHARE**	Previous year's SHARE ratio
	MVP	Equals 1 if player is a previous MVP, 0 otherwise
	CY*	Equals 1 if player is a previous CY Young winner, 0 otherwise
	AGE	Age of player
	AGE2	Age of player squared
	YP	Years player has played
	YP2	Years player has played squared
Player Quality	**WS**	Current season win shares
	HR	Current season home runs
	ERA*	Current season earned run average
	WINS*	Current season pitching wins
Position	**1B**	Equals 1 if first baseman, 0 otherwise
	2B	Equals 1 if second baseman, 0 otherwise
	3B	Equals 1 if third baseman, 0 otherwise
	SS	Equals 1 if short stop, 0 othersiwe
	C	Equals 1 if catcher, 0 otherwise
	OF	Equals 1 if outfielder, 0 otherwise
	P	Equals 1 if pitcher, 0 otherwise
	DH	Equals 1 if designated hitter, 0 otherwise
	SP*	Equals 1 if starting pitcher, 0 otherwise
Race	**US**	Equals 1 if US born, 0 otherwise
	BLACK	Equals 1 for Blacks, 0 otherwise
	LATINO	Equals 1 for Latinos, 0 otherwise
	WHITE	Equals 1 for Whites, 0 otherwise
	OTHER	Equals 1 for Other, 0 otherwise

League	**AL**	Equals 1 if player's team in American League, 0 otherwise
	EAST	Player's team is in the Eastern Division
	CENT	Player's team is in the Central Division
	WEST	Player's team is in the Western Division
Team	**MKT**	Equals 1 if team is in a large market, 0 otherwise
	TWPCT	Current season team win percentage

*Variables used only in Cy Young regression

Observations (MVP) - 505
Observations (CY) - 131

Therefore, the first independent statistic I would like to focus on is Win Shares (WS). The statistic measures the number of a team's wins that are directly attributed to a single player's performance. It seems quite intuitive then that Win Shares should ultimately determine the winner of the MVP and Cy Young award. Due to the fact a player's Win Shares are divided between teams if traded mid-season, it was necessary to make adjustments for those players under analysis falling into this category. In these few circumstances, I simply added the player's Win Shares together for his season total. In this study I would hope to find the win shares statistic to be statistically significant and very highly correlated with a player's voting share.

It is not always the case, however, that the players with the most Win Shares receive these awards. Therefore other statistics should be taken into consideration when examining player specific bias in MVP voting. In my opinion, the most recognizable player quality statistic that could be a significant factor in determining an MVP winner is a player's home run total (HR).[1] Home runs are arguably the most exciting and influential event in a baseball game. As a result, the media focuses much of its attention on players' home runs; whether it is a highlight on SportsCenter or headlined in the

Sports page. Although HRs will be correlated with Win Shares I have still decided to include them for the reasons above.

In the case of the Cy Young award, other player specific statistics should be taken into consideration when examining voting bias as well. In this part of the analysis I will examine a pitcher's earned run average (ERA) and their total number of wins (WINS). A pitcher's ERA, although its telling qualities can be muddled if he receives no run support, has become a major statistic when differentiating among pitchers. Even though wins are not a very telling statistic, people like the number because they come with bragging rights. Each year the media lists baseball's 20-game winners, not pitchers with exceptional winning percentages. I believe these to be the two most recognizable pitching quality statistics, which could factor into Cy Young voting share.

On top of performance, another common player specific factor of how voters tend to judge is on reputation. Many previous studies have analyzed reputation effects by including a variable for previous awards received. My research suggests differences in the perception that a player formerly recognized with an award tends to receive more attention when it comes time again for voting. In this study I will include a dummy variable "MVP" in the MVP regression and "CY" in the Cy Young regression, which will equal one if a player has in the past received the MVP or Cy Young award. This dummy variable will control for any reputation effects in conjunction with the award. I further control for player recognition by including a variable "PSHARE" to represent the share of votes the player received in the previous year. This variable will control for any type of carryover effect voters have from players they voted for in the previous year's ballot. I further control for name recognition by including variables "AGE" and "AGE2"

for the age of a player (and its quadratic) and "YP" and "YP2" for the number of years the player has played (and its quadratic).

A further player specific area to study in this analysis is positional discrimination. The truth is that some positions may enjoy different levels of notoriety. For one, it is common baseball knowledge that certain sportswriters will simply not vote for a pitcher to win the MVP. Furthermore, catchers typically do not produce eye-catching offensive numbers; they are sound hitters distinguished rather for their defensive capabilities. It would thus be interesting to see if a catcher with the same or slightly less Win Shares is more likely to get the MVP. In the MVP analysis I will therefore create position dummy variables.[1]

Along the same lines, there might be bias by sportswriters towards relief pitchers in Cy Young award voting. The rationale here is that relievers do not throw more than 100 innings in a season, while starters pitch average 200+. Therefore how can relievers possibly be considered as important to their teams? The only stipulation is that relievers are usually a part of more games in the season as they can be counted on throwing a couple of innings in consecutive games. As a result there have not been many relievers over the past 20 years to win the Cy Young. In the MVP analysis I therefore create a dummy variable "SP" which will equal one if the player is a starting pitcher.

The next topic of media voting discrimination I chose to examine was that of player race. Although it seems quite evident today that racial discrimination is not a part of baseball as it once used to be in the 1950s, I still wanted to see if the color of ones skin would significantly affect the number of votes a player received in the MVP and Cy Young balloting. I controlled for race by using dummy variables "WHITE", "BLACK",

"LATINO", "OTHER", with white being the omitted category in the empirical specifications. The method I used to determine what players fit into each category is a standard method used by most sporting analysts. The idea is to look at the player's picture under consideration and make a purely visual perception of their race. This method, I believe, is better than going by their actual nationality because nationality simply is not a true reflection of skin color. A black player like Manny Ramirez, for example, can be very deceiving. If categorized by his Dominican nationality he would most definitely fit under Latino. However, most categorizations of people are based on their appearance and under that assumption Manny certainly fits better into Black. I think this approach is the best way to determine if there is any racial discrimination in media voting because looks have a first and lasting impression.

The last type of player specific factor considered was a nationality bias. Is there, in fact, any type of discrimination towards players not born within the United States? It seems possible the American media would like to see its country's baseball awards go to an American personality that it could connect with. In this study I will include a dummy variable "US" in the MVP and Cy Young regression, which will equal one if a player was born within the United States. This bias might be towards giving the country a face of its own someone to identify with.

Moving on to team specific variables, it is first necessary to include a league dummy variable to account for the many differences between the American and National leagues. Thus, a dummy variable "AL" will be included that equals one if the team is in the American League.

Furthermore, evidence from James (2002) and Verducci (2003) suggests that it is not only how good a player is that determines the number of MVP votes received, but also how good others are around him. Therefore, team winning percentage (TWPCT) most certainly needs to be included in this analysis. By doing so we will see if the media truly does hold bias against those players who are on sub .500 teams.

The next team specific variable to examine possible bias in award voting is market size. When differentiating between large and small market teams I chose the dividing factor to be a population of 5 million people. The statistical areas associated with a team's city at or above 5 million people were classified as large market teams and those below were considered small market teams. In this analysis I used the dummy variable "MKT" which equaled one if the team was in a large market. See Table 3 below for the separation of teams. Additionally, see Appendix A for exact Metropolitan Statistical Areas/Combined Statistical Areas population number.

Table 3. Market Size Chart (2000)

Large Market Cities (> 5 Million people)
NYY, NYM, ANA, LAD, CHC, CHW, BAL, OAK, SFG, PHI, BOS, DET, TEX, ATL*
* Atlanta the exception here with a population < 5 Million.

Small Market Cities (< 5 Million people)
FLA**, TOR, HOU, SEA, MON, ARI, MIN, CLE, SDP, STL, COL, TBD, PIT, CIN, KCR, MIL
** Florida the exception here with a population > 5 Million.

Two exceptions to this method of classification, however, were made. The Florida Marlins, with a market size slightly over 5 million people are the one exception in the Small Market group. It is understood in the baseball world that the Marlins are perennial poor draws for their games and that the other sports teams in the area take

precedence. Even during winning seasons the team still cannot fill its stands and gets very little attention in the press. The Atlanta Braves, with a market size slightly below 5 million people are the one exception in the Large Market group. The Braves, a contender year after year, have every single game broadcast nationally on TBS and are arguable as recognizable as the Yankees and other top large market teams. Therefore I believed this team fits in much more under the large market classification.

The last type of discrimination to be considered was a regional bias. While Depkin and Ford (2003) based their analysis on the four distinct regions of the United States, I decided to go solely by the separation of divisions within the American and National leagues, which certainly describe what part of the country a team is located. Regional differences in voting are controlled for with dummy variables "EAST", "CENTRAL", and "WEST", with East being omitted in empirical conditions. In addition, by measuring the regional bias this way we will also see if any divisions tend to receive more of the votes.

IV. RESULTS:

The MVP models are estimated using OLS on a sample of 505 observations; the results are reported in Table 4. Model I controls for the American League, Model II controls for the National League, Model III controls for AL positional dummy variables, and Model IV controls for NL positional dummy variables. Models I and II were needed to take into account the Designated Hitter in the American League and used as the basis for coefficient analysis as they produced the highest R^2. Models III and IV are necessary

to test for positional bias since I could not factor in pitchers into the main regressions because the HR variable would skew results.

Table 4. MVP Robust Regression Results
(Dependent Variable: Ratio of MVP votes received divided by the max points)

Variable	Model I	Model II	Model III	Model IV
(Constant)	-1.273	-0.172	-1.502	-0.332
	(0.664)	(0.407)	(0.691)	(0.440)
PSHARE	-0.079	0.080	-0.031	0.114
	(0.065)	(0.051)	(0.066)	(0.054)
WS	0.019	0.016	0.022	0.020
	(0.002)*	(0.002)*	(0.002)*	(0.002)*
TW_PCT	0.772	0.909	0.718	0.799
	(0.189)*	(0.159)*	(0.195)	(0.169)
HR	0.006	0.007	---	---
	(0.001)*	(0.001)*		
1B	-0.186	-0.207	0.018	0.008
	(0.061)	(0.049)	(0.037)	(0.035)
2B	-0.190	-0.176	-0.075	-0.058
	(0.066)	(0.044)	(0.060)	(0.041)
3B	-0.249	-0.123	-0.089	0.083
	(0.063)	(0.050)	(0.054)**	(0.040)*
SS	-0.101	-0.207	0.037	-0.093
	(0.053)	(0.058)	(0.042)	(0.054)**
C	-0.119	-0.106	0.028	0.080
	(0.063)	(0.055)	(0.055)	(0.050)
OF	-0.174	-0.211	---	---
	(0.052)	(0.045)		
DH	-0.138	---	0.032	---
	(0.070)		(0.061)	
P	---	---	0.028	0.013
			(0.038)	(0.035)
CENT	-0.013	0.054	-0.021	0.082
	(0.036)	(0.027)*	(0.037)	(0.028)
WEST	0.033	0.083	0.051	0.110
	(0.030)	(0.026)*	(0.031)	(0.028)
US	0.023	-0.009	0.033	-0.055
	(0.035)	(0.031)	(0.036)	(0.032)
BLACK	0.079	0.059	0.100	0.025
	(0.033)*	(0.029)*	(0.034)	(0.031)
LATINO	0.084	-0.046	0.114	-0.072
	(0.044)**	(0.042)	(0.045)	(0.045)
OTHER	0.041	---	-0.042	---
	(0.091)		(0.061)	

AGE		0.036	-0.045	0.045	-0.036
		(0.044)	(0.027)**	(0.045)	(0.029)
AGE2		-0.0006	0.001	-0.0007	0.001
		(0.001)	(0.000)**	(0.001)	(0.000)
MKT		-0.038	-0.020	-0.044	-0.006
		(0.030)	(0.021)	(0.031)	(0.023)
MVP		0.021	-0.062	0.011	-0.050
		(0.042)	(0.032)**	(0.043)	(0.034)
	R^2	.508	.621	.471	.561
	Adj R^2	.461	.592	.422	.529
	F-value	10.677	21.210	9.702	17.536

adjusted standard errors reported in parentheses below each coefficient.
* (**) Indicates significance at the 5% (10%) level in a two-tailed test.

The empirical results show that when controlling for various types of potential bias factors, the media for the large part did vote for the most valuable players. As seen in Model I and Model II, Win Shares, did in fact have a positive coefficient and was statistically significant at a 5% level. The positive effect on an unbiased measure of player performance like Win Shares (WS) proves the media is voting for its league's best players. Nevertheless, home runs (HR), the other player quality statistic, were also positively correlated and statistically significant at a 5% level. While the coefficient on Win Shares was greater than the coefficient on home runs, this positive effect still reiterates the fact that one receives increased recognition from hitting the ball far.

Confirming James (2002) and Verducci (2003), I found noteworthy evidence of team specific bias towards those ball clubs with high winning percentages. In both Model I and Model II, the estimated coefficient on team winning percentage (TWPCT) is positively correlated to voting share and statistically significant at a 5% level. While James discovers a player on a pennant-winning team is more likely to win the MVP Award than the same player playing on a non-championship team, I was able to further extend upon this. The evidence shows that any player who is on a winning team is more

likely to receive more MVP votes; it does not just apply to those who are on pennant-winning teams. The other team specific variable thought to have a positive bias was a team's market size. Market size is one of those topics continually discussed on sports TV shows and talk radio describing the disadvantages players have playing for teams in small markets. While the general consensus is that players on these teams do not get recognized for their talents and are hurt in MVP voting, the evidence suggests this is not true. Market size (MKT) was negatively correlated with MVP voting share. Perhaps this new age of readily available access to player statistics on the internet has solved this problem.

Reputation effects, although only National League variables are statistically significant at a 10% level, did produce opposite results in the two models. Model I, although not statistically significant at a 10% level, essentially agrees with the research of Zillante (2005) confirming a positive reputation bias. In the American League, previously receiving the MVP award (MVP) along with player age[1] (AGE) does positively affect MVP voting share. The only discrepancy is from the negative coefficient for MVP votes received in the previous year (PSHARE). Model II on the other hand, seems to follow the analysis of James (2002) that previously established players are not as likely to win the MVP as players who have their first outstanding season. The National League regression shows opposite results where the effect of PSHARE is positive versus MVP and AGE being negative.

Racial and Nationalistic bias also had varying results. In the American League, nonwhites do not receive less MVP votes. Actually a case could be made that Whites are being discriminated over the past decade in MVP voting share. However, the National

League does show some negative effects on Latinos and non US-born players.[1] As a result, these findings cannot fully put to rest the question whether media-based discrimination extends to perceived race and actual nationality. Nevertheless, this data in conjunction with Jewell, Brown, and Miles (2002) contrast Desser, Monks, and Robinson (1999) who proposed a strong media voting bias against both African Americans and Latin Americans.

As far as regional bias the only statistically significant indication exists in the National League. Model II suggests that there is a positive MVP voting bias towards teams in both the Central and West divisions, with both WEST and CENT having a positive coefficient and being statistically significant at a 5% level. This finding is consistent with James (2002) when he disproves a bias in MVP voting for players who play on New York teams. Even though some may argue there still might be an east-coast bias on the east coast, it certainly does not play into the MVP voting results.

The Cy Young models are estimated using OLS on a sample of 131 observations; the results are reported in Table 5. Model V controls for the American League and National League, Model VI controls for the American League, and Model VII controls for the National League.

Table 5. Cy Young Robust Regression Results
(Dependent Variable: Ratio of Cy Young votes received divided by the max points)

Variable	Model V	Model VI	Model VII
(Constant)	2.238	1.853	1.849
	(1.376)	(0.898)	(1.303)
PSHARE	-0.079	-0.074	-0.135
	(0.080)	(0.080)	(0.100)
WS	0.044	0.045	0.046
	(0.008)*	(0.008)*	(0.009)*
WINS	0.027	0.026	0.022
	(0.008)*	(0.008)*	(0.010)*
TW_PCT	0.013	0.063	0.728
	(0.319)	(0.317)	(0.473)
ERA	-0.044	-0.023	-0.091
	(0.045)	(0.044)	(0.059)
SP	-0.341	-0.335	-0.250
	(0.100)*	(0.097)*	(0.118)*
CENT	-0.065	-0.082	-0.026
	(0.058)	(0.058)	(0.082)
WEST	-0.014	-0.023	0.054
	(0.051)	(0.050)	(0.081)
US	-0.092	-0.098	-0.068
	(0.107)	(0.090)	(0.129)
BLACK	-0.060	-0.017	0.110
	(0.118)	(0.102)	(0.167)
LATINO	-0.081	-0.060	-0.045
	(0.111)	(0.099)	(0.144)
OTHER	0.123	---	0.070
	(0.202)		(0.180)
AGE	-0.183	-0.161	-0.194
	(0.091)	(0.056)	(0.080)
AGE2	0.003	0.003	0.003
	(0.001)	(0.001)	(0.001)
YP	0.016	---	---
	(0.026)		
YP2	-0.001	---	---
	(0.001)		
MKT	-0.096	-0.080	-0.061
	(0.046)	(0.044)	(0.057)
CY	0.102	0.088	0.015
	(0.061)	(0.059)	(0.089)
AL	0.095	---	---
	(0.042)		
R^2	.682	.667	.774
Adj R^2	.628	.623	.699
F-value	12.548	15.347	10.298

adjusted standard errors reported in parentheses below each coefficient.
* (**) Indicates significance at the 5% (10%) level in a two-tailed test.

The empirical results show that when controlling for various types of potential bias factors in the Cy Young award, the media did vote for the worthy Cy Young candidates. As seen in all three models, Win Shares had a large positive coefficient and was statistically significant at a 5% level. Again, the positive effect on an unbiased measure of player performance like Win Shares (WS) proves the media is voting for its league's best pitchers. The other player specific statistics WINS and ERA did produce some notable results as well. WINS, much like home runs in the MVP award, were also positively correlated and statistically significant at a 5% level. This proves that the media focus on 20-game winners is taken into account come time to vote. I was a little surprised though by ERA. Although not statistically significant and not originally thought to be as significant as winning games, the negative coefficient connected with ERA does not seem like it fits.

Like in the MVP analysis team winning percentage does positively affect Cy Young voting share. However, none of the three models produce a statistically significant outcome of 10%. Parallel to the MVP analysis again, the population in proximity to a team does not affect Cy Young voting either.

Unlike in the MVP award, there is a positive bias in voting to being a previous winner of the Cy Young award. However, young players seem to be discriminated against in comparison to older pitchers. Additionally, the amount of votes one receives in the previous year for Cy Young voting does positively effect his voting share the next year.

Considering racial bias, Models V and VI did show a negative correlation to nonwhite pitchers. Model VI only showed negative bias towards those pitchers who are Latino. Although none were statistically significant, this does raise the point that many of the Cy Young award winners in this past decade have been white. Upon closer review however, the pitcher who does end up winning the reward is more often than not the player with the highest Win Share total.

The most surprising aspect of the Cy Young regression was the effect of status as a starting or relief pitcher. Common notion suggests that a starting pitcher would have a large bias towards them, however in all three models SP had a negative coefficient and was statistically significant at a 5% level.

Models I and II show regional bias does seem to be in favor of pitchers in the East. In reality however, some of the great pitching staffs in the last decade (as calculated in Win Shares) have been on teams like the New York Yankees, the Boston Red Sox, and Atlanta Braves, which all reside in the East division.

V. CONCLUSIONS:

I began this thesis by attempting to create my own model like Zech and Woolway to determine a measure of player rating to use in an MVP award analysis. However, I soon realized that the data provided from Win Shares could be of much greater use. As long as we can trust Win Shares, and not another statistic like home runs as a player measure, everything else has an indication of true bias. That being said I truly believe in Win Shares as the best measure of player quality.

This paper adds to the empirical literature investigating media-based discrimination in professional sports. Using evidence of recent research by Jewell, Brown, and Miles (2002) on the effects of race on player induction into the National Baseball Hall of Fame and James (2002) on MVP voting bias, I have extended analysis of media discrimination to Major League Baseball's most prestigious end of the season awards. Keeping the similar the concept of voting shares as the dependant variable, I investigate many potential bias factors in the MVP and Cy Young award, awards solely voted on by the BBWAA. After controlling for various player and team attributes, I found evidence of media-based discrimination in both the MVP and Cy Young voting during the past ten year period after the strike in 1995. In both awards the media does vote in favor of the leagues best pitchers; however other player quality statistics do factor in. Surprisingly, findings did differ in some cases relevant to each award. Perhaps the most notable finding was the discrimination on nonwhite pitchers in Cy Young voting. In the MVP voting on the other hand, no true evidence was found that nonwhites receive a lesser share of MVP votes. Instead, Whites were being discriminated against over the past decade in MVP voting share.

Furthermore, perhaps other player and team specific variables should have been included. Conceivably, a variable should have been included in the MVP analysis for a player's RBI total. As James (2002) noted, the history of voting for MVP usually goes to the RBI leader if there is any way that it can. Additionally, it might make sense to look at players on surprisingly good teams. This might help propel their value in the media's mind. Nevertheless, the true weakness in my approach may lie in the actual design of the model. While this regression model allows one to analyze a broad array of

discriminatory factors, basing the dependent variable on the share of votes received only takes into account players who actually received votes for these awards in the first place. In other words, perhaps the proper method should encompass all Major League players. By doing so one could additionally study discrimination using Win Shares to determine the players worthy of receiving votes for these awards but do not – James (2002) thus maybe correct in his approach.

For a complete analysis of media bias in MLB Awards, further research should also be conducted using Rookie of the Year voting results. This is the last MLB award voted on by the BBWAA and it could provide other vital information of bias among the young stars in baseball.

"In God we trust, all others must show data" - **Mets pitching coach Rick Peterson**

VI. APPENDIX

Appendix A. Metropolitan Statistical Areas/Combined Statistical Areas (2000)

City	Population
New York-Newark-Bridgeport, NY-NJ-CT-PA (CSA)	**21,404,951**
Los Angeles-Long Beach-Riverside, CA (CSA)	**16,438,558**
Chicago-Naperville-Michigan City, IL-IN-WI (CSA)	**9,333,801**
Washington-Baltimore-Northern Virginia, DC-MD-VA-WV (CSA)	**7,603,314**
San Jose-San Francisco-Oakland, CA (CSA)	**7,114,984**
Philadelphia-Camden-Vineland, PA-NJ-DE-MD (CSA)	**5,839,997**
Boston-Worcester-Manchester, MA-NH (CSA)	**5,729,912**
Dallas-Fort Worth, TX (CSA)	**5,413,219**
Detroit-Warren-Flint, MI (CSA)	**5,366,154**
Miami-Fort Lauderdale-Miami Beach, FL (MSA)	**5,028,727**
Toronto	**4,747,200**
Houston-Baytown-Huntsville, TX (CSA)	**4,840,989**
Atlanta-Sandy Springs-Gainesville, GA-AL (CSA)	**4,583,927**
Seattle-Tacoma-Olympia, WA (CSA)	**3,614,199**
Montreal	**3,471,300**
Phoenix-Mesa-Scottsdale, AZ (MSA)	**3,277,946**
Minneapolis-St. Paul-St. Cloud, MN-WI (CSA)	**3,285,046**
Cleveland-Akron-Elyria, OH (CSA)	**2,946,894**
San Diego-Carlsbad-San Marcos, CA (MSA)	**2,824,591**
St. Louis-St. Charles-Farmington, MO-IL (CSA)	**2,757,596**
Denver-Aurora-Boulder, CO (CSA)	**2,464,811**
Tampa-St. Petersburg-Clearwater, FL (MSA)	**2,404,323**
Pittsburgh-New Castle, PA (CSA)	**2,523,884**
Cincinnati-Middletown-Wilmington, OH-KY-IN (CSA)	**2,055,174**
Kansas City-Overland Park-Kansas City, MO-KS (CSA)	**1,908,039**
Milwaukee-Racine-Waukesha, WI (CSA)	**1,691,294**

VII. REFERENCES:

Depken, Craig A. and Jon M. Ford (2003). "Customer-Based Discrimination against Major League Baseball Players: Additional Evidence from All-star Ballots". Department of Economics, University of Texas at Arlington. (December 2003).

Desser, Arna, James Monks, and Michael Robinson (1999). "Baseball Hall of Fame Voting: A Test of the Customer Discrimination Hypothesis." *Social Science Quarterly*. V80, n3 (September 1999): 591-603

Heyman, Jon (1999). "Baseball writers also rate an award for silliness". *Sporting News* (November 29, 1999): Vol. 223, Issue 48, p71

James, Bill and Jim Henzler (2002). "Biases in MVP Voting". *Win Shares*. Morton Grove, Ill. STATS Publishing. p194

James, Bill and Jim Henzler (2002). "Win Shares and MVP Voting". *Win Shares*. Morton Grove, Ill. STATS Publishing. p237

Jewell, R. Todd, Robert W. Brown, and Scott E. Miles (2002). "Measuring Discrimination in Major League Baseball: Evidence from the Baseball Hall of Fame." *Applied Economics* v34, n2 (January 2002): 167-177

Leeds, Michael, and Peter von Allmen (2005). *The Economics of Sports*. Pearson Education, Inc. 2nd edition

Nardinelli, Clark and Curtis Simon (1990). "Customer Racial Discrimination in the Market for Memorabilia: The Case of Baseball*". *Quarterly Journal of Economics*. (August 1990): Vol. CV, Issue 3

Streit, Al (2005). "Baseball Markets". *Baseball Almanac*. http://www.baseball-almanac.com

Verducci, Tom (2004). "Baseball Preview 2004: Welcome to the New Age of Information: Baseball's conventional wisdom is taking a beating from a new stat-crazy culture that turns the numbers inside out". *Sports Illustrated* (April 5, 2004): 50

Verducci, Tom (2003). "The Way it Looks From Here". *Sports Illustrated* (September 29, 2003): Vol. 99, Issue 12

Woolway, Mark David. "Empirically Estimating the Production Function for Major League Baseball and Examining Worker Disincentives in Response to Multi-Year Contracts." *American Economist* v41, n2 (Fall 1997): p77-83.

Zech, Charles E. (1981). "An Empirical Estimation of a Production Function: The Case of Major League Baseball." *American Economist* v25, n2 (Fall 1981): p19-23.

Zillante, Arthur (2005). "Reputation Effects in Gold Glove Award Voting". ICES and George Mason University. (February 6, 2005).

VIII. STATISTICS:

James, Bill and Jim Henzler. "Win Shares by Team". *Win Shares*. Morton Grove, Ill. STATS Publishing. 2002, p194

Internet http://sports.espn.go.com/mlb/index
Internet http://sportsillustrated.cnn.com/baseball
Internet http://www.baseball-reference.com
Internet http://www.mlb.com

Internet http://www.bea.gov (Bureau of Economic Analysis)
Internet http://www.statcan.ca/start.html (Canada's National Statistical Agency)