7-2011

Comparing Hall of Fame Baseball Players Using Most Valuable Player Ranks

Paul Kvam
University of Richmond, pkvam@richmond.edu

Recommended Citation

Kvam, Paul H. "Comparing Hall of Fame Baseball Players Using Most Valuable Player Ranks." *Journal of Quantitative Analysis in Sports* 7, no. 3 (July 2011): Article 19, 1-20. doi:10.2202/1559-0410.1337.

Journal of Quantitative Analysis in Sports

Volume 7, *Issue* 3 2011 *Article* 19

Comparing Hall of Fame Baseball Players Using Most Valuable Player Ranks

Paul H. Kvam, *Georgia Institute of Technololgy*

Recommended Citation:
Kvam, Paul H. (2011) "Comparing Hall of Fame Baseball Players Using Most Valuable Player Ranks," *Journal of Quantitative Analysis in Sports*: Vol. 7: Iss. 3, Article 19.

DOI: 10.2202/1559-0410.1337

Comparing Hall of Fame Baseball Players Using Most Valuable Player Ranks

Paul H. Kvam

Abstract

We propose a rank-based statistical procedure for comparing performances of top major league baseball players who performed in different eras. The model is based on using the player ranks from voting results for the most valuable player awards in the American and National Leagues. The current voting procedure has remained the same since 1932, so the analysis regards only data for players whose career blossomed after that time. Because the analysis is based on quantiles, its basis is nonparametric and relies on a simple link function. Results are stratified by fielding position, and we compare 73 Hall of Fame players up to 2010. We also consider the players on the 2011 Hall of Fame ballot as well as other potential Hall of Fame candidates. The analysis is based on the method of maximum likelihood, and results are illustrated graphically.

KEYWORDS: maximum likelihood estimators, quantiles, censoring, order statistics

Author Notes: This work depended on the aid of Georgia Tech student Heeseung Moon, who compiled all of the vote data for this analysis.

1 Introduction

In this paper, we propose a rank-based statistical estimator for comparing the performances of top major league baseball (MLB) players who played in different eras, from 1931 to present. Rather than addressing popular batting and pitching statistics and trying to predict how those statistics could be compared across different eras (see Berry, et al., 1999), we instead focus on a simple overall rank measure: The Baseball Writers Association of America (BWAA) Most Valuable Player (MVP) votes. This analysis will be limited to position players, and not pitchers. Although some pitchers are considered for the MVP award, many baseball writers consider the MVP a position-player award, leaving the Cy Young award as its pitching equivalent.

There are valid reasons to use MVP votes for dictating how we rank players from different eras in baseball. Mainly, baseball statistics have evolved dramatically throughout the years, and critics who compare players using such statistics risk confounding the effects of player value with the effects of his baseball era. From 1974 through 1976, for example, Mike Schmidt led the league in home runs, though never hitting more than 38. In 1998, 17 players hit 38 or more home runs. Although the voting procedure has its drawbacks (discussed later), it serves as a consistent metric for how the public perceives player performance and value across a wide range of years.

Starting in 1931, the Baseball Writers Association of America began awarding the MVP award to players in both the American League (AL) and National League (NL) using a weighted scoring system. This is more or less the same form of voting that is used today, and replaced previous awards (Chalmers Award from 1911-1915, League Award from 1922-1929) which used an inconveniently changing criteria for awarding the MVP. For example, the League Award listed only one player per team on the ballot and American League players could only win the award one time. For this reason, our analysis of MVP voting is used only for players whose career blossomed after 1930, and as a consequence, leaves out arguably the greatest player of all time, Babe Ruth, who played his best years between 1915 and 1933.

Table 1 shows the MVP voting for the 2010 season. There are 32 votes in the NL, which is two for every franchise, and 28 for the AL. Josh Hamilton won the AL MVP award with 358 points based on 22 (out of 28) first place votes, four second place votes and two fourth place votes. In the NL, Joey Votto won with 443 points based on 31 (out of 32) first place votes and one second place vote. A first place vote is worth 14 points, a second place vote is worth 9 points, and values decrease point by point after that.

The thesis of this research is based on a simple assumption: no matter how many different statistics can be used to assess the value of a player on a year-to-year basis, their ultimate value is best interpreted through the votes of experts, and the BWAA MVP award represents the best available ranking procedure of that ilk. The point total, based on a weighted ranking system, could be modeled using nonparametric rank statistics, but we focus only on the actual rank, treating the outcome as an observed percentile of a value statistic whose distribution is arbitrary outside its ability to correctly order players according to this unobservable measure of value. Because of the large variability of productivity between fielding positions, we also focus on comparisons by position.

The data are comprised of 73 players who were elected into the Hall of Fame and includes players selected by the Veterans Committee. All of the data was made available at *Baseball-reference.com*, which includes a comprehensive list of inducted players and information about how each player ranked in MVP voting in any given year. We do not include members who played in the 1920s and earlier.

The writers who vote rely on several sources of information, and mainly lean on batting statistics such as runs batted in, batting percentage, on-base average and slugging percentage. Other factors are considered, as well, including fielding statistics, apparent leadership abilities, and perceived ability to perform well in clutch plays. However, rankings can be biased with the affects of media exposure, home team favoritism, player reputation and writer ignorance. Because of these biases, some past MVP awards have been controversial, although most critics seem to be in agreement about award winners in most years. Discrepancies due to bias become more obvious in lower rankings such as candidates who score 50 or fewer points. This is partly due to the attention the writers heap on their first or second choice, and it might also be due to the difficulty in distinguishing two similarly qualified players who have different strengths and weaknesses. For example, Ryan Howard, the power-hitting first-baseman for the Philadelphia Phillies received only eight votes from the 32 writers, but got an unexpected second and third place vote from Philadelphia writers. This placed him one point shy of Martin Prado, the Atlanta Braves second-baseman, who received 17 votes, but none higher than fifth. Prado and Howard are dramatically different in terms of skills they offer their team, and its hard to compare their 2010 player values using head-to-head statistics.

DOI: 10.2202/1559-0410.1337

American League		National League	
Josh Hamilton, Texas	358	Joey Votto, Cincinnati	443
Miguel Cabrera, Detroit	262	Albert Pujols, St. Louis	279
Robinson Cano, New York	229	Carlos Gonzalez, Colorado	240
Jose Bautista, Toronto	165	Adrian Gonzalez, San Diego	197
Paul Konerko, Chicago	130	Troy Tulowitzki, Colorado	132
Evan Longoria, Tampa Bay	100	Roy Halladay, Philadelphia	130
Carl Crawford, Tampa Bay	98	Aubrey Huff, San Francisco	70
Joe Mauer, Minnesota	97	Jayson Werth, Philadelphia	52
Adrian Beltre, Boston	83	Martin Prado, Atlanta	51
Delmon Young, Minnesota	44	Ryan Howard, Philadelphia	50
Vladimir Guerrero, Texas	22	Buster Posey, San Francisco	40
Rafael Soriano, Tampa Bay	21	Matt Holliday, St. Louis	32
CC Sabathia, New York	13	Brian Wilson, San Francisco	28
Shin-Soo Choo, Cleveland	9	Scott Rolen, Cincinnati	26
Alex Rodriguez, New York	8	Ryan Braun, Milwaukee	19
Felix Hernandez, Seattle	6	Ryan Zimmerman, Washington	18
Ichiro Suzuki, Seattle	3	Carlos Ruiz, Philadelphia	12
Jim Thome, Minnesota	2	Dan Uggla, Florida	12
Joakim Soria, Kansas City	1	Adam Wainwright, St. Louis	12
Mark Teixeira, New York	1	Jason Heyward, Atlanta	11
		Brian McCann, Atlanta	9
		Adam Dunn, Washington	9
		Ubaldo Jimenez, Colorado	7
		David Wright, New York	3
		Corey Hart, Milwaukee	2
		Josh Johnson, Florida	2
		Heath Bell, San Diego	2

Table 1: 2010 Most Valuable Player Ranks and Points.

2 The Model

Our goal is to compare top position players from different eras using their MVP rankings. We treat the rank as the quantile outcome out of N players that compete for rankings on a yearly basis. Although the number of teams and players have increased through the years, we will treat N as a constant, based on how we perceive the population of candidates. For example, if we consider only players who achieve 500 or more plate appearances (which is the minimum number required to qualify for a batting title), we are limited to about 150 players from the combined leagues. However, if we consider all the players throughout the system who are competing for a major league position, N will be in the thousands. We will focus on this problem more in the next section.

Most position players in the MLB Hall of Fame (after 1931) have received MVP votes in ten or more years of their career, and achieved top-5 rankings in four or more of those years. Hank Aaron, as the extreme, was voted MVP only once, but landed in the top 5 an amazing 13 times, and received votes in 17 years of his long, historic career. At the other end of the spectrum, second-baseman Bill Mazeroski, a controversial Hall of Fame selection, received MVP votes in only two years, placing 8th and 23rd.

In our model, we consider a maximum $m = 20$ year career for any given player, knowing that few players could actually perform at the major league level for even half that time. Suppose a player receives votes in k different years of his career. For any of the k years in which the player receives MVP votes, we treat his MVP rank as a quantile for some unobserved quality statistics that could be used to judge the merit of the player's overall performance. For any of the $m-k$ years in which the player does not receive votes, we treat the quality statistic as censored - specifically, it would be *type-II left censoring*, since only the upper percentiles of the quality statistic are observed.

To illustrate, we will use a classic example of two baseball legends and their legion of fans who have debated the same question for fifty years: Who is better, Willie or Mick? Willie Mays, who played center field for the New York Giants came up to the major leagues in 1951, winning the NL rookie-of-the-year award before entering military service during the Korean War. He returned for his first full season in 1954, winning the NL MVP. His cross-town rival Mickey Mantle also broke in with the New York Yankees in 1951 and was constantly compared to Mays (even after the Giants moved to San Francisco in 1958) because both outfielders were considered to be the best players in their respective league. Mays won two NL MVP awards and received votes in 15 years between 1954 and 1971. Mantle won the AL MVP three times and received votes in 14 years of his spectacular but injury-plagued career between 1951 and 1968.

For a given player, let $v_1, \cdots, v_k$ represent the set of MVP ranks achieved during his career. These ranks correspond to the top k unobserved quality measurements

$$X_{m-k+1:m}, \cdots, X_{m:m}$$

from a player's $m = 20$ year career. The order statistic $X_{i:m}$ designates the i^{th} smallest observation from the sample $(X_1, \cdots, X_m)$. The remaining order statistics $(X_{1:m}, \cdots, X_{m-k:m})$ correspond to years in which the player did not receive MVP votes from the BWAA, possibly because the player was not even in the major leagues. The observed quantiles of $X_{m-k+1:m}, \cdots, X_{m:m}$ are a simple function of the ordered MVP ranks:

$$r_{i:m} = \frac{N - v_{i:m} + 1}{N - 1}, \quad i = m - k + 1, \cdots, m. \tag{1}$$

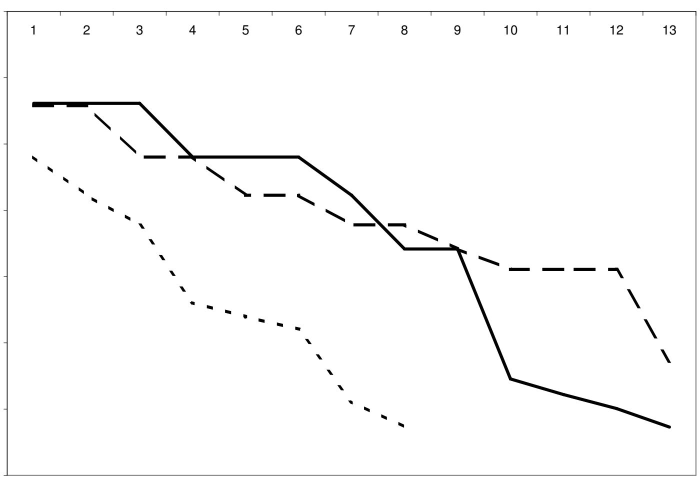

Figure 1: Ordered values of -log(v+1), where v is player MVP rank. Three players are plotted across 15 years: Willie Mays (dashed line), Mickey Mantle (solid line) and Duke Snyder (dotted line).

Figure 1 graphically shows how Mantle, Mays and the other cross-town Hall of Fame center fielder (for the Brooklyn Dodgers) Duke Snyder compare in terms of MVP ranks. To make the differences clear, the figure plots the ordered values of $-\ln(v+1)$, so each player's highest rank appear at the left side of the graph and the logarithm causes the top ranks to appear more separated compared to lower ranks. The graph suggests that Mickey Mantle had a brilliant run over ten years, but Willie Mays proved to be better over a longer period of time. Duke Snyder, an already-established center field all star when Mantle and Mays broke into the big leagues, had several productive years with the Dodgers, but his career began to fade in 1956, just when the careers of Mays and Mantle started booming.

Without further refining the model, the player statistics we propose to use $(r_{m-k+1:m}, \cdots, r_{m:m})$ are insufficient for effective comparisons between players (even if N is specified). To make such comparisons possible, we assign each Hall of Fame player a parameter θ to represent a quality characteristic. Suppose the distribution of the quality statistic for a baseline player is designated $F(x)$. For a player with parameter θ, the distribution $G(x)$ of the quality statistic is related to the baseline player using the link function

$$G(x) = F(x)^{\theta}. \tag{2}$$

This is a common link used in semi-parametric analysis and accelerated life testing, although it would be more common to link survival functions $(1 - G(x)) = (1 - F(x))^{\gamma}$, which makes it easier to relate hazard functions. In this case, the link based on the cumulative distribution function is more convenient in forming maximum likelihood estimators in the following section.

Because we are only observing quantile values, without loss of generality, we can assign $F(x) = x$, with $0 \le x \le 1$, so that $G(x)$ is the cumulative distribution function for a Beta(θ,1) distribution. Larger values of parameter θ reflect higher quality characteristics for the player. This is a semi-parametric analysis, because the distribution of the quality measurement is arbitrary, but the link function is specified with an unknown parameter θ.

3 Estimating the Quality Characteristic

For a given player we observe his highest k quantiles $(r_{m-k+1:m}, \cdots, r_{m:m})$ out of m years, and $m-k$ of the smallest quantiles are left censored. In most years, there are no more than 30 players in either league that receive MVP votes, and sometimes fewer than 20. With the type-II left censored values,

DOI: 10.2202/1559-0410.1337

we essentially have upper bounds for the missing quantiles, with $r_{i:m} \le (N-30+1)/(N+1)$, where $i = m-k+1, \cdots, m$.

To estimate θ, we treat the k observed quantiles as observations from the Beta(θ,1) distribution, and $m-k$ unobserved quantiles as right-censored (corresponding to the left-censored quality statistic) at the upper quantile $q = (N-29)/(N+1)$. Finally, our choice of N will affect model fit, although it will not affect the orderings of the different values of θ between different Hall of Fame players. For example, we could choose N so that q would be at approximately 0.95, meaning that the top 5% of major league players receive MVP votes in any given year. This is satisfied with $N = 600$. If q increases to 0.99, we need $N = 3000$. By choosing $m = 20$, we are considering a large population of ballplayers at various stages of their potential career, including time in the minor leagues, or even early retirement. As a consequence, this choice requires that q will be over 0.95, so N will be larger than 600.

Out of the 73 Hall of Fame position players, the average number of years a player receives MVP votes is 9.6, or roughly half of their potential 20-year career. Our goal is to choose q (and N) so that the estimate for θ corresponds to a distribution G_θ with median $\hat{x}_{0.50}$ which is as close to q as possible. That is, when averaged across θ for the 73 players, a player has a 50% chance of receiving MVP votes in any given year. This happens to match up approximately at $q = 0.99$ ($N = 3000$) where the average of these 73 medians is 0.987.

For purposes of this study, then, we will treat $q = 0.99$ and $N = 3000$. Implicit with this model assumption is that, on average, 50% of the player rankings will be greater than 0.99, or in other words, for a randomly chosen Hall of Fame player, they are expected perform in the top 1% during their best 10 years in baseball.

To estimate θ, we apply the method of maximum likelihood. The likelihood function, based on a $m = 20$ year career of a player with quality characteristic θ can be expressed as

$$L(\theta) = \prod_{j=1}^{m-k} q^\theta \prod_{i=1}^{k} \theta r_{m-i+1:m}^{\theta-1}, \tag{3}$$

where $r_{i:m} = (N - v_{i:m} + 1)/(N-1)$. In (3), $L(\theta)$ is based on k observations from a Beta(θ,1) distribution along with $m-k$ left censored observations at q. The maximum likelihood estimator (MLE) for θ has a convenient closed form:

$$\hat{\theta} = \frac{-k}{(m-k)\log(q) + \sum_{i=1}^{k} \log(r_{m-i+1:m})}. \tag{4}$$

Based on the Fisher information, we can estimate the standard deviation of $\hat{\theta}$ as

$$\hat{\sigma} \approx \left(-E\left(\frac{\partial^2}{\partial\theta^2} \log L(\theta)|_{\hat{\theta}} \right) \right)^{-1/2} = \frac{k}{\hat{\theta}}. \tag{5}$$

With a gage of uncertainty, we have a compelling way to rank Hall of Fame baseball players according to their estimated quality characteristic. As an example, if we use $N = 3000$, the Mays versus Mantle debate is summarized from (4) and (5) as

	$\hat{\theta}$	$\hat{\sigma}$
Willie Mays	185.2	12.3
Mickey Mantle	139.4	10.0

According to our model, Willie Mays is the superior player, mostly because Mantle shows a quicker decline in performance after their careers peaked. Both players ranked in the top six in MVP voting in their best nine years, but Mays continued in the top six for another three years. It is well known, too, that Mantle suffered from debilitating injuries that hampered him throughout his career, while Mays was relatively injury free. The only debate left in this case is how good Mantle would have been if he had not tore the cartilage in his right knee while fielding a ball in the 1951 World Series game against the New York Giants. Ironically, the ball was hit by Willie Mays.

4 Goodness of Fit

Before giving further credence to the quality estimator, we will investigate the goodness of fit for the proposed model in (2), based on the Beta distribution. We test this model in two steps. The first step is to see if the observed value of k is consistent with what the model predicts. If the discrepancies are small enough, overall, we will assume the k observed quantiles are from the truncated Beta(θ,1) distribution, with the $m-k$ missing observations being less than $q = 0.99$ and not included in the truncated distribution. By doing the test in two steps, we make the goodness-of-fit test simple to compute and easier to understand.

For all the players in the data, the model does well in matching the actual number of times the player received MVP votes (k) and the expected number according to the model (E_k). Out of 73 players, only four players have a difference $|E_k - k|$ greater than one: Joe DiMaggio (k=12, E_k=13.7), Willie Mays (k=15, E_k=16.9), Mike Schmidt (k=12, E_k=13.5), and Eddie Murray (k=9, E_k=10.2). In each of these cases, the player was expected to receive votes in more years than they actually did, due to the high rankings they received during their prime years. In the next step, by treating the observed k as fixed for each player, we examine the model fit for the truncated Beta(θ,1) distribution.

For a given player with quality parameter θ and $m = 20$ years of service, we have m independent quantile ranks $\boldsymbol{r} = (r_1, \cdots, r_m)$, of which only the top k are observed: $(r_{m-k+1:m}, \cdots, r_{m:m})$. Based on the model in (2), the m variables in $\boldsymbol{r}$ are generated from a Beta(θ,1) distribution, and the elements of $\boldsymbol{r}$ in the set $(r_{m-k+1:m}, \cdots, r_{m:m})$ represent the k largest order statistics from $\boldsymbol{r}$.

To test the proposed link model in (2), we make a convenient assumption that the k upper quantiles are actually the observations from the *truncated* Beta(θ,1) distribution, where only quantiles greater than q are observed. To do this, we employ the Cramér-von-Mises test statistic

$$= \int (H_k(t) - H_0(t))^2 \, dH_0(x). \tag{6}$$

In this case, H_k is the empirical distribution function based on the upper k observed player quantiles, and H_0 is the left-truncated Beta(θ,1) distribution:

$$H_0(t) = \frac{t^\theta - q^\theta}{1 - q^\theta}, \quad q \leq x \leq 1. \tag{7}$$

For every Hall of Fame player, we computed the MLE for θ and the Cramér-von-Mises test statistic ψ. We approximated ψ by treating θ as a known parameter, so the computation of the test statistic is more straightforward (see Chapter 6.3.2 in Kvam and Vidakovic (2008), for example). For players who appeared in six or more MVP lists, Csörgo and Faraway (1996) showed the approximation to the Cramér-von-Mises test statistic

$$= \frac{1}{12k} + \sum_{j=m-k+1}^{m} \left(H_0(r_i) - \frac{2j-1}{2k} \right)^2 \qquad (8)$$

is sufficiently precise, for all practical purposes.

Out of the 73 players in the data, eight goodness of fit tests produced a p-value smaller than 0.05, and two players in particular are associated with test statistics with p-values close to 0.01: Eddie Murray and Joe DiMaggio. Thirteen out of 73 tests have a p-value smaller than 0.10, which is not unexpected if the model is correct. This suggests that overall, the model fit is adequate.

5 Comparing Hall of Fame Players

To compare Hall of Fame players, we have stratified players into the fielding positions they were most typically associated during their career. In Figure 2, the quality statistic $\hat{\theta}$ (in terms of a 95% confidence interval) is plotted for nine Hall of Fame catchers from the modern era. The same 95% confidence intervals are plotted for Hall of Fame members who played first base (Figure 3), second base (Figure 4), third base (Figure 5), shortstop (Figure 6) and outfield (Figure 7). This section contains only a brief commentary on the results shown in these figures, and also comments on how rankings are affected by the model in 2.

5.1 Catcher

In Figure 2, the estimate for the quality statistic shows that Yogi Berra, who played for the New York Yankees from 1946 to 1963, is unrivaled among catchers. Berra received MVP votes in 15 years, winning the award three times, and having a top five rank in seven of those 15 years. In the next section, we will consider the merits of great players who are on the ballot for Hall of Fame voting. Because there are so few catchers on this list (at least catchers who stand a reasonable chance of election) we will mention Mike Piazza, who stands out among this group. His score of $\hat{\theta} = 56.3$ clusters him with Bill Dickey and Johnny Bench, among the top catchers of all time. The graph does not include Gabby Harnett, also among the top all-time catchers, who played for the Chicago Cubs between 1922 and 1940.

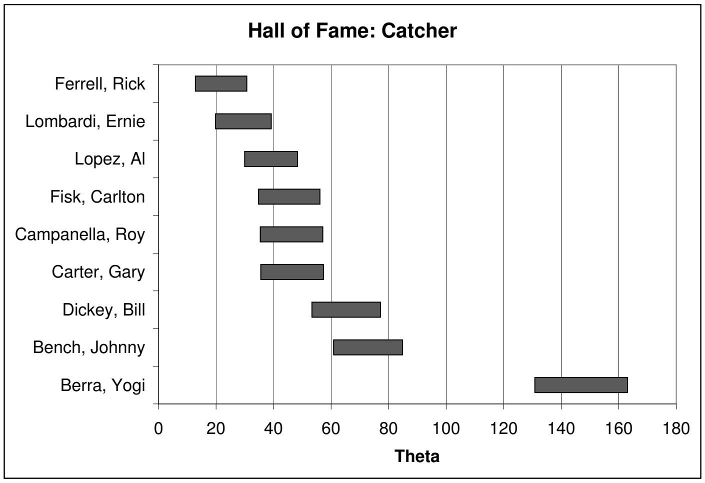

Figure 2: 95% confidence intervals for quality parameter θ of Hall of Fame Catchers

5.2 First Base

Figure 3 shows the interval estimates of quality for seven out of eight Hall of Fame first basemen. It does not include the estimate for Stan Musial ($\hat{\theta} = 297.5$). Because Stan Musial's career statistics overshadow all of the other 1B players, it is difficult to compare the other players when the graph is scaled to include Musial's score. Musial not only won MVP three times, but was runner-up four times and received MVP votes in 18 seasons. Since the Hall of Fame voting took its current form after 1930, we are not able to include two of the greatest 1B to play the game: Jimmie Foxx and Lou Gehrig.

5.3 Second Base

Figure 4 shows interval estimates of quality for eleven Hall of Fame second basemen. While the high score of Charlie Gehringer is noteworthy, the figure shows in a clear way how Bill Mazeroski (previously mentioned as a controversial selection for the Hall of Fame by the Veteran's Committee in 2001) scores below the other second basemen in this group. Mazeroski was never seriously considered by the sportswriters when it came to MVP voting, and he landed in the top 10 only once in his career, ranking 8th in 1958 NL voting. The Veteran's Committee placed more weight on Mazeroski's fielding

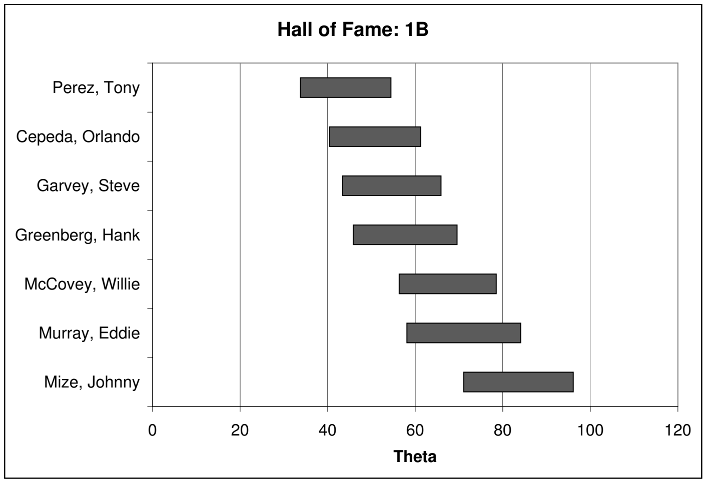

Figure 3: 95% confidence intervals for quality parameter θ of Hall of Fame First Basemen

abilities and his post-season performance, and specifically on his dramatic game-ending home run that won the 1960 World Series for the Pittsburgh Pirates.

Because the data analysis is limited to players from the past 80 years, other great 2B in the Hall of Fame are left out, including Rogers Hornsby, Frankie Frisch, Nap Lajoie and Eddie Collins.

5.4 Third Base

Figure 5 shows interval estimates of quality for eleven Hall of Fame third basemen. Mike Schmidt, who played for the Philadelphia Phillies between 1972 and 1989, was not recognized among statistical leaders because he was a home run leader in an era when not many home runs were hit. He led the NL in home runs eight times, but hit more than forty in only one of those eight years. During the steroid era (1998 - 2003) it was not uncommon for 10 or more major league batters to hit 45 or more home runs in a given year. This aspect of the ranking data shows an advantage over using batting statistics, which can vary from year to year and make player comparisons difficult.

DOI: 10.2202/1559-0410.1337

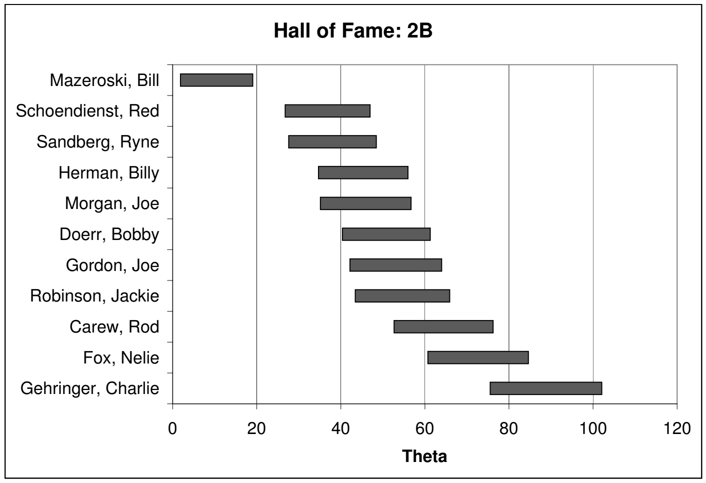

Figure 4: 95% confidence intervals for quality parameter θ of Hall of Fame Second Basemen

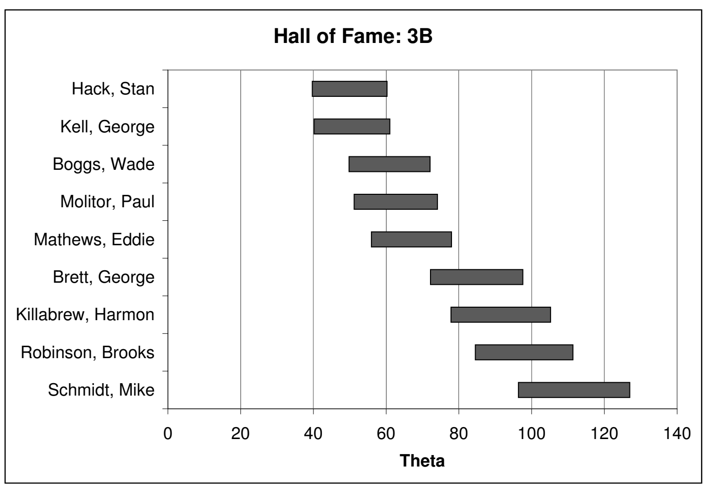

Figure 5: 95% confidence intervals for quality parameter θ of Hall of Fame Third Basemen

5.5 Shortstop

Figure 6 shows interval estimates of quality for eleven Hall of Fame shortstops. In contrast to the 3B graph, the two weakest shortstops on this list are recent inductees: Ozzie Smith and Robin Yount. Cal Ripken, who also played at 3B and holds the consecutive game playing record, actually has a career that goes beyond 20 years with the Baltimore Orioles, and was a 19-time all-star. Despite this longevity, his MVP data is adequately modeled using the $m = 20$ year career upperbound (with goodness-of-fit test statistic p-value larger than 0.10). In terms of future ballots, we will see that Derek Jeter will rank higher than Smith, Yount and Ripken after he is inducted as a Hall of Fame shortstop.

Honus Wagner was the first shortstop inducted into the Hall of Fame, and played between 1897 and 1917, so is not included in this analysis.

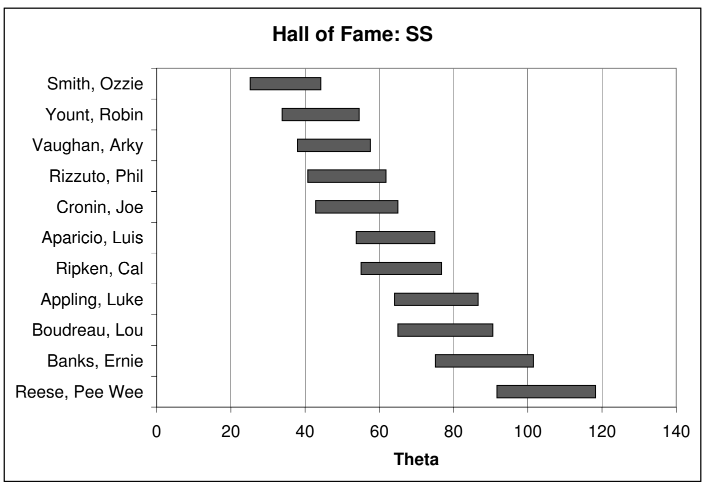

Figure 6: 95% confidence intervals for quality parameter θ of Hall of Fame Shortstops

5.6 Outfield

Figure 7 shows interval estimates of quality 27 Hall of Fame outfielders (not distinguishing between left, center or right field). The graph shows off the great career achievements of Hank Aaron and Ted Williams, two of the greatest players in baseball history. Undoubtedly, the particular order implied by each outfielder's $\hat{\theta}$ statistic will not be agreeable to all experts, and serves to illustrate how our model might perform differently compared to other models used for ranking player achievement. As an example, Tony Gwynn, who played for the San Diego Padres from 1982 to 2001, ranks higher than Rickey Henderson, who many consider to be the greatest lead-off hitter in baseball history. Henderson won the AL MVP in 1990 when he led the Oakland Athletics to the World Series, but Gwynn received MVP votes in more years than Henderson (12 versus 8). This suggests that longevity is more highly valued in this model.

The lowest ranked OF here is Larry Doby, who was the first black player allowed to play in the AL, following Jackie Robinson who played in the NL. They both debuted in 1947, often playing in front of hostile crowds. His major league career was solid if not spectacular, and he received MVP votes in four of his years with the Cleveland Indians. He was inducted into the Hall of Fame by the Veterans Committee in 1998.

Barry Bonds, who will be eligible for Hall of Fame voting in a future year, won the NL MVP seven times. His quality estimate of $\hat{\theta} = 204$ will rank him above all the modern Hall of Fame outfielders except Ted Williams and Hank Aaron. Along with Babe Ruth, there are other exceptional Hall of Fame outfielders who played in an earlier era and are not included in this analysis, including Ty Cobb, Tris Speaker, Al Simmons and Hack Wilson.

6 Discussion

This model based on MVP ranks has several advantages over traditional regression models that employ a profusion of batting data. Batting statistics can be misleading because so many factors that influence them are ancillary to the actual player's performance, especially the year-to-year variability in league offense. There are valid criticisms of MVP balloting, but despite the apparent flaws in the system, this model produces a useful link between voting outcomes from different years, and implies a fair manner in which to make player comparisons across different eras.

By sorting players into fielding positions, we can see that the threshold for greatness is different for different position players. The median value for $\hat{\theta}$ for the positions are 46.24 for Catcher, 50.83 for 2B, 64.37 for SS, 66.96 for 3B, 67.42 for 1B, and 69.13 for OF. The data suggest that catchers and second basemen can be voted into the Hall of Fame without the same frequency of

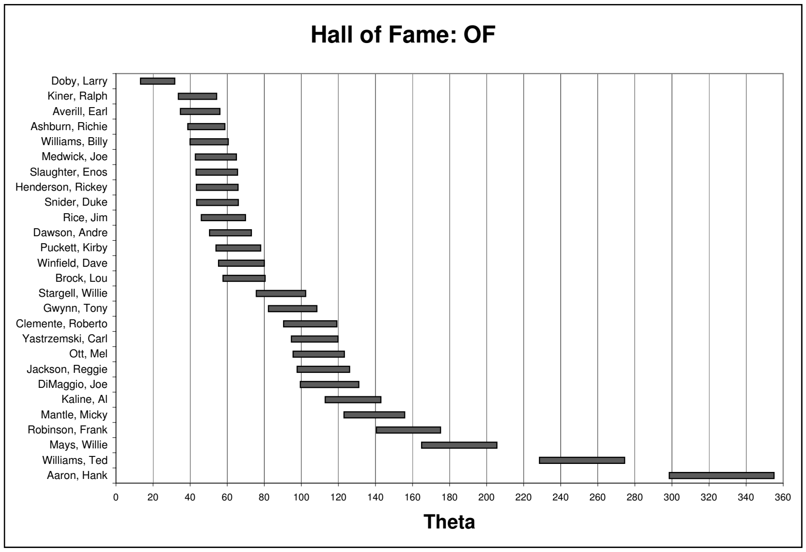

Figure 7: 95% confidence intervals for quality parameter θ of Hall of Fame Outfielders

MVP votes received by other position players. This is important to note in comparing current Hall of Fame players as well as players that will qualify for Hall of Fame election in the future.

Figure 8 shows the quality estimate for the top 15 players on the 2011 Hall of Fame ballot. This is a helpful graph to rank potential inductees, especially 1B and OF because they comprise a majority of the players. As an example, we can gage Herold Baine's MVP votes by noting how far behind he lags from other outfielders such as Larry Walker or Dave Parker. For other positions such as 2B, SS and 3B, it will be helpful to compare their quality estimates to some recent Hall of Fame inductees, which can be used to benchmark quality scores by position. Currently, there are no catchers ranked high on the 2011 ballot, but we discussed Mike Piazza's statistics in the previous section.

Figure 9 plots the quality estimates for six shortstops, including three current Hall of Fame members (Smith, Yount and Ripken) along with two shortstops on the 2011 ballot (Larkin and Trammell) and current player Derek Jeter. Note, Jeter's value can only rise if he receives MVP votes in his remaining active years, but few expect him have future years in which he will be highly ranked. If Ozzie Smith represents a benchmark for gaining

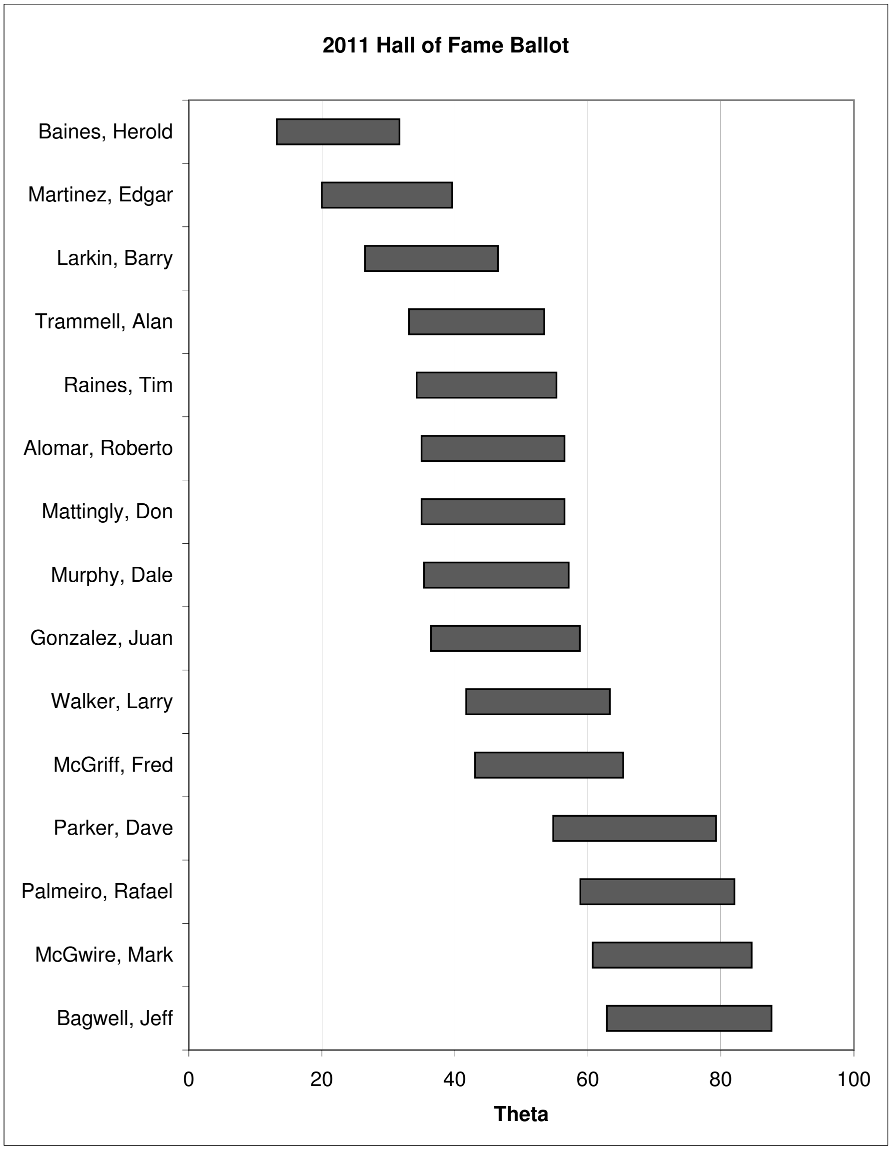

Figure 8: 95% confidence intervals for quality parameter θ of players on 2011 Hall of Fame Ballot

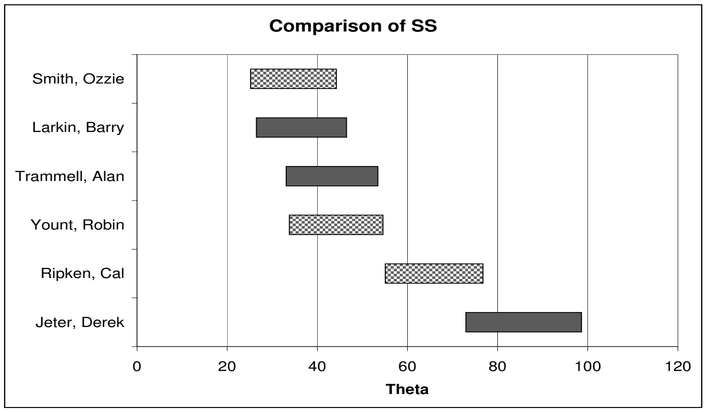

Figure 9: 95% confidence intervals for quality parameter θ of Shortstops. Hall of Fame players are plotted with checkerboard pattern, and other players with solid bar.

entry into the Hall of Fame for shortstops, then Larkin, Trammell and Jeter compare well. This need not be the case, of course. It is not clear that Larkin or Trammell belong in the Hall of Fame group in Figure 6, but Jeter's inclusion is hardly debatable using these metrics.

Figure 10 shows the quality estimates of Hall of Fame 2B Sandburg, Morgan and Carew along with Roberto Alomar (on 2011 ballot), Jeff Kent and Craig Biggio (on future ballots). From Figure 4, we saw that Sandberg represents one of the lowest 2B ratings of any Hall of Fame 2B outside Bill Mazeroski. This fact does not support Biggio's case for Hall of Fame induction. Kent and Alomar, however, compare well to Joe Morgan, who was inducted easily in 1990.

Figure 11 shows quality estimates for 3B, including two current Hall of Fame players, Wade Boggs and George Brett. Alex Rodriguez, still playing and likely to receive MVP votes in the next years, has a quality rating that far exceeds the rest of the group and would place him at the top of Hall of Fame 3B in Figure 5. Ron Santo is included only because his recent death has helped catalyze discussion about his Hall of Fame chances, though his voting period is long passed and induction would require getting sufficient votes from the Veterans Committee. Edgar Martinez, known primarily as a designated hitter for the Seattle Mariners, is often thought to be a 1B, since he played there sporadically during the second half of his career. For the purposes of Hall of Fame voting, however, he will be considered by the BWAA as 3B.

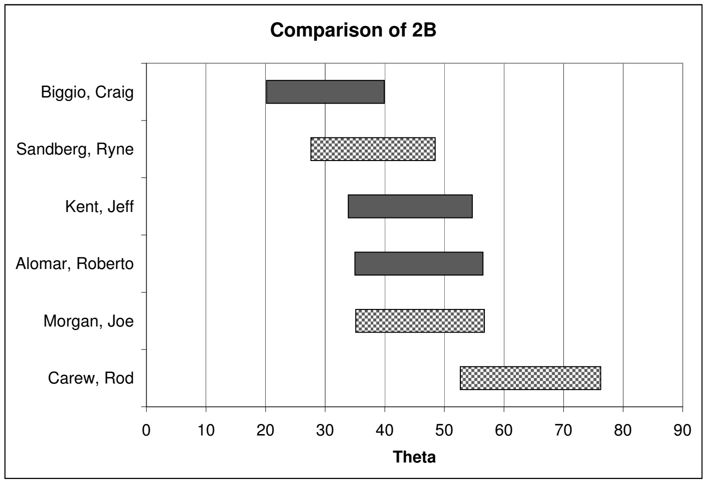

Figure 10: 95% confidence intervals for quality parameter θ of Second Basemen. Hall of Fame players are plotted with checkerboard pattern, and other players with solid bar.

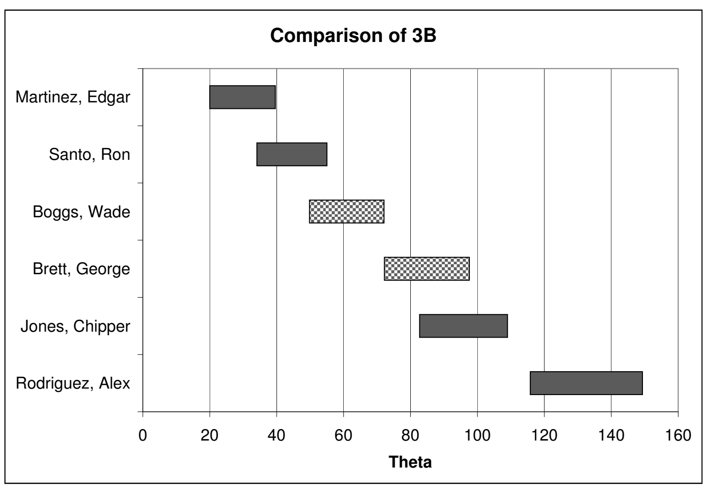

Figure 11: 95% confidence intervals for quality parameter θ of Third Basemen. Hall of Fame players are plotted with checkerboard pattern, and other players with solid bar.

While no single player ranking method can be definitive, these results may help to clear up debates and controversies surrounding player comparisons, especially for if the players being compared did not have overlapping careers. The graphs in this paper serve as the best summary of player evaluation that this analysis can glean from 80 years of BWAA voting data.

References

[1] Berry, S. M., Reese, S. P. and Larkey, P. D. (1999) "Bridging Different Eras in Sports" *Journal of the American Statistical Association*, 94, No. 447, 661-676.

[2] S. Csörgo, S. and J.F. Faraway, J. F. (1996) "The exact and asymptotic distributions of Cramér-von Mises statistics", *Journal of the Royal Statistical Society, Ser. B* 58 (1), 221-234.

[3] Kvam, P. H. and Vidakovic, B. (2008) *Nonparametric Statistics with Applications to Science and Engineering*, Wiley, New York.

DOI: 10.2202/1559-0410.1337