Reputation Effects in Gold Glove Award Voting

Arthur Zillante
ICES and George Mason University

February 6^{th}, 2005

Abstract

Reputation effects have been thought to influence how candidates in an election are viewed by the electorate. Using data from Major League Baseball, I attempt to quantify the effect that reputation plays in voting for the Gold Glove award. While the award is designed to reflect current-year defensive accomplishments, two other hypotheses have been suggested to explain voting behavior. The first is that voters use current-year offensive accomplishments in lieu of defensive accomplishments. The second hypothesis is that voters rely on the past performance of the players when casting their ballots, implying that reputation effects exist in the minds of voters. Results from probit estimation show that while reputation effects appear to have a significant effect on the outcome of the election, current-year offensive accomplishments do not.

JEL Codes: D72

Keywords: Voting behavior, baseball

1 Introduction

Award voting, such as for end of the season sports awards or motion picture awards like the Oscars, differs from voting for public office in that the award voters generally do not have a stake in the outcome of the election. Perhaps the only stake award voters have is that their voting privileges may be revoked if it is decided that they have done poorly, whereas larger stakes (economic and social concerns) abound in the political arena. Despite the lesser impact on society as a whole, award voting may have substantial impact on individuals. In the case of sports, many players have incentive clauses based upon the outcome of the award voting, where the incentives may be in the million dollar range. Or consider economists vying for the Nobel Prize – not only do they receive the award money if they win, but they then also have the benefit of increasing their appearance fees. Thus, determining the information and decision rules that award voters use when casting their ballots may be of great importance to the individuals associated with these awards, and may also uncover which information people use when voting.

The award under study is the Gold Glove award, given annually in each of the two leagues in Major League Baseball. The award is supposed to be given to the best defensive player at each position, although there is no specific criteria which defines the term "best". There is some speculation that voters rely on offensive, or batting, performance when casting their ballots for the Gold Glove award. In addition, given that there have been a number of players who have won the award for their position multiple times, it is possible that voters rely on a rule of thumb where they choose the current year winner based upon who has previously won the award. The goal of this paper is to determine which information is actually used by voters when they cast their ballots.

There is very little published research on award voting. One paper, Ginsburgh [3], discusses the relationship between economic success and award success for movies and books. However, he does not provide a detailed statistical study of the possible criteria used in the award voting. Another paper, Coupe [1], discusses best paper awards in economics and finance journals, and finds that the papers chosen as best papers are not necessarily the ones that will be most cited. Finally, Hamermesh and Schmidt [4] discuss the determinants of being elected as an Econometric Society Fellow. While this is likely the most closely related study in terms of technique and data, the elections in [4] are once and for all elections; thus, there is no possibility for repeated selection as a winner.

The topic is tangentially related to fields in other disciplines. There is a vast literature in political science discussing incumbency advantages in political elections. Given that players who win multiple Gold Glove awards typically win them consecutively, it is possible that an incumbency advantage exists for the player who won the award the previous year. It is also possible that halo effects surround certain players. A halo effect can be described as generalizing from one outstanding trait of a person or object to an overly favorable evaluation of the entire person or object. The fields of sociology, labor economics, and marketing all discuss halo effects with respect to people or products. Halo effects are related to incumbency advantage in that a successful performer one year may have a halo effect bestowed upon him that carries over into the next year.

Section 2 of the paper presents background information on the Gold Glove award, its winners, and the amount of turnover in the electorate throughout the time period. Section 3 provides a detailed description of the data used, and section 4 presents the empirical models used to test the hypotheses. Section 5 discusses the results of the estimation, and section 6 concludes.

2 Award History and Facts

The Gold Glove award was created in 1957 by Rawlings Sporting Goods to award defensive excellence at each fielding position and was promoted by *The Sporting News*[1]. In the first year of the award, 1957, only one winner per fielding position from all of Major League Baseball (MLB) was chosen. From 1958 to 1961, one winner at each of the nine fielding positions is chosen from each of the two leagues, the American League (AL) and the National League (NL). In 1962, the process of selecting outfielders changed. Prior to 1961 one outfielder at each outfield position (left-fielder, center-fielder, and right-fielder) was chosen from each league. From 1961 to the present, the position specific requirement for outfielders was dropped from the voting process, meaning that three outfielders who played the same position could all win the Gold Glove award[2].

The award winner for each position is determined by a plurality vote. In 1957, a committee of 19 sportswriters voted for the award winner. In 1958, active players voted for the award winner. The process of players voting for the award lasted until 1965. Since 1965 MLB managers and coaches have comprised the electorate. The voting structure is relatively simple, as managers and coaches vote for one player in their league whom they feel has been the best defensive player at each position. The vote is conducted near the end of the baseball season and the only stipulation is that voters cannot vote for players on their own team. The final balloting is covered in a cloud of secrecy – complete vote totals are not available for the Gold Glove award, which is unlike most other MLB awards. Most other recognized awards in MLB, including the Most Valuable Player, the Cy Young, the Rookie of the Year, and Hall of Fame voting, have complete vote totals available for each year.

The only rules governing the voting process are that each voter selects only one player at each position and that a voter cannot vote for players from his own team. Thus, the minimum standard that must be met for eligibility is that the player must be in the league at the time of the vote. Historically this has not led to positional inconsistencies, such as a shortstop winning the first base award, but the recent selection of one award winner caused a considerable amount of controversy. The AL Gold Glove winner at first base in 1999 was Rafael Palmeiro of the Texas Rangers. The controversy was not about Palmeiro's fielding ability, as he had won the award in the past, but about the fact that Palmeiro played in 158 games during the season and only 28 of them (18%) were at first base. In fact, he won the Gold Glove award for first base while winning another award, the Silver Slugger award[3], as the AL's best designated hitter, which is a non-fielding position. While the selection of Palmeiro is perhaps the most extreme example, other past selections suggest that voters for the Gold Glove award rely more on reputation than current performance when casting their ballots.

	Catcher	1^{st} base	2^{nd} base	3^{rd} base	Shortstop
# different winnners	27	22	28	23	30
# different AL	13	12	12	12	13
# different NL	16	11	16	13	17
# both leagues	2	1	0	2	0
Player with most	Bench[a]	Hernandez	Sandberg[b]	Robinson	Smith
# most	10	11	9	16	13

a Ivan Rodriguez passed Bench with 11 total in 2004
b Roberto Alomar passed Sandberg with 10 total in 2001

Table 1: Statistical information on award winners

2.1 Award Winner Information

This section provides background information on the number of Gold Glove award winners at the five infield positions (catcher, 1^{st} base, 2^{nd} base, 3^{rd} base, and shortstop). From 1957-1999 there were 85 total winners at each position, one each year from each league from 1958-1999 and one winner at each position in 1957. Table (1) breaks down the number of different winners for each of the five infield positions in various manners. There have been between 22 and 30 different winners at each position, and for each league there have been between 11 and 17 different winners. Only five players have ever won a Gold Glove in both leagues: Tony Pena and Bob Boone as catchers; J.T. Snow as a 1^{st} baseman; and Matt Williams and Robin Ventura as third basemen. No player has ever won a Gold Glove award at multiple infield positions in his career.

The last two rows show the player who has won the most Gold Gloves at a position and the number of Gold Gloves that player won. Note that 9 Gold Gloves is a little more than 10% of the total number of awards given at a position, while 16 Gold Gloves is a little less than 19%. Thus, these players have won a large share of the total Gold Gloves awarded, and an even larger share considering that players are only eligible to win one of the two Gold Gloves awarded at their position each year. It should also be noted that each of the five players that have won the most Gold Gloves at their position also won those awards consecutively[4]. Thus, it may be that the players were consistently better than their peers over the time period during which they won the Gold Glove, or it may be that the players developed a reputation early in their careers and a halo effect was bestowed on them as voters overlooked some small declines in ability.

2.2 Voter Turnover

While the individual choice of each voter will likely never be known, it is possible to approximate voter turnover for the time period since the voters for the Gold Glove award are the managers and coaches in MLB. From this point on, I will use the term coaches to refer to both coaches and managers. Calculating voter turnover is important because it creates a starting point for the discussion of reputation effects. If the identities of the voters for the Gold Glove award change rapidly then it may be difficult to hypothesize that reputation effects exist. However, if the identities of the voters change slowly, then it is plausible that the voters may use some measure of previous performance when casting their ballot.

Using data from Thorn [6], from 1965-2000 there were 857 total coaches, with 583 different American League coaches and 535 different National League coaches. Figure 1 shows the average decay rate for all coaches and managers in MLB, as well as by individual league. The decay rate is calculated in the following manner. Begin by fixing a year, say 1965, and then determine the percentage of coaches who were still coaching in MLB each of the 34 years later. Call this the decay rate for the 1965 coach cohort. Call the percentage change in coaches from 1965 to 1966 the 1-year percentage change in the 1965 cohort, the percentage change in coaches from 1965 to 1967 the 2-year percentage change in the 1965 cohort, etc. Then fix the second year, 1966, and calculate the percentage of coaches still coaching in MLB in each of the 33

[1] This section is based on Deane [2], who provides a more detailed history of the award as well as some anecdotal evidence on voting behavior.

[2] Prior to 2001, this had occurred 3 times in the NL and 11 times in the AL. All of these occurrences involve 3 center-fielders winning the award.

[3] The Silver Slugger is given to a player at each position in each league based on their offensive contributions.

[4] The two cases footnoted in the table, Rodriguez and Alomar, did NOT win all of their awards consecutively.

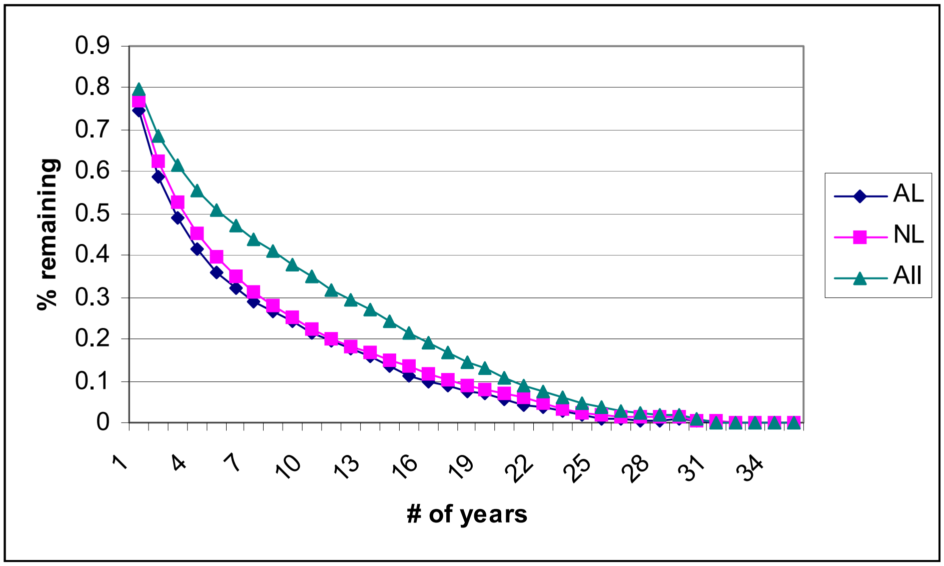

Figure 1: Average percentage of coaches remaining after a given number of years has passed

following years. The percentage change from 1966 to 1967 is the 1-year percentage change in the 1966 cohort. Continue the process for the remaining years from 1967-1998. The decay rate in figure 1 is calculated as the average of each of the t-year percentage changes. Thus, the 1-year percentage change in the figure has 34 observations (one for each year from 1965-1998), while the 34-year percentage change uses only the result from 1965. An identical process was used to calculate the decay rates for coaches in the AL and NL. Note that the coaches in the AL and NL decline at a faster rate than those in MLB as a whole, as some coaches will change leagues.

Suppose a player wins a Gold Glove in time period zero. From figure 1 we can see that approximately 44% of the coaches who were active in MLB that year will still be active in MLB 7 years later. For the individual leagues about 30% of the coaches are still active in the same league 7 years later. At year 12 the percentage of coaches who were active in MLB falls to 30%, while the percentage who are still active in each league is around 20%. Given these numbers, it is possible that some players who have been active in a league for a few years benefit from a reputation effect built up in an earlier year. Additionally, many coaches were former players and may have formed opinions about the ability of current players while the coach was an active player. Thus, it is unlikely that a first-time coach is just beginning a career in MLB and is watching the players for the first time.

3 Data

The data used in the analysis is taken from the Lahman database [5] and spans the years 1957-1999. A variety of quantitative and qualitative independent variables are used[5]. Quantitative variables can be broken into defensive and offensive variables. The defensive variables are total games played at the position, putouts, assists, errors, double plays participated in, and fielding percentage. All of these variables are counting variables except for fielding percentage. Fielding percentage is calculated as the ratio of successful defensive chances to total defensive chances[6]. The defensive variables are position specific per player per year in the database, so that if a player plays multiple positions throughout the year only those defensive statistics recorded at that position are tabulated. As an example, Jose Oquendo appeared at every position for the St. Louis Cardinals in 1988. He has seven different entries[7] for his defensive statistics, one for each

[5] Definitions of the offensive and defensive measures can be found at MLB's website, www.mlb.mlb.com.
[6] Technically, I calculate this as (putouts + assists)/(putouts + assists +errors).
[7] The three outfield positions are listed as one position in the data until 1996, when they are listed as three separate positions.

Position	**#, games ≥ 1**	**#, games ≥ 25**	**#, games ≥ 50**	**# diff winners**
Catcher	3262	2173	1564	27
1^{st} base	4749	1808	1292	22
2^{nd} base	3828	1776	1278	28
3^{rd} base	4266	1792	1248	23
Shortstop	3475	1727	1251	30

Table 2: Number of players who have appeared at a position in the data

position he played. Thus, Oquendo appears as seven "different" players in 1988, at least according to his defensive statistics.

The offensive statistics used in the study are runs scored, hits, homeruns, runs batted in (RBI), walks, stolen bases, batting average, slugging percentage, and on-base percentage. The first six variables listed are counting variables, while the last three variables are rates. The offensive statistics are not position specific; that is, if a player appears in the data as both a second baseman and a shortstop during one season then his offensive statistics for that season are the same for each entry in the data. In the aforementioned case of Jose Oquendo, the offensive statistics for each of his seven defensive entries would be the offensive statistics he accumulated over the course of the entire season.

Some may argue that there may be better measures of offensive and defensive performance, usually termed sabermetric measures. However, the sabermetric measures are complex and most of these measures had not been developed until the 1970s and 1980s; thus these measures were unavailable to voters in the early part of the study. Although the creation of these measures may have impacted the award voting, it is unlikely given that the measures have just recently begun to gain favor in the majority of the baseball community. Since the purpose of this study is to determine what factors impact the probability of winning a Gold Glove award, and not to determine if the voters actually voted for the "best" defender at the position, the traditional offensive and defensive statistics are used instead of the newer measures.

The qualitative variables used typically reflect some form of reputation that the player may have developed. There are two seasonal binary variables. The first takes the value 1 if the player's team reached the postseason during a given year and 0 otherwise. It is commonly thought that players whose teams make the postseason receive a boost in end of the year award voting. The second seasonal binary variable takes the value 1 if the player makes the All-Star team during a given year and 0 otherwise. Selection to the All-Star team, while typically based on offensive achievement, may push some players names ahead of others in the voters' minds. Although players may have appeared at more than one position during any given year, the All-Star binary only registers as a 1 for the player's primary position, as determined by the position at which he played the greatest number of games. As an example, Kirby Puckett appears in 2 games as a second baseman and 149 games as an outfielder in 1992. He was also selected to the all-star game in that season. Since his outfield games are larger than his second baseman games, his All-Star binary as a second baseman in 1992 registers as a zero.

The final variables used in the study are the reputation variables pertaining to prior Gold Glove award voting outcomes. The first variable used is a count of Gold Glove awards that a player has won prior to the current season. Thus, it can be seen if winning more Gold Glove awards increases one's chances of winning another Gold Glove. A second variable used is a binary that registers as a 1 if a player has ever won a Gold Glove award in the past.

The positions evaluated in this study are limited to the infield positions (first base, second base, third base, shortstop and catcher). Pitchers are not evaluated because there is very little variation in the defensive and offensive data for pitchers as most pitchers do not participate in more than 250 innings per year[8]. Outfielders are not used in the analysis for two reasons. First, there is the aforementioned change in the voting process in 1961, which removed the outfield position specific voting. Second, there are some data aggregation issues with outfielders.

Table 2 breaks down the number of players who have appeared at each position according to different games played criteria. Note that this games played criteria is defensive position specific, so that a player who appears in 150 games as a shortstop and 2 games as a second baseman will be included in the "#, games

[8] Typically, regular players at the other positions will participate in more than 1000 innings per year.

$\geq$ 1" column for both positions but will only be included in the columns for "games $\geq$ 25" and "games $\geq$ 50" for the shortstop position. The columns for 25 games and 50 games are included because it will be necessary to remove those players who only appeared in a handful of games at a position when considering offensive statistics, as some players may have played a few games at a position but many games overall, increasing their offensive totals[9]. While the choices of 25 and 50 are arbitrary, at the 25 game level all of the Gold Glove winners still remain in the data set as the lowest total games played at a position for which a Gold Glove was won is the 28 games at 1^{st} base by Rafael Palmeiro. All results reported in the following sections are for the 50 game criteria, since there are little quantitative and qualitative changes in the results if the 25 games criteria is used.

4 Empirical Models

Various empirical models are used to test the validity of the competing hypotheses. The simplest model incorporates fielding statistics for the current season, which is what the guidelines of the award suggest should be used. Let Y_{ipt} be a binary taking the value 1 if a player i won a Gold Glove at a particular position (p) in a given year (t) and 0 otherwise. Let D_{ipt} be a vector comprised of the defensive statistics described above and δ be a corresponding vector of parameters. Let X_i denote a vector of league and time dummy variables, and ν be its associated vector of parameters. The league and time dummies will be present in all models. The term ε_{ipt} is the error term in all models. The "defense-only model" is then:

$$Y_{ipt} = f\left(D_{ipt}\delta + X_i\nu + \varepsilon_{ipt}\right) \tag{1}$$

The defensive statistics initially used in the defense-only model will be games fielding at the position, putouts, assists, errors, double plays participated in, and fielding percentage. Due to possible multicollinearity among the variables, some variables may be removed from the equation if they do not meet commonly accepted levels of significance[10].

For the defense-only model, it is expected that the signs of all the defensive variables will be positive except for errors. The rationale behind this expectation is that making more successful defensive plays (putouts, assists, and double plays) should lead to a greater probability in winning the Gold Glove, while making more unsuccessful plays (errors) should lead to a lower probability. The coefficient on fielding percentage should also be positive, as having a higher ratio of successful plays to total plays should also increase one's chances of winning the award. Different positions may have different significant regressors. This is particularly true in the first base and third base model, as first basemen typically do not have many assists and third basemen typically do not have many putouts.

The second model incorporates the offensive statistics, which are denoted by the vector H_{it}. Letting η be its vector of parameters, the "offense-defense model" is:

$$Y_{ipt} = f\left(D_{ipt}\delta + H_{it}\eta + X_i\nu + \varepsilon_{ipt}\right) \tag{2}$$

The offensive statistics that will initially be used are: runs, hits, homeruns, RBI, stolen bases, walks, batting average, slugging percentage, and on base percentage[11]. Again, some variables may be removed from the equation if they do not meet commonly accepted levels of significance.

The estimated coefficients of the defensive variables in the offense-defense model are expected to have the same signs as in the defense-only model. Given that all of the offensive variables have positive value towards winning games, it is expected that the offensive variables will have positive signs. This is particularly true if the hypothesis that voters are relying on offensive measures when casting their ballots is true.

Finally, let R_{ipt} be a vector representing the reputation parameters and ρ be its vector of parameters. The "full model" is:

[9] See the previous example regarding Kirby Puckett appearing as a second baseman in 1992.

[10] The time and league dummies will not be removed regardless of their levels of significance. Excluding those variables does not alter the results.

[11] Due to limitations in the data, on-base percentage is calculated as (walks plus hits) divided by (at bats plus walks), neglecting hit by pitches, sacrifice flies, and catcher's interference. The small number of occurrences of these events should not impact on-base percentage for a large percentage of the players.

$$Y_{ipt} = f\left(D_{ipt}\delta + H_{it}\eta + R_{ipt}\rho + X_i\nu + \varepsilon_{ipt}\right) \tag{3}$$

The reputation measures consist of binaries for being selected to the all-star team in that year, the player's team reaching the postseason, and winning a previous Gold Glove award. In addition, a count of previous Gold Glove awards is also used as a measure of reputation. It is expected that the signs on the estimated coefficients of the reputation measures will all be positive, as reputation effects as they are measured here reflect positive aspects of reputation.

5 Results

The results of the probit estimation for the restricted version of the full model[12] described in equation (3) are contained in table 3, while results for the restricted models described in equations (1) and (2) are contained in tables 5 and 6, respectively. In addition, results for an offense-only model are contained in table 4. All models used data sets where the number of games played at the position was greater than or equal to 50. Using an alternative cutoff of 25 games played at a position yields little difference in both the qualitative and quantitative coefficient estimates. All models also contained dummy variables for year and league, which are not reported here due to space limitations. Most of the dummy variables were not statistically different from zero at the 10% level.

The middle three columns of table 3 show the coefficient estimates, standard errors, and p-values for the regressors that were significantly different from zero at the 10% level for the individual positions. The only common variable across all 5 positions is the binary representing whether or not the player had previously won a Gold Glove award. All models include either putouts or assists (or both), as these variables seem to act as a censoring mechanism; the more putouts or assists a player amasses, the better his chance of winning the Gold Glove. The rather large coefficients on fielding percentage and on base percentage can be explained by the fact that these variables are rate statistics measured in thousandths, with a maximum value of 1; thus, increases in these variables should be viewed as a one-one thousandth unit increase rather than a one-unit increase.

Perhaps the most startling result is the lack of statistical significance of offensive statistics in the restricted version of the full model, as a common belief among baseball professionals and aficionados is that voters use offensive statistics in lieu of defensive statistics when casting their ballots. In the catcher, first base, and second base models, no offensive variables are statistically significant at the 10% level. In the third base model, the only offensive variable that is statistically significant is RBI, which suggests that voters put some weight on production at the plate when choosing a third base award winner[13]. The shortstop model is the only model which includes multiple offensive statistics: runs scored, hits, homeruns, stolen bases, walks, and on base percentage. However, three of these offensive statistics (hits, homeruns, and walks) have negative coefficients, suggesting that there are some offensive events which detract from a player's possibility of winning the award. It is possible that the negative coefficients on walks and hits are offsetting the positive value of a higher on base percentage, although a likelihood ratio[14] test suggests that these variables are jointly significant. The positive coefficient on stolen bases may reflect an unobserved individual effect, speed and quickness, that most observers deem necessary to perform well as a defensive shortstop. Thus, stolen bases may act as a proxy for speed and quickness, which may not be captured by the defensive statistics. The final offensive variable, runs scored, appears to be similar to RBI for third basemen.

The coefficients for the defensive variables all have the correct sign, and the results can be explained from the viewpoint of a baseball observer. The variables total games played at a position and double plays participated in are not significant in any models. The failure of total games played to appear in any of the models is unsurprising, as there is likely a collinear relationship between games played and the performance based counting measures (putouts, assists, and errors). It was typically the case that the removal of the games played variable from the probit equation substantially altered the significance of the remaining counting measures, which suggests this collinearity. However, it is surprising that double plays

[12]Results for all unrestricted models are available from the author upon request.

[13]This is not unusual as voters for most awards in baseball, particularly the Most Valuable Player Award, tend to be enamored with the RBI statistic.

[14]This test is discussed in detail below.

	Regressor	Coefficient	Std. Error	P-value	χ^2	P-value
Catcher	Putouts	0.00231	0.00097	0.01714	9.9594	0.7651
	Assists	0.03204	0.00813	0.00008	$df = 14$	
	Errors	−0.05044	0.02542	0.04743		
	Allstar	0.76705	0.21718	0.00043		
	Won before	1.17040	0.19174	< 0.00001		
1^{st} base	Putouts	0.00307	0.00063	< 0.00001	20.8028	0.1432
	Errors	−0.11277	0.03222	0.00048	$df = 15$	
	Won before	1.48298	0.31409	< 0.00001		
	Prev Amt	0.12907	0.06704	0.05444		
2^{nd} base	Putouts	0.00603	0.00332	0.06974	18.5342	0.1383
	Assists	0.00785	0.00265	0.00316	$df = 13$	
	Fielding pct	72.31809	24.47603	0.00319		
	Postseason	0.71085	0.25585	0.00555		
	Allstar	0.93797	0.23087	0.00005		
	Won before	1.30177	0.21699	< 0.00001		
3^{rd} base	Assists	0.01091	0.00307	0.00039	13.528	0.5616
	Fielding pct	25.91087	10.55147	0.01420	$df = 15$	
	RBI	0.02134	0.00568	0.00018		
	Won before	1.54079	0.22696	< 0.00001		
Shortstop	Putouts	0.01505	0.00495	0.00240	12.91696	0.2039
	Fielding %	81.53809	14.25952	< 0.00001	$df = 9$	
	Runs	0.03028	0.01140	0.00804		
	Hits	−0.01839	0.00919	0.04554		
	Homeruns	−0.03789	0.02003	0.05873		
	Stolen bases	0.01445	0.00690	0.03652		
	Walks	−0.01964	0.01175	0.09492		
	On base pct	12.95638	6.38070	0.04252		
	Allstar	0.47378	0.26613	0.07529		
	Won before	0.81414	0.23908	0.00068		

Table 3: Probit estimation results for the restricted models

	Regressor	Coefficient	Std. Error	P-value	χ^2	P-value
Catcher	Hits	0.02763	0.00442	< 0.00001	1.6812	0.9754
	Batting avg	−8.67360	4.59873	0.05947	$df = 7$	
1^{st} base	Runs	0.01788	0.00886	0.04380	5.0598	0.5362
	Hits	0.01430	0.00447	0.00142	$df = 6$	
	Slugging pct	−6.45482	1.62830	0.00008		
2^{nd} base	Homeruns	0.05834	0.01274	0.00001	3.5042	0.6228
	Walks	0.07482	0.01858	0.00006	$df = 5$	
	Batting avg	69.07822	15.98684	0.00002		
	On base pct	−69.54434	17.43661	0.00007		
3^{rd} base	Hits	0.00660	0.00412	0.10981	7.9632	0.3359
	RBI	0.02749	0.00489	< 0.00001	$df = 7$	
Shortstop	Runs	0.01429	0.00632	0.02389	1.0132	0.9078
	Homeruns	−0.06275	0.02410	0.00934	$df = 4$	
	RBI	0.01740	0.00966	0.07187		
	Stolen bases	0.02103	0.00609	0.00057		
	Walks	0.00924	0.00560	0.09898		

Table 4: Probit estimation results for the offense-only restricted models

	Regressor	Coefficient	Std. Error	P-value	χ^2	P-value
Catcher	Games	0.01549	0.01043	0.13767		
	Putouts	0.00207	0.00123	0.09318	1.2878	0.5252
	Assists	0.02751	0.00820	0.00082	$df = 2$	
	Errors	−0.06900	0.02276	0.00247		
1^{st} base	Putouts	0.00280	0.00052	< 0.00001		
	Assists	0.00716	0.00431	0.09698	3.3388	0.3423
	Errors	−0.13394	0.02654	< 0.00001	$df = 3$	
2^{nd} base	Games	0.03025	0.01158	0.00911		
	Assists	0.00501	0.00226	0.02712	2.632	0.4519
	Fielding pct	77.93951	19.10462	0.00005	$df = 3$	
3^{rd} base	Assists	0.01500	0.00202	< 0.00001	1.0196	0.9068
	Fielding pct	33.01677	8.93300	0.00023	$df = 4$	
Shortstop	Putouts	0.00798	0.00397	0.04476		
	Assists	0.00323	0.00194	0.09567	2.3438	0.5042
	Fielding pct	74.03203	12.71675	< 0.00001	$df = 3$	

Table 5: Probit estimation results for the defense-only restricted models

participated in does not have as large an effect, particularly for the second base and shortstop positions, as turning a double play is typically regarded as an important skill for these positions.

In the catcher model, putouts and assists are both significant and positive, while errors made is significant and negative. Catchers are rewarded for making plays in which an out is recorded and penalized when they fail to make plays where an out should be recorded. The coefficient on assists is ten times as large as the coefficient on putouts, which likely reflects the belief that a good defensive catcher is one who throws out a large quantity of would-be base stealers, as about 50%-60% of catcher assists are from throwing out runners attempting to steal. The absence of fielding percentage at the catcher position may be because the counting measures absorb most of the effect of fielding percentage or because catchers tend to be measured more on level effects rather than rate effects. One variable which was not available in the data set which would likely be a useful measure for catchers is the percentage of runners thrown out attempting to steal. While assists may act as a proxy for this measure, it is possible for some poor catchers to have a larger assist total than their outstanding counterparts due to the fact that more stolen bases are attempted against the poor catchers.

Only two defensive variables are significant in the first base model, putouts and errors. First baseman, unlike the other positions, is generally not regarded as a skill position; most fielding chances for a first baseman are likely to be receiving throws from other infielders. Thus, a good defensive first baseman is typically regarded as one who can catch throws from other players, regardless of where the throws may be. In most cases, if a throw is errant a first baseman is not charged with an error, so there tends to be little variation in fielding percentages among first baseman. The absence of assists as a significant regressor is not surprising given that first baseman are typically not required to make many assists. Oftentimes if a ball is batted towards a first baseman he will field the ball himself and then record an unassisted putout by touching first base.

The significant defensive variables for the remaining positions have similar explanations. At second base, putouts, assists, and fielding percentage are the significant variables. At third base, only assists and fielding percentage are significant. Third basemen are essentially the opposite of first basemen, in that most of the value from the third basemen comes from his ability to throw the ball across the diamond to record an assist. Third basemen typically have the fewest putouts among the infield position, which is why that regressor is not significant. For shortstops, putouts and fielding percentage are the only significant defensive variables. This is quite surprising as most baseball observers would like choose assists as the more relevant of the counting measures for shortstops.

As previously mentioned, the only variable that is common across all 5 positions is the binary reputation variable that records if a player has previously won a Gold Glove award. Among the 5 positions, the estimated coefficient is lowest among shortstops, which may reflect the fact that more different players have won Gold Glove awards at shortstop than at any other infield position.

The allstar variable is significant for catchers, second basemen, and shortstops. While this seems to suggest that voters rely on allstar reputations when casting their votes, it may also reflect the belief of baseball professionals that it is important to be defensively sound in the middle of the field (catcher, second base, shortstop, and centerfield). Thus, the causal relationship between allstar selections and Gold Glove awards is debatable, although selection to the allstar team occurs prior to Gold Glove voting. The fact that the allstar variable is not significant in the first base and third base models supports the belief that teams look primarily for players who can contribute offensively at these positions, and that fans (who vote for the allstar game starters) also emphasize offensive over defensive contributions from these positions.

The first base model is the only one that includes the amount of Gold Gloves previously won as a significant regressor. The significance of the amount of awards previously won in that model may be attributable to two factors. The first factor is that because first base is viewed as the least demanding position, good defensive first basemen may remain competitive in the award voting even after they have passed the primes of their careers. At other positions, once players "lose a step or two" it is likely to be more noticeable to observers, and they may vote for a younger, more athletic player. Also, it may take more time for voters to form an opinion of a first baseman as a defensive player, which may cause them to select more veteran players. Thus, once a first baseman gains inertia as an award winner, the only thing that may slow him down is retirement.

The final two columns contain the χ^2 statistic for a likelihood ratio test, the degrees of freedom for the χ^2, and the p-value for the statistic. The likelihood ratio test can be used to test for the joint significance of regressors. The test statistic is computed by taking the difference of the likelihood functions of a restricted and unrestricted model and multiplying that difference by negative two. The statistic is distributed χ^2 with degrees of freedom equal to the number of restrictions imposed. The unrestricted model for table 3 corresponds to the "full model" in equation (3), and contains the time and league dummy variables as well as the 19 regressors (6 defensive statistics, 9 offensive statistics, 4 reputation measures) for all positions. The number of restrictions imposed differs in most of the models, as different positions had different regressors which were deteremined to be significant.

Let r be the number of restrictions imposed on the model and β_i be the coefficient for the i^{th} regressor for which a restriction is imposed. The null hypothesis of the likelihood ratio test is then $\beta_i = 0 \forall i = 1, .., r$. The alternative hypothesis is that at least one $\beta_i \neq 0$. Note that the lowest p-value, for the second base model, is 0.1383, which means that the null hypothesis is not rejected at the 10% level. Thus, the omitted regressors are not jointly significant at the 10% level for any of the models. The next section shows that removing those regressors does not substantially alter the predictive power of the model.

Tables 4-6 show the results for three restricted models: offense-only, defense-only, and defense-offense. In these models, only variables in the categories that correspond to the model name are used. Thus, the offense-only unrestricted model has 9 variables, defense-only has 6 variables, and offense-defense has 15. Results are shown for those models where only significant regressors are used in the estimation, and likelihood ratio tests are performed to test for the joint significance of the removed regressors. In all models, the regressors that were removed from the equation were not jointly significant. For all of the models, the signs on the estimated coefficients in the restricted version of the full model (table 3) are identical to the signs in the restricted versions of the other three models. Thus, adding the reputation measures does not alter the signs of those regressors that are significant, although the reputation measures may cause some regressors to become insignificant. In addition, the quantitative value of the regressors that remain significant when the reputation effects are added change only slightly.

5.1 Predictions

In addition to the t-statistics and likelihood ratio tests, a goodness-of-fit measure is used to test the predictive power of the models. A standard goodness-of-fit measure involves calculating the predicted value ($\hat{y}_i$) of the dependent variable for each observation, and then classifying that observation as a 0 if $\hat{y}_i < 0.5$ and 1 if $\hat{y}_i > 0.5$. The predicted classifications can then be compared to the observed value of the dependent variable, and percentages can be obtained to see how many times the model yields correct predictions.

In this study, I use a modified version of the goodness-of-fit measure. Given that only one player at each

	Regressor	Coefficient	Std. Error	P-value	χ^2	P-value
Catcher	Putouts	0.00242	0.00089	0.00653	8.565	0.6620
	Assists	0.03275	0.00779	0.00003	$df = 11$	
	Errors	-0.06912	0.02371	0.00361		
	RBI	0.01538	0.00423	0.00028		
1^{st} base	Putouts	0.00268	0.00071	0.00018	7.8692	0.5474
	Errors	-0.12204	0.02768	0.00001	$df = 9$	
	Runs	0.02226	0.00959	0.02049		
	Walks	-0.00758	0.00498	0.12774		
	Batting avg	7.82634	4.63850	0.09181		
	Slugging pct	-6.29683	2.16575	0.00371		
2^{nd} base	Putouts	0.00792	0.00300	0.00844	5.0818	0.5334
	Assists	0.00671	0.00246	0.00644	$df = 6$	
	Fielding pct	85.08968	21.19704	0.00006		
	Hits	-0.02582	0.01223	0.03498		
	Homeruns	0.06290	0.01662	0.00016		
	Stolen bases	0.01352	0.00828	0.10301		
	Walks	0.06700	0.03709	0.07112		
	Batting avg	86.91510	37.75753	0.02151		
	On base pct	-69.11532	34.14888	0.04319		
3^{rd} base	Assists	0.01393	0.00264	< 0.00001	2.3688	0.9927
	Field pct	39.04854	9.08772	0.00002	$df = 10$	
	Hits	-0.01616	0.00903	0.07386		
	RBI	0.02852	0.00636	0.00001		
	Batting avg	12.82103	6.93071	0.06458		
Shortstop	Putouts	0.01378	0.00484	0.00447	2.1383	0.9065
	Field pct	96.85161	14.45526	< 0.00001	$df = 6$	
	Runs	0.03411	0.01260	0.00689		
	Hits	-0.02230	0.01130	0.04871		
	Homeruns	-0.08378	0.03629	0.02113		
	RBI	0.02371	0.01475	0.10819		
	Stolen bases	0.02433	0.00667	0.00027		
	Walks	-0.01879	0.01188	0.11408		
	On base pct	15.13292	6.62312	0.02249		

Table 6: Probit estimation results for the offense-defense restricted models

	Catcher	1^{st} base	2^{nd} base	3^{rd} base	Shortstop
Models using all regressors					
Defense, offense, & reputation	55	52	62	64	56
All offense only	31	29	33	31	32
All defense only	41	32	34	40	42
All defense & offense	42	31	45	46	55
Models using only significant regressors					
Defense, offense, & reputation	53	50	59	63	55
Offense only	29	26	32	31	29
Defense only	40	34	36	41	37
Defense & offense	42	32	48	47	51
Lagged	53	56	50	50	44

Table 7: Number of correctly predicted winners of the Gold Glove award at each position by various model specifications

position per league can win a Gold Glove award each year, I calculate the predicted value for each player and then compare that value with the predicted values of the other players at that position who were in the league that year. The player with the highest predicted value per league per season is then selected as the predicted winner of the Gold Glove award among that league/season cohort.

The results of the prediction analysis for nine models are contained in table 7. The first four rows of table 7 show the number of correctly predicted winners by a particular model containing all of the variables in a given category. For instance, the "All defense only" row shows how many correct predictions were made using the coefficient estimates from the model where all 6 defensive variables were used as regressors, while the "All offense only" row shows the number of correct predictions made using the coefficient estimates where all 9 offensive variables were used in the estimation process. The second set of rows shows the number of correct predictions made when only the significant regressors are present. Note that each of the positions may have a different set of regressors in these models, which are listed in tables 3-6. The prediction results for the models with only reputation effects are not reported due to the fact that 3 of the 4 variables are binary, leading to a large number of ties for the predicted winners[15]. The final row uses a simple measure for predicting the current year winner, which is to assume that the player who won the award the previous season will win it again in the current year. While this measure is simplistic, it provides a baseline to which the probit estimation predictions can be compared.

Table 7 shows that the models with reputation effects outperform the other models in predictive ability, and in some cases by substantial amounts. The offense only models are particularly poor predictors, with a maximum number of 33 correct predictions out of 84 winners (39%). The best predictors are the models using all 19 variables, except in the case of the 1^{st} base model where the simple lagged player model correctly predicts four additional winners. While it may be expected that the model with the most regressors predicts the best, note that the models only including the significant regressors for defense, offense, and reputation have the second highest amount of correct predictions, and that these are not much lower than the amount of correct predictions by the models using all 19 variables[16]. In fact, the simple lagged model outperforms every model except the ones including reputation effects for all positions but shortstop. Thus, it does not appear as if it is the inclusion of meaningless additional regressors that is driving the prediction results, but that the reputation effects have a significant effect on the voters' decisions. This supports the results of table 3, which suggest that reputation effects play a large role in determining the voters' decisions.

[15] As an example, in the 1^{st} base model, there were 8 two-way ties and 3 three-way ties.

[16] The maximum difference is 3 in the 2^{nd} base model, and this is still 11 more correct predictions than the next closest 2^{nd} base model, not including the lagged model.

6 Conclusion

Of the three hypotheses discussed at the outset, it appears as if voters use a combination of defensive measures and reputations when making their decision about the Gold Glove award winner. The models that include reputation effects are the best predictors, and in some cases substantially so, of Gold Glove award voting. This suggests that the voters may be relying more on reputation than on actual defensive performance, while the guidelines of the award state that the winner should be determined based on defensive performance for the current season. Although there is speculation that improvement in offensive performance also increases a player's probability of winning a Gold Glove award, there is little evidence to suggest that offensive performance impacts the voting process on a regular basis. While there certainly may be individual instances where offensive performance enters into voting decisions, these appear to be isolated instances.

There is additional information that may add to the ability of the models to correctly predict the winner. One piece of information that could be included is a measure of record-breaking performance. Many baseball observers pay considerable attention to consecutive game streaks, so that a player who commits no errors in many consecutive games receives more attention than a player who makes less errors, but has them spread more evenly over the course of the season. Thus, the focus on the player with the record-breaking performance may enhance that player's reputation at the expense of a better defensive player. Another possibility is to determine the league leaders among the different offensive and defensive categories, as league leaders may gain more notoriety than those who do not lead the league in a category.

The results from Gold Glove award voting show that voters may use a rule of thumb involving reputation effects when casting their ballots, instead of relying on available information that is relevant according to the guidelines of the award. While this study only focuses on a small subset of voters, baseball coaches, and one award, it does suggest that future studies of voting behavior incorporate measures of candidate reputation in their analysis.

References

[1] Tom Coupe. The Prize is Right? An Analysis of Best Paper Prizes, 2004. Working Paper.

[2] Bill Deane. Awards and Honors. In John Thorn, Pete Palmer, Michael Gershman, and David Pietrusza, editors, *Total Baseball V: The Official Encyclopedia of Major League Baseball*, pages 207–241. Penguin Group, New York, 1997.

[3] Victor Ginsburgh. Awards, Success and Aesthetic Quality in the Arts. *The Journal of Economic Perspectives*, 17:2:99–112, 2003.

[4] Daniel Hamermesh and Peter Schmidt. The Determinants of Econometric Society Fellows Elections. *Econometrica*, 71:1:399–408, 2003.

[5] Sean Lahman. The Baseball Archive, 1999. CD-ROM.

[6] John Thorn. *Total Baseball V: The Official Encyclopedia of Major League Baseball*. Penguin Group, New York, 1997.