Ilmenau University of Technology
Institute of Economics

Ilmenau Economics Discussion Papers, Vol. 27, No. 170

Differences in the Voting Patterns of Experts, Peers, and Fans - Analyzing the NFL's All-Star Team Selections

Philipp Kunz-Kaltenhäuser, Sophia Gaenssle, Oliver Budzinski

January 2023

Institute of Economics
Ehrenbergstraße 29
Ernst-Abbe-Zentrum
D-98693 Ilmenau
Phone 03677/69-4030
Fax 03677/69-4203

ISSN 0949-3859

https://www.tu-ilmenau.de/iedp

Differences in the Voting Patterns of Experts, Peers, and Fans - Analyzing the NFL's All-Star Team Selections

Philipp Kunz-Kaltenhäuser[*], Sophia Gaenssle[+], Oliver Budzinski[°∇]

Abstract: Experts' voting behavior is conjectured to be more objective than peer voting (own group/ peers) and public voters (everyone interested), who are supposedly influenced by all sorts of subjective aspects. We examine differences in voting behavior between these groups by analyzing the voting outcomes for all-star teams in American Football. This paper analyzes the impact of performance as well as non-performance markers and team effects on the voting outcome. It contains a comparative analysis across the mentioned groups to elaborate on differences.

The econometric analysis uses unbalanced panel data of All-Pro and Pro Bowl player selections over 78 seasons (1951-2019). It applies panel probit regression to assess the impact of the markers on the outcome probability of winning one of the All-Star awards.

We find that expert, peer, and public voting show similarities and are partially driven by the same performance and non-performance markers. However, none of the three analyzed voting systems is free from the influence of non-performance markers. We find exposure effects as well as effects from team affiliation in all of them, including in fact expert voting. Positive effects of team success are found in expert and, to a lesser extent, in peer voting. Team-specific effects are found in public voting, providing evidence for partisanship voting by fans.

Our results shed doubt on the suspected objectiveness of expert voting. Furthermore, they fortify the notion of public voting being inefficient at identifying objective quality and extend the literature on voting biases among experts, peers, and the general public.

Keywords: Voting Behavior, Voting Bias, Expert Voting, Public Voting, Sports Economics, National Football League, American Football

JEL: D72, Z20, L83

[*] Chair for Economic Theory, Institute of Economics, Ilmenau University of Technology, Germany. Corresponding author: philipp.kunz-kaltenhaeuser@tu-ilmenau.de

[+] Assistant Professor of Creative Industries and Digitalization, Erasmus School of History, Culture and Communication, Erasmus University Rotterdam, Netherlands, email: gaenssle@eshcc.eur.nl

[°] Professor of Economic Theory, Institute of Economics, Ilmenau University of Technology, Germany.

[∇] We thank Thomas Grebel for providing valuable help collecting the data set; Benedikt Schmal, Stefan Heyder, and the participants of 56. Hohenheimer Oberseminar (Düsseldorf, Germany), as well as the participants of NAASE Sessions at WEAI Conference 2022 (Portland, OR, USA) for helpful comments on earlier versions of this paper.

Table of Contents

1. Introduction: Motivation and Related Literature

Experts and the general public are likely to differ in their voting behavior. While experts are believed to be knowledgeable and vote according to the matter-of-fact goals of the vote, laypersons are expected to compensate for their lack of expertise by relying on secondary criteria that deviate from the original goal of the vote (Franck & Nüesch, 2008). Thus, expert panels where each expert votes for herself and the result is the sum of the votes are conjectured to be determined by objective criteria, whereas public assessments are supposed to be heavily influenced by all sorts of secondary non-performance markers. These conjectures may influence the allocation of voting rights between experts and laypersons in so different areas of society as law proceedings (judges versus jurors), political systems, and entertainment contests. Therefore, the empirical determinants of different groups' voting behavior provide valuable knowledge about the likely effect of vote designs. It is difficult to observe "true" voting behavior in a natural setting without biases of stated preferences and social desirability/standards. Voting for all-star teams (i.e., the team composed of the most favored players of the season) serves as a natural experiment, where it is possible to collect information on all voting parties. Moreover, it is possible to compare the voting results for thousands of players to analyze if non-performance markers bias voting behavior and whether the jury solely votes for the objectively "best" players of the season. As such, all-star team voting is similar to music contest voting (which is used extensively in the literature; see below), but with the advantage that the true qualities of sports players can be measured much better than musical or entertainment talent. This facilitates the identification of voting biases.

The voting behavior of experts and the general public has been empirically and comparatively analyzed in the literature in the context of music contests like the Eurovision song contest where both expert juries and the television audience vote for the winner (inter alia, Haan et al., 2005; Ginsburgh & Noury, 2008; Spierdijk & Vellekoop, 2009; Budzinski & Pannicke, 2017, 2022). This literature predominantly confirms a so-called voting bias (i.e., deviation of voting from the original goals of the vote), which is stronger with fans and public audiences than with experts. However, even the experts are often not completely unbiased. Further empirical analyses of expert voting systems include instrument-playing-related music competitions (Flôres & Ginsburgh, 1996; Glejser & Heyndels, 2001; Ginsburgh & van Ours, 2003).

There has been a body of literature that analyzes sports disciplines where expert juries determine the result like figure skating, ski jumping, gymnastics, and diving (Campbell & Galbraith, 1996; Popović, 2000; Zitzewitz, 2006; Emerson et al., 2009; Pope & Pope, 2015; Lyngstad et al., 2020; Scholten et al., 2020; Heiniger & Mercier, 2021). All of them find versions of biased voting in experts: same-group racial preferences, nationalistic voting, cultural similarities, or protective voting against competitors in favor of their preferred athletes. Furthermore, Johnson & McCarthy (2022) showed evidence of cultural bias in the FIFA's World's Best Male Football Player award, which uses a mix of experts and peers voting. Here, media expert voters were found less biased than peer players which in turn were less based than coaches. However, closest to our research is the analysis of voting biases in American sports. Coleman et al. (2010) analyzed Associated Press College Football polls where a group of 100+ local and national sports experts vote on college football teams to determine their power rankings and future schedule. They found (i) familiarity effects in jury voting where teams that were televised more received *ceteris paribus* significantly more votes than their less televised competitors and (ii) team performance effects were found in expert voting. [2] Campbell et al. (2007) found the same biases as Coleman et al. in expert voting when correcting for their own and their opponent's on-field performance. Analyzing the same voting poll, Logan (2007) found the margin-of-victory of a team to be irrelevant to expert voting. Walther et al. (2002) found the evaluation of sports games even stronger biased by team success when assessed by experts than by non-experts.

The literature altogether establishes two main insights: firstly, expert juries are indeed less influenced by secondary markers than audiences but, secondly, expert juries are not unbiased but are also influenced by non-performance aspects of contestants like familiarity or affiliation. Our analysis goes beyond this literature by analyzing not only objective quality (performance markers) and bias variables (non-performance marker and team effects) but also comparing them to one another across voting systems to conclude their relative strength of influence on voting outcomes in these groups. To the best of our knowledge, the literature so far has not approached this area of research and has not taken advantage of voting behavior in all-star selections of American team sports in general.

In the National Football League[3], the term all-star team refers to a hypothetical team of players from the squads of different clubs, namely those players who have performed best throughout a given season. In the context of the NFL, two such teams are compiled through voting processes of different styles: the All-Pro Awards and the NFL Pro Bowl. Choosing this approach allows us to investigate the voting behavior of three different groups: (i) industry pundits (experts), (ii) insider professionals (internal peers; players, and coaches), and (iii) the general public (fans/audience). Thus, we add to the literature by providing more differentiated insights into the voting determinants of the different voting groups.

In an empirical analysis of data from the annual All-Pro Awards and NFL Pro Bowl votes, we tackle two research questions. First, do non-performance markers matter in all-star voting? Second, what are the differences between expert, peer, and public voting in all-star selection? We employ panel probit regressions to assess the impact of performance markers of every player's quality on the probability of the binary outcome of winning one of the mentioned awards. These performance markers are

[2] The positive effect of familiarity on voting behavior was previously studied by Zajonc (1968) and furthered by Harmon-Jones & Allen (2001), Bornstein (1989), and Verrier (2012). These mere exposure effects occur when subjects systematically prefer familiar entities (words, objects, journal papers, people). This familiarity effect has been shown to be robust across voting systems, from infants' food choices (Houston-Price et al., 2009) to public election voting behavior (Verhulst et al., 2010)

[3] With an average viewership of more than 17mio viewers across all 272 regular games in 2021 Associated Press (2022) and an estimated total revenue of more than US$12bn every year for the last five years (Forbes.com (2020)), the National Football League (NFL) is the main professional league of American Football. Furthermore, it ranks among the world's most popular and commercially successful sports leagues across disciplines.

touchdowns (rushing, passing, receiving), interceptions (thrown and caught), sacks (and outstanding defensive plays, so-called “solos”). We understand these performance markers to represent the “true” goals of the polls, i.e., an unbiased vote should follow these markers (although their respective weight may differ depending on which performance elements are valued most by the different voting groups). Additionally, we examine the influence of non-performance markers, which are specific to the individual, yet are independent of their performance. We introduce exposure (measured in years in the league), playtime (number of games played, number of games as a starter), as well as team performance (win-to-loss ratio, margin-of-victory), and general team-specific effects. All these markers do not directly relate to the performance of the *individual* player; hence, they should not influence his individual all-star probability. Thus, we interpret any significant influence of these factors on the voting results as voting bias. This enables us to assess the impact factors and possible biases as a deviation from the objective markers of quality in the respective voting groups.

The paper is structured as follows. Section 2 explains the two all-star votes in more detail and provides a conceptual framework for the analysis. The description of the data and the explanation of the employed methods follow in section 3 before section 4 reports the results of the empirical analysis according to the respective voting parties. Section 5 compares the results and shows the sizes of effects between groups. Section 6 concludes.

2. Structure and Issues of All-Pro and Pro Bowl Voting

2.1. All-Pro

The honor of an All-Pro designation is awarded to NFL players by press organizations. All-Pro teams are lists of players, consisting of the best player at his position in a given season, creating a hypothetical All-Star team. There is no physical game to be played by these All-Pro teams. They usually consist of 11 offensive players, 11 defensive players, and special teams (depending on the press organization). Most press organizations select a first team, listing the players which received the most votes, and a second team, consisting of the runner-ups. The first team and second team selections are considered All-Pro players for this study. We observe the voting behavior of eight press organizations over all seasons from 1951 until 2019: Associated Press, United Press International, Pro Football Writers, Pro Football Focus, Pro Football Weekly, The Sporting News, NY Daily News, Newspaper Enterprise Association.

The process of voting differs between the mentioned news organizations, yet it shares similarities that make it suitable for empirical analysis. The voting result is comprised of individual sports writers and other experts of the sport. Their votes are tallied and published by the news organization. By their characteristics they are qualified members of a board of examination, therefore, they match the literature's idea of an expert jury (Ginsburgh & van Ours, 2003). Selection to this jury and the fairness of this voting process are not without controversy, as some home markets are over- or underrepresented[4].

2.2. Pro Bowl

The Pro Bowl[5] is the NFL's official All-Star designation. It markedly differs from the All-Pro all-star team because the selected players participate in an actual physical contest: the Pro Bowl Game. All players that are selected for the Pro Bowl receive the title of being "Pro Bowler", even if they do not participate in the game. All players who received this title (independent of their actual participation in the game) are considered award winners for this paper. An increasing percentage of players selected for the Pro Bowl do not attend the Pro Bowl due to injuries, risk of injury, or Super Bowl preparations, allowing the Pro Bowl honor to be passed on to the next player.[6] In addition, top-tier players and their teams do not want to risk preparations for the following season by participating in an additional showcase game where the sporting result has no bearing on the national championship. Therefore, the larger number of spots and the existence of substitutes make the honor of an invitation to the Pro Bowl less exclusive than the All-Pro (Kunz-Kaltenhäuser, 2021).

2.3. Differences in Voting

There are major differences in the voting processes which allow us to draw conclusions about voting behavior. In voting for the All-Pros, groups of media members e.g., broadcasters and writers, sports journalists as well as other experts covering topics around the NFL cast their votes on the best player for any single position. These experts select an All-Pro team by voting on their perceived best player for a certain position in the season. Votes are tallied and published by the press organization at the end of the regular season (end of December/beginning of January). Each voter only gets one vote per position. We, therefore, use this to estimate expert voting behavior.

Voting for the Pro Bowl follows a different process than voting for the All-Pro. Until 1995, only players and coaches had voting rights (peer voting). In 1995, in an attempt to raise the legitimacy of the All-Star selections, the NFL introduced public voting (fan voting). From that year onwards, all three ballots, (1) players, (2) coaches, and (3) public, each count for one-third of the votes (NFL Players Association, 2020). For this paper, we follow the notion of Haan et al. (2005) to understand the voting behavior of the general public as anybody that takes an interest in the outcome of the vote and is willing to participate in the process. This helps us account for the fact that voters in the so-called fan voting process are

[4] The most prominent All-Pro team is published by Associated Press and is selected by 50 expert journalists across the U.S. Some cities with a NFL team do not have a vote in the balloting, e.g. Detroit (O'Hara (2013)); which is why All-Pro voting is not without criticism. Unfortunately, this is not a question that can empirically be examined with this data set as there is no systematic information on who is voting and what their associations and specific voting outcomes are.

[5] The official title of the game is AFC-NFC Pro Bowl.

[6] After 2009, not a single Pro Bowl has exceeded 75% of Original Ballot Pro Bowlers participation rate (Kunz-Kaltenhäuser, 2021).

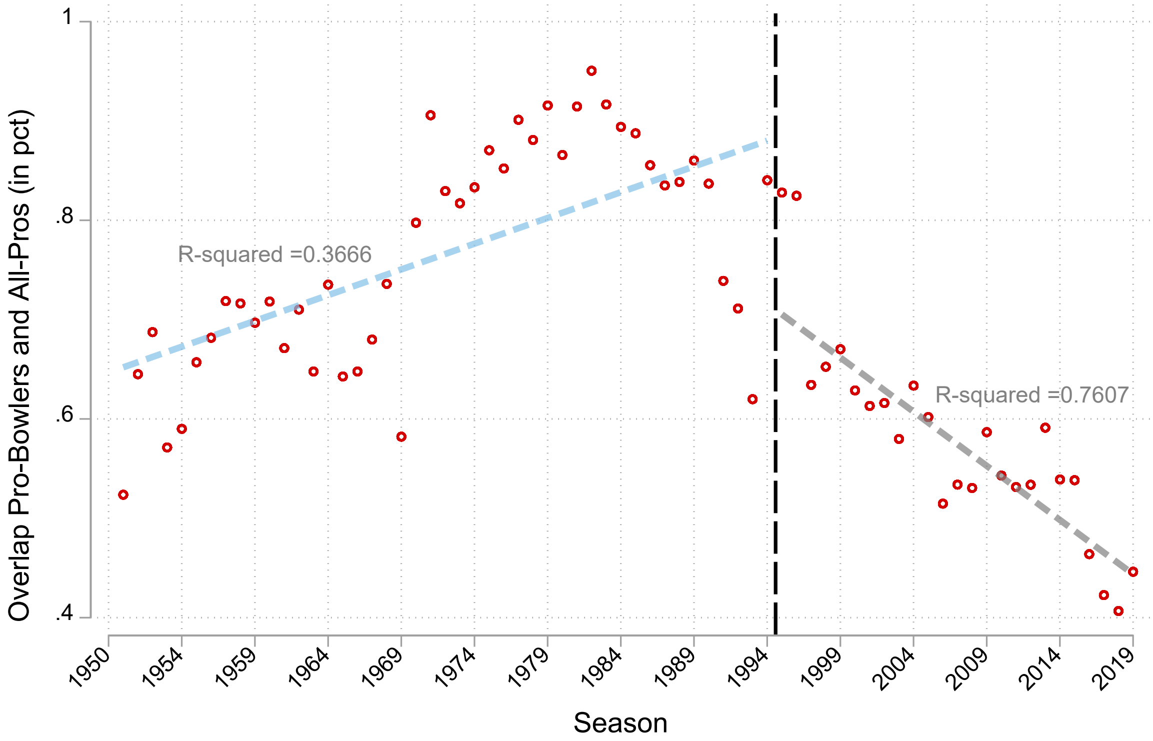
heterogeneous. They include casual viewers, hardcore fans of teams, supporters of single superstar players, general enthusiasts of the sport, and national and international viewers. Voting is open from mid-November until mid-December, and results are published by the NFL by the end of the regular season (end of December). We can therefore argue that the results of the voting in these two awards are independent of each other in any given year and do not influence one another. Table 1 shows an overview of the voting systems:

Table 1 Voting system comparison All-Pro versus Pro Bowl

<table>
<tr><th>Characteristic</th><th>All-Pro Awards</th><th colspan="2">Pro Bowl Awards</th></tr>
<tr><td>No. of players</td><td>11 Offense, 11 Defense,
Special Teams
(2nd Team)</td><td colspan="2">53 + Substitutes</td></tr>
<tr><td>Voting</td><td>1951-now:
Juries of specialists,
journalists, and other
experts</td><td>Until 1995:
Players and coaches</td><td>Since 1995:
1/3 Public
1/3 Players
1/3 Coaches</td></tr>
</table>

Figure 1 shows the overlap of the All-Pro and Pro Bowl players selected in any year. A value of 1 would indicate that every Pro Bowler also made it into an All-Pro team (100% overlap), whereas 0 would be interpreted as no intersection of athletes between the two awards. The introduction of public voting to Pro Bowl voting is marked as a vertical dotted line. In the first decades of the observed period, the overlap of players selected for both awards increased rather consistently. Starting with the introduction of fan voting in 1995, the voting results of the two awards started to diverge. The linear fitted regressions lead us to believe that there are systematic differences between the Pro Bowl selection with and without public voting included.

Figure 1 Overlap of players selected as All-Pro and Pro Bowlers [in percent]



Drawing from the literature and the structure of the voting systems, we expect to find differences in voting behavior between the groups. In expert voting, only performance markers are expected to be significant to the selection process. Due to the easier measurability of athletic performance compared to the assessment of e.g., artistic performance (as studied in the literature), we expect the assessment by experts to be more objective compared to other areas. By contrast, non-performance markers are expected to be insignificant for experts. In our study, we differentiate experts from the group of peers, which is novel to the literature. Both of these groups are qualified in the specific field, however, peers are actively involved in the sport, whereas experts are not. Peers are an internal part of the competition, and experts are external to the competition. Due to being internal, i.e., members of and/or affiliated with NFL teams, peers should be less objective and experience more incentives to deviate from pure performance-related voting behavior. Generally, peers may benefit from voting for their associates emotionally (sympathy votes). Although we cannot measure sympathy (or antipathy) in our data, we assume that this indirectly shows a lower relevance of players' performance markers. Moreover, affiliation effects of the insiders (peers) and, thus, a stronger influence of team success markers can be expected. Also, low-effort voting, i.e., just going with the perceived general consensus, should be visible in team success markers. While experts and the interested public are likely to be concerned about their voting choices, low-effort voting may be more prominent among peers.

The theory on public voting behavior predicts comparatively less objectivity for this group. Drawing from the literature, we expect to find some insignificant results for performance markers in public voting, with significant results for non-performance markers due to various biases. Here, a set of biases would be expected as compensation for the lack of expertise: mere exposure effects, familiarity effects, and team affiliation effects. We examine those by analyzing the influence of the number of games that players played (exposure), the number of previous awards (familiarity), and team success markers (affiliation effects). However, the voting behavior of sports fans may differ from audiences in music contests. Music audiences are comparatively less able to judge a performance accurately and therefore compensate by relying on secondary criteria/markers (see section 2). Yet, while audiences vote on (musical) talent, all-star voting is voting on athletic performance. The results of athletic performance are measurable (individual markers), broadly accessible, and easily comprehensible. To exacerbate this effect, fans of the sport may even gain an information advantage compared to some experts and peers by investing more in getting to know the statistics than experts and peers. Thus, the public may have considerably more expert knowledge than audiences in music competitions. Therefore, the gap in evaluation between experts and public opinion may decrease in All-star voting. Altogether, experts should be most suited to identifying objective athletic performance, peers less so due to deviating incentives, and fans even less due to a lack of specific knowledge.

3. Data and Model

3.1. Data and Descriptive Statistics

All data was provided by pro-football-reference.com. In our sample, we analyze the voting outcomes of Pro Bowl selections over 68 seasons (1951-2019).[7] We analyze the voting outcomes of All-Pro selections and Pro Bowl selections as balanced panel data on the player level to differentiate between the performance markers of a player and biases in voting.

The dependent variable is a binary outcome variable for each award. This variable is equal to 1 if a player is selected for this specific award in a season, and 0 if not. It is separated into two awards, so a player that is selected to the Pro Bowl is not necessarily selected to an All-Pro team and the other way around; however, players can be selected for both awards individually within one year. Since players can get multiple awards over many seasons, the number of observations is higher than the number of total players receiving awards.

Table 2 Overview number of awards

Data	All-Pros	Pro Bowlers	No Award
No. of Players	2,242	2,211	16,269
No. of Observations	5,969	6,184	66,630

[7] In 1951, the NFL's changed the format of its All-Star game considerably from a league champions vs. All-Stars game (1939-1942) to the conference-based All-Star game that it is today.

We generate three subsamples of this data set: players that made an All-Star team (expert voting), players that made the Pro Bowl before 1995 (peer voting), and players that made the Pro Bowl after 1995 (peer and public voting combined, see Tab. 3).

Table 3 Overview of subsamples per voting party

Expert (Jury) Voting	Peer Voting	Peer & Public Voting
All-Pro	Pro Bowl before 1995	Pro Bowl after 1995

The independent variables are (i) performance markers (individual performance in either offense or defense; positive/success and negative/mistakes) and (ii) non-performance markers (independent of athletic performance). The performance markers in our methodology consist of offensive as well as defensive markers. On the offensive side, we use touchdowns (rushing, passing, receiving) as well as offensive yards to capture the offensive quality of a player. On the defensive side, we use defensive touchdowns and yards, sacks, interceptions caught, and outstanding defensive plays and tackles, so-called solos. A player's probability to be selected for one (or both) award(s) should increase with an increase in these markers. Also included are objective errors by players (interceptions thrown, fumbles). A player's probability to win an award should *ceteris paribus* decrease with an increase in mistakes, as long as the accompanying positive marker is included in the model, e.g., controlling for rushing yards when examining the number of fumbles.

Bias variables are related to the player himself (non-performance markers), as well as the team (team success markers). To examine exposure effects and the effect of familiarity, we include the age of a player, the number of games in the season ($GamesInSeason$) as well as the number of times he has been named a starter ($GamesStarted$) in our model. The influence of past all-star selections on the probability of a player being selected is modeled by the number of previous awards awarded to a player in t-1, as captured in the numerical variable (*PreviousAwards*). Team performance effects of affiliation from the team that a player is on, the win-to-loss ratio ($WinToLoss$) as well as average margin of victory ($MarginofVictory$[8]) are included.

[8] Win-to-loss is the standard win-to-loss percentage commonly used in the NFL, normalized for values between 0 and 1. Average margin of victory of a team is calculated as follows: $Margin\ of\ Victory = \frac{Points\ Scored - Points\ Allowed}{Number\ of\ Games}$.

Table 4 Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
PassingTouchdowns	72599	.496	2.96	0	55
RushingTouchdowns	72599	.325	1.367	0	28
ReceivingTouchdowns	72599	.495	1.511	0	23
PassingYards	71820	26.083	276.628	-3	5477
RushingYards	72599	44.494	166.057	-33	2105
ReceivingYards	72599	80.332	203.378	-20	1964
DefensiveYards	72599	5.996	20.237	-14	376
InterceptionsThrown	72599	.435	2.335	0	35
Fumbles	72597	.603	1.607	0	23
DefensiveTouchdowns	72599	.034	.203	0	4
Sacks	54497	.813	2.06	0	22.5
Solos	54325	14.01	22.973	0	214
InterceptionsCaught	25984	1.216	1.661	0	14
Age	72599	26.524	3.295	20	48
revAwards	72599	.817	2.343	0	27
GamesInSeason	72599	12.501	4.133	0	17
GamesStarted	72564	6.854	6.229	0	16
WL	61925	1.824	36.125	0	1000
MoV	61925	.602	6.371	-20.4	19.7

To make comparative claims about the effect size across voting systems, we standardize all independent variables of interest. Since they vary widely on means and variances, we standardize the variables to their deviations (β-Coefficients; comparable to a z-Score). We rescale them to have a mean of zero and a standard deviation of one so that each observation's value indicates its difference from the mean of the original variable as the standard deviations. The value is then expressed in standard deviations above resp. below the mean, enabling us to compare coefficients across covariates and models (Newman & Browner, 1991). All coefficients for the variables of expressed interest that are reported in this paper are standardized and are interpreted as such.[9]

3.2. Model

We employ panel probit regressions to assess the impact of the aforementioned variables on the probability of the binary outcome of winning one of the awards. First, the influence of the aforementioned factors on the probability of winning an All-Pro award is examined (Models 1 & 2). $ALLPRO_{ij}$ is the probability of a player i in season j to win an All-Pro award. This allows for an estimation of what drives expert voting. We estimate panel probit regressions with fixed effects that take the following form.

$$\begin{aligned} ALLPRO_{ij} = {} & \alpha + \beta PerformanceMarkers_{ij} + \sigma NonPerformanceMarkers_{ij} \\ & + \gamma TeamSuccessMarkers_{ij} + P_i + T_i + S_j + \varepsilon_{ij} \end{aligned} \tag{1}$$

Our models include player-fixed effects P_i, team-fixed effects T_i, and season-fixed effects S_j. This enables us to assess the impact of influence factors and possible biases as a deviation from the

[9] Variables that cover season- and team-fixed effects are not standardized.

performance markers in the respective voting groups. We assume that ε_{ij} is a mean zero, constant variance random variable, while β, σ, γ are parameters to be estimated to examine the influence of $PerformanceMarkers_{ij}$ as well as non-performance markers $NonPerformanceMarkers_{ij}$ and $TeamSuccessMarkers_{ij}$.

Secondly, the same model is estimated for the probability of being selected for the Pro Bowl, prior to 1995 (Models 3 & 4). This gives us an estimation of what drives peer voting outcomes. Analog to the first estimation model (*ALLPRO*), the second dependent variable $PROBOWL_{ij}$ is a binary outcome variable of whether a player i is selected for the Pro Bowl Team or not.

$$\begin{aligned} PROBOWL_{ij} = \alpha + {} & \beta PerformanceMarkers_{ij} \\ & + \sigma NonPerformanceMarkers_{ij} + \gamma TeamSuccessMarkers_{ij} \\ & + P_i + T_i + S_j + \varepsilon_{ij} \end{aligned} \tag{2}$$

Thirdly, the impact of fan voting on Pro Bowl voting outcomes is assessed as the difference in the influence of covariates before and after 1995. We create interaction terms between the inclusion of fan voting and the covariates to estimate their influence in the before and after periods. This enables us to estimate λ, θ, δ. This enables conclusions on how the results in Pro Bowl voting changed through the inclusion of public voting. The margins of these models are calculated to allow for a discussion of the differences (Models 5 & 6).

$$\begin{aligned} PROBOWL_{ij} = \alpha + {} & \beta PerformanceMarkers_{ij} \\ & + \sigma NonPerformanceMarkers_{ij} + \gamma TeamSuccessMarkers_{ij} \\ & + \delta PublicVoting_i * PerformanceMarkers_{ij} \\ & + \lambda PublicVoting_i * NonPerformanceMarkers_{ij} \\ & + \theta PublicVoting * TeamSuccessMarkers_{ij} + P_i + T_i + S_j \\ & + \varepsilon_{ij} \end{aligned} \tag{3}$$

4. Empirical Results

4.1. Expert Voting (All-Pro)

This section is sorted by the voting parties (experts, peers, and public/fans) and displays the results of the different regression estimations respectively. It gives an overview of the results, however, the in-depth comparison between voting groups and effect sizes follows in Section 5. Models (1) and (2) show the estimated coefficients for expert voting; the dependent variable being the probability of gaining an All-Pro award. We use touchdowns in model (1) and yards in model (2) as independent variables because touchdowns and yards are positively correlated, esp. of the same kind (e.g., $r_{PassingYards,PassingTouchdowns} = 0.72$; see variance inflation factors and further robustness checks in chapter 4.4). Hence, we estimate separate models. All types of touchdowns, as well as all yards, are positive and significant. Therefore, offensive performance markers are found to be positive on the probability to win an All-Pro award. The effect size is rather small (see the comparison in Figure 3). Defensive performance markers are ambiguous: the effect of sacks is positive (0.320***), yet defensive touchdowns are insignificant and outstanding defensive plays (solos) are significantly negative for the probability to win in expert voting -0.149***). Other defensive indicators are insignificant or show very low robustness (e.g., *DefensiveYards*).

Table 5 Results probit regression estimations: Expert voting

VARIABLES	(1) All-Pro (touchdowns)	(2) All-Pro (yards)
PassingTouchdowns	0.172***	
	(0.0336)	
RushingTouchdowns	0.217***	
	(0.0250)	
ReceivingTouchdowns	0.313***	
	(0.0272)	
PassingYards		0.0960***
		(0.0330)
RushingYards		0.230***
		(0.0255)
ReceivingYards		0.396***
		(0.0333)
InterceptionsThrown	-0.283***	-0.201***
	(0.0614)	(0.0778)
Fumbles	0.145***	0.0884**
	(0.0403)	(0.0409)
DefensiveTouchdowns	0.0144	
	(0.0112)	
DefensiveYards		0.0317*
		(0.0184)
Sacks	0.320***	0.338***
	(0.0205)	(0.0213)
Solos	-0.146***	-0.115***
	(0.0327)	(0.0334)
InterceptionsCaught	0.118***	0.109***
	(0.0331)	(0.0407)
Age	0.0629*	0.0908***
	(0.0332)	(0.0345)
GamesInSeason	0.766***	0.749***
	(0.0618)	(0.0621)
GamesStarted	0.322***	0.290***
	(0.0365)	(0.0378)
PrevAwards	-0.0160	-0.0301
	(0.0254)	(0.0260)
Win-To-Loss	0.251***	0.250***
	(0.0476)	(0.0480)

Margin-of-Victory	0.129***	0.151***
	(0.0495)	(0.0498)
TEAM-FE	YES	YES
SEASON-FE	YES	YES
Constant	0.0599	0.0648
	(0.218)	(0.220)
Pseudo R^2	0.4701	0.4794
Observations	16,405	16,405
Number of Players	4,835	4,835

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Non-performance markers are also significant: age significantly increases the outcome probability (0.0629*). More participation significantly raises the probability to be selected by expert voting, as the number of games in a season and the number of games a player started has a robust significant positive effect (0.766***; 0.322***). The number of previous awards shows no significant influence on the probability to be selected by expert voting. We also find team success markers significant to expert voting: the margin-of-victory of a team that a player is on significantly increases his probability to be selected by expert voting; the same is not true for the win-to-loss ratio. The results are largely robust in size and significance across models. We, therefore, reject the notion that experts' decision is not influenced by non-performance markers.

4.2. Peer Voting (Pro Bowl)

Results from peer voting show our estimations regarding the Pro Bowl award. The following models (3) and (4) analyze Pro Bowl voting before the introduction of public voting in 1995 (see (II) above); again, touchdowns and yards are estimated in different models. For the purpose of this study, we use this as an estimation of peer voting behavior, as coaches and players are voting on their peers. The coefficients for individual performance markers (touchdowns, yards, etc.) are positive and significant, with a larger effect size (compared to expert voting; models 1 & 2). Negative individual performance (*Fumbles*) shows a low robustness and positive significance. Dropping the football while running has a positive impact on a player's probability to be selected (0.184***), even when controlling for the overall production that the player had as a runner. However, we attribute this to our model controlling for the majority of objective performance markers. The attributed negative coefficient is measuring residual effects, comparable to the negative number of rooms in the housing market, when controlling for square meters (inter alia, Wittowsky et al., 2020).

Defensive performance markers have an ambiguous influence: the coefficients for defensive touchdowns and yards are insignificant but solos and interceptions caught are negative with high robustness (-0.201***). Age does not show a robust significant negative impact. However, the number of previous awards displays a robust significant positive impact (0.317***; 0.293***) on the probability of being selected for the Pro Bowl. Team performance effects are found in peer voting: the win-to-loss ratio, as well as the margin-of-victory, have a robust positive significant effect on a player's probability to be voted for (0.170**; 0.179**). This leads us to believe that peers' decisions are, in fact, influenced by non-performance markers. The pseudo-R2 is higher in this model than in the other models, which is most likely due to the smaller number of observations (n = 4,011 versus n= 16,405 in models (4) & (5)); yet it is still at a reasonable level to induce statistical significance.

Table 6 Results probit regression estimations: Peer voting

VARIABLES	(3) Pro Bowl Pre95 (touchdowns)	(4) Pro Bowl Pre95 (yards)
PassingTouchdowns	0.488***	
	(0.186)	
RushingTouchdowns	0.155***	
	(0.0575)	
ReceivingTouchdowns	0.462***	
	(0.0630)	
PassingYards		0.553***
		(0.146)
RushingYards		0.303***
		(0.0552)
ReceivingYards		0.570***
		(0.0642)
InterceptionsThrown	-0.316*	-0.613***
	(0.188)	(0.223)
Fumbles	0.184***	0.00835
	(0.0605)	(0.0683)
DefensiveTouchdowns	0.0219	
	(0.0200)	
DefensiveYards		0.0506
		(0.0327)
Sacks	0.197***	0.226***
	(0.0287)	(0.0308)
Solos	-0.0632*	-0.0337
	(0.0341)	(0.0351)
InterceptionsCaught	-0.201***	-0.193**
	(0.0647)	(0.0755)
Age	-0.102*	-0.0446
	(0.0600)	(0.0640)
GamesInSeason	0.438***	0.453***

	(0.0974)	(0.103)
GamesStarted	0.479***	0.432***
	(0.0644)	(0.0667)
PrevAwards	0.317***	0.293***
	(0.0443)	(0.0479)
Win-To-Loss	0.170**	0.179**
	(0.0811)	(0.0837)
Margin-of-Victory	0.179**	0.194**
	(0.0868)	(0.0895)
TEAM-FE	YES	YES
SEASON-FE	YES	YES
Constant	-0.182	-0.258
	(0.282)	(0.304)
Pseudo R^2	0.7349	0.7430
Observations	4,011	4,011
Number of Players	1,485	1,485

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

4.3. Public Voting (Pro Bowl)

Unfortunately, data for public voting is not available individually since we only have the voting results combining peer and public voting from the Pro Bowl post-1995. Thus, we must identify the public voting influence from the comparative estimation of public voting into the Pro Bowl voting system. Therefore, this section shows the results for pre and post-1995 Pro Bowl votes with an interaction term to interpret the effect of public voting. Models (5) and (6) estimate the β_i's for Pro Bowl voting, just like Models (3) and (4). The difference between (3)/(4) and (5)/(6) is that in the latter we introduce the interaction terms to the regression as explained in section 3.2. To this end, we implemented a dummy variable of public voting that changes value from 0 to 1 with the introduction of fan voting. Table 7 depicts the estimated coefficients for Pro Bowl voting after the inclusion of the public voting process, showing only the statistical contrast caused by fan voting in Pro Bowl voting. It is interpreted as the difference with the inclusion of public voting, therefore reflecting the preferences of the general public. The coefficients of performance markers are ambiguous since most types of offensive touchdowns and yards lose their significance compared to the other models. Some defensive performance markers are negatively influential to the selection of players to the Pro Bowl (*Solos* = -0.451***), whereas others like defensive touchdowns and yards are insignificant. The coefficients for the number of games in a season (*GamesInSeason*) are positive (0.286**; 0.262**), and the coefficients for games as a starter (*GamesStarted*) are insignificant. Therefore, we find some exposure effects in public voting. The number of previous awards has a negative influence (-0.248***). The coefficients for win-to-loss and margin-of-victory are insignificant, hence the influence of team success factors is found to be weaker compared to models (1)/(2). This is likely due to the team-fixed effects covering more of the explanatory power than the objective team performance (for further examination, see chapter 5). This fits the narrative that fans voting for their teams irrespective of their performance for reasons of fandom are prevalent in Pro Bowl voting and account for a significant part of the result. In general, models (5) and (6) show comparatively less significant and less robust results than the other models.

Table 7 Results probit regression estimations: Public voting

VARIABLES	(5) Pro Bowl All Seasons (touchdowns)	(6) Pro Bowl All Seasons (yards)
PublicVoting = 1	-1.297***	-1.338***
	(0.208)	(0.221)
PassingTouchdowns#PublicVoting	-0.151	
	(0.242)	
RushingTouchdowns#PublicVoting	0.0493	
	(0.0877)	
ReceivingTouchdowns#PublicVoting	0.00637	
	(0.0843)	
PassingYards#PublicVoting		-0.345**
		(0.172)
RushingYards#PublicVoting		-0.1000
		(0.0698)
ReceivingYards#PublicVoting		0.0267
		(0.0818)
InterceptionsThrown#PublicVoting	0.218	0.601**
	(0.265)	(0.276)
Fumbles#PublicVoting	0.200**	0.253**
	(0.101)	(0.107)
DefensiveTouchdowns#PublicVoting	-0.00141	
	(0.0258)	
DefensiveYards#PublicVoting		-0.00630
		(0.0421)
Sacks#PublicVoting	0.0558	0.0506
	(0.0346)	(0.0359)
Solos#PublicVoting	-0.451***	-0.436***
	(0.0609)	(0.0628)
InterceptionsCaught#PublicVoting	0.451***	0.412***
	(0.0743)	(0.0878)
Age#PublicVoting	0.222***	0.185**
	(0.0744)	(0.0764)
GamesInSeason#PublicVoting	0.286**	0.262**
	(0.125)	(0.130)
GamesStarted#PublicVoting	0.0649	0.0932

	(0.0819)	(0.0850)
PrevAwards#PublicVoting	-0.248***	-0.249***
	(0.0523)	(0.0538)
Win-To-Loss#PublicVoting	-0.0172	-0.0289
	(0.106)	(0.109)
Margin-of-Victory#PublicVoting	-0.103	-0.0896
	(0.111)	(0.114)
TEAM-FE	YES	YES
SEASON-FE	YES	YES
Constant	-0.551**	-0.675***
	(0.239)	(0.255)
Pseudo R^2	0.5127	0.5210
Observations	16,405	16,405
Number of Players	4,835	4,835

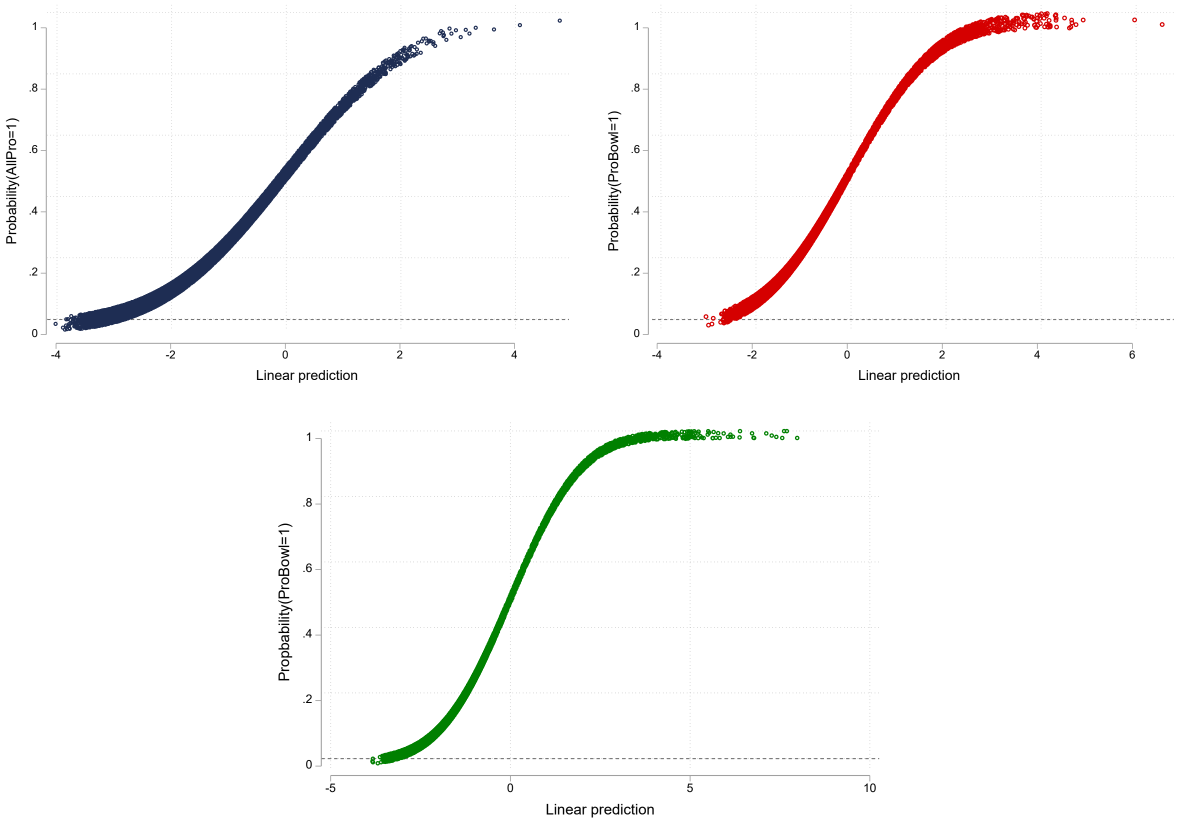
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

4.4. Robustness Checks

We estimate separate models for touchdowns and yards but include biased variables and controls in both. The suggested correlation between yards and touchdowns ($r_{PassingYards,PassingTouchdowns} = 0.51$; $r_{RushingYards,RushingTouchdowns} = 0.45$; $r_{ReceivingYards,ReceivingTouchdowns} = 0.86$; $r_{DefensiveYards,DefensiveTouchdowns} = 0.86$) would raise concerns of multicollinearity. We check for overinflation of variance. E.g., the number of games that a player plays and his status as a starter might be suspected to be multicollinear to his performance. However, there are no variables in the data set that show high multicollinearity. Mean Variance Inflation Factors are below $vif = 3.0$ with $\text{mean vif} = 1.92$ for models (1), (3) & (5) and $\text{mean vif} = 2.096$ for models (2), (4) & (6) respectively. As this is well below the accepted econometric standard value, we do not support the concern of multicollinearity in the model. We have verified the absence of strong collinearity of our independent variables (see vif-values in the appendix).

Figure 2 shows the deviance residuals for the estimated models (1), (3) & (5). The deviance of the models is computed as the (negative double) logarithm of the difference between a data point's predicted probability and the complement of its actual value. Therefore, a small deviance residual indicates a well-fitting data point, and a large value a poorly fitting data point. The residuals for model (1) are symmetrically distributed and closely grouped around 0. For model (3), the residuals show a slightly larger deviance, especially on the positive outcome ($Pr = 1$). Model (5) has the least robust fit when considering this metric, and shows some signs of heteroscedasticity. Yet, since the overall log-likelihood of the model is within the accepted standard ($Pseudo\ R^2 = 0.5127$) and therefore the overall fit of the model is fungible, we use it to draw comparisons.

Figure 2 Deviance Residuals for Models (1), (3) & (5)



It may be suspected that the individual-specific effects are correlated with the independent variables. A Hausman specification test confirmed our approach by showing significant differences between the random effects estimator and fixed effects estimator ($p = 0.000$), therefore the random effects assumption is rejected (Hausman, 1978) and a fixed effects model is chosen. For further robustness checks, we calculated the coefficients with position fixed-effects instead of player fixed-effects. The results are displayed in Tables 12-14 in the appendix. They support our claims by providing robust coefficients. This would be expected as players mostly keep their position over the course of their NFL career.[10]

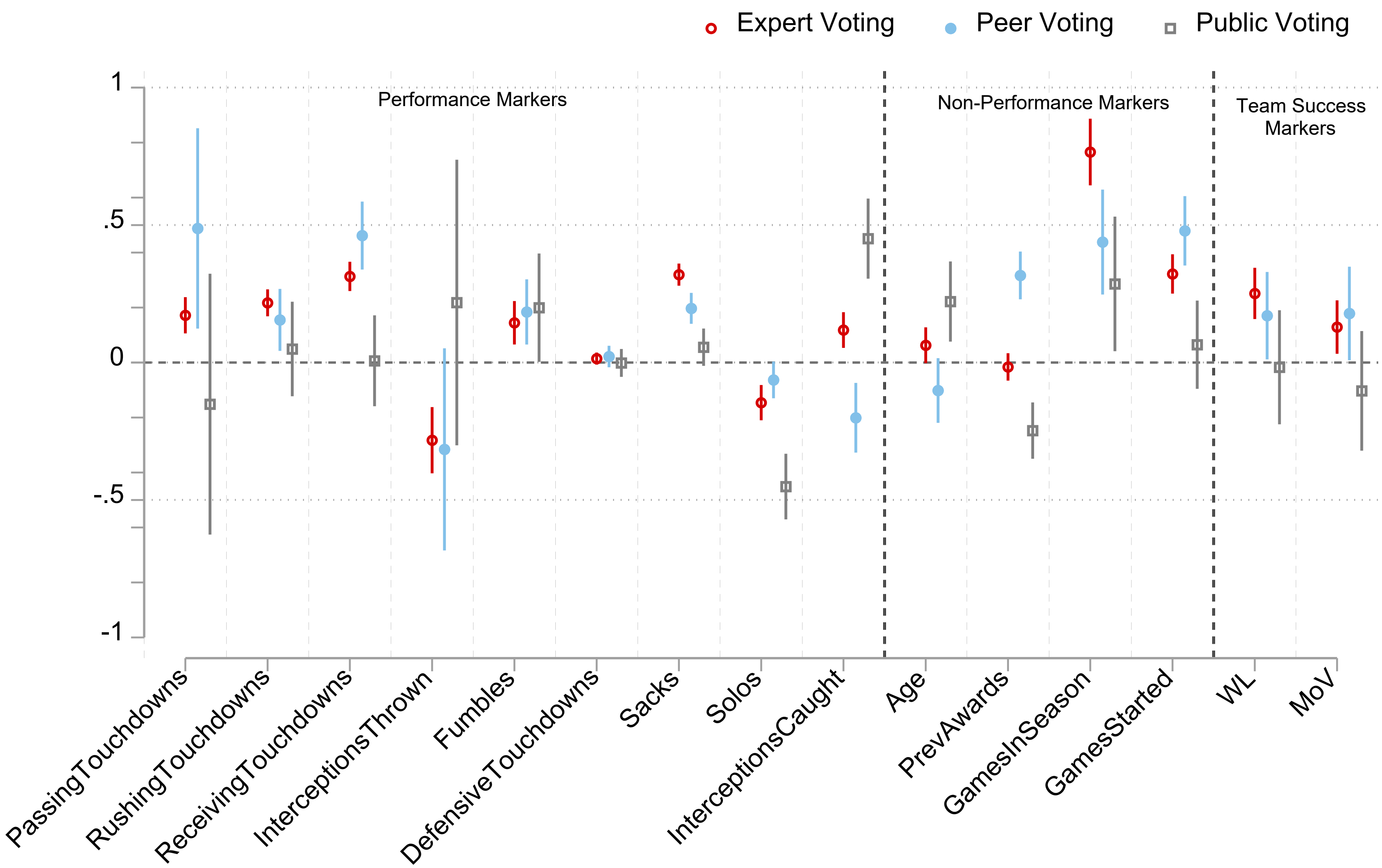
5. Interpretation and Discussion

Section 4 displayed the results for all voting parties individually. However, the most interesting conclusions can be drawn from the comparison of the groups. This section illustrates the differences between the groups. The standardization of independent variables as explained in section 3.1 allows for

[10] Out of the 9843 players that we have data on their position in the data set, 669 (~6.8%) changed positional groups over the course of their career (e.g., going from linebacker to defensive back).

a direct comparison of the results. The following graph plots the coefficients for expert, peer, and public voting (models 1;3;5). For legibility, we include common labels for performance markers, non-performance markers, and team-success-effects. All box-whiskers that are distinctly different from zero indicate a 95% confidence interval of the coefficient that is distinctly different from zero and therefore significant within the model. The independent variables are grouped for experts/peers/fans to show a direct comparison of the results.

Figure 3 Box-Whisker-Plot of beta-coefficients, Models (1), (3), and (5).



1. *Performance markers:* overall, positive performance markers seem to influence the probability of becoming an All-Star positively. This effect is stronger for peers than for experts. Interestingly, the results for public voting are not significant and seem to scatter in comparison to the other groups. The general public does not seem to vote as homogenously as peers and experts (results for the 95% confidence interval get very wide, e.g., for passing touchdowns). Apart from being a more heterogenous group anyway, this might be due to fans sticking to their favorite team (and its best players) rather than taking objectively the best overall player into consideration (see team-related effects). Through all three groups, defensive performance markers are less valued than offensive ones. On the one hand, this may not be surprising given that offensive actions may be more exciting, more eye-catching, and, image-wise, more flamboyant than the naturally more destructive defensive performance. However, also experts and peers – who should know and value the importance of defensive action – do not value defensive achievements as much as offensive ones, although they perform a little better than public voting here. Negative performance markers display strongly differing effects; some are insignificant while others counterintuitively display a positive effect[11].
2. *Non-performance markers:* Non-performance markers can be interpreted as specific bias, since they do not address the individual quality of a player, but non-performance related aspects. We find that non-performance markers (such as the age of the player and previous awards) have comparably similar effects on the results than performance markers across voting systems. There are significant results showing that some markers have an influence, e.g., the age of players has a positive influence on expert and peer voting. For the peers and experts, games in the season and games as a starter are positively significant, indicating preferences for players with more playtime and, *ceteris paribus*, starter status. This is in line with the findings of Coleman et al. (2010) who also found positive effects of exposure on voting results in expert voting. Interestingly, for peers, previous awards are also significantly positive. The fame and recognition among their own peers seem to be reflected in these results, whereas the effect on general public opinion is contrary to that. In public voting, previous awards negatively influence the evaluation of a player. Experts are not influenced by the recognition of previous awards.
3. *Team-success markers:* Team success factors (win-to-loss and margin-of-victory) are found to be relevant in expert and peer voting. Especially experts seem to be driven by the success of the overall team, which contradicts our expectations of objective expert voting. Interestingly, team-success markers not influencing public voting fit well with the theory-driven predictions from section 2.3: fans may be driven by individual club favoritism rather than by team success. Therefore, team-specific effects may overshadow the explanatory power of these covariates. To assess the impact of the team-specific effects and their influence on public voting, we utilize

[11] This is probably measuring residual effects after the other coefficients account for the positive influence of other markers, as explained in section 4.2.

several post estimations: a likelihood-ratio test was conducted to assess the goodness of fit of the empirical models based on their ratio of likelihood. Constraining the team fixed-effects to zero lowers the fit of the model for public voting significantly; the hypothesis that the data is equally likely under the full model and the restricted model is rejected with $p = 0.0171$. Adding the team effects increases the explanatory power of the model, the data is 49.9 times more likely ($LR\ X^2 = 49.90$) under the unconstrained model. For peer voting (Pro Bowl PRE95), by contrast, the hypothesis of the likelihood ratio test was not rejected ($p = 0.3670$; $LR\ \mathrm{X}^2 = 33.05$). There is no significant increase in the explanatory power of the model by including team-fixed effects in peer voting. These two findings give a strong indication that there is a relationship between the team that a player plays for in any given season and the public voting outcome. The same effect did not exist in Pro Bowl voting prior to the introduction of public fan voting. The results of a likelihood-ratio test indicate that the statement is valid. Furthermore, choosing the most popular team by social media numbers 2022 (Dallas Cowboys)[12] as a base for the team-fixed effects gives negative significant coefficients to the team-fixed effects.[13] The likelihood-ratio tests and the base group analysis show significant effects of the team affiliation. Thus, we find team effects in public voting that are related to the "brand" of a team, not their current sporting success.[14] Overall, in expert and peer voting, we find significant biases towards players of successful teams; while for fans not the overall success of the team, but the individual favorite team seems to matter, i.e. they are not primarily voting for players of the strongest team, but for players of their home/favorite team. At the same time, peers do not seem to bias their votes according to their team affiliation but, like experts, are rather influenced by team success.

6. Conclusion

Do non-performance markers matter in all-star voting? What are the differences between expert, peer, and public voting in all-star selections? This paper set out to answer research questions of differences in voting behavior between those groups by analyzing outcomes for all-star teams in American Football. To answer the first research question, all groups are susceptible to biases and non-performance-driven qualities. None of the analyzed voting systems is free from the influence of non-performance markers. Therefore, we can answer the first research question and post that non-performance markers do matter in all-star voting.

[12] According to an auxiliary data set that contains social media data on NFL teams, as well as Molski (2022).
[13] For example, players on the New York Jets or (then) Oakland Raiders had a significantly lower chance to be selected by fans, ceteris paribus.
[14] For Expert voting: reject likelihood ratio test p=0.0708; LR X^2=46.79

Concerning the differences, we find that expert voting, peer voting, and public voting share some similarities, yet differ systematically. There are in fact differences in the outcomes produced by the three voting systems. All three groups seem to value the quality of players since a lot of performance markers were found to influence results significantly and positively. For public fan voting, the majority of performance markers are found to be insignificant or fail to show robust significant effects. While the general public does not vote as homogenously as experts and peers, the latter two value especially positive performance markers and, thus, successful athletic performance.

The results of this study lead us to be doubtful about the suspected objectiveness of expert voting systems. Their assessments are found to be influenced by non-performance markers and affiliation with success. Furthermore, public voting is not consistently found to be influenced by the non-performance variables that were included in the model. The results fortify the notion of public voting being inefficient at identifying objective quality. Neither objective performance nor previous achievements are the driving forces, but rather the general "brand" of a team that a player is associated with seems to be driving the outcome. In this case, the specific affiliation seems to impact the assessment of the individual. This indication of partisanship voting may be used to draw implications for the usage of public voting systems when affiliations are involved, e.g., political parties, teams, and companies.

References

Associated Press (2022). NFL regular-season ratings increase 10% over last season. Retrieved from https://www.espn.com/nfl/story/_/id/33050695/nfl-regular-season-ratings-increase-10-percent-last-season

Bornstein, R. F. (1989). Exposure and affect: Overview and meta-analysis of research, 1968–1987. *Psychological Bulletin*, *106*(2), 265–289.

Budzinski, O., & Pannicke, J. (2017). Culturally biased voting in the Eurovision Song Contest: Do national contests differ? *Journal of Cultural Economics*, *41*(4), 343–378.

Budzinski, O., & Pannicke, J. (2022). Does popularity matter in a TV song competition? Evidence from a national music contest. *Creative Industries Journal.* Advance online publication.

Campbell, B., & Galbraith, J. W. (1996). Nonparametric Tests of the Unbiasedness of Olympic Figure-Skating Judgments. *The Statistician*, *45*(4), 521.

Campbell, N. D., Rogers, T. M., & Finney, R. Z. (2007). Evidence of Television Exposure Effects in AP Top 25 College Football Rankings. *Journal of Sports Economics*, *8*(4), 425–434.

Coleman, B. J., Gallo, A., Mason, P. M., & Steagall, J. W. (2010). Voter Bias in the Associated Press College Football Poll. *Journal of Sports Economics*, *11*(4), 397–417.

Emerson, J. W., Seltzer, M., & Lin, D. (2009). Assessing Judging Bias: An Example From the 2000 Olympic Games. *The American Statistician*, *63*(2), 124–131.

Flôres, R. G., & Ginsburgh, V. (1996). The Queen Elisabeth musical competition: How fair is the final ranking? *Journal of the Royal Statistical Society*, *45*(1), 97–104.

Forbes.com (2020). Sports Money: 2020 NFL Valuations. Retrieved from https://www.forbes.com/nfl-valuations/list/

Ginsburgh, V., & Noury, A. G. (2008). The Eurovision Song Contest. Is voting political or cultural? *European Journal of Political Economy*, *24*(1), 41–52.

Ginsburgh, V., & van Ours, J. C. (2003). Expert Opinion and Compensation: Evidence from a Musical Competition. *American Economic Review*, *93*(1), 289–296.

Glejser, H., & Heyndels, B. (2001). Efficiency and Inefficiency in the Ranking in Competitions: the Case of the Queen Elisabeth Music Contest. *Journal of Cultural Economics*, *25*(2), 109–129.

Haan, M. A., Dijkstra, S. G., & Dijkstra, P. T. (2005). Expert Judgment Versus Public Opinion ? Evidence from the Eurovision Song Contest. *Journal of Cultural Economics*, *29*(1), 59–78.

Harmon-Jones, E., & Allen, J. J. B. (2001). The Role of Affect in the Mere Exposure Effect: Evidence from Psychophysiological and Individual Differences Approaches. *Personality and Social Psychology Bulletin*, *27*(7), 889–898.

Hausman, J. A. (1978). Specification Tests in Econometrics. *Econometrica*, *46*(6), 1251.

Heiniger, S., & Mercier, H. (2021). Judging the judges: evaluating the accuracy and national bias of international gymnastics judges. *Journal of Quantitative Analysis in Sports*, *17*(4), 289–305.

Houston-Price, C., Burton, E., Hickinson, R., Inett, J., Moore, E., Salmon, K., & Shiba, P. (2009). Picture book exposure elicits positive visual preferences in toddlers. *Journal of Experimental Child Psychology*, *104*(1), 89–104.

Johnson, M. R., & McCarthy, I. P. (2022). Cultural similarity and impartiality on voting bias: The case of FIFA's World's Best Male Football Player Award. *PloS One*, *17*(7), e0270546.

Kunz-Kaltenhäuser, P. (2021). Is the NFL's Pro Bowl broken? Considering the players' perspective. (Ilmenau Economics Discussion Papers).

Logan, T. (2007). Whoa, Nellie! Empirical Tests of College Football & Conventional Wisdom NBER Working Paper No. w13596. (National Bureau of Economic Research).

Lyngstad, T. H., Härkönen, J., & Rønneberg, L. T. S. (2020). Nationalistic bias in sport performance evaluations: An example from the ski jumping world cup. *European Journal for Sport and Society*, *17*(3), 250–264.

Newman, T. B., & Browner, W. S. (1991). In defense of standardized regression coefficients. *Epidemiology (Cambridge, Mass.)*, *2*(5), 383–386.

NFL Players Association (2020). Collective Bargaining Agreement. Retrieved from https://nflpaweb.blob.core.windows.net/media/Default/NFLPA/CBA2020/NFL-NFLPA_CBA_March_5_2020.pdf

O'Hara, M. (2013). O'Hara's Monday Countdown: Why Calvin Johnson wasn't a unanimous All-Pro selection. Retrieved from http://www.detroitlions.com/news/article-1/OHaras-Monday-Countdown-Why-Calvin-Johnson-wasnt-a-unanimous-All-Pro-selection/5e8873fb-7fac-4b88-a8a3-679b8e9905e4

Pope, B. R., & Pope, N. G. (2015). Own-Nationality Bias: Evidence from the UEFA Champions League Football Referees. *Economic Inquiry*, *53*(2), 1292–1304.

Popović, R. (2000). International Bias detected in judging Rhythmic Gymnastics Competition at Sydney 2000 Olympic Games. *Physical Education and Sport*, *1*(7), 1–13.

Scholten, H., Schneemann, S., & Deutscher, C. (2020). The Impact of Age on Nationality Bias and Cultural Proximity Bias: Evidence from Ski-Jumping. *Journal of Institutional and Theoretical Economics*, *176*(4), 708.

Spierdijk, L., & Vellekoop, M. (2009). The structure of bias in peer voting systems: lessons from the Eurovision Song Contest. *Empirical Economics*, *36*(2), 403–425.

Verhulst, B., Lodge, M., & Lavine, H. (2010). The Attractiveness Halo: Why Some Candidates are Perceived More Favorably than Others. *Journal of Nonverbal Behavior*, *34*(2), 111–117.

Verrier, D. (2012). Evidence for the influence of the mere-exposure effect on voting in the Eurovision Song Contest. *Judgement and Decision Making*, *7*(5), 639–643.

Walther, E., Fiedler, K., Horn, K., & Zembrod, A. (2002). To play championship soccer by flowing with the ball or to go down by trying to hard: Constructive illusions of memory in soccer experts and novices. *Psychologie Und Sport*, *9*(4), 159–171.

Wittowsky, D., Hoekveld, J., Welsch, J., & Steier, M. (2020). Residential housing prices: impact of housing characteristics, accessibility and neighbouring apartments – a case study of Dortmund, Germany. *Urban, Planning and Transport Research*, *8*(1), 44–70.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, *9*(2, Pt.2), 1–27.

Zitzewitz, E. (2006). Nationalism in Winter Sports Judging and Its Lessons for Organizational Decision Making. *Journal of Economics & Management Strategy*, *15*(1), 67–99.

Appendix

Table 8 Variance Inflation factors

	VIF (Model 1, 3 & 5)	VIF (Model 2, 4 & 6)
PassingTouchdowns	5.63	-
RushingTouchdowns	1.519	-
ReceivingTouchdowns	1.227	-
PassingYards	-	8.488
RushingYards	-	1.615
ReceivingYards	-	1.317
InterceptionsThrown	5.834	8.174
Fumbles	1.434	2.993
DefensiveTouchdowns	1.202	-
DefensiveYards	-	2.289
Sacks	1.211	1.216
Solos	1.504	1.524
InterceptionsCaught	1.37	2.521
PrevAwards	1.44	1.45
Age	1.092	1.094
GamesInSeason	1.255	1.256
GamesStarted	1.685	1.714
Win-To-Loss	1.042	1.04
Margin-of-Victory	1.18	1.177
Mean	1.924	2.096

Table 9, correlations

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
(1) PassingTouchd~s	1.000																		
(2) RushingTouchd~s	0.140	1.000																	
(3) ReceivingTouc~s	-0.023	0.104	1.000																
(4) PassingYards	0.979	0.142	-0.024	1.000															
(5) RushingYards	0.104	0.925	0.114	0.101	1.000														
(6) ReceivingYards	-0.025	0.182	0.916	-0.028	0.205	1.000													
(7) DefensiveYards	-0.064	-0.074	-0.094	-0.065	-0.077	-0.104	1.000												
(8) Interceptions~n	0.902	0.131	-0.022	0.935	0.094	-0.025	-0.062	1.000											

(9) Fumbles	0.641	0.496	0.165	0.662	0.502	0.220	-0.063	0.654	1.000										
(10) DefensiveTou~s	-0.037	-0.042	-0.055	-0.038	-0.044	-0.062	0.641	-0.036	-0.040	1.000									
(11) Sacks	-0.069	-0.080	-0.105	-0.070	-0.085	-0.119	-0.100	-0.067	-0.122	-0.019	1.000								
(12) Solos	-0.136	-0.161	-0.204	-0.139	-0.169	-0.227	0.229	-0.133	-0.206	0.148	0.209	1.000							
(13) Interception~t	-0.088	-0.105	-0.130	-0.090	-0.110	-0.144	0.749	-0.086	-0.095	0.404	-0.146	0.285	1.000						
(14) Age	0.118	-0.038	0.015	0.119	-0.045	0.009	-0.031	0.101	0.025	-0.013	0.051	-0.028	-0.028	1.000					
(15) PrevAwards	0.134	0.071	0.140	0.129	0.073	0.148	-0.026	0.111	0.106	-0.001	0.103	-0.063	-0.042	0.485	1.000				
(16) GamesInSeason	0.048	0.055	0.082	0.049	0.059	0.093	0.084	0.046	0.080	0.060	0.160	0.251	0.096	0.054	0.048	1.000			
(17) GamesStarted	0.109	0.110	0.143	0.112	0.113	0.158	0.156	0.105	0.119	0.094	0.233	0.379	0.201	0.200	0.236	0.420	1.000		
(18) WL	0.022	-0.002	0.031	0.014	-0.004	0.016	0.002	0.005	0.006	0.007	0.008	0.005	0.001	0.017	0.022	0.007	0.010	1.000	
(19) MoV	0.103	0.061	0.080	0.096	0.048	0.069	0.046	0.081	0.082	0.043	0.035	-0.057	0.039	0.058	0.081	0.079	0.053	0.106	1.000

Table 10 Expert Voting with Position Fixed-effects

VARIABLES	(1) All-Pro Yards wPosition	(1) All-Pro TD wPosition
PassingTouchdowns		0.0568***
		(0.0138)
RushingTouchdowns		0.0525***
		(0.0168)
ReceivingTouchdowns		0.0349**
		(0.0155)
PassingYards	0.000148	
	(0.000160)	
RushingYards	0.000387**	
	(0.000171)	
ReceivingYards	0.000340**	
	(0.000147)	
InterceptionsThrown	-0.0506**	-0.0589**
	(0.0246)	(0.0245)
Fumbles	0.0117	0.0149
	(0.0195)	(0.0199)
DefensiveTouchdowns		0.0606
		(0.0517)
DefensiveYards	0.00126	
	(0.000809)	
Sacks	0.178***	0.178***
	(0.0101)	(0.0101)
Solo	0.00767***	0.00756***
	(0.00111)	(0.00111)
InterceptionsCaught	0.176***	0.188***
	(0.0214)	(0.0182)
Age	-0.0379***	-0.0380***
	(0.00855)	(0.00853)
PreviousAwards	0.0619***	0.0606***
	(0.00860)	(0.00853)
Games	0.0757***	0.0743***
	(0.0159)	(0.0159)
GamesStarted	0.0644***	0.0655***
	(0.00726)	(0.00720)
WinToLoss	0.000273	0.000225

	(0.000444)	(0.000456)
MarginOfVictory	0.0512***	0.0499***
	(0.00382)	(0.00382)
POSITION-FE	YES	YES
TEAM-FE	YES	YES
SEASON-FE	YES	YES
Constant	0.728**	-3.246***
	(0.351)	(0.310)
Observations	14,337	18,215
Number of Players	3,881	4,767

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 11 Peer Voting with Position Fixed-effects

VARIABLES	(3) Pro Bow Pre95 Yards wPosition	(3) Pro Bowl Pre95 TD wPosition
PassingTouchdowns		-0.00963
		(0.0429)
RushingTouchdowns		-0.00626
		(0.0372)
ReceivingTouchdowns		0.148***
		(0.0339)
PassingYards	0.000122	
	(0.000592)	
RushingYards	0.00132***	
	(0.000403)	
ReceivingYards	0.00162***	
	(0.000324)	
InterceptionsThrown	-0.146**	-0.153**
	(0.0633)	(0.0662)
Fumbles	-0.0155	0.0311
	(0.0386)	(0.0347)
DefensiveTouchdowns		0.104
		(0.0972)
DefensiveYards	0.00280*	
	(0.00152)	
Sacks	0.153***	0.149***
	(0.0175)	(0.0173)
Solo	0.00318**	0.00281**
	(0.00135)	(0.00133)
InterceptionsCaught	0.0811**	0.0996***
	(0.0396)	(0.0360)
Age	-0.0454***	-0.0521***
	(0.0168)	(0.0166)

PreviousAwards	0.135***	0.138***
	(0.0160)	(0.0157)
Games	-0.00576	-0.0102
	(0.0309)	(0.0322)
GamesStarted	0.107***	0.123***
	(0.0171)	(0.0185)
WinToLoss	0.748*	0.784*
	(0.449)	(0.447)
MarginOfVictory	0.0239*	0.0218
	(0.0140)	(0.0139)
POSITION-FE	YES	YES
TEAM-FE	YES	YES
SEASON-FE	YES	YES
Constant	0.619*	-1.946***
	(0.369)	(0.613)
Observations	14,337	3,616
Number of Players	3,881	1,270

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 12 Public voting with Position Fixed-effects

VARIABLES	(1) Public voting TD wPosition	(2) Public voting Yards wPosition
PublicVoting =1	-2.867***	
	(0.670)	
PassingTouchdowns	-0.0120	
	(0.0418)	
RushingTouchdowns	0.0425	
	(0.0402)	
ReceivingTouchdowns	0.0103	
	(0.0372)	
PassingYards		-0.000469
		(0.000350)
RushingYards		-0.000194
		(0.000325)
ReceivingYards		0.000104
		(0.000282)
InterceptionsThrown	0.0393	0.124*
	(0.0670)	(0.0706)
Fumbles	0.0634	0.0972*
	(0.0534)	(0.0590)
DefensiveTouchdowns	0.0158	
	(0.120)	

DefensiveYards		-0.000525
		(0.00187)
Sacks	-0.000388	-0.00335
	(0.0152)	(0.0154)
Solo	-0.00652***	-0.00621***
	(0.00225)	(0.00226)
InterceptionsCaught	0.131***	0.127***
	(0.0361)	(0.0433)
Age	0.0563***	0.0520***
	(0.0198)	(0.0199)
PreviousAwards	-0.0708***	-0.0711***
	(0.0203)	(0.0204)
Games	0.0974***	0.0958***
	(0.0319)	(0.0322)
GamesStarted	-0.0351**	-0.0319**
	(0.0138)	(0.0138)
WinToLoss	-0.607	-0.613
	(0.451)	(0.456)
MarginOfVictory	0.00538	0.00661
	(0.0147)	(0.0148)
POSITION-FE	YES	YES
TEAM-FE	YES	YES
SEASON-FE	YES	YES
Constant	0.263	-0.364
	(0.622)	(0.652)
Observations	14,334	14,334
Number of Players	3,881	3,881

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1