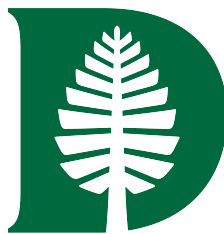# Team Versus Player?
A Study of Baseball Salary Arbitration and the
Arbitrator Exchangeability Hypothesis

DARTMOUTH

Devan Fink

Honors Thesis

Michael Herron, Advisor

Program in Quantitative Social Science

Dartmouth College

June 4, 2023

Abstract

I assess the decisions of arbitrators in Major League Baseball (MLB) salary arbitration hearings under the arbitrator exchangeability hypothesis. Salary arbitration occurs when a player, typically one with more than three, but fewer than six, years of major league service, cannot reach an agreement with his team on a contract for a given year. When this happens, the player and team go to an arbitration hearing. In a hearing, each side presents oral arguments in front of a panel of three independent arbitrators, proclaiming why the arbitrators should rule in their favor. The arbitrators then either decide to award the player his request or the team's offer as his salary for the upcoming season. Arbitrators are not permitted to issue compromises. Historically, teams have won roughly 60 percent of hearings, suggesting that arbitrators might have a pro-team bias. However, my research demonstrates that teams should have won approximately 70 percent of hearings, indicating that arbitrators might actually favor the players. Because there is a statistically-significant difference between the 60 percent observed team win rate and the 70 percent expected team win rate, my results suggest the baseball arbitrators behave in a manner that is inconsistent with the arbitrator exchangeability hypothesis, with a resulting pro-player skew.

Acknowledgements

When I applied to Dartmouth in November 2018, my goal was to study Quantitative Social Science and to write an honors thesis on Major League Baseball. Nearly five years later, as I graduate, I can happily report that I have accomplished both objectives exactly as planned. I would not have been able to do so, however, without the incredible support system that I have had in my life.

To the Dartmouth faculty, I can say without a doubt that you live up to the expectations set by the college. Dartmouth prides itself on its undergraduate teaching focus, and I feel as if I have received an excellent education over the last four years. I feel so fortunate to have been a direct recipient of your knowledge, passion, and expertise.

To Professor Herron, I first met you in February 2018 on a college visit to Hanover. You made time for me as a high school junior who was interested in Dartmouth, a gesture that really resonated with me. Our meeting confirmed that the QSS program was the perfect fit for my interests and that Dartmouth had a unique community feel; it is one main reason why I applied to Dartmouth through Early Decision. Since my first day on campus, your guidance has been so appreciated. I have felt so fortunate to have taken many of your classes, and I always knew that your interest in sports and statistics would make for the perfect advisor. I have cherished all of your enthusiasm and advice as I have worked through this project, and I know I would not have been able to complete it without your valuable insights. Thank you for all you do to help your students succeed.

To my friends, thank you for being so kind, loyal, and fun. You have all played such a crucial role in my college experience, and I have appreciated your genuine interest in my research from the day I started working on this project.

And, finally, to my family, there is nothing I can say that appropriately captures how grateful I am for you. I know I can always turn to you when I need anything. To my mom and dad, you both are my number one fans. Thank you for always being there for me through everything. You are the best.

Contents

1   Introduction

In this thesis, I assess the decisions of arbitrators in the Major League Baseball (MLB) salary arbitration process under the arbitrator exchangeability hypothesis. The salary arbitration process allows qualifying players — typically those with between three and six years of major league experience[1] — to file for final-offer arbitration (FOA) that determines their salaries for an upcoming season.

When a player files for arbitration, both the player and his team submit salary offers and orally present their cases in front of a panel of three independent arbitrators.[2] Unlike in conventional arbitration, where arbitrators may issue any ruling they deem appropriate, the framework of FOA requires that MLB arbitrators must award the player a salary equal to one of the two competing offers, either the player's final request or the team's final offer. In other words, arbitrator compromises are not allowed.

Historically, when a hearing is necessary, the team has typically defeated the player. Per my data, teams have won 57 percent of salary arbitration hearings since 1974, suggesting that arbitrators may be partial toward teams in their rulings. However, that might not necessarily be true based on the win rate alone. In my work, I assess the performance of baseball arbitrators under the arbitrator exchangeability hypothesis. The exchangeability hypothesis implies that one must analyze the quality of each of the two sides' offers in order to appropriately assess the performance of the arbitrators. If one side consistently makes fairer offers more than 50 percent of the time, then that side *should* be winning more than 50 percent of hearings, and arbitrators are performing as expected. In this thesis, I analyze baseball arbitrators' performance in the context of offer quality, rather than on the team-player win-loss record alone.

---

[1] Certain players known as Super Two players — defined as those who rank in the top 22 percent of Major League service time among players between two and three years of service at the end of any given season — earn the right to salary arbitration prior to their third year.

[2] Prior to 1995, salary arbitration hearings were decided by one arbitrator. From 1995-99, some cases were assigned one arbitrator, while others were assigned three. By 2000, all cases included three arbitrators (Edmonds, 2018).

Though arbitration hearings are the main focus of this thesis, not all players who are arbitration eligible file for hearings. Instead, players may reach a salary settlement with their teams to avoid arbitration. A hearing only takes place when negotiations between a player and his respective team break down, with the sides unable to reach an agreement. Most players, in fact, avoid arbitration; in 2022, just 13 of the 240 arbitration-eligible players went to hearings (Cot's Baseball Contracts, 2023a).

To better understand how the baseball arbitration process typically unfolds, consider an example. In mid-January 2019, starting pitcher Gerrit Cole and his team, the Houston Astros, failed to reach an agreement on Cole's salary for the upcoming season (Anderson, 2019). Since he was not yet a free agent, Cole could not sign with any team but the Astros. Nonetheless, Cole was unable to agree with the team on how much he should earn. Thus, he filed for salary arbitration (Rome, 2019). Cole was coming off a 2018 season in which he ranked near the top of the pitching leaderboards in many statistical categories.[3] Cole requested $13,500,000, while the Astros offered $11,425,000. In mid-February 2019, Cole and the Astros orally presented their arguments on his compensation in front of a panel of three independent arbitrators. This panel had two options: award Cole his $13,500,000 request or award Cole the Astros' $11,425,000 offer. The arbitrators ruled in favor of Cole, and he thus earned $13,500,000 during the 2019 season. Cole was one of ten players to go to a salary arbitration hearing in 2019.

The salary arbitration process has existed in baseball for nearly 50 years. In 1973, MLB and the Major League Baseball Players' Association (MLBPA) incorporated salary arbitration into the collective bargaining agreement (CBA) that governs the economics of baseball (Dworkin, 1986). By offering players opportunities to arbitrate their salaries, ownership hoped to quell the players' desire for an end to the reverse clause, which prohibited them from ever choosing a new team on their own prerogative as a free agent (Dworkin, 1986; Scully, 1974). With the onset of salary arbitration, players earned the right to some bar-

---

[3] Cole ranked 14th in the major leagues in wins, 11th in innings pitched, 10th in earned run average, third in strikeouts, and fourth in FanGraphs' version of wins above replacement (FanGraphs, 2018).

gaining power over their compensation, but they still did not yet enjoy free market benefits. However, just three years later, in 1976, MLB players earned the right to free agency, but only after their sixth season in the majors (Baseball Prospectus, 2022). With both arbitration and free agency available at different points during their careers, players now had two points of negotiating leverage that they did not have prior to 1973.

A tiered compensation structure in MLB ultimately developed. Such tiers represent the amount of negotiating power available to the players within a specific range of major league experience. In a player's first three seasons in the majors, their salary can be unilaterally determined by their organization, provided that they are paid at least the major league minimum.[4] In a player's fourth through sixth seasons, they earn the right to salary arbitration. Then, following year six, a player becomes a free agent. Salaries settled through arbitration do not reach the levels as those paid via the open market, in part because a player is only able to negotiate a contract with his current organization (Marburger, 2004). However, arbitration salaries remain considerably higher than players' earnings over their first three seasons in professional baseball — when their salaries tend to fall close to the major league minimum. Therefore, arbitration salaries typically represent a middle-ground between the major league minimum salary and the average free agent salary (Marburger, 2004).

2 Motivation

In this section, I provide motivation for my research. This is broken down into three parts: 1) an overview on how salary arbitration hearings work, with some analysis from players who went to hearings in 2023; 2) criticisms of the salary arbitration process, which provide some insight into factors that may deter teams and players from going to hearings; and 3) a short introduction into the arbitrator exchangeability hypothesis and how it applies to baseball salary arbitration. The arbitrator exchangeability hypothesis is what underpins this thesis;

---

[4] This minimum salary is outlined in the Collective Bargaining agreement between Major League Baseball and the Major League Baseball Players Association (2017-21 Major League Baseball Collective Bargaining Agreement, 2022).

more details on the topic are covered in the subsequent literature review.

2.1 How salary arbitration hearings operate

The process for salary arbitration hearings is outlined in the CBA (2017-21 Major League Baseball Collective Bargaining Agreement, 2022). All hearings are conducted on a confidential basis (p. 20). Case materials are exchanged between the player and team at the start of each hearing (p. 20). The player and team are limited to one hour for their initial presentation and one-half hour for a rebuttal (p. 20).

There are four main rounds within a hearing: 1) player initial presentation, 2) team initial presentation, 3) player rebuttal and summation, and 4) team rebuttal and summation (pp. 20-21). The player and team, respectively, can then respond to new issues raised in rebuttal with very brief surrebuttals to finish the hearing (p. 21). The player and club equally divide the cost of the hearing, and each party is responsible for their own expenses, which may include outside counsel or other representatives (p. 21).

The criteria on which the arbitrators rule is also outlined in the CBA (p. 21). These include: the quality of the player's contribution in the previous season, the "length and consistency" of the player's career contribution to date, the record of the player's past compensation, comparative baseball salaries, the existence of any injuries, and the performance of the team (p. 21). The only admissible statistics in an arbitration hearing are those that are publicly available (p. 22). The arbitrators are also provided with a table showing the minimum major league salary as well as the salaries for all players in the preceding season (p. 23). The arbitrators consider the salaries of all comparable players to determine the fair award for the player in the case at hand (p. 23). No reason for decision is provided to either side (p. 23).

In a Twitter thread posted in February 2023, Tampa Bay Rays pitcher Ryan Thompson shined additional light on salary arbitration by providing an in-depth analysis of his hearing loss (Thompson, 2023). Thompson requested $1,200,000 for his 2023 salary, while the Rays

offered $1,000,000. Throughout his thread, Thompson emphasized the importance of the $1,100,000 midpoint in his case. If Thompson was able to successfully prove how he was just as good or better than similar pitchers who earned $1,100,000 or more, then he should have won his hearing. Thompson laid out his case as to why he was nearly as good as Brusdar Graterol, who earned $1,225,000 in an arbitration settlement in 2023, and better than Cam Bedrosian, who earned $1,100,000 via an arbitration settlement in 2018. In not winning the hearing, Thompson argued, the arbitrators "must have chosen the other side for reasons not stated in the criteria." The process Thompson describes – in which players and teams compare the on-field performance of a hearing-bound player to historical comparative players who earned similar salaries – guides my modeling and research techniques throughout this thesis.

2.2   Criticisms of the salary arbitration process

This salary arbitration process has faced scrutiny. Ryan Thompson mostly expressed concerns with the arbitrators themselves and said later that he held no "ill will" towards the Rays (Topkin, 2023). Not all players feel the same way. Milwaukee Brewers pitcher Corbin Burnes, who also lost his arbitration hearing in February 2023, voiced disappointment with his team post-hearing. He said that his relationship with the Brewers was "definitely hurt" as a result of the hearing, and that, during the hearing, the Brewers put him "at the forefront" of why they did not make the postseason in 2022 (Haudricourt, 2023). Rick Shapiro, who represented Houston Astros outfielder Kyle Tucker in his arbitration hearing in 2023, said that the team told Tucker, "[N]one of the tools you bring to the game, none of the athleticism, none of the base-running, none of the Gold Glove defense, none of that matters in salary arbitration" (Rosenthal, 2023).

This uncomfortable setup – where players have to listen to their teams explain why they think the player is not as valuable as the player believes – has created tension between players and the league as to whether salary arbitration should continue to exist going forward.

According to Rosenthal (2023), in the latest round of negotiations over baseball's newest CBA, the league proposed to eliminate salary arbitration in favor of a formula-based approach that determines salaries based on on-field performance. The MLBPA declined this request, arguing that the right to arbitration provides players with legitimate negotiating power that would not exist in a formula-based salary model. At the same time, however, when players and teams do decide to engage in a hearing, the "three-person panels' decisions often come off as random" to both sides (Rosenthal, 2023).

2.3   Baseball arbitrators and the exchangeability hypothesis

In my research, I assess arbitrators against the "arbitrator exchangeability hypothesis" as the null, in accordance with Ashenfelter (1987). In his work, Ashenfelter shows, using examples from arbitration among New Jersey public safety officers and Iowa public sector employees, that a simple evaluation of the historical case record is not enough to assess the performance of arbitrators. In particular, a 50-50 win split between labor and management is not necessarily the long-term outcome of impartial arbitrators accurately performing their duties. Instead, Ashenfelter shows, arbitrators must be judged based on how they decide between the two competing offers. If either the labor side or management side continually makes the fairer offer, then that side should win more arbitration hearings. In the presence of truly unbiased arbitrators, if this is the case, then decisions *correctly* stray from 50 percent labor, 50 percent management in their rulings.

In baseball salary arbitration, teams frequently defeat the players in hearings. There have been 606 hearings since 1974.[5] Teams have won 344 hearings, equal to a 57 percent win rate, while players have won just 262 hearings, representing a 43 percent win rate. As mentioned earlier, baseball arbitrators are only able to select one of the two offers and cannot compromise. Therefore, I assume that if the midpoint (or mean) of the player and team offers

---

[5] Data for all arbitration hearings was the result of combining three individual data sources: (1) Ed Edmonds' annual hearings charts (Edmonds, 2018), (2) Doug Pappas' salary arbitration database (Pappas, 2004), and (3) Cot's Baseball Contracts arbitration database (Cot's Baseball Contracts, 2023a)

was precisely equal to the "true" salary figure for that particular player, arbitrator decisions would be split uniformly. If this were the case, then I would expect an unbiased arbitrator to select the team's (or, similarly, player's) offer in roughly 50 percent of cases. Thus, assuming a 50-50 split as the null, I conduct a one-proportion z-test on the historical case record, with null hypothesis probability $p = 0.5$. And, I can reject this null that arbitrators rule in favor of teams (or players) at a 50 percent rate an $\alpha = 0.05$ significance level, with $p < 0.001$.[6] Due to data limitations in my modeling discussed later, my analysis only focuses on arbitration cases since 1990. In this timeframe, there have been 327 cases, with 190 team victories (58 percent win rate) and 137 player victories (42 percent win rate). This also statistically differs from a 50-50 split, with $p = 0.003$.[7]

With teams winning nearly 60 percent of hearings since 1974, it is perhaps easy to believe that arbitrators have a pro-team bias when issuing their rulings. In February 2023, one player agent told Ken Rosenthal of *The Athletic* that he believes management beats labor more often than not in arbitration in every industry and that baseball is no different (Rosenthal, 2023). And, if arbitrators do unfairly favor teams with their decisions, major league players may be losing out on millions of dollars every year. Since 1974 and through 2022, based on my data, arbitrator decisions have determined $1.64 billion in player salaries (in 2022 constant dollars). Even a small bias away from a 50-50 decision split means arbitrators have potentially cost players millions over the nearly 50-year time frame of baseball arbitration. Despite players earning considerable bargaining power with the advent of the salary arbitration process, it remains possible that they are still not receiving their fair share.

However, while arbitrators have ruled in favor of teams nearly 60 percent of the time, it is also possible that they actually have a pro-player bias. Arbitrators have an incentive to rule under the perception of being unbiased. So, even if the teams consistently make a fairer offer, the arbitrators may vote for a less deserving player request to maintain an approximately

---

[6] In my two-tailed test, I compared the 57 team win rate against a 50 percent null win rate at a sample size of $n = 606$. The test statistic $z = 3.33$.

[7] Same process as in Footnote 6. This time, I compared the 58 percent win rate against a 50 percent null across a sample size of $n = 327$. The test statistic $z = 2.93$.

equal win-loss split in the overall case record between players and teams. The theory that the arbitrators try to balance the case record rests in two main assumptions. First, this theory assumes that such individuals would prefer to keep their jobs as baseball arbitrators in future years and that predictably biased arbitrators (e.g., those with a known propensity to vote for the player or team more often than they should) would be vetoed by either MLB or the MLBPA when they determine the list of arbitrators for a given year (Faurot and McAllister, 1992). Indeed, there is precedent for this. In 2019, the league fired all three arbitrators when they awarded Cleveland Indians pitcher Trevor Bauer $13,000,000 rather than the team's offer of $11,000,000 (Rosenthal, 2023). And, second, this theory of case record balance also assumes that arbitrators believe their performance is judged solely on the total number of cases won by teams and players, as it often is in the media (Passan, 2019; Rosenthal, 2023). It is possible, then, for arbitrators to have a pro-player bias even though players have only won 43 percent of cases historically.

Therefore, determining whether baseball arbitrators are biased is not as simple as an evaluation of the historical case win-loss record; this is where the application of the arbitrator exchangeability hypothesis is useful. In accordance with the exchangeability hypothesis, baseball arbitrator decisions should be analyzed only after assessing which side made the closer offer to the player's "true" value. If the offers are of equal distance to the true value, such that the true player salary is precisely the midpoint between each offer, then an unbiased arbitrator would be expected to rule in favor of each side roughly 50 percent of the time. However, given that they have ruled in favor of teams roughly 60 percent of the time, it is possible that teams have made fairer offers than players approximately 60 percent of the time. Simply, the exchangeability hypothesis proposes the existence of what I refer to as "calibrated" arbitrators, such that long-run labor-management win rates should be reflective of each side's historical offer quality. In baseball, if it is indeed the case that teams make the fairer offer in roughly 60 percent of cases, then I can conclude that arbitrators have produced calibrated rulings. Such a result would indicate that baseball arbitrators do

behave in a manner that is consistent with the exchangeability hypothesis.

As I show below in the literature review, papers analyzing baseball arbitration tend to assume the arbitrator exchangeability hypothesis before conducting their specific research, but none have tested whether this assumption is valid.

3   Literature review

Having discussed the motivation for this thesis, I now turn to a short review of prior academic work on the topic of arbitration. This research is broken down into four parts. In the first subsection, I review the literature on final-offer arbitration, the style used in Major League Baseball. Here, I discuss the rules of final-offer arbitration, how final-offer arbitration differs from conventional arbitration, and in what contexts final-offer arbitration is used beyond baseball. In the second subsection, I further describe the arbitrator exchangeability hypothesis, showing how an unbalanced case record, as there is in baseball arbitration decisions, may not necessarily suggest bias on the part of the arbitrators. In the third subsection, I cover past analyses of baseball arbitrator behavior. Finally, in the fourth, I turn to prior statistical models of baseball salaries; these models provide grounding for the modeling I do in my research.

3.1   Final-offer arbitration (FOA)

Final-offer arbitration was first formulated by Stevens (1966). Initially referred to as "one-or-the-other" arbitration, Stevens argues that, relative to other forms of arbitration, FOA is "well designed to encourage genuine pre-arbitration negotiation" (p. 46). This is in contrast to compromise arbitration, in which an arbitrator is able to select any award he or she deems to be appropriate (pp. 44-45). The result of compromise arbitration depends on the difference to be split between two parties (p. 45). Rather than encourage genuine pre-arbitration negotiation, which would increase the probability of a pre-arbitration settlement,

compromise arbitration does the opposite (p. 45). Both sides are incentivized to make large demands and few concessions, which may push the expected arbitration award in the direction of either party (p. 45).

FOA, on the other hand, encourages parties to make more reasonable offers by imposing a cost on disagreement (p. 50). Under the threat of FOA, both sides are incentivized to move toward the center during pre-hearing negotiations. In maintaining an unreasonable offer during pre-arbitration negotiation, either party risks the arbitrator choosing the opposing offer; moderating towards the middle increases the likelihood one side's offer is selected by an arbitrator. These reasons are why Stevens argues that FOA is the most optimal of the multiple arbitration styles he examines in his work (p. 50). Given that the vast majority of players and teams settle prior to a hearing, the use of FOA in baseball salary arbitration supports Stevens' argument that the FOA structure promotes authentic pre-arbitration negotiation. In 2022, just 13 of 240 arbitration-eligible players went to a salary arbitration hearing; the remaining reached settlements. In 2013, not a single hearing was needed, and only three times since 1990 did more than 20 hearings occur, out of hundreds of arbitration-eligible players each year.[8] In baseball, the incentive created by FOA to settle prior to arbitration seems to work.

Farber (1980) analyzes the main principles of FOA. In the FOA framework, the Nash equilibrium pair of final offers is the pair in which neither party can achieve a higher expected utility by changing its final offer (p. 690). However, when presented with a hearing, a side can sacrifice some of its position in order to increase its win probability. Therefore, in FOA, the more risk-averse party submits the offer that is closer to the arbitrator's notion of the true award. In doing so, this risk-averse party will win more cases, assuming an unbiased arbitrator's tendency to select the offer that is closer to the true award (p. 692). Over the set of arbitration cases, if one party has a tendency to win more than 50 percent of hearings, this may not reflect bias on the part of the arbitrators. It may simply illustrate that one

---

[8] This is per my data. In 1990, 24 hearings were held, in 1992, 20 hearings were held, and in 2018, 22 hearings were held.

side tends to be more risk-averse in its final offer submissions than the other (p. 697).

Though FOA has earned the nickname of "baseball arbitration" given its prominence in Major League Baseball, it is not solely used to settle salary disputes between baseball players and their clubs (Bazerman and Kahneman, 2016). As of 2013, fourteen U.S. states included some form of FOA in the collective bargaining processes for select groups of public sector employees (Carrell and Bales, 2013, pp. 23-24). For example, laws in Michigan, Wisconsin, and New Jersey prohibit firefighters and police officers from striking; FOA is provided as an alternative to striking — while still allowing for bargaining — to maintain these public services. In Indiana, a 2011 law instituted FOA to settle labor issues among K-12 teachers. Maine, meanwhile, utilizes FOA to limit strikes with employees in its most important industry: agriculture. Other states, such Iowa, Connecticut, and Minnesota, mandate FOA to settle labor disputes among all state employees (Carrell and Bales, 2013, p. 23-24). Thus, implications for my research expand beyond the MLB context alone — my work could inspire future scholars to conduct new research into the arbitrator exchangeability hypothesis in non-baseball FOA contexts.

3.2 The arbitrator exchangeability hypothesis

In FOA, arbitrators determine the fair award based on their judgment, calculate the distance from this fair award to each party's offer, and then select the offer that is closest to the fair award (Farber and Bazerman, 1984). Ashenfelter (1987) argues that arbitrator decisions are "statistically exchangeable," meaning that these decisions follow a common probability density function with an unpredictable component that randomly varies from case-to-case. Arbitrators attempt to "preserve an image of impartiality" to sustain their employment (Hadley and Ruggiero, 2006). In doing so, they make decisions in a "copycat fashion," such that individual arbitrator-to-arbitrator differences are unpredictable. Accordingly, predictably biased arbitrators are vetoed by the adversely-impacted party (Faurot and McAllister, 1992). The arbitrator's decision is then a random draw from a probability

distribution of outcomes anchored by each side's offer (Hadley and Ruggiero, 2006).

This idea of statistically exchangeable arbitrators forms the basis for the arbitrator exchangeability hypothesis, which posits that arbitrators are not systematically biased in favor of either labor or management in the context of a labor negotiation. Ashenfelter (1987) examines the arbitrator exchangeability hypothesis in two non-baseball FOA contexts: New Jersey public safety officers from 1978 to 1980 and Iowa public sector employees from 1976 to 1983. Labor had won two-thirds of arbitration cases in New Jersey, while management had won two-thirds of cases in Iowa.

In both examples, Ashenfelter concluded that the winning side in a majority of cases made the more reasonable offers. He drew this conclusion based on a comparison of the labor and management offers under FOA to awards rendered when the sides used conventional arbitration instead (pp. 344-345). Thus, despite unbalanced win-loss records in both New Jersey and Iowa, Ashenfelter concludes that arbitrators in these two states were not violating the arbitrator exchangeability hypothesis. Instead, they were accurately determining the deserving winner. In Iowa, Ashenfelter showed that, across a period of eight years from 1976 to 1983, management won an observed 65.5 percent of cases compared to an expected 61.1 percent. He uses this as an "impressive confirmation" that arbitrators in Iowa behaved according to the arbitrator exchangeability hypothesis (p. 345). Ashenfelter writes, "[T]he arbitrator exchangeability hypothesis provides a benchmark that can serve as a convenient null hypothesis in the search for predictable deviations to it that might be (profitably) exploited by the parties" (p. 346). Thus, the arbitrator exchangeability hypothesis is the null hypothesis to guide my research in baseball arbitration.

3.3   Past analysis of baseball arbitrators

There is limited research in the area of baseball arbitration. Scully (1978) is the closest model for my methodology. He analyzed arbitration cases involving the Oakland Athletics and their players from 1974 to 1975. Scully used a regression model to estimate the fair award for

the Athletics' 13 cases. He then evaluated whether the arbitrator made the correct decision based on which offer fell closer to his projected award. According to Scully, arbitrators should have ruled in favor of players ten of 13 times (23 percent team win rate). However, in reality, they ruled in favor of the Athletics seven of 13 times (54 percent team win rate). Due to this stark contrast (23 percent versus 54 percent), Scully concluded that fairness in salary arbitration must mean "rendering one-half of the decisions to the players and one-half to the owners without regard to the merits of the case" (p. 447).

This conclusion has three limitations. First, all of the cases analyzed are within one organization, reducing the ability to extrapolate to baseball arbitration cases in general. Second, the difference between expected (ten) and actual (six) number of rulings in favor of players is not statistically significant ($p = 0.107$).[9] Furthermore, the claim made by Scully that arbitrators award 50 percent of cases to each side without considering the merits of the case has not been supported by the comprehensive analysis of cases since 1974, as shown in the motivation section of this thesis. In reality, teams have consistently won a significantly higher number of cases than would be expected in a simple 50-50 split. This indicates that the actual outcomes do not align with Scully's claim. If Scully's assessment were still accurate, the arbitrator decisions should be indistinguishable from a 50-50 distribution. However, the evidence suggests otherwise. Regardless, through his work, Scully does provide a framework that I can replicate in order to test the existence of arbitrator bias on a larger set of hearings.

Burger and Walters (2005) build on Scully and evaluate baseball arbitrator behavior across 391 cases. Using "competitive market principles," they implement a regression model to estimate each player's marginal revenue product to predict the "fair" award in each case. They estimate probit model to predict arbitrator decisions, revealing that arbitrators tend to favor teams 61 percent of the time when faced with equally distant team and player offers from the "true" award. Burger and Walters do not determine the proportion of the time that team and player offers actually are equidistant from their estimation of the "true" award, as

---

[9] These are my calculations based on Table 3 in Scully (p. 445).

I do. Though they find a statistically-significant pro-team bias, the statistical significance of this general pro-team bias disappears in models that account for player race and ethnicity. That is, after accounting for a player being Black or Latin American, the effect of a general club bias driving team arbitration victory was no longer statistically distinguishable from zero. This result is consistent with Fizel (1996), who found racial bias in arbitrator decisions. Studying cases from 1985 to 1990, his work demonstrates that Latin American and Black players were 33 percent and 19 percent less likely to win a case relative to a White player, respectively. Regarding arbitration cases in general, Fizel supports Scully's observation that arbitrators assess fairness based on achieving a balanced win-loss record (p. 261).

In general, scholars are mixed on whether baseball arbitrators typically rule in an attempt to achieve a balanced win-loss record. Scully (1978), Faurot and McAllister (1992), and Fizel (1996) make this claim. Burger and Walters (2005), on the other hand, argue that there may be bias. For the scholars who have previously suggested that arbitrators appear to balance the record, that appears to be incorrect, with there being a statistically-significant deviation from 50-50 among the set of all cases since 1974. Meanwhile, Burger and Walters (2005), while finding the existence of "bias," do so using a model that is not trained on previous salary data but is instead derived from "fundamental competitive principles" established by Scully (1974). My research builds on and improves prior approaches by directly applying the market-setting salaries among players and teams to develop a model for hearing-bound players, before then assessing the potential for bias among arbitrators. I do so against the backdrop of the arbitrator exchangeability hypothesis, which emphasizes that 50-50 does not necessarily represent fairness, particularly when one party consistently makes fairer offers than the other.

3.4   Models of baseball salaries

To test the arbitrator exchangeability hypothesis in baseball, I build statistical models to estimate what the player should receive in salary. My models are grounded in Scully (1974),

who is believed to be among the first to model baseball player salary using on-field performance statistics (Weber, 2009). In his analysis, Scully estimated professional baseball players' net marginal revenue product using each player's performance, length of career, team market size, team performance, and other variables. Using his results, Scully calculated the impact of non-competitive baseball market practices on the suppression of player salaries; he concluded that "economic analysis points to the exploitation of the professional baseball player under the reverse clause through the introduction of monopsony power" (p. 929). Scully's work is considered by many to be influential in the professional baseball players' organization against the owners in the early-1970s, leading to the end of the reverse clause with the creation of free agency in 1976 (Weber, 2009).

Following Scully's lead, other scholars have studied baseball arbitration salaries using regression models to estimate player salaries. These scholars typically build their model with a combination of independent variables, including on-field player performance and team quality statistics. These variables are in line with the baseball arbitration criteria as outlined in the CBA (2017-21 Major League Baseball Collective Bargaining Agreement, 2022, p. 21). Consistent with the arbitration criteria, Faurot and McAllister (1992) find that there are statistically significant effects on the arbitrator's fair settlement for four player-specific variables that arbitrators consider in making their decisions, including: player's performance in the previous season, length and consistency of the player's career performance, previous compensation, and player position. Additionally, other authors, including Fizel (1996), Marburger (2004), Hadley and Ruggiero (2006), Swartz (2011), Dolinar and Chamberlain (2015b), and Rieders (2015) use statistical models to estimate player salaries. In all cases, different models are used for pitchers and position players because pitchers and position players represent entirely separate classes of baseball player; performance is measured differently for a pitcher than it is for a position player. The best of these models find that player- and team-specific variables explain roughly 80 percent of the variability in salary (that is, $R^2 \approx 0.80$), though most models have a R^2 value between 0.7 and 0.8. The aforementioned authors then use

their models to either investigate some features of baseball arbitration (Faurot and McAllister, 1992; Fizel, 1996; Hadley and Ruggiero, 2006; Marburger, 2004), to make predictions about future arbitration salaries (Swartz, 2011), or to simply evaluate which performance statistics best explain salary in general (Dolinar and Chamberlain, 2015b; Rieders, 2015).

4 Methods

In this section, I describe my research methods, which use Scully (1978) as a guide. Consistent with Scully, in order to evaluate the performance of baseball arbitrators, I determine which of two salary offers – either the team's offer, which I refer to as S_T, or the player's request, which I refer to as S_P – is a fairer representation of the player's true value W^*. I use statistical models to estimate the W^*, i.e. what each player "should" earn, for each player who went to an arbitration hearing, based on their on-field performance. The statistical models estimating W^* are known as "salary models" throughout the rest of this thesis.

In a given arbitration case, if $|W^* - S_P| < |W^* - S_T|$, then I expect the player to win. If this inequality holds, then the player's requested salary is closer to what he should receive than the team's offer. Conversely, if $|W^* - S_T| < |W^* - S_P|$, then I expect the team to win the arbitration hearing by the same logic. Using these inequalities across all players who had an arbitration hearing, I evaluate the performance of the arbitrators by comparing the observed team win rate to the expected team win rate based on the estimated fairness of each of the two sides' offers.[10]

My work can be contrasted with Scully's in three dimensions. First, Scully evaluated 13 arbitration cases, while I assess more than 300, a greater than 20 times increase. Second, I take a different approach to estimating player salaries. While Scully used a model to estimate the marginal revenue product for each player, I train my models based on observed salary data of the other arbitration-eligible, major league players who settled prior to hearings.

---

[10] Without loss of generality, I am focusing on the team win rate because the team win rate and player win rate sums to one.

Third, I employ a modeling strategy called k-nearest neighbors regression, which involves finding the nearest data points to a given observation and using those to make predictions (Yildirim, 2020). To the best of my knowledge, I have not observed this technique used in published baseball arbitration research.

There are three main steps in developing my salary models: Step 1, filtering the salary data set on major league players to the set of salaries for past arbitration-eligible players; Step 2, training salary models for past arbitration-eligible position players and pitchers based on their on-field performance; and Step 3, applying the salary models to players who went to arbitration hearings, allowing me to evaluate the performance of the arbitrators.

In Section 4.1, I describe Step 1. I explain how I filter my data set of all major league salaries to just those for players who were arbitration eligible; after this process is complete, I have a data set of salaries exclusively among players who were arbitration eligible but instead settled with their team through pre-arbitration negotiation. Because the baseball arbitration criteria specifically recommends that arbitrators focus their attention on salaries among players with similar major league experience to the player in the case at hand, these are the most important salaries on which to train my model (2017-21 Major League Baseball Collective Bargaining Agreement, 2022, p. 22). Previous settlements (i.e., salaries determined by agreements between players and teams rather than via a hearing) are considered the backbone of arbitration in determining salaries for new players (Passan, 2019). Prior settlements represent "admissions from both sides that this is the actual market value and not just a number chosen by three people [the arbitrators]" (Passan, 2019). Because arbitration-eligibility data is not available prior to 2007, I train a logistic regression to estimate the probability that each player was arbitration eligible in a given year back to 1990.

In Section 4.2, I describe Step 2. I explain how I train salary models on the filtered data set of arbitration settlements (from Step 1). I first describe my modeling strategy for position players (in subsection 4.2.1), then pitchers (in subsection 4.2.2). Certain independent variables are included for both position players and pitchers. For both player subgroups, I

trained two type of models: linear regressions and k-nearest neighbor regressions.

Finally, in Section 4.3, I describe Step 3. I apply the trained and tested salary models to all players who had an arbitration hearing since 1990. I evaluate whether the arbitrators made the correct decision in each case by analyzing whether they selected the offer that was closer to the model-estimated salary for each player.

4.1 Identifying arbitration-eligible players

A player's arbitration eligibility is determined by his major league service time. Players who have recorded more than three, but fewer than six, years of major league service are eligible for arbitration (select players who have more than two years but fewer than three years are also eligible).

Determining which players are arbitration eligible in any given year – particularly in older years – is challenging. As of the most recent CBA, 172 days on a major league active roster equaled one full year of service (2017-21 Major League Baseball Collective Bargaining Agreement, 2022, p. 104). It is challenging to identify which players are arbitration eligible because players may play in more than three major league seasons before procuring three years of major league service. This could happen for multiple reasons, but one leading factor is that teams can hold certain players in the minor leagues to begin a season so as to keep them from reaching 172 days of service as a rookie. This delays their arbitration eligibility and eventual free agency (Kahrl, 2019). Having service time data since 1990 would allow me to identify which players were arbitration eligible in any given season, and thus I could train a model on the full set of historical arbitration settlements. However, complete service time data only exist back to 2007 (Cot's Baseball Contracts, 2022a).

To identify which players were arbitration eligible, I used data from Pappas (2004) and Cot's Baseball Contracts (2022b) to compile a list of all Opening Day major league player salaries from 1985 to 2022. Ideally, I could identify for the salaries specifically representing arbitration-eligible contracts (either from settlements or from a hearing), but without com-

plete service time data, I need to estimate which players were arbitration eligible in a given season. However, knowing that players who sign free agent contracts are no longer arbitration eligible, I eliminated any salary determined via free agency from the major league salary data. Again combining data from Pappas (2004) and Cot's Baseball Contracts (2023b), I assembled a list of every free agent signing in major league history. Within the salary data set, if a player's salary was one determined by a free agent contract, I filtered it out. Because players become eligible for free agency once they are no longer arbitration eligible, any salaries determined on a free agent contract do not represent arbitration settlements.

With free agent contracts filtered out, I then needed to estimate which of the remaining salaries represented arbitration salaries. With service time data from 2007 onward, I used a logistic regression model to assign an estimate of a player's first year of arbitration eligibility to each season he was in the major leagues on a non-free-agent salary; this model was trained and tested on the post-2007 data before being applied to all salaries since 1985. The model takes each player-season in a given year and estimates the probability that that individual player-season is said player's first year of arbitration eligibility. For each player-season, the two independent variables in this model were a player's salary in the previous season (adjusted for inflation) as well as the multiplicative increase in year-over-year salary from the previous year to the current year. Arbitration-eligible players cannot return to non-arbitration-eligible status. Therefore, I used the estimate of their first year of eligibility and considered their next three seasons to also be arbitration-eligible seasons, because players cannot be eligible for arbitration more than four times.

How these two independent variables – previous salary and multiplicative salary increase – can capture a player's arbitration eligibility is shown in Table 1, with New York Yankees outfielder Aaron Judge serving as an example. From 2017 to 2019, Judge's first three full seasons in the major leagues, he earned close to the league-minimum salary. Then, Judge's salary increased to $9,777,573 in 2020, his first year of arbitration-eligibility. This represented a 12.3 times increase over Judge's 2019 salary, the largest multiplicative increase in

salary Judge has experienced over his entire career, significantly larger than his jump from $19,000,000 to $39,518,942 (a 2.08 times increase) that he earned from going from his final season of arbitration eligibility to his first season on a free agent contract.

Table 1 demonstrates the three-tiered salary system for major league players over their career. Judge earned very little while he had no bargaining power (rows 1-3, "team controlled"), then earned more once he gained arbitration eligibility (rows 4-6, "arbitration-eligible"), and then earned the most once he gained the right to the open market forces of free agency (row 7, "free agent"). Judge re-signed with the Yankees on a nine-year, $360,000,000 contract in December 2022 that pays him $40,000,000 annually (Hoch, 2022), after earning roughly $40,000,000 over his entire career prior to this new deal.

Table 1: Aaron Judge's salary over time

Year	Type	MLS	Salary	Times increase
2017	Team-Controlled	0.051	661,323	—
2018	Team-Controlled	1.051	737,795	1.11
2019	Team-Controlled	2.051	796,863	1.08
2020	**Arbitration-Eligible**	**3.051**	**9,777,573**	**12.27**
2021	Arbitration-Eligible	4.051	11,179,132	1.14
2022	Arbitration-Eligible	5.051	19,000,000	1.70
2023	Free Agent	6.051	39,518,942	2.08

Note: all salaries in 2022 dollars. "MLS" refers to the amount of major league service Judge had at the beginning of each season. At the beginning of 2018, for example, Judge had 1 year, 51 days of service time.

Table 1 shows how a model using a player's multiplicative salary increase can serve as an informative input variable in a model determining arbitration eligibility. Figure 1 shows the receiving operator characteristic (ROC) curve for the logistic regression model estimating player arbitration eligibility. This model was trained and tested on labeled, post-2007 data. The ROC curve shows the false and true positive rates for the classifier at different probability thresholds. For example, at a probability threshold of 0.00, every salary was assigned to be first-time-arbitration-eligible, irrespective of the probability the model estimated; at a probability threshold of 0.95, only when the model estimated the probability

Figure 1: ROC curve for logit estimating first year of arbitration eligibility (AUC = 0.97)
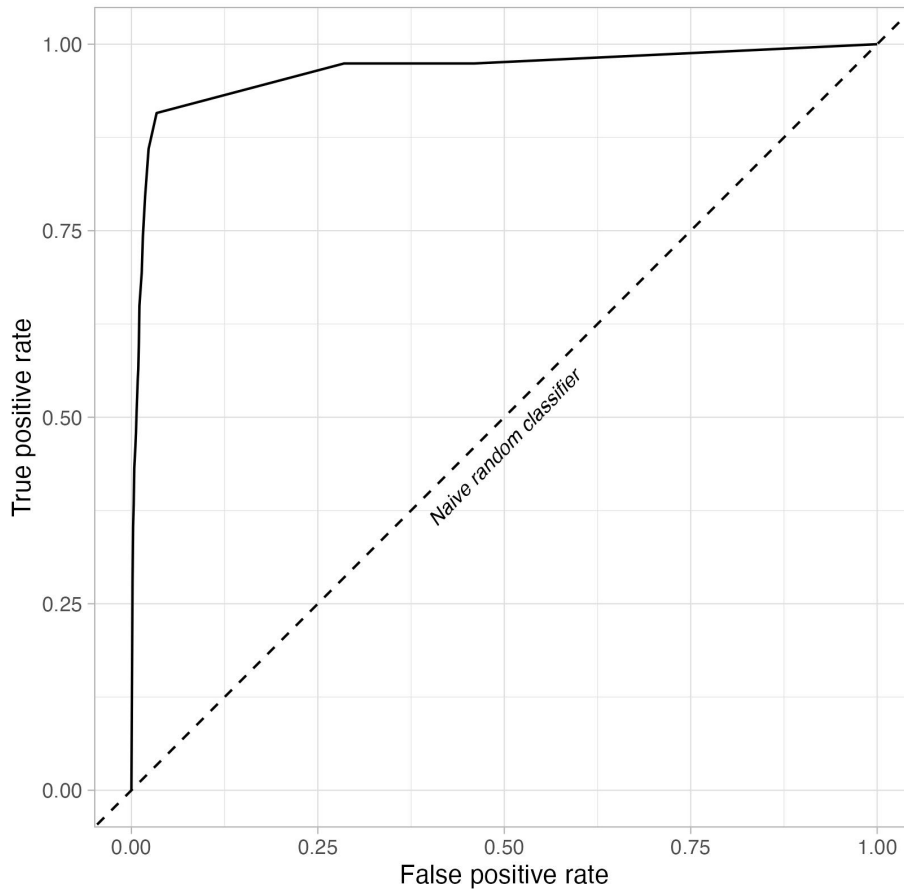
of first-time-arbitration-eligible to be greater or equal to 0.95 did such a player-season get assigned to be first-year-arbitration-eligible. The total area under the curve (AUC) is an common statistic that represents a classifier's overall accuracy (Fogarty et al., 2005). A perfect classifier has an AUC of 1.00, so a 0.97 AUC, as I have here, suggests that the logit model is rather successful at classifying players' first year of arbitration eligibility.

For my analysis, I used a 0.25 probability threshold to determine each player's first year of arbitration eligibility. At this threshold, the logit model had a true positive rate of 0.80 and a false positive rate of 0.02. This means that, among the total true positives in the data set, 80 percent were classified as true positive by the model at this threshold. More importantly, however, the false positive rate is just 0.02, meaning that only two percent of

cases classified as a player's first year of arbitration eligibility were not such player's first year of arbitration eligibility. When used to estimate arbitration eligibility on the pre-2007 data, the logit model should result in limited data contamination of player-seasons that were not arbitration eligible being classified as those that were. My threshold maintains a large sample size in my settlement data set. Higher thresholds of eligibility, though they would eliminate more false positives, would significantly reduce my sample size for settlements on which to build the salary models.[11]

As an example of how the logit model estimates first-time-arbitration-eligibility, consider San Diego Padres third baseman Manny Machado's salary path, as shown in Table 2. Machado became arbitration eligible for the first time in 2016. The model correctly identifies 2016 as being the most likely season to be his first year of arbitration eligibility, with a probability greater than 0.99; the model also is able to clearly distinguish this season from Machado's other seasons, as no other Machado season is assigned a probability greater than 0.11.

Table 2: Manny Machado's salary over time

Year	Type	MLS	Salary	Times increase	Y1 arb. probability
2013	Team-Controlled	0.056	632,593	—	—
2014	Team-Controlled	1.056	652,676	1.03	0.11
2015	Team-Controlled	2.056	688,329	1.05	0.10
2016	**Arbitration-Eligible**	**3.056**	**6,202,132**	**9.01**	**> 0.99**
2017	Arbitration-Eligible	4.056	13,967,349	2.25	0.00
2018	Arbitration-Eligible	5.056	18,969,507	1.36	0.00

Note: all salaries in 2022 dollars.

After analyzing the performance of the logit model on the labeled post-2007 data, I then applied it to the 1985-2022 data and estimated the likelihood of each season being an individual player's first year of arbitration eligibility. After grouping the data by player, I found the maximum probability that any individual season was a player's first year of arbitration eligibility. If this probability was above my chosen threshold of 0.25, I assigned

---

[11] Even just increasing the probability threshold from 0.25 to 0.50 reduces my sample size from nearly 8,000 salaries to just over 5,000; I believe this trade-off is not worth reducing the false positive rate from 0.02 to 0.01.

said season to be that player's first year of arbitration eligibility. Then, after determining which season was each player's first year of eligibility, I considered the next three seasons of player salaries to also be arbitration eligible, since players can be arbitration eligible for a maximum of four times.

If a player had a maximum first-year arbitration probability exceeding 0.25 for any of his seasons, the player was retained in the salary data set, with his salary data spanning up to four seasons. I disregarded players for whom the model could not determine a clear first year of arbitration eligibility. Among the 32,652 Opening Day salaries in the 1985 to 2022 data set, 7,957, or about 24 percent of all salaries, were identified as being player salaries for arbitration-eligible players. This is close to the proportion of arbitration-eligible players on Opening Day rosters in 2022, which was 29 percent (Cot's Baseball Contracts, 2022b). Among the 7,477 arbitration-eligible player salaries in my data set, 4,179 were salaries for position players and the remaining 3,778 were for pitchers.

4.2 Training and testing salary models

With the salary data set filtered for arbitration-eligible players, I completed Step 1 of my outlined methodology. Now, in Step 2, I trained and tested models that regressed arbitration-eligible players' salaries based on their on-field performance statistics.

For both position players and pitchers, I developed two types of regression models: a linear model and a k-nearest neighbors model. For all players, I adjusted their salaries into constant dollars as of November 2022, using the average annual consumer price index (US Inflation Calculator, 2023). I then transformed each salary to the natural logarithmic scale to make the salary data more closely follow a Normal distribution, as the non-adjusted salaries in my data set are skewed right (shown in Figure 2). I conduct cross validation using a test-train split of 70 percent train, 30 percent test to estimate my models.

Salary models that rely on linear regression regress player salary on a vector of on-field statistics, which makes this type of model easily interpretable. For example, for position

Figure 2: Arbitration-eligible player salaries, before and after logarithmic transformation

Salary (in millions)   **Natural Log of Salary**
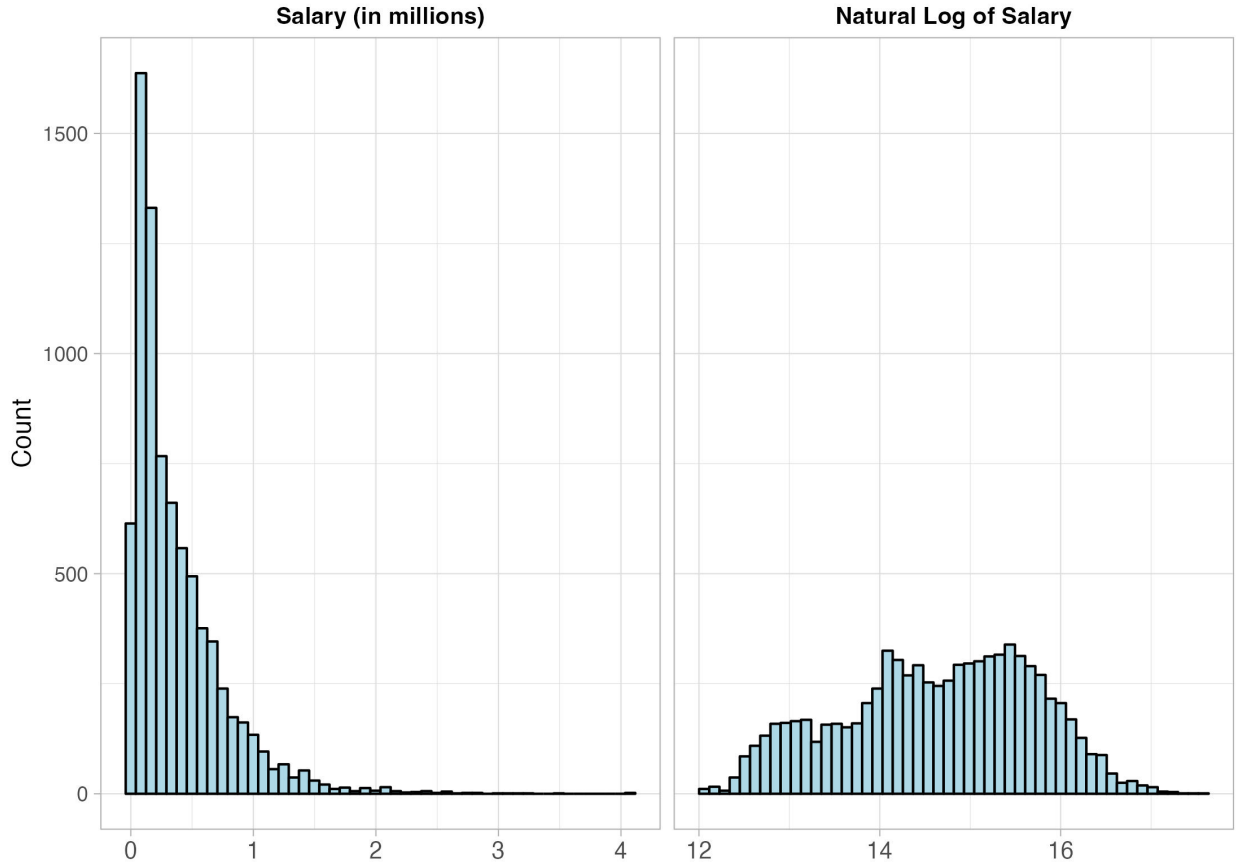
Note: all salaries in 2022 dollars.

players, the linear regression model provides a dollar estimate for the marginal value of each
home run hit, each run batted in, and every point of on-base plus slugging percentage (OPS).
The k-nearest neighbors model, however, more closely models the arbitration process, despite
the model being less interpretable than linear regression. For each observation in the test
data set, this model finds the k-nearest observations in the training set, minimizing the total
Euclidean distance among each independent variable. The k-nearest neighbors model then
calculates a weighted average of the salaries (based on distance) for each of the k neighbors,
before assigning that average to the new observation. The k-nearest neighbors model might
be more effective for salary arbitration because it more closely follows the process of finding
comparative salaries among players with similar statistics before using those comparative

players to determine the new player's expected salary. This non-parametric method of k-nearest neighbors is more flexible and performs better at finding non-linear trends, if any, in the data (Steorts, 2017). However, as noted, k-nearest neighbors is less interpretable than the linear model due to its lack of an explicit functional form.

Because players, teams, and arbitrators only have the settlement information available to them at the time the hearing was heard, I partitioned my data to account for the fact that there is no knowledge of future settlements. For example, in 1997, Tim Wakefield and the Boston Red Sox went to an arbitration hearing. At the time of that hearing, all parties involved would have knowledge of arbitration settlements that occurred in years up to and including 1997. Neither Wakefield nor the Red Sox were able to compare Wakefield to players who settled after 1997. Thus, for all my models, I partitioned the data by year, so that the model would only be trained on settlements in a given year or in prior years.

As a result, I tested the model on different window sizes, ranging from zero years to four years. For example, for players in 1995, a window size of zero means that the model was trained on settlements in 1995 alone, while a window size of four means that the model was trained on settlements up to four years in the past, including settlements from 1991 to 1995. Results as a function of window size are shown in the appendix Figures 5, 6, 7, and 8. Due to sample size constraints – with few arbitration-eligible players in the earliest years within my data set – I trained and tested my models beginning in 1990 and only analyze arbitration cases from 1990 to 2022.

Within each model, I accounted for time with year fixed effects. I also included variables to account for each player's previous salary. However, despite previous salary being a significant predictor of each player's current salary, not all players had equal amounts of bargaining power in the prior year. As mentioned earlier, before a player becomes arbitration eligible for the first time, they have limited negotiating power over their team. Many players make the league minimum salary, but teams can choose to pay them slightly more than the minimum if they wish. To account for this behavior, I classify all players whose previous salary was

within five percent of the league's minimum in the previous year as being on the minimum. And, therefore, to incorporate each player's previous salary in my models, I included two interaction terms: 1) if the player was on the league-minimum salary in the previous year × their previous salary or 2) if the player was not on the league-minimum salary in the previous year × their previous salary.

For example, to demonstrate the within-five-percent-of-minimum-salary buffer, if a player like Joe Musgrove is making just slightly above the minimum salary, he is still considered to have made the minimum salary in the previous year for the purposes of the model. Musgrove's salary path is shown in Table 3.

Table 3: Joe Musgrove's salary over time

Year	Type	MLS	Salary	% Above Minimum	Minimum?
2017	Team-Controlled	0.063	659,988	1.57	Yes
2018	Team-Controlled	1.063	676,974	4.77	Yes
2019	Team-Controlled	2.063	678,318	4.95	Yes
2020	Arbitration-Eligible	3.063	3,220,848	396.89	No
2021	Arbitration-Eligible	4.063	4,889,154	680.02	No
2022	Arbitration-Eligible	5.063	8,625,000	1,132.14	No

Note: all salaries in 2022 dollars.

The two interaction terms – 1) on minimum in previous year × previous salary or 2) not on minimum in previous year × previous salary – create separate coefficients on previous salary for players making the league-minimum salary in the previous year versus those who were not, so that different coefficient estimates are calculated for previous salary depending on its minimum versus non-minimum classification.

4.2.1 Salary models for position players

Linear regression. I considered the following prior season performance statistics as independent variables: plate appearances, batting average, on-base percentage, slugging percentage, hits, runs scored, extra-base hits, home runs, stolen bases, runs batted in, on-base plus slugging percentage, and FanGraphs wins above replacement. I also included the following statistics for each player's career to date: plate appearances, batting average, on-base per-

centage, slugging percentage, home runs, runs batted in, on-base plus slugging percentage, and FanGraphs wins above replacement. These statistics are based on a combination of approaches described in the literature review. Marburger (2004) uses at bats (which I switch to plate appearances), on-base percentage, slugging percentage, and home runs in his models. Rieders (2015) highlights the importance of extra-base hits and runs scored. Dolinar and Chamberlain (2015a) used FanGraphs' version of WAR. Other statistics included in my models, such as batting average and stolen bases, are also common ways to evaluate players (Baseball Reference Bullpen, 2022). All statistics were scraped from FanGraphs using the baseballR package (Petti, 2021). With each player's on-field statistics, I trained and tested the model with window sizes of zero to four, creating a different model for each year at each window size.

To find the best combination of variables at predicting arbitration settlements, I used a statistical technique called subset regression. To do this, I needed to test a large number of possible models. For each year of settlements, I loop over all possible combinations of independent variables, with each variable either included or excluded. Because I had 20 independent variables, this resulted in testing $2^{20} - 1$ unique models for each year.[12] For each year, I identified the combination of variables that produced the lowest root mean squared error when predicting unseen settlements. The model with the lowest root mean squared error on unseen settlements is called the "winning model." A different model (i.e., a model with a potentially-unique set of independent variables) is determined to be the winner in any given year. Table 4 shows the results for each of the winning models with a window size of zero. For each year, the winning model's R^2 and root mean squared error is reported. The third column shows the standard deviation of the log of position player salaries for that year, while the fourth column calculates the ratio of root mean squared error to that year's standard deviation in log salaries. The root mean squared error ranges from 0.19 in the 2013

---

[12] I tested this large number of models using Dartmouth Discovery, a 3000-core Linux cluster that is available to the College's research community. I ran each year's models in parallel to maximize the speed of this task. Thank you to Dartmouth's research systems engineer, Richard Brittain, for his assistance with this process.

model to 0.45 in the 2001 model, while the R^2 ranges from 0.68 to 0.92. The average root mean squared error across all years was equal to 0.35 standard deviations in salaries; this ranged from 0.24 standard deviations in 2012 to 0.56 standard deviations in 2009. I repeated the subset regression process for models at each window size to test the sensitivity of my results. Tables for window sizes one to four are not reported here.

As mentioned above, all models included year fixed effects and one of two interaction variables depending on if the player's previous salary was on (or near) the league minimum. For example, for the 1990 model, the "winning model" included the following variables: year, previous salary $\times$ minimum, previous salary $\times$ not minimum, plate appearances recorded in the preceding season, on-base percentage in the preceding season, and on-base plus slugging percentage in the preceding season. The 1991 model, meanwhile, included the following variables: year, previous salary $\times$ minimum, previous salary $\times$ not minimum, plate appearances recorded in a player's career to date, batting average in the preceding season, on-base percentage in the preceding season, on-base percentage in a player's career to date, on-base plus slugging percentage in the preceding season, and FanGraphs wins above replacement in the preceding season.

K-nearest neighbors regression. I considered the same statistics as independent variables as I did with the linear regression. For each observation in the test data set, I used the 10 nearest neighbors and calculated a distance-based weighted average (with closer points receiving larger weights) salary of those 10 neighbors' salaries as the predicted salary for the unseen data point. I again trained and tested a k-nearest neighbors regression model at window sizes of zero to four. Because my process for the k-nearest neighbors model does not specify different independent variables for each year, I aggregated the test data from all years into a single data set and calculated an overall R^2 at each window size: 0.78 at window size zero, 0.84 at one, 0.83 at two, 0.83 at three and 0.81 at four. This performance is qualitatively equivalent to that of the linear model.

Table 4: Position player salary models - window size of zero

Year	R^2	RMSE	Salary SD	RMSE / SD
1990	0.91	0.21	0.81	0.26
1991	0.88	0.24	0.86	0.28
1992	0.92	0.37	0.94	0.39
1993	0.85	0.37	1.01	0.37
1994	0.91	0.38	1.10	0.35
1995	0.88	0.36	1.07	0.34
1996	0.89	0.39	1.13	0.35
1997	0.84	0.29	1.12	0.26
1998	0.81	0.40	1.06	0.38
1999	0.87	0.40	1.00	0.40
2000	0.76	0.28	0.97	0.29
2001	0.89	0.45	1.03	0.43
2002	0.86	0.34	0.95	0.36
2003	0.79	0.42	0.96	0.44
2004	0.91	0.38	0.94	0.40
2005	0.91	0.25	0.82	0.30
2006	0.92	0.24	0.78	0.31
2007	0.80	0.22	0.71	0.31
2008	0.68	0.26	0.77	0.34
2009	0.71	0.40	0.71	0.56
2010	0.77	0.34	0.76	0.45
2011	0.79	0.31	0.76	0.41
2012	0.88	0.21	0.88	0.24
2013	0.87	0.19	0.78	0.24
2014	0.83	0.25	0.76	0.33
2015	0.86	0.21	0.71	0.30
2016	0.69	0.28	0.72	0.39
2017	0.75	0.24	0.73	0.33
2018	0.91	0.28	0.72	0.39
2019	0.73	0.29	0.71	0.41
2020	0.83	0.32	0.75	0.43
2021	0.73	0.20	0.79	0.25
2022	0.88	0.33	0.80	0.41

Note: these models produce the lowest RMSE for each year. "Salary SD" refers to the standard deviation of logged salary for each year. "RMSE / SD" is the ratio of root mean squared error to the annual standard deviation of logged salary.

4.2.2 Salary models for pitchers

Linear regression. The pitcher salary models are structured similarly to the salary models for the position players. For pitchers, I considered following prior season performance

statistics as independent variables: innings pitched, games started percentage, wins, losses, saves, earned run average, strikeouts, walks, and FanGraphs wins above replacement. I also included the following career to date statistics as potential input variables: innings pitched, games started percentage, wins, losses, saves, earned run average, strikeouts, walks, and FanGraphs wins above replacement. These statistics are based on a combination of approaches described in the literature review. Marburger (2004) used innings pitched, earned run average, and saves in his analysis. Dolinar and Chamberlain (2015a) include wins and WAR. Rieders (2015) cites strikeouts. Games started percentage is used to denote that some pitchers are exclusively starting pitchers, some pitchers are exclusively relievers, and others are somewhere in-between. Losses and walks are also common ways to evaluate pitchers (Baseball Reference Bullpen, 2022).

I trained and tested linear models with window sizes of zero to four. With 18 independent variables, $2^{18} - 1$ models were trained and tested for each year and at each window size. The by-year results for the linear model with the lowest root mean squared error at window size zero are shown in Table 5. Columns 2 and 3 report the R^2 and root mean squared error for each model, while column 4 reports each year's standard deviation in salary and column 5 reports the root mean squared error to standard deviation ratio. The root mean squared error on unseen arbitration settlements ranges from 0.22 in the 2021 model to 0.50 in the 2003 model, while the R^2 ranges from 0.41 to 0.87. The mean root mean squared error to standard deviation ratio is 0.41 standard deviations; this ranges from 0.30 to 0.58 standard deviations. The pitcher models do not fit to salaries as well as the position pitcher models, though this could mostly be random variation, as the mean RMSE-to-standard-deviation ratio is just 0.06 standard deviations higher for pitchers than for position players.

As with the position players, for each year and within each window size, a different suite of variables was selected as the winning model. For example, for the 1990 model, the following variables minimized the root mean squared error: year, previous salary $\times$ on minimum, previous salary $\times$ not on minimum, innings pitched for the pitcher's career to

Table 5: Pitcher salary models - window size of zero

Year	R^2	RMSE	Salary SD	RMSE / SD
1990	0.85	0.24	0.72	0.33
1991	0.76	0.32	0.82	0.39
1992	0.87	0.35	0.99	0.35
1993	0.77	0.42	0.98	0.43
1994	0.68	0.48	1.02	0.47
1995	0.83	0.42	1.06	0.40
1996	0.87	0.36	1.10	0.33
1997	0.82	0.36	1.00	0.36
1998	0.76	0.41	1.00	0.41
1999	0.74	0.36	0.98	0.37
2000	0.85	0.38	0.94	0.40
2001	0.71	0.45	0.77	0.58
2002	0.56	0.40	0.79	0.51
2003	0.70	0.50	0.99	0.51
2004	0.73	0.32	0.94	0.34
2005	0.81	0.26	0.86	0.30
2006	0.75	0.27	0.73	0.37
2007	0.74	0.25	0.78	0.32
2008	0.70	0.24	0.78	0.31
2009	0.64	0.38	0.79	0.48
2010	0.68	0.36	0.75	0.48
2011	0.74	0.30	0.84	0.36
2012	0.80	0.35	0.79	0.44
2013	0.81	0.34	0.74	0.46
2014	0.41	0.36	0.71	0.51
2015	0.68	0.29	0.75	0.39
2016	0.76	0.29	0.66	0.44
2017	0.73	0.26	0.69	0.42
2018	0.80	0.31	0.65	0.48
2019	0.76	0.26	0.69	0.38
2020	0.51	0.42	0.74	0.57
2021	0.80	0.22	0.63	0.35
2022	0.76	0.30	0.70	0.43

Note: these models produce the lowest RMSE for each year. "Salary SD" refers to the standard deviation of logged salary for each year. "RMSE / SD" is the ratio of root mean squared error to the annual standard deviation of logged salary.

date, games started percentage for the pitcher's career to date, losses for the pitcher's career to date, saves for the pitcher's career to date, earned run average in the preceding season,

strikeouts in the preceding season, strikeouts for the pitcher's career to date, and wins above replacement in the preceding season.

K-nearest neighbors regression. I considered the same statistics as independent variables as I did with the linear regression. For each unseen data point, I again used the 10 nearest neighbors and calculated a distance-based weighted average (with closer points receiving larger weights) average salary of those 10 neighbors' salaries as the predicted salary for the unseen data point. I trained and tested this model at window sizes of zero to four, with the R^2 of each of those models being 0.75, 0.78, 0.79, 0.80, and 0.81 across all years. These R^2 values are qualitatively similar to the results from linear regression.

4.3 Using model-estimated salaries to evaluate arbitrators

With salary models trained and tested for both position players and pitchers, I turn to Step 3: using salary models to assess the performance of the arbitrators. I first estimate the true salary W^* for each player who went to an arbitration hearing since 1990. I then assess how this true salary estimate compares to the player's requested salary, the team's offered salary, and the midpoint (defined as M) in each hearing. Each estimate falls into one of four buckets: 1) below S_T, 2) between S_T and M, 3) between M and S_P, and 4) above S_P. If the estimated true salary is in buckets 1 or 2, my model predicts a team victory in the case. Conversely, if the estimated true salary is in buckets 3 or 4, my model predicts a player victory.

I use my model to estimate an expected player and team win rate and compare this win rate – which is now based on the fairness of the two offers relative to the player's true value – to the actual arbitration hearing win rate for players and teams since 1990. This is the test of the arbitrator exchangeability hypothesis. If the expected win rate for each side is not statistically different from the actual win rate, then I am able to conclude that baseball arbitrators behave in accordance with this theory. However, if there is a significant difference between those two rates, then I conclude that baseball arbitrators do not adhere

to the arbitrator exchangeability hypothesis.

5 Results

In Table 6, I use the linear regression model with a window size of zero to report predictions for all arbitration hearings in 2022. "Winner" represents the observed winner in a hearing, while "Expected" represents the expected winner. Whether the arbitrator made the "correct" decision, based on model output, is shown in the final column.

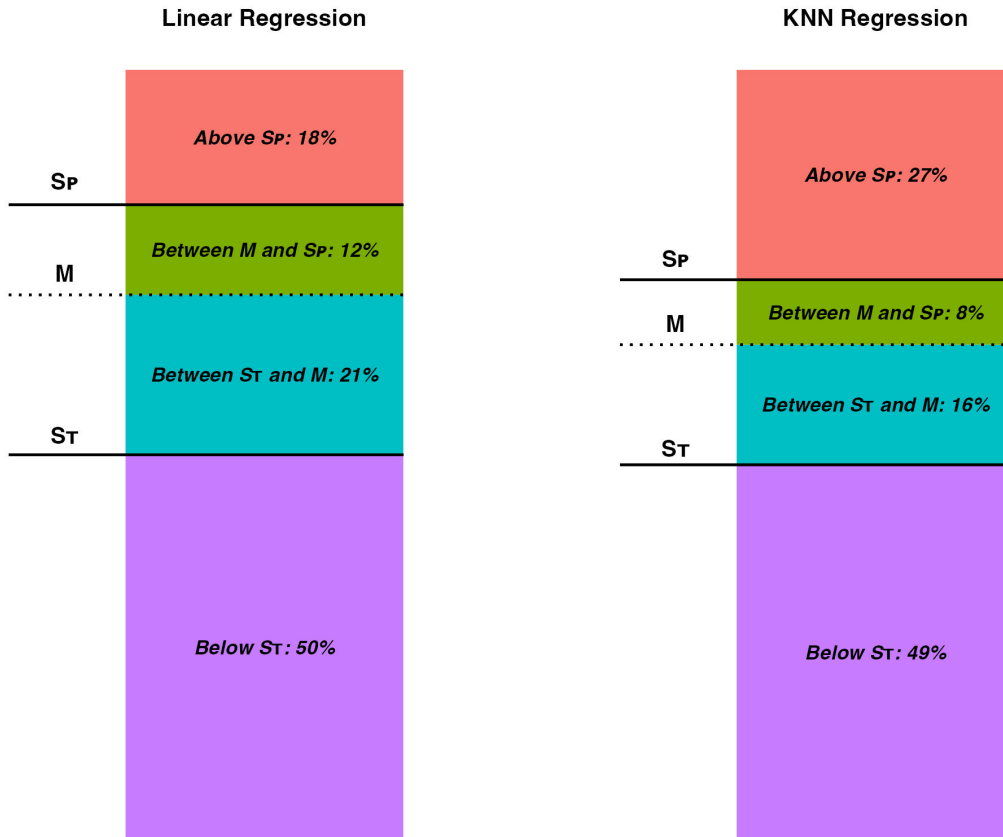Table 6: Baseball arbitration cases in 2022

Player	Team	S_P	S_T	W^*	Winner	Expected	Correct?
Adam Duvall	ATL	10,275,000	9,275,000	7,438,884	Team	Team	Yes
Dansby Swanson	ATL	10,000,000	9,200,000	7,249,509	Player	Team	No
Andrew Benintendi	KCR	8,500,000	7,300,000	6,182,832	Player	Team	No
Adam Frazier	SEA	8,000,000	6,700,000	9,011,782	Player	Player	Yes
Austin Riley	ATL	4,200,000	3,950,000	3,744,538	Team	Team	Yes
Tyler O'Neill	STL	4,150,000	3,400,000	3,351,502	Team	Team	Yes
Jacob Stallings	MIA	3,100,000	2,450,000	2,530,433	Team	Team	Yes
Nicky Lopez	KCR	2,950,000	2,550,000	2,968,247	Team	Player	No
Max Fried	ATL	6,850,000	6,600,000	7,244,530	Player	Player	Yes
Luke Jackson	ATL	4,000,000	3,600,000	2,034,804	Team	Team	Yes
Pablo Lopez	MIA	3,000,000	2,450,000	2,044,757	Team	Team	Yes
Adrian Houser	MIL	3,000,000	2,425,000	1,695,089	Team	Team	Yes
Lucas Sims	CIN	1,600,000	1,200,000	1,730,662	Team	Player	No

Note: position players above the dashed line; pitchers after.

In 2022, teams won nine of 13 hearings, and my model expected teams to win 10 of 13. My model assessed the arbitrators as having made the correct decision in nine of 13 cases, including in two of the four hearings players won. The model's estimate varies in its proximity to the player and team salaries. In some cases, like in Adam Frazier vs. the Mariners, the model predicted a salary higher than the player's request. In other cases, like in Adam Duvall vs. the Braves, the model predicted a salary lower than the team's offer. And, in others, like in Jacob Stallings vs. the Marlins, the model estimate lands between each of the two offers. In accordance with this pattern, for all hearings I assess how often (1) my model estimates a salary above S_P, (2) between M and S_P, (3) between S_T and M, and

(4) below S_T. If the model estimate falls in Group 1 or Group 2, then the model projects a player victory; if the model estimate falls in Group 3 or Group 4, the model projects a team victory. The results for the linear position player model with window size zero can be seen in Figure 3 (left panel).

Figure 3: Position player hearing estimates with window size zero

Linear Regression

Above S_P: 18%

S_P

Between M and S_P: 12%

M

Between S_T and M: 21%

S_T

Below S_T: 50%

KNN Regression

Above S_P: 27%

S_P

Between M and S_P: 8%

M

Between S_T and M: 16%

S_T

Below S_T: 49%

With 50 percent of model estimates coming below the team offer, and an additional 21 percent falling between the team offer and the midpoint, this model expects a total of 71 percent of position player hearings to have been won by teams (and, correspondingly, 29 percent by players). Even when varying the window size from zero up to four, the results are not qualitatively different. In Figure 5, in the appendix, the team's expected win rate falls between 71 percent (zero-year window size) and 78 percent (two- and four-year window

sizes).

I conducted the same analysis, evaluating how often the estimate falls into each group, using the k-nearest neighbors salary model, with similar results. Figure 3 (right panel) shows the results from the k-nearest neighbors model with window size zero. With this model and at a window size of zero, the expected team win rate is 64 percent. In the appendix, Figure 6 shows the results from the k-nearest neighbors model with varying window size; the expected team win rate varies from 64 percent (zero-year window size) to 72 percent (four-year window size). These are the same results qualitatively across window sizes.

Table 7 summarizes the above results, showing the expected team win rate based on model type and window size. Teams are expected to have won between 64 percent (k-nearest neighbors, window size zero) and 78 percent (linear regression, window size two and four) of position player arbitration hearings, depending on the model and window size chosen.

Table 7: Expected position player hearing results by model type and window size

Salary Model	Window Size	Expected Team Win %
K-Nearest	0	64
K-Nearest	1	68
K-Nearest	2	69
K-Nearest	3	69
Linear	0	71
Linear	3	72
Linear	1	72
K-Nearest	4	72
Linear	4	78
Linear	2	78

I conducted this same sensitivity analysis for pitcher arbitration hearings, evaluating how the results change depending on the window size. The results from the linear model and at a window size of zero are shown in Figure 4 (left panel), while the results with varied window size are shown in Figure 7 in the appendix.

Similar to the results from the position players, I find that teams are expected to have won between 72 and 80 percent of pitcher arbitration hearings when using the linear model. The results from the k-nearest neighbors regression model for pitchers is shown in Figure 4

Figure 4: Pitcher hearing estimates with window size zero

Linear Regression

Above S$_P$: 19%

S$_P$

Between M and S$_P$: 11%

M

Between S$_T$ and M: 17%

S$_T$

Below S$_T$: 53%

KNN Regression

Above S$_P$: 20%

S$_P$

Between M and S$_P$: 11%

M

Between S$_T$ and M: 16%

S$_T$

Below S$_T$: 54%

(right panel). Figure 8, in the appendix, shows the results of the k-nearest neighbors model with a varied window size. Depending on choice of window size, the k-nearest neighbors models expect teams to have won between 70 and 71 percent of arbitration hearings.

Table 8 summarizes these results, showing that, depending on the model choice and window size, teams are expected to have won between 70 and 80 percent of arbitration hearings for pitchers.

I applied the position player and pitcher linear regression models with window size of zero years to 323 of 327 arbitration hearings since 1990.[13] I estimate that teams should have won 227 of 323 hearings (70 percent) compared to the 186 observed team wins (58 percent). I reject the null hypothesis that these two win rates are identical, with $p <$

---

[13] Four cases were removed due to data limitations.

Table 8: Expected pitcher hearing results by model type and window size

Salary Model	Window Size	Expected Team Win %
K-Nearest	0	70
K-Nearest	4	71
K-Nearest	1	71
K-Nearest	2	71
K-Nearest	3	71
Linear	0	72
Linear	1	76
Linear	3	76
Linear	4	78
Linear	2	80

0.001. This result would suggest that Major League Baseball arbitrators are not behaving according to the arbitrator exchangeability hypothesis. Namely, I find that teams should be winning significantly more hearings than they are. Thus, the directionality of my results indicates that arbitrators actually have a *pro-player* bias, despite teams winning nearly 60 percent of hearings; per my results, players have won 41 more hearings than they should have. This 12 percentage point gap in expected versus observed win rates is also far larger than Ashenfelter's four-point gap between expected and actual employer win rates in Iowa (Ashenfelter, 1987, p. 345).

My model also assess the arbitrators as having made 166 correct decisions since 1990, representing 51 percent of all 323 hearings. Of the 166 correct decisions, teams were the winners in 128 (77 percent). Of the 157 incorrect decisions the arbitrators made, the teams won just 58 (37 percent).

6   Discussion

My results show that baseball arbitrators may be violating the arbitrator exchangeability hypothesis. With this established, I now evaluate (1) the assumption of my model's accuracy and (2) why baseball arbitrators may behave in a manner that is inconsistent with the arbitrator exchangeability hypothesis.

6.1   Evaluating the assumption of model accuracy

The most important question facing this analysis is whether the salary models are accurately estimating the player's true value. Each salary model is meant to capture the factors evaluated by the players, teams, and arbitrators at the time of the hearing. They are trained on past arbitration settlements, but only those known at the time the hearing was held. Even when varying the number of years of past precedent included (i.e., the different window sizes), the results do not change qualitatively. Additionally, the strength of the model predictions is in line with other models used to estimate player salaries; for example, Scully (1974) reports an R^2 of 0.81 for hitters and 0.78 for pitchers, while my average linear R^2 is 0.83 for hitters and 0.74 for pitchers. Lastly, my results do not change qualitatively even when using two different modeling techniques; if the assumptions of linearity are violated in the linear regression model, a non-parametric approach – such as the k-nearest neighbors regression – could potentially work more effectively. But this, too, did not yield qualitatively different results. No matter the specifications, all models expected teams to win far more hearings than they did; the most player-friendly model was the k-nearest neighbors model for position players at a window size of zero, and this model still expected teams to win 64 percent of hearings.

Increasing my confidence in the accuracy of these models is the fact that position player and pitcher models yield similar results despite using unique input statistics (reflecting the fact that position player and pitcher performance is measured differently). Because each of the salary models show that teams should win approximately 70 percent of cases *irrespective* of player type, this would suggest that the models may be more likely accurately capturing true player value than if the models yielded significantly different results based on player type.

Beyond predicting teams to have won 70 percent of arbitration cases, the models also suggest that teams made an offer that was *above* what estimated true value roughly 50 percent of the time. In these instances, this suggests that both teams *and* players valued the

player as being more valuable than what the model expected. On the surface, the results seem to indicate that teams are being generous and are offering well above what the player was actually worth. However, this may not be the case. I analyze three possibilities: 1) that teams are actually accurately valuing the players, and these results are consistent with that, 2) that there may be omitted variable bias driving down the estimates, and 3) that teams are actually indeed being generous with their final offers.

6.1.1 Teams might be making accurate offers

While the results of the model suggest that teams were offering above what the player was worth about half the time, the results also suggest that the teams offer the player *below* what he was worth the other half of the time. If anything, these results might be indicating that teams are virtually *nailing* the correct valuation of the player. Though Figures 3 and 4 demonstrate the four main regions in each case, I could evaluate the proportion of the time the model's estimate came in above or below the team's offer. Using the window size zero, linear regression models as an example, teams were above and below the model's estimate 50 percent of the time for position players and above the estimate 47 percent and below the estimate 53 percent of the time for pitchers.

Given that teams have large front office and analytical staffs (Carleton and Morrison, 2016), there is presumably a gap in statistical prowess between them and the player agents. Thus, it is far more likely that teams (rather than players and their agents) are making the most accurate offers relative to the player's intrinsic value. Baseball agents themselves have commented on this discrepancy in knowledge, noting that this "information gap" is particularly a problem in salary arbitration (Anderson, 2017). If this is the case, the salary models might not be underestimating player value at all; they might simply be reflecting a difference in understanding of the player's true value between the team and the player over the long term. Teams, with their statistical and analytical advantage, might be consistently offering salaries at the player's true value exactly, which is why the salary estimates are

above and below their offers roughly half the time. The players would appear to be asking for salaries that are too high.

6.1.2   Results may suffer from omitted variable bias

However, the theory that teams are making precisely accurate offers may not be the case; it remains true that both the team and player offered a salary above the estimate roughly 50 percent of the time, which could imply that the estimates are too low on average. This underestimation could be due to omitted variable bias. The models could be missing *something* about the players who go to arbitration hearings, that the group of players that go to arbitration hearings are significantly different in some meaningful and unknown way from the players who settle. However, given the fact that arbitration hearings are collectively bargained as being based on salary precedent, the models are being trained on the correct salaries, assuming that arbitrators actually do look to past precedent in their rulings. The model is fitting those salaries with a high degree of accuracy.

6.1.3   Teams may move toward more moderate offers

Another theory explaining why the model comes in below both offers 50 percent of the time could simply be that teams are, in fact, being "generous" and are raising their final offers above the player's true value. It is possible that teams prefer to offer a little more to the player in exchange for an increase in the probability of winning the hearing. As mentioned above, Farber (1980) argued that labor and management do not necessarily make offers that are equidistant from the true value. Farber argued that the more risk-averse party submits the offer that is closer to the arbitrator's notion of the fair award (p. 692). Perhaps, then, in the baseball case, teams are the more risk-averse party. Because teams and players are required to split the cost of an arbitration hearing equally, irrespective of which side wins the hearing, the player likely has more to lose from moderating his offer towards the middle. He is just an individual facing off against a valuable business that has far more financial

resources. This theory is supported by evidence in the media as well, with *ESPN* writer Jeff Passan detailing how teams "know and leverage" the fact that going to a hearing is expensive for the player's agent (Passan, 2019). "When the spread, or the difference between the sides, is minimal and the 5 percent fee on the difference won't come close to covering the attorney fees, the incentive [for the player] is clearly to settle," Passan writes. This would potentially indicate that when players *do* decide to go to hearings, they could be less likely to budge on their offer than the teams, given the fact that hearing costs are relatively high for them in comparison to the teams.

Teams might also potentially be more willing to take a loss on raising their offer in exchange of an increased win probability, since each player's next-year salary is based at least partly on his preceding year's salary. Keeping the player's salary lower in his earlier years of arbitration may yield a cascading cost savings over the entire arbitration-eligible period than if the player won with his higher ask. Because of this potential imbalance of power between the team and player, the team may have more of an incentive to submit a more risk-averse offer. Even still, however, teams appear to not be winning hearings as often as they should be.

6.2   Analyzing baseball arbitrators' pro-player bias

Assuming that my findings are accurate, they would suggest that arbitrators have a pro-player bias. There are three potential explanations for this phenomenon, beyond just a general affinity for players: 1) a small number of extremely pro-player arbitrators skews the overall result, 2) more-popular players are more likely to win hearings even when they should not, and player popularity drives this bias, or 3) arbitrators are motivated by appearing under the guise of being 50-50, which in turn results in them ruling in favor of the players even when they should not. I investigate all three of these possible theories and conclude that the third theory, a pro-50-50 bias, is the likeliest explanation for the pro-player bias.

6.2.1 Pro-player bias is consistent across arbitrators

In analyzing an overall pro-player result, it is imperative to understand if this bias is the result of one (or few) extremely pro-player arbitrators or if these effects are relatively uniform across all. Using data from Edmonds (2018), I evaluated the performance of individual arbitrators in cases from 1991 through 2018. As mentioned in Footnote 2, Major League Baseball has varied in the number of arbitrators assigned to each case over time. Even within certain years, some cases were assigned one arbitrator, while others were assigned three. Thus, to analyze whether any individual arbitrator is responsible for this overall bias, I simply analyzed the *presence* of each arbitrator on the overall case record. To evaluate bias, I evaluated whether the case result for each arbitrator fell into one of three groups, based on the window size zero, linear regression salary model estimate: 1) correct decision, 2) pro-team incorrect decision (team won when player should have), or 3) pro-player incorrect decision (player won when team should have). I compared the pro-team incorrect decisions to the pro-player incorrect decisions to determine each arbitrator's raw skew conditional on there being an incorrect decision made in the case.

Most arbitrators are not assigned to many cases, with just 22 arbitrators having been assigned to at least 10 cases from 1991 through 2018. Of these, 17 made more incorrect decisions in favor of the players than they did for the teams, three had an equal number of incorrect decisions in favor of players and teams, and just two favored the teams. Conducting a frequentist difference-in-proportions test to compare each arbitrator's incorrect team versus incorrect player rates, I found that six of the 22 arbitrators had a bias that was statistically significant at an $\alpha = 0.05$ confidence level, and all six of these arbitrators made more incorrect rulings in favor of players than they did in favor of teams. Table 9, reviewing the performance of the 22 arbitrators assigned to at least 10 cases from 1991 to 2018, is located in the appendix.

6.2.2   Pro-player bias is not due to player popularity

Having demonstrated that the presence of a few very-pro-player arbitrators is not what explains the overall bias, I turn to player popularity as the next explanation for why players win more hearings than expected. One possible omitted variable in my research is popularity. A potential argument here is that players with higher levels of popularity might be more likely to win a hearing even if their on-field performance does not indicate as such, that arbitrators see a player who is well-known or well-recognized and are therefore partial to that particular individual. Measuring popularity is not straightforward, but the size of one's social media following has been used to proxy popularity in other work (Garcia et al., 2017; Vergeer and Mulder, 2019). Some celebrities themselves have suggested that social media following is the 21st century version of the QScore, a metric developed in the 1960s to measure brand and individual familiarity and popularity (Andrews, 2022).

Thus, to briefly evaluate whether popularity might explain why there could be a bias, I analyzed players' Twitter followings among players who went to arbitration hearings from 2013 to 2022. I compared two groups: the size of the player's Twitter following in cases where the player won the hearing when he should not have and the size of the player's Twitter following in all other hearings (which include two groups: 1) correct decisions and 2) cases where the player should have won but the team did). From 2013 to 2022, there were 69 cases for which I could identify the player's Twitter account and follower count. The average number of followers for players who won when they should not have ($n = 20$ cases) was 95,884 ($\sigma = 117,676$). The average number of followers for players in all other cases ($n = 49$ cases) was 90,573 ($\sigma = 152,095$). A histogram showing players' Twitter following by group is located in appendix Figure 9. A difference-in-means z-test for these two groups was not statistically significant at an $\alpha = 0.05$ confidence level ($p = 0.877$). Therefore, when using Twitter follower count as a proxy to measure player popularity, I could not conclude that more-popular players won hearings when they were not expected to based on my model. Based on Twitter followers alone, I also could not conclude that this

omitted variable, popularity, explains why arbitrators might be biased toward players. I did not evaluate, however, if players who choose to file for hearings are more popular on average than those who choose to settle.

6.2.3   A pro-player bias may instead be a pro-50-50 bias

The third potential explanation presents what I refer to as a pro-50-50 bias. Though it is true that arbitrators have ruled more in favor of players than my models would expect, it is possible that this is not necessarily pro-player bias. Arbitrators could have a pro-50-50 bias, meaning that baseball arbitrators are only focused in ruling as close to 50-50 as possible across the population of cases. In this case, a pro-50-50 bias would manifest itself as a pro-player bias, given the fact that my model expects a team victory significantly more often than it does a player victory. While arbitrators might recognize that teams make better offers a large percent of the time, they could rule for players even when they probably should not according to the player's true value. This image of impartiality, then, ensures their future employment, as baseball arbitrators are typically judged based on the team-player win-loss record as opposed to how they ruled relative to the quality of the two offers.[14] As noted above, even Scully (1978) concluded in his early analysis of 13 cases that fairness in salary arbitration meant "rendering one-half of the decisions to the players and one-half to the owners without regard to the merits of the case" (p. 447). Given that arbitrators have ruled in favor of teams significantly more than half of the time (with a 344-262 record all time), it is possible that the spirit of Scully's statement remains true even though arbitrators do significantly rule in favor of one side more than half the time. What I am suggesting, then, is that arbitrators perhaps recognize that teams make fairer offers than players significantly more often, but in an attempt to *appear* unbiased, they keep the team's win rate from approaching a figure that would make them seem partial. Arbitrators

---

[14] As a brief example of how baseball arbitrators are judged in the media, the Associated Press cited the team-player win-loss record in each of the past three years without mentioning relative offer quality (Associated Press, 2022, 2021, 2020).

blend these two competing factors – that teams are making better offers 70 percent of the time but that their performance is judged based on the overall win-loss record – with the resulting outcome being a slight (and statistically significant) partiality toward the teams but one that does not fully account for how frequently teams make the better offer.

In his 13-case example, Scully shows that arbitrators made the correct decision five times, or in 39 percent of cases. My modeling technique suggests that arbitrators made the correct decision in arbitration cases 51 percent of the time, which suggests that arbitrators perform better than in Scully's assessment but still no better than flipping a coin. Both agents and team executives have shared that the arbitrators' decisions often seem to come off as random, and perhaps this is evidence to support that claim (Rosenthal, 2023). However, there must be some recognition on the part of the arbitrators that teams are making consistently fairer offers than the players; if the decisions were entirely random, then there would not be such an imbalance in the overall case record that we see, something that I have already shown is statistically-distinguishable from 50-50. My results indicate that there is likely some type of blend between true 50-50 randomness and a nod to the fact that teams make more fair offers.

Regardless of *why* arbitrators behave in this manner, the result of my research suggests a pro-player skew. One could speculate, then, that the league knows this behavior, given that MLB pushed heavily in the last round of CBA negotiations to replace arbitration with a formula-based salary system using FanGraphs' calculation of wins above replacement (Drellich and Rosenthal, 2021; Rosenthal, 2023). MLB has argued that eliminating salary arbitration would help to reduce the uncomfortable hearing process that pits players against their own organizations (Rosenthal, 2023), but it is fair to at least wonder if they are cognizant of the players' underlying advantage that is not evident from the overall win-loss record. Despite its "apparent" disadvantage in cases, the union has preferred to keep the arbitration system in place (Rosenthal, 2023). Perhaps these stances reflect a deeper understanding of the arbitrator behavior and who is actually benefiting from the current system.

Perhaps the union should be encouraging *more* players to file for hearings if there is such an advantage as my research suggests.

7 Areas for improvement and further research

This thesis is not the last word on baseball salary arbitration. While the k-nearest neighbors modeling approach and 323-case sample size both expand on the existing literature, other choices could result in a more successful salary model, which in turn could improve assessment of the baseball arbitrators.

For example, my thesis used a proxy for player arbitration-eligibility to select for the largest possible sample size of cases; future research might instead focus on arbitration cases in years where major league service time data is known. This could improve the accuracy of the modeling. It is possible that my data set of arbitration-eligible salaries includes salaries among players who were not arbitration eligible in a given year, which could bias the coefficients on the linear model or provide counterproductive neighbors for the k-nearest neighbors model. I assume that any data contamination of arbitration-eligible players who were not actually arbitration eligible is random and does not bias the results in any direction. Additionally, in using years where precise service time is known, a future salary model might use a player's precise service time as an independent variable – that is, if a player is at three years, four years, or five years of service – rather than my approach, which did not include a player's service time explicitly but instead used his previous salary (which would be correlated with service time but not perfectly associated).

This thesis only used arbitration settlements as the training for the model. Future models could incorporate both previous settlements as well as case precedent in the salary estimations. I did not do this because it would have reduced my ability to analyze as many cases as I did; if I was to include cases from 1990 to 2000, for example, then I would only be able to analyze cases from 2001 to 2022. In the future, another researcher may opt to analyze fewer

cases and include case precedent as well as arbitration settlements within their training set. However, if one was to take this approach, they could not include a blanket "had hearing" independent variable within their model, because it would not provide more information to explain the player's inherent value. If players made more money just from going to arbitration, or if arbitrators ruled in such a way that rewarded players for going to an arbitration hearing, then one would expect all players to file for arbitration in equilibrium. Because they do not, it must be the case that going to a hearing alone does not result in earning more money. That is a brief explanation as to why I did not include past cases into the training of the model, but it is a future consideration as long as the training and testing data remain separate and the scholar is not training a model that includes cases they are also analyzing.

This thesis leaves various questions unanswered. In my research, there was no analysis as to why certain players opt to file for arbitration; more research is necessary to determine if players who go to arbitration hearings are meaningfully different from players who settle. My research assumes that there are no differences between players who settle and players who go to hearings, but this assumption may not be valid. A violation of that assumption could be why the models consistently estimate salaries below the team's offer. This might be the result if players who go to arbitration hearings are better in some meaningful way from players who settle, while my model only estimates their true value based on settlements. One would assume, however, that this distinction would appear in their on-field performance statistics, all of which were included in the models. If nothing else, a further investigation into the game theory behind the negotiating process could be useful to develop a deeper understanding into the selection forces at play. Simply speaking, why do certain players go to arbitration hearings while others do not? This question is challenging to answer, however, because offers prior to the hearing are made in secret; the only information that is made public are the final two salary recommendations for the arbitrator.

Other research questions persist: Do certain teams go to hearings more often than others? Are some more stingy than others in the offers that they make? Do certain arbitrators rule

in favor of teams more often than others? Within a particular year, do arbitrators try to balance the case record; that is, if they rule in favor of one side early within a specific year, do they then try to rule for the other side even if those offers are of significantly worse quality? This analysis assumes no arbitrator-to-arbitrator differences and does not evaluate the differences between individual teams in hearing frequency or offer quality. It also does not propose a mechanism for *how* arbitrators may attempt to rule closer to 50-50. It simply posits that, because arbitrators tend to be judged in the press by the overall case record, they are cognizant of how far the case record strays from 50 percent to each side. Many of these other questions could be researched by future authors who are interested in the topic of baseball salary arbitration.

8   Conclusion

In this thesis, I have investigated whether the arbitrators in Major League Baseball's salary arbitration process behave according to the arbitrator exchangeability hypothesis. In MLB, players who have more than three, but fewer than six, years of major league service are typically eligible for salary arbitration. Most players and teams reach settlements to avoid arbitration, but in the event that salary negotiations do not conclude in an agreement, the two sides each submit salary offers to the league. Then, a panel of three independent arbitrators votes whether to award the player his salary request or the team's offer. This process, in which the arbitrators are not permitted to issue compromises, is known as final-offer arbitration and has been used to determine baseball salaries since 1974. Since that year, teams have won roughly 60 percent of all hearings.

Ashenfelter (1987) presented the arbitrator exchangeability hypothesis as a null hypothesis against which one can test the potential for bias among the arbitrators. The arbitration exchangeability hypothesis assumes that such arbitrators are statistically exchangeable, such that they do not have a consistent and predictable pro-labor or pro-management bias. Since

teams have historically won more hearings than players, it has been a bit of an open question as to whether baseball arbitrators adhere to this theory. In his foundational work on the topic, Scully (1978) argued that baseball arbitrators appeared to rule as closely to 50-50 as possible, without actually analyzing the facts of each case, rather than necessarily being directionally biased. Other authors, too, have suggested that baseball arbitrators judge fairness as a case record that reflects a 50 percent team, 50 percent player split. This is consistent with how arbitrators are typically judged in the baseball media, where a primary focus is placed on the case record between players and teams (Rosenthal, 2023; Associated Press, 2022, 2021, 2020).

The exchangeability hypothesis, however, dismisses the notion that a 50-50 case record indicates fairness in arbitrators' decisions. Rather, arbitrators should be assessed based on the quality of the offers. If one side consistently makes fairer offers more than 50 percent of the time, then that side *should* be winning more than 50 percent of hearings; the arbitrators would be carrying out their job responsibility appropriately. So, it is possible, then, that teams have made the better of the two offers more than 50 percent of the time, and that is why the historical case record reflects a tilt in favor of teams over the players.

Thus, to answer the question of potential arbitrator bias, I developed salary models for both position players and pitchers, using historical arbitration settlements as my training data. I used both linear regression to parameterize players' salaries based on their on-field statistics as well as k-nearest neighbors regression to find comparative players and assign a salary estimate based on the comparative players' salaries. Then, I applied these models to players who went to arbitration hearings and assessed the accuracy of arbitrator decisions. My models indicate that teams should have won roughly 70 percent of arbitration hearings since 1990, compared to the observed 58 percent they did win. This would suggest that arbitrators *do* have some bias and are not following the arbitrator exchangeability hypothesis. Additionally, my models assessed the arbitrators as having made the correct decisions in 51 percent of cases.

My models also project salaries that are below what the team offered and player requested roughly half the time. This might indicate that teams are correctly valuing the players. Or, this could be an underestimate due to some type of selection bias in the types of players who opt into arbitration hearings. Or, lastly, this could be due to teams being the more risk-averse party, offering salaries that are above the player's true value in an attempt to increase their case win probability.

If the model estimates are accurate, why arbitrators might violate the exchangeability hypothesis is unclear. It could be because arbitrators are attempting to preserve their employment. Arbitrators know they are judged publicly by nothing other than the case win-loss record, so they may feel pressure to rule as close to 50-50 as possible even while being cognizant of the fact that teams make better offers in the aggregate. The result might be a blend of the two factors arbitrators use in their rulings, that teams make better offers significantly more often than players (inducing them to rule more in favor of the teams) but that these decisions are closely followed and analyzed by those in the sports media, and that a large stray from 50-50 would result in them losing their employment. The end result is a significant benefit to the players, who have won more than 40 additional cases since 1990 than the linear regression model with window size zero expected. Perhaps players would benefit from more aggressively utilizing the arbitration system in future years, if baseball arbitrators do continue to operate in this manner.

References

2017-21 Major League Baseball Collective Bargaining Agreement (2022). https://www.mlbplayers.com/cba.

Anderson, R. (2017). How baseball agents are negotiating the gap in information and analytics in MLB. *CBS Sports.* https://www.cbssports.com/mlb/news/how-baseball-agents-are-negotiating-the-gap-in-information-and-analytics-in-mlb/.

Anderson, R. (2019). Everything you need to know from MLB's 2019 salary arbitration filing deadline. https://www.cbssports.com/mlb/news/everything-you-need-to-know-from-mlbs-2019-salary-arbitration-filing-deadline/.

Andrews, T. M. (2022). Whatever happened to the almighty Q Score? *The Washington Post.* https://www.washingtonpost.com/arts-entertainment/2022/10/07/q-score-what-is-it/.

Ashenfelter, O. (1987). Arbitrator Behavior. *The American Economic Review,* 77(2):342–346. Publisher: American Economic Association.

Associated Press (2020). Archie Bradley wins last arbitration case of the year; teams beat players 7-5 overall. *NBC Sports.* https://mlb.nbcsports.com/2020/02/21/bradley-wins-last-arbitration-case-of-the-year-teams-beat-players-7-5-overall/.

Associated Press (2021). Happ beats Cubs in arbitration; teams finish with 5-4 record. *Associated Press.* https://apnews.com/article/sports-mlb-baseball-ian-happ-mike-soroka-9c3aa2a62029ac7d33f31566203afb86.

Associated Press (2022). Max Fried beats Braves in arbitration, gets $6.85 million. *Associated Press.* https://apnews.com/article/mlb-sports-new-york-atlanta-baseball-a58766395dc08b7a9c3c01455b918690.

Baseball Prospectus (2022). CBA History. *Cot's Baseball Contracts.* https://legacy.baseballprospectus.com/compensation/cots/league-info/cba-history/.

Baseball Reference Bullpen (2022). Baseball statistics. *Baseball Reference.* https://www.baseball-reference.com/bullpen/Baseball_statistics.

Bazerman, M. and Kahneman, D. (2016). How to Make the Other Side Play Fair. *Harvard Business Review.* Section: Strategy.

Burger, J. D. and Walters, S. J. K. (2005). Arbitrator bias and self-interest: Lessons from the baseball labor market. *Journal of Labor Research,* 26(2):267–280.

Carleton, R. A. and Morrison, K. (2016). The Perils of MLB's Sorting System, Part 1: Where Do Front Office Workers Come From? *Baseball Prospectus.* https://www.baseballprospectus.com/news/article/29576/the-perils-of-mlbs-sorting-system-part-1-where-do-front-office-workers-come-from/.

Carrell, M. and Bales, R. (2013). Considering Final Offer Arbitration to Resolve Public Sector Impasses in Times of Concession Bargaining. *Ohio State Journal on Dispute Resolution*.

Cot's Baseball Contracts (2022a). Major League Service, 2007-22.

Cot's Baseball Contracts (2022b). Opening Day Salaries, 2000-22.

Cot's Baseball Contracts (2023a). MLB Arbitration, 2018-23.

Cot's Baseball Contracts (2023b). MLB Free Agency, 1991-2023.

Dolinar, S. and Chamberlain, A. (2015a). Modeling Salary Arbitration: Introduction.

Dolinar, S. and Chamberlain, A. (2015b). Modeling Salary Arbitration: Stat Components. https://blogs.fangraphs.com/modeling-salary-arbitration-stat-components/.

Drellich, E. and Rosenthal, K. (2021). Before free agency, MLB proposes paying players based on Fangraphs' calculation of WAR. *The Athletic*. https://theathletic.com/2950283/2021/11/11/before-free-agency-mlb-proposes-paying-players-based-on-fangraphs-calculation-of-war/.

Dworkin, J. B. (1986). Salary Arbitration in Baseball: An Impartial Assessment After Ten Years. *Arbitration Journal*, 41(1):63–69. Publisher: American Arbitration Association Inc.

Edmonds, E. (2018). Baseball Salary Arbitration. https://scholarship.law.nd.edu/baseball_arb_hearings_annual/.

FanGraphs (2018). 2018 major league pitching statistics.

Farber, H. S. (1980). An Analysis of Final-Offer Arbitration. *The Journal of Conflict Resolution*, 24(4):683–705. Publisher: Sage Publications, Inc.

Farber, H. S. and Bazerman, M. H. (1984). The General Basis of Arbitrator Behavior: An Empirical Analysis of Conventional and Final-Offer Arbitration.

Faurot, D. J. and McAllister, S. (1992). Salary Arbitration and Pre-Arbitration Negotiation in Major League Baseball. *ILR Review*, 45(4):697–710. Publisher: SAGE Publications Inc.

Fizel, J. (1996). Bias in salary arbitration: the case of major league baseball. *Applied Economics*, 28(2):255–265. Publisher: Routledge _eprint: https://doi.org/10.1080/000368496328894.

Fogarty, J., Baker, R., and Hudson, S. (2005). Case Studies in the use of ROC Curve Analysis for Sensor-Based Estimates in Human Computer Interaction. *Proceedings of Graphics Interface*, pages 129–136.

Garcia, D., Mavrodiev, P., Casati, D., and Schweitzer, F. (2017). Understanding Popularity, Reputation, and Social Influence in the Twitter Society. *Policy and Interest*, 9(3):343–364.

Hadley, L. and Ruggiero, J. (2006). Final-offer arbitration in major league baseball: A nonparametric analysis. *Annals of Operations Research*, 145(1):201–209.

Haudricourt, T. (2023). Corbin Burnes sounds off on Milwaukee Brewers' loss in arbitration. *Milwaukee Journal Sentinel*. https://www.jsonline.com/story/sports/mlb/brewers/2023/02/16/corbin-burnes-sounds-off-on-milwaukee-brewers-loss-in-arbitration/69910224007/.

Hoch, B. (2022). Judge's record deal with Yanks Official. https://www.mlb.com/news/aaron-judge-signs-nine-year-contract-with-yankees.

Kahrl, C. (2019). What you need to know about service-time practices in MLB. https://www.espn.com/mlb/story/_/id/26514863/what-need-know-service-practices-mlb.

Marburger, D. R. (2004). Arbitrator Compromise in Final Offer Arbitration: Evidence from Major League Baseball. *Economic Inquiry*, 42(1):60–68. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1093/ei/cbh044.

Pappas, D. (2004). Downloadable Data & Documents. http://roadsidephotos.sabr.org/baseball/data.htm.

Passan, J. (2019). Why record-setting money will fly in MLB salary arbitration. https://www.espn.com/mlb/story/_/id/25722707/jeff-passan-wild-wonky-world-mlb-salary-arbitration.

Petti, B. (2021). baseballr: Acquiring and Analyzing Baseball Data.

Rieders, R. (2015). A New Predictive Model for Determining Arbitration Salaries. https://community.fangraphs.com/a-new-predictive-model-for-determining-arbitration-salaries/.

Rome, C. (2019). Gerrit Cole wins salary arbitration case with Astros. *Houston Chronicle*. https://www.chron.com/sports/astros/article/Gerrit-Cole-wins-salary-arbitration-case-Astros-13613825.php.

Rosenthal, K. (2023). MLB salary arbitration a potential sticking point in CBA negotiations. *The Athletic*. https://theathletic.com/4230583/2023/02/20/mlb-salary-arbitration-cba-tension/.

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

Scully, G. W. (1978). Binding Salary Arbitration in Major League Baseball. *American Behavioral Scientist*, 21(3):431–450. Publisher: SAGE Publications Inc.

Steorts, R. (2017). Introduction to K-Nearest Neighbors. *Duke University*. http://www2.stat.duke.edu/ rcs46/lectures_2017/03-lr/03-knn.pdf.

Stevens, C. M. (1966). Is Compulsory Arbitration Compatible With Bargaining? *Industrial Relations: A Journal of Economy and Society*, 5(2):38–52. https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-232X.1966.tb00450.x.

Swartz, M. (2011). MLB Trade Rumors' Arbitration Projections. *MLB Trade Rumors*. https://www.mlbtraderumors.com/2011/10/mlb-trade-rumors-arbitration-projections.html.

Thompson, R. (2023). My thoughts concerning arbitration - Rays reliever Ryan Thompson on Twitter. https://twitter.com/R_Thompson15/status/1628585201973460992.

Topkin, M. (2023). Rays' Ryan Thompson learns about arbitration on Twitter thanks to Brusdar Graterol. *Tampa Bay Times*. https://www.tampabay.com/sports/rays/2023/02/23/rays-ryan-thompson-arbitration-twitter-brusdar-graterol/.

US Inflation Calculator (2023). Consumer Price Index and Annual Percent Changes From 1913 to 2008. https://www.usinflationcalculator.com/inflation/consumer-price-index-and-annual-percent-changes-from-1913-to-2008/.

Vergeer, M. and Mulder, L. (2019). Football Players' Popularity on Twitter Explained: Performance on the Pitch or Performance on Twitter? *International Journal of Sport Communication*, 12(3):376–396.

Weber, B. (2009). Gerald W. Scully, Who Wrote Landmark Baseball Analysis, Dies at 67. *The New York Times*. https://www.nytimes.com/2009/06/07/sports/baseball/07scully.html.

Yildirim, S. (2020). K-Nearest Neighbors (kNN) — Explained. *Towards Data Science*. https://towardsdatascience.com/k-nearest-neighbors-knn-explained-cbc31849a7e3.

Appendix

The following graphs show the impact of varying window size on linear and k-nearest neighbor regression models for both pitchers and position players.

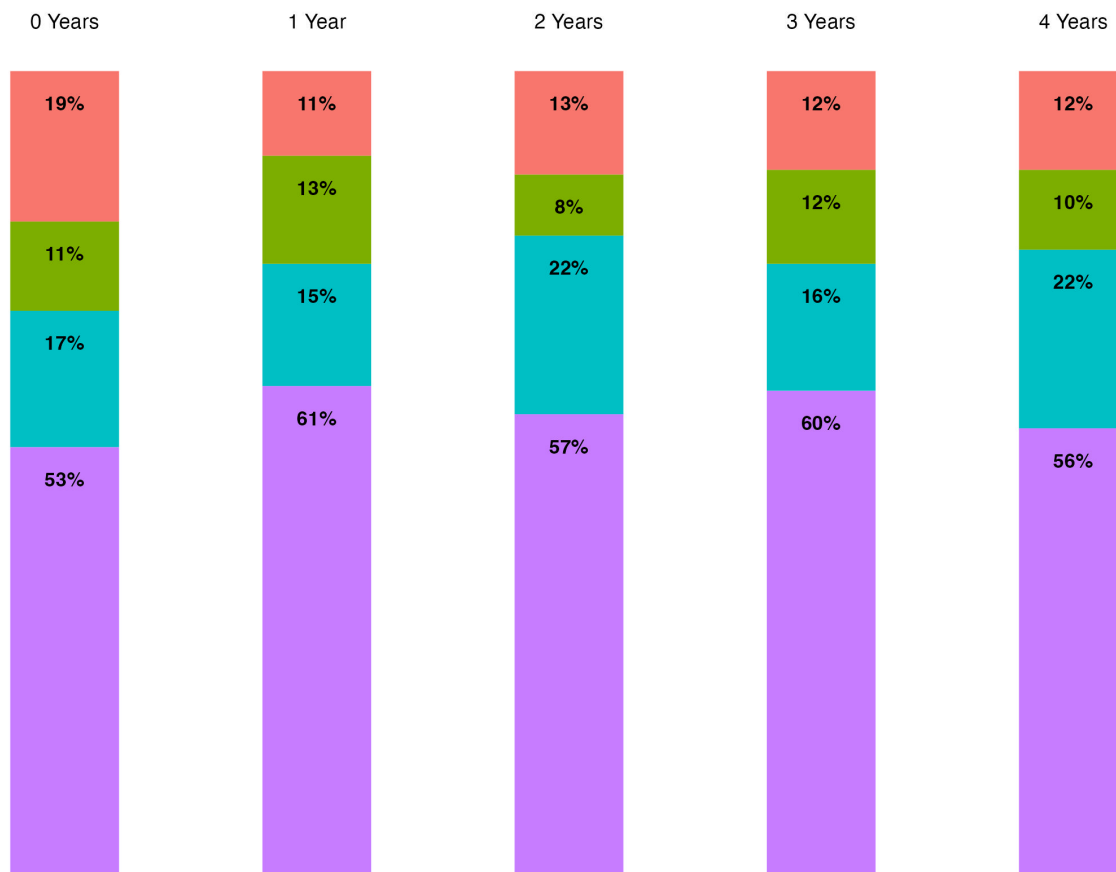Figure 5: Position player hearing estimates from linear model with varying window sizes
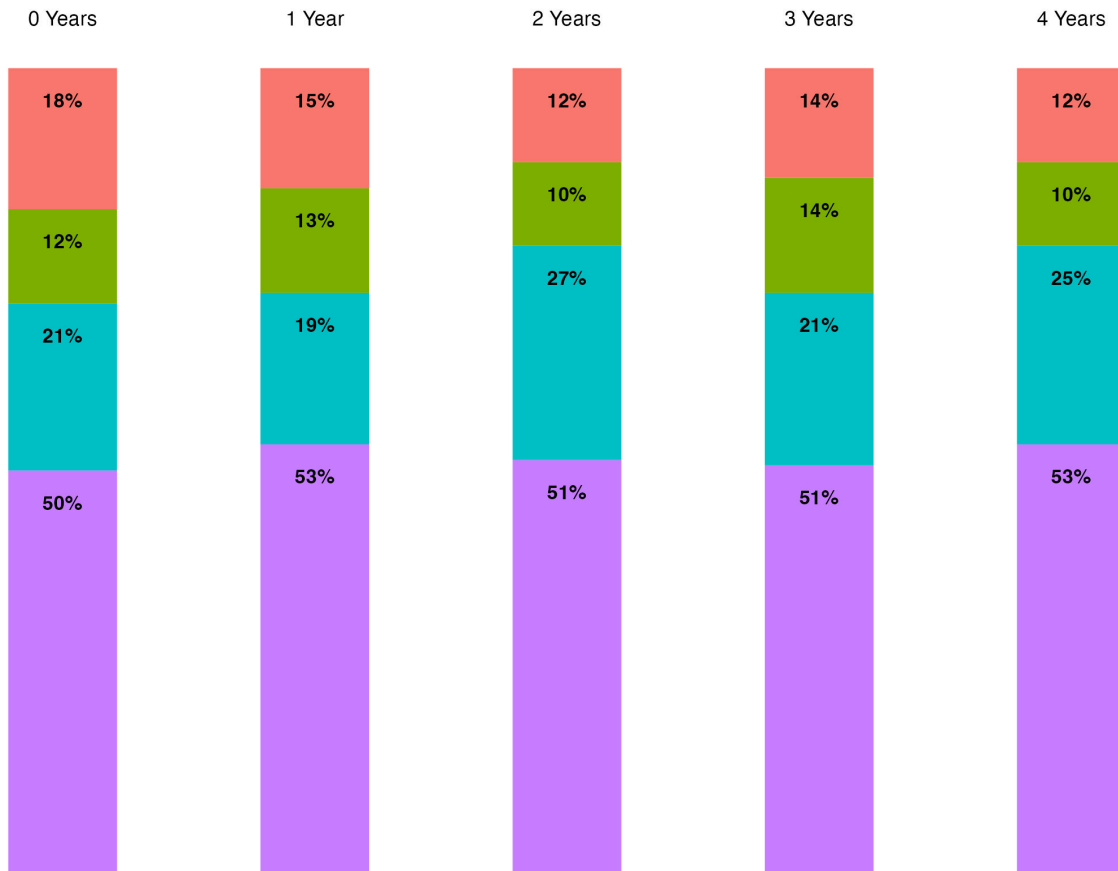
Note: The purple region represents the proportion of salary estimates below the team's offer, the blue region represents the proportion of salary estimates between the team's offer and the midpoint, the green region represents the proportion of salary estimates between the midpoint and the player request, and the red region represents the proportion of salary estimates above the player's request.

Figure 6: Position player hearing estimates from KNN model with varying window sizes
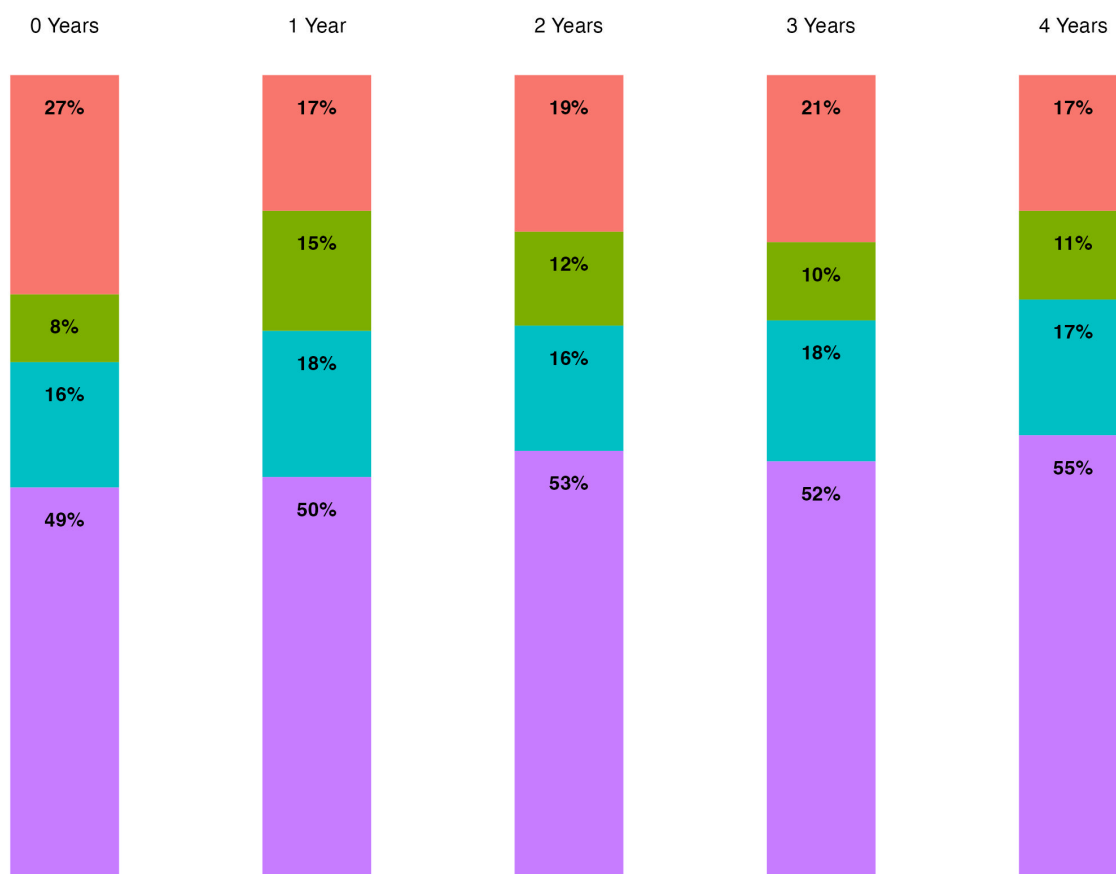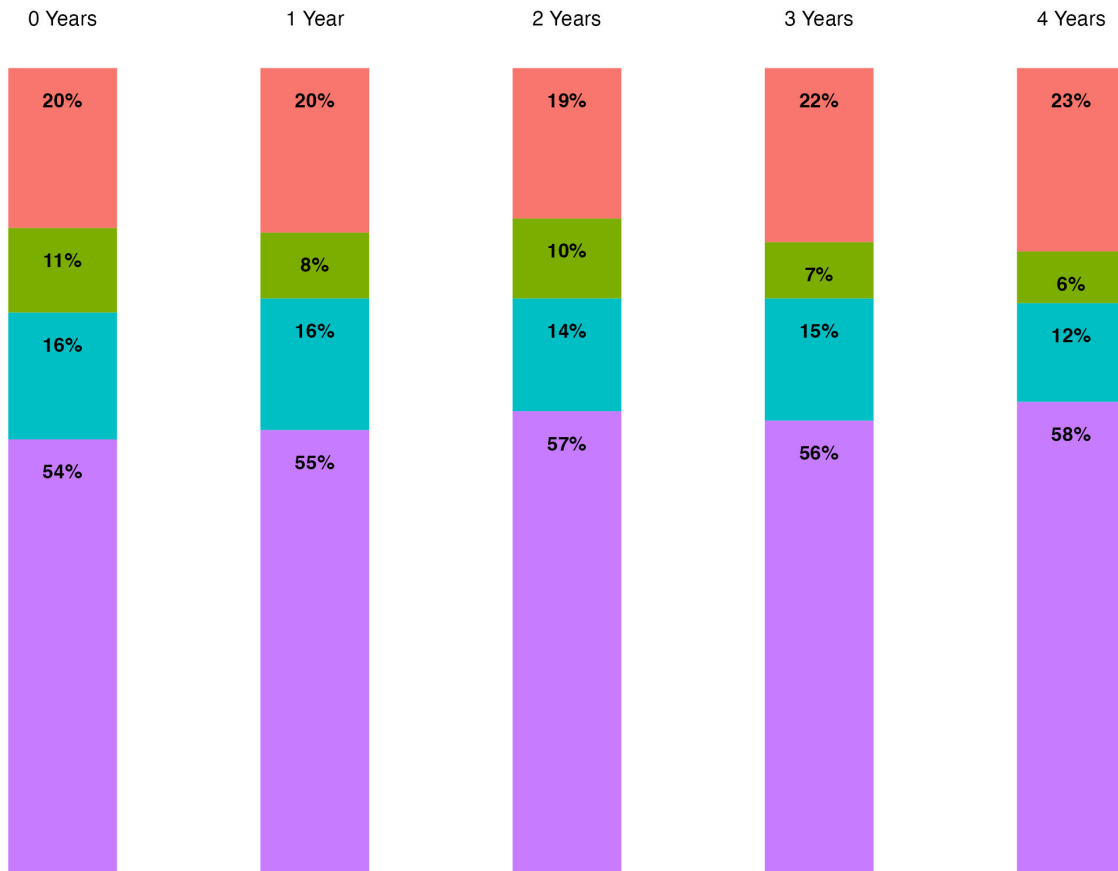

Note: The purple region represents the proportion of salary estimates below the team's offer, the blue region represents the proportion of salary estimates between the team's offer and the midpoint, the green region represents the proportion of salary estimates between the midpoint and the player request, and the red region represents the proportion of salary estimates above the player's request.

Figure 7: Pitcher hearing estimates from linear model with varying window sizes

Note: The purple region represents the proportion of salary estimates below the team's offer, the blue region represents the proportion of salary estimates between the team's offer and the midpoint, the green region represents the proportion of salary estimates between the midpoint and the player request, and the red region represents the proportion of salary estimates above the player's request.

Figure 8: Pitcher hearing estimates from KNN model with varying window sizes


Note: The purple region represents the proportion of salary estimates below the team's offer, the blue region represents the proportion of salary estimates between the team's offer and the midpoint, the green region represents the proportion of salary estimates between the midpoint and the player request, and the red region represents the proportion of salary estimates above the player's request.

This table shows the biases among individual arbitrators, with the case outcome being used in the assessment of their performance. "IP-IT" represents pro-player incorrect decisions minutes pro-team incorrect decisions, for a raw skew total.

Table 9: Skew by arbitrator, among arbitrators with 10 cases from 1991-2018

Arbitrator	Cases	Correct	Incorrect-player	Incorrect-team	IP-IT	p-value
Elizabeth Neumeier	39	21	12	6	6	0.046
Stephen Goldberg	28	12	8	8	0	1.000
Steven Wolf	28	14	10	4	6	0.023
Dan Brent	25	15	7	3	4	0.074
Robert Herzog	21	8	9	4	5	0.050
Elliott Shriftman	19	8	6	5	1	0.670
Gil Vernon	19	7	8	4	4	0.102
Jack Clarke	17	8	7	2	5	0.018
James Oldham	17	6	7	4	3	0.201
Margaret Brogan	16	6	6	4	2	0.371
Sylvia Skratek	16	7	8	1	7	0.001
Richard Bloch	15	9	3	3	0	1.000
Anthony Sinicropi	14	8	4	2	2	0.248
Mark Burstein	14	7	5	2	3	0.109
Roger Kaplan	14	6	6	2	4	0.046
Howard Block	12	7	4	1	3	0.058
Jerome Ross	12	10	2	0	2	0.046
Phillip LaPorte	12	6	3	3	0	1.000
Reginald Alleyne	11	6	3	2	1	0.527
Gary Kendellen	10	9	0	1	-1	0.157
I.B. Helburn	10	7	2	1	1	0.414
Nicholas Zumas	10	6	1	3	-2	0.157

Figure 9: Twitter following for 69 players who went arbitration cases from 2013-22
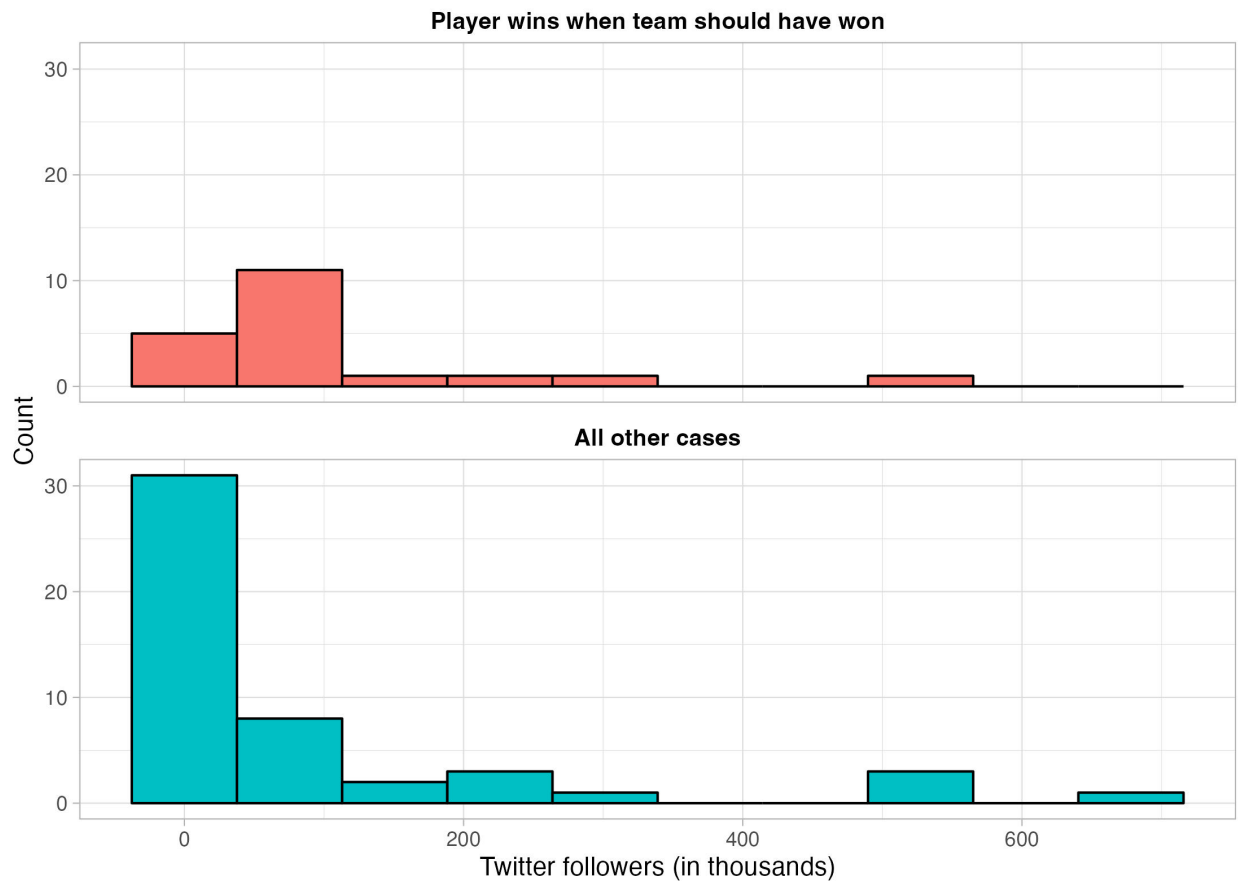
Note: *The two groups represent (1) arbitration hearings for which the player won when he should not have and (2) all other cases.*