8-2019

Essays on Pricing Mechanisms in Sport Economic Markets

Jeremy Mitchell Losak
Clemson University, jmlosak@syr.edu

ESSAYS ON PRICING MECHANISMS IN SPORT ECONOMIC MARKETS

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Economics

---

by
Jeremy Mitchell Losak
August 2019

Accepted by:
Dr. Raymond Sauer, Committee Chair
Dr. Paul W. Wilson, Committee Co-Chair
Dr. F. Andrew Hanssen
Dr. Scott Baier

Abstract

In the first chapter, I provide an empirical approach to answering a legal question in daily fantasy sports. A major legal question the past few years is whether daily fantasy sports is a game of chance or skill, and therefore whether it constitutes gambling. Results show that there are mispricings in DraftKings pricing mechanism, providing skilled players an opportunity to take advantage of this mispricing and improve their chances of winning. This provides evidence that certain elements of daily fantasy sports involve skill, and a long-run strategy exists for participants to win money. Results also show evidence of heterogeneity in skill level, and that skilled players target contests with particular settings.

In the second and third chapters, I switch focus to baseball arbitration. Over the past two decades, Major League Baseball teams have adjusted how they value certain player attributes based on how those functions aid in winning games. Salaries doled out in the free agent market have adjusted to better compensate for these attributes. For example, today power is valued less relative to the ability to reach base than it was in the pre-Moneyball era. However, the arbitration market, which uses previous arbitration cases to determine player salaries, has not seen this adjustment. Using the non-parametric free disposal hull estimator, the second chapter estimates the upper and lower bounds of contract zones for arbitration-eligible position players, and then uses a second-stage regression to determine which attributes improve players' relative contract position on their contract zone.

This chapter finds that power hitters are overcompensated and on-base hitters are undercompensated in arbitration.

In chapter three, I analyze the consequences of the arbitration mispricing in the context of a theoretical model. I expand current theoretical work in arbitration modeling by addressing steps of the arbitration process not previously explored. I then analyze how risk aversion, negotiation breakdown, arbitration panel valuation uncertainty, and release fee percentage impacts outcomes. Results support conclusions from previous research into baseball arbitration, while providing a model framework for future research.

DEDICATION

To those who have assisted me both professionally and personally during my time at Clemson University. I could not have gotten to this point without you.

ACKNOWLEDGEMENTS

I am truly grateful to all of those who have aided in my research. My advisors, Dr. Raymond Sauer and Dr. Paul Wilson, have been instrumental, and their research expertise and academic interests are sprinkled throughout these chapters. Dr. Sauer has taught me how to think like a sport economist, which will carry me in my future professional endeavors. Dr. Wilson exposed me to a wide-ranging set of econometric tools. I had the pleasure of taking four of his classes during my time at Clemson University and the tools I learned in those courses play crucial roles in these chapters.

I am appreciative to everyone else at Clemson University who provided feedback on my research. I am especially thankful to Dr. F. Andrew Hanssen who has provided excellent feedback and has been a great sounding board throughout my time at Clemson. I am also thankful to Aspen Fryberger and Shannon Graham. Aspen's contributions were instrumental to the first chapter, providing coding guidance and instructing me on how to operate the Palmetto Cluster. Shannon provided copy edits for every draft of my three chapters and consistently offered thoughtful feedback. Finally, I would like to thank Benjamin Posmanick, Jacob Walloga, and Jason Long for their copy edits and friendship along the way.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

DAILY FANTASY SPORTS: A GAME OF CHANCE OR SKILL?

1.1 Introduction

Over 59.3 million people from the United States and Canada have participated or are currently participating in a fantasy sport contest, with the average fantasy player spending $653 annually.[1] Professional leagues are incentivized to help grow the fantasy sport sector as fantasy sport participants consume 40% more content after joining, 64% watch more live events, and 61% read more about sports.[2] With team revenues driven partially by ads and sponsorship dollars, higher participant engagement equates to higher revenue totals.

Daily fantasy sports (DFS) is a subset of the fantasy sport industry, with fiscal year 2017–18 worldwide annual handle—the total amount of money wagered—at $3.19 billion.[3] There are over 10 million users with DFS accounts, 4.7 million of which are active users. Per Vardhman (2017), the top daily fantasy sites, DraftKings and FanDuel, hold 90% market share in DFS. Those sites are backed by major corporations, leagues, media companies, and more.[4] This paper focuses exclusively on DraftKings' Daily Fantasy Football product.

[1] https://fsta.org/research/industry-demographics/

[2] https://betting-sites.me.uk/unstoppable-growth-fantasy-sports-infographic/

[3] New York State Gaming Commission 2017 Interactive Fantasy Sports Report

[4] Some major DraftKings or FanDuel investors include: Disney, Revolution Growth, The National Football League (NFL), Fox Sports, Time Warner, Comcast Ventures, KKR, capitalG, NBC Sports, Major League Baseball (MLB), and Shamrock

For an entry fee, DFS contestants are given equal amounts of virtual currency used to purchase players and compile lineups. Figure 1.1 shows a typical layout contestants see when selecting their lineups on DraftKings. Each player has a distinctive price or salary, which is assigned by the DFS organizer based on the player's expected point contributions. Players score points based on their real-life performances, and those points accumulate for the participants. The contestant wins a predetermined payout if his or her lineup outperforms a specific percentage of other contestant lineups.[5]

Prices are set after they are posted; they do not adjust for shocks. Some examples of shocks include: an injury or suspension announced, a report about the expected usage of a player, injuries to key members of the opposing defense, etc. This is unique in sport gambling markets as most gambling lines and spreads are impacted by consumer demand.[6] This heightens the importance of Draftkings' pricing mechanism. For Draftkings' pricing mechanism to be efficent, assuming their objective is to produce a mechanism that matches player prices with expected point contributions, the initially assigned player prices should fully reflect

[5]There are multiple game types where the winning threshold is different. For example, in double-up matchups, the top half of performing lineups receive the payoff. In triple-up, only the top third win.

[6]There are various methods to gamble on a sporting event. One can bet on the spread, line, odds, over/under, and more. When a shock happens, the prices for these gambling methods adjust to account for changes in expected results. They also adjust when consumers have strong preferences for single particular outcomes. For example, one of the more popular gambling methods is to bet against the spread. A gambling spread sets the handicap for a particular real-world event, and consumers bet on either a team to cover or to not cover it. Suppose the New York Giants were +7 against the Patriots, and the Patriots won 20-14; the Giants would have covered the spread since $20 < 14 + 7$. If there is heavy consumer demand for one side of the spread, odds makers may adjust the line to even out the money coming in on both sides. Under an efficient market, the probability of covering a spread should be .5.

the players' true values. If this is a game of chance, a player's expected returns would be perfectly computable based on their prices, and thus there is no strategy to consistently beat the market. Alternatively, if this is a game of skill, there are opportunities for skilled players to improve their probabilities of winning by taking advantage of mechanism mispricing.

The legal landscape for daily fantasy games varies dramatically by state. Nineteen states have laws on the books specifically allowing DFS. Eight states either have laws on the books banning DFS or have attorney generals who have effectively outlawed them. Three states have active legislation looking to regulate DFS. The remaining states are currently in a gray area where DFS has neither been expressly outlawed nor legalized.[7][8] Despite the uncertainty, DraftKings and FanDuel currently operate in 41 and 40 states, respectively.[9] The industry is very much in flux and those numbers are likely to have changed by the time this paper is published. Most of the legal debate is around the concept of chance versus skill. If DFS is a game of skill, then there must exist some strategy to 'beat the market'—win more than lose. Otherwise, a chance-based game would be mostly random, where the expected long-run winnings are zero (or negative). If such a strategy to beat the market does not exist, the outcome is completely based on chance, and thus constitutes gambling.

Cabot and Miller (2011) opined that answering the chance versus skill question should adhere to the following methodologies:

---

[7] http://abcnews.go.com/Sports/daily-fantasy-sports-state-state-tracker/story?id=48138210
[8] https://www.legalsportsreport.com/dfs-bill-tracker/
[9] https://www.legalsportsreport.com/daily-fantasy-sports-blocked-allowed-states/

> *First, the effects-based analysis should compare the experience of average persons, without augmentation through experience or practice, with that of the most highly skilled players to determine the skill levels of the game. Second, the game should not be reviewed in isolation, but in the way it is being offered. For example, a single game of poker may be predominately chance-based, but a tournament may be skilled-based. Third, the results of a mathematical analysis of play does not need to result in the more skilled person winning virtually every time, but instead only a statistically relevant number of times in order to show that overall, in the particular game or format offered, skill is the predominate factor.*[10]

I address the chance versus skill question using a two-part procedure. First, I analyze 2016–2018 player pricing for DraftKings to see if its pricing mechanism is efficient compared to a market where market forces would drive prices such that a player's expected point contribution matches the price. I show evidence that certain player attributes are mispriced, providing opportunities for skilled players to increase their lineup's expected performance. Second, I analyze 2018 DraftKing contest results using two different non-parametric tests. The first is a test for stochastic dominance, which I use to analyze point score distributions between contest types. The purpose of this is to see if there are different skill levels and different strategies implemented for different contest settings. The second test is a test for multimodality. The existence of multiple modes in lineup score densities might suggest the existence of multiple player 'types', which could be interpreted as players with different skill levels. While I do not observe which participants are 'skilled', I can identify whether skilled players exist. Identifying the existence of skilled players and potential strategies for them to implement in contest selec-

[10] This excerpt from Cabot and Miller (2011) was pulled from Meehan *2015).

tion would support the idea that DFS are games of skill. I show there are skilled players that target certain contests to maximize expected profits.

Multiple legal journal articles have discussed the theory behind whether DFS is a game of chance or skill. Trippiedi (2014) differentiates between season-long contests and daily contests in that players in season-long contests can only be owned by one team so there is significant strategy in acquiring players that does not exist in the daily game. He argues that many of the elements that make season-long contests games of skill, such as acquiring sleepers and other pickups with the long-term in mind, are not prevalent in DFS. On the other hand, Meehan (2015) argues that DFS is a game of imperfect information where utilizing game theory can lead to consistent profits over time. Ehrman (2015) identifies managing one's bankroll as a skill in DFS. In particular, skillful players will have an understanding of their profit margins at various buy-in levels and league sizes. Both papers agree that DFS is a game of skill.

To my knowledge, the only paper that has applied an empirical approach to the chance versus skill question is Easton and Newell (2019). They use a linear programming lineup-based approach where they randomly compile a roster (for both football and baseball), and analyze how that lineup performs against a team selecting players based on a particular strategy (in football) and against other real-life teams (in baseball). They find convincing evidence that rejects the null hypothesis that DFS is a game of chance. However, they set a low standard for what constitutes as skill, thus rejecting the possibility that some baseline level of skill exists where chance becomes the deciding factor in whether a participant wins. If skill level is homogeneous among participants, chance becomes the predominant factor since nobody would benefit from any one particular strategy. In contrast, the player pricing approach used in this paper sets a higher standard for rejecting the null that it is a game of chance, looking at the market pricing mechanism to determine if there exists a strategy (skill) to beat the market. I look at particular selection strategies and player attributes and utilize a maximum likelihood econometric model to determine whether these strategies are properly incorporated into player prices. Finally, I check for heterogeneity in skill level. Even with a more stringent definition of what constitutes a skill-based contest, this paper ultimately supports the conclusions from Easton and Newell (2019) that DFS is a game of skill and not chance.

Section 2 discusses gambling legislation and how the chance versus skill question plays a role. Section 3 introduces both the logistic regression approach to determine DraftKing pricing efficiency and the non-parametric tests looking for the existence of skilled players. Section 4 discusses the data and the particular variables used in the models. Sections 5 and 6 identify key results from the two empirical approaches. Section 7 provides summarizing and concluding remarks.

1.2 Legal Background

Gambling legislation was first introduced as part of Chapter 50 of the Federal Wire Act of 1961.

> *Whoever being engaged in the business of betting or wagering knowingly uses a wire communication facility for the transmission in inter-*

> *state or foreign commerce of bets or wagers or information assisting in the placing of bets or wagers on any sporting event or contest, or for the transmission of a wire communication which entitles the recipient to receive money or credit as a result of bets or wagers, or for information assisting in the placing of bets or wagers, shall be fined under this title or imprisoned not more than two years, or both.*[11]

This law essentially banned interstate gambling and national online gambling operations. For example, a person in New York would not be allowed to receive payment from a gambling operation in New Jersey without physically travelling to that state to place the wager and collect the winnings. Thirty years later the Professional and Amateur Sports Protection Act of 1992 (PASPA) would restrict intrastate gambling.

> *It shall be unlawful for—*
>
> *(1) a government entity to sponsor, operate, advertise, promote, license, or authorize by law or compact, or*
>
> *(2) a person to sponsor, operate, advertise, or promote, pursuant to the law or compact of a government entity, a lottery, sweepstakes, or other betting, gambling, or wagering scheme based, directly or indirectly (through the use of geographical references or otherwise), on one or more competitive games in which amateur or professional athletes participate, or are intended to participate, or on one or more performances of such athletes in such games.* [12]

Essentially, PASPA compelled states to outlaw sports betting. Nevada was grandfathered into the law, giving Vegas a monopoly on legal sport betting for over 25 years. In May 2018, the United State Supreme Court ruled that PASPA violated

[11]Federal Wire Act of 1961. Pub. 18 U.S. Code 1084—Transmission of wagering information penalties. Legal Information Institute https://www.law.cornell.edu/uscode/text/18/1084

[12]Professional And Amateur Sports Protection Act of 1992. Pub. 28 U.S. Code § 3702—Unlawful Sports Betting. Legal Information Institute

the 10th Amendment since it compelled states to pass (or in this case not pass) particular legislation.

The overturn of PASPA has opened the door for states to pass their own rules and regulations regarding sports gambling, including DFS. With that said, states still have to pass legislation and the Federal Wire Act still prohibits interstate gambling (including interstate online gambling). In absence of state legislation, the legality of such operations is unclear and up to judicial interpretations of state and federal laws already on the books.

In 2006, Congress passed the Unlawful Internet Gambling Enforcement Act (UIGEA) creating a carve-out for fantasy sports in the Federal Wire Act. Specifically, the law specifies that the term ‘bet’ or ‘wager’ does not include:

> *Participation in any fantasy or simulation sports game or educational game or contest in which (if the game or contest involves a team or teams) no fantasy or simulation sports team is based on the current membership of an actual team that is a member of an amateur or professional sports organization.*[13]

According to the UIGEA, this exception holds only if prizes and awards are made known before joining the contest, and if “all winning outcomes reflect the relative knowledge and skill of the participants.” The rational behind this carve-out was that fantasy sports are games of skill and not synonymous with gambling that constitutes chance.

While this exemption is well accepted for season-long fantasy contests, its

[13] Unlawful Internet Gambling Enforcement Act of 2006. 31 U.S. Code § 5362—Definitions. Legal Information Institute

interpretation is less clear for DFS. In season-long contests, players act like team managers by drafting rosters, making player transactions, and setting lineups over a full season. These leagues generally have entry fees and the money is kept in a prize pool where the person with the best performing team wins. DFS contests also have entry fees with a predetermined prize pool. However, these contests operate under a much shorter time frame and only involve building a single (per entry) lineup. While season-long player performance is relatively predictable,[14] a player's performance in a particular contest is much more variable. The obvious question then is whether there exists a strategy to beat the market, or if the large variation in actual player performance is simply due to chance. A follow-up question is whether there is an opportunity for skilled players to exist. If there are strategies for skilled players to incorporate and there is evidence skilled players exist, it can be concluded that DFS is a game of skill and should not be subject to the Federal Wire Act under the UIGEA exemption. The uncertainty in the answers to these questions has led multiple states to challenge the legality of DFS contests.

A more notable legal dispute took place in New York, one of the largest paid-entry DFS markets, where Attorney General Eric Schneiderman issued cease-and-desist orders to DraftKings and FanDuel. This set off a chain reaction of other state attorney generals looking into DFS. Following back-and-forth court proceedings that essentially maintained the status quo, and heavy lobbying efforts by the

[14]The word relatively should be emphasized here. While season-long player projections do a better job predicting season-long outcomes compared to daily projections predicting daily outcomes, there is plenty of variation in both.

industry, New York legislators passed legislation legalizing DFS.[15]

1.3 Empirical Approach

This study addresses the chance versus skill question in DFS in two stages. First, I examine the efficiency of DraftKings' player pricing mechanism using a player-versus-player logistic model approach. Second, I look for the existence of player skill "types" by examining the distribution of lineup scoring outcomes using non-parametric tests for stochastic dominance and modality.

1.3.1 Player-vs-Player Approach

When picking a linuep on DraftKings, participants select one quarterback (QB), two running backs (RB), three wide receivers (WR), a tight end (TE), a flex player (FLEX), and a Defense (Def).[16] For the flex position, participants can chose to start an additional RB, WR, or TE. Participants are given 50,000 units of virtual currency to buy players, with better players costing more to buy.[17] Since prices are assigned in increments of 100, many players are assigned the same price which is the basis of this player-versus-player model. If two players cost the same in a given week, then conditional on playing the same position, their expected point contributions should be equal.

[15] https://www.legalsportsreport.com/ny/

[16] Due to the uniqueness of the Def compared to the other positions, the position is excluded from the analysis.

[17] $\frac{dPrice}{dE(Performance)} > 0:$ the price of a player increases as their expected performance increases.

The outcome variable, y_i, is defined as whether a player scores more points than another player,

$$y_i = \begin{cases} 1 & \text{if player } m \text{ scores more than player } n \\ 0 & \text{otherwise} \end{cases}. \tag{1.1}$$

Matchup *i* is a combination of two players *m* and *n* that play the same position, $po_n = po_m$, and have the same salary in the same week, $s_n = s_m, w_n = w_m$. By conditioning on equal salaries, the efficient market hypothesis suggests that the participant should be indifferent between selecting either player *m* or *n* to their lineup.

Whether a player scores more is only known ex-post, so the expected probability that player *m* scores more than player *n* when selecting a lineup, $P(y_i = 1)$, is a latent variable. The latent variable, y_i^*, is interpreted as the log odds ratio of the probability of player *m* scoring more than player *n* given by

$$log\left(\frac{P(y_i = 1)}{P(y_i = 0)}\right) = y_i^* = \beta_{m,po}X_{i,m} + \beta_{n,po}X_{i,n} + \varepsilon_{i,m} + \varepsilon_{i,n}. \tag{1.2}$$

The variables of interest are contained in $X_{i,m}$ and $X_{i,n}$. These covariates include measures of experience, game-time weather conditions, team quarterback ability, strength of opposing defense, team offensive line ability, injury status, and position (for the FLEX model). Since each position has different objectives and roles, coefficient estimates are allowed to vary by position. For example, a running back is less likely to be impacted by an opposing pass defense than a wide receiver, so the adjustments made in their pricing equations should differ. Assuming that the covariates impact y_i^* equally for players *m* and *n*, I can set $\beta_{m,po} = -\beta_{n,po} = \beta_{po}$. After setting $\varepsilon_{i,m} + \varepsilon_{i,n} = \varepsilon_i$, Equation 1.2 becomes

$$y_i^* = \beta_{po}\left(X_{i,m} - X_{i,n}\right) + \varepsilon_i, \tag{1.3}$$

where $X_{i,m} - X_{i,n}$ is the difference in the covariates for players *m* and *n*. The error has a logistic distribution (Λ) given by

$$\Lambda(\varepsilon_i) = \frac{1}{1 + e^{-(\varepsilon_i)}}. \tag{1.4}$$

Using Equations 1.2 and 1.4 and setting $X_{i,m} - X_{i,n} = X_i$ yields the following probability model,

$$P(y = y_i) = (\Lambda[\beta_{po}X_i])^{y_i}(1 - \Lambda[\beta_{po}X_i])^{1-y_i}.[18] \tag{1.5}$$

Finally, maximum likelihood estimation is used to estimate the parameters from Equation 1.5. The log-likelihood function is

$$LLF = \sum_{i=1}^{n} y_i log(\Lambda[\beta_{po}X_i]) + \sum_{i=1}^{n}(1 - y_i)log(1 - \Lambda[\beta_{po}X_i]). \tag{1.6}$$

To allow for variation in the dependent variable, player *m* having a higher point total than player *n*, identifying which of the players is *m* and which is *n*

[18]Team fixed effects are not necessary since much of the team specific variation is contained in the covariates.

is chosen at random. This disperses point differentials ensuring the model coefficients are estimating the effects of the covariates on the general outcome of scoring more points. In a particular matchup, $y_i = 1$ if player *m* scores more points than player *n*. This holds true regardless of the scoring margin; player *m* can score one more point than player *n* or 30 more points than player *n*, but y_i still equals one. If the majority of the higher point differential observations were assigned such that $y_i = 1$, while the majority of the low point differential observations were assigned such that $y_i = 0$, estimates would no longer be estimating their impact on the probability of winning, but rather their impact on the probability of having a high or low scoring margin. Randomly assigning *m* and *n* within pairs eliminates this concern.

The null hypothesis is that none of the covariates impact the probability of winning since they are already accounted for in player salaries. The alternative hypothesis is that at least one covariate does impact the probability of winning and thus is not properly accounted for in the player salary. Under the null, DFS is a game of chance since there does not exist a strategy that allows for increased profit opportunities. Under the alternative, DFS is a game of skill since a skilled participant could take advantage of market mispricing.

A follow-up question from this approach is whether participants adjust their strategies to account for any mispricings. The more adjustments take place, the less of an advantage there is to exploit by skilled participants. However, participants adjusting would suggest they are incorporating skill in their lineup decisions. I test whether participants make adjustments using an ordinary least squares model regressing the dependent variable, usage percentage, on the covariates used in the previous maximum likelihood model.

1.3.2 Non-parametric Tests

Stochastic Dominance

DraftKings offers participants contests that vary in size, entry fee, and win condition. For size, the smallest contest in the sample has just 97 participants while the largest has 676,700 participants. For entry fee, the cheapest contest in the sample costs one dollar to enter, while the most expensive is $25. For win condition, there are multiplier contests, tournaments (also known as progression contests or guaranteed prize pool contests), and winner-take-all contests. Multiplier contests pay off to the top percentage of lineup scores. For example, the top half (payout 2x entry fee), third (payout 3x entry fee), tenth (payout 10x entry fee), etc, disperse about 90% of the pot with the remainder going to DraftKings. Tournaments pay more the better you finish, with the biggest prize going to the person that finishes in first place.[19] Winner-take-all contests are exactly what they sound like.

Better or more experienced players may play in different types of contests compared to beginners or newer players. They also may have a sense of where newer players may play, and thus where they may have a better opportunity to win.

[19]For example, in Week 11 there was a contest called ‘Casual NFL $3K First Down’ (contest ID: 64153809). This was a $1 entry fee tournament with 3,563 participants and the following payoff breakdown: 1st $150.00, 2nd $100.00, 3rd $70.00, 4th $50.00, 5th $40.00, 6th $30.00, 7th $25.00, 8th $21.00, 9th $18.00, 10th $16.00, 11th-15th $14.00, 16th-20th $12.00, 21st-30th $10.00, 31st-40th $8.00, 41st-50th $7.00, 51st-100th $6.00, 101st-150th $5.00, 151st-300th $4.00, 301st-450th $3.00, 451st-700th $2.00, 701st-3,567th $0.00.

More-skilled players may also be more likely to play in higher stakes contests.

Contests with higher-skilled participants should have higher contest scores on average. Using the test for stochastic dominance, I test to see if contests with certain settings produce higher scores than other contest types. For example, I test if higher entry fee contests have lineup score distributions that stochastically dominate lower entry fee contests.

Econometrically, I am interested in weak first order stochastic dominance of G over F, where G and F are lineup score cumulative distribution functions, which corresponds to testing if

$$G(z) \leq F(z) \ \forall z \in \mathbb{R}. \tag{1.7}$$

Stochastic dominance means that at any point total, z, the probability of scoring at most that many points will be greater in F than G. For the test for first order stochastic dominance, the only noteworthy necessary assumptions are that F and G are continuous functions, and the samples are independent and randomly drawn from the population distributions F and G. When testing for stochastic dominance, the null hypothesis is that G does stochastically dominate F for all points z, while the alternative is that G does not stochastically dominate F for at least one value of z.

Barrett and Donald (2003) propose a first order stochastic dominance test statistic,

$$\hat{s}_j = \left(\frac{nm}{n+m}\right)^{1/2} \sup_z(\hat{G}_m - \hat{F}_n), \tag{1.8}$$

where n is the sample size for distribution F, m is the sample size for distribution G, and $\hat{G}_m$ and $\hat{F}_n$ are the empirical distribution functions given by,

$$\hat{F}_n = n^{-1}\sum_{i=1}^{n} I(X_i \leq z) \quad and \quad \hat{G}_m = m^{-1}\sum_{i=1}^{m} I(Y_i \leq z). \tag{1.9}$$

In the empirical distribution function, I is the indicator function and $\{X_i\}_{i=1}^{n}$ and $\{Y_i\}_{i=1}^{m}$ are independent random samples from distributions with cumulative density functions F and G, respectively. Using results from Billingsley (1968), and as described in McFadden (1989), it can be shown that

$$\hat{p} = exp(-2\hat{s}_j^2), \tag{1.10}$$

where $\hat{p}$ is the p-value from the test. These tests are conducted comparing the individual contests against each other, as well as using aggregated data by contest type. Figures 1.2a and 1.2b show examples of what the stochastic dominance test looks like when applied to the data. The vertical lines in both figures illustrate the point values (z) that maximize the difference between $\hat{G}_m$ and $\hat{F}_n$ in Equation 1.8. The top figure provides an example comparing lineups in Week 9 across two entry fees and rejects stochastic dominance in favor of no stochastic dominance. The bottom figure provides an example comparing contest types in Week 6 and rejects stochastic dominance in tournament contests over double-up contests, but does not reject stochastic dominance in double-up over tournaments.

The final test in the section uses the stochastic dominance test to compare the actual contest point distributions for Week 1 of the 2018 NFL season against a subsample from the set of potential lineups from the population. There are 319,551,313,684 unique ways to have constructed a DraftKings lineup in 2018 Week 1 to add to a total salary of 50,000. Due to computational burdens, a random subsample of about 1.1 million combinations is drawn from the set of possible population combinations and used for this test. If DraftKings contests are truly games of chance, than one would expect participant lineups to be random draws from the population set of potential lineups. In that case, the population distribution would be the equivalent of a distribution of draws made by players with no skill. However, if these contests do require at least some baseline level of skill, the contest distributions should stochastically dominate the population distribution.

Bimodal Test

Another relevant feature of the lineup score densities is the number of modes. In other applications, such as in Haruvy, Stahl, and Wilson (2001), the existence of multimodality has been used to suggest evidence of multiple "types" in population data. In this application, there may be participants of differing skill levels (high and low type), experience (beginner and expert), or knowledge level (avid fan and casual fan). The bootstrap test for multimodality could provide evidence that these types exist in the population.

For any given contest, I test the null that the lineup score density is unimodal versus the alternative that the density has more than one mode. Silverman (1981) proposes a methodology that produces a p-value to test this hypothesis.

First, I identify h_0, the smallest bandwidth for the epanechnikov kernel estimate such that $\hat{f}(\cdot|h_0)$, the sample density, has one mode. Unimodality is determined by examining the slopes of the sample density along a grid and determining the frequency at which the sign changes from positive to negative, or vice versa. In the case of unimodality, the sign should only change once from positive to negative.

Next, I draw a new sample of size n from the sample density, and find h_j, the smallest bandwidth for sample j such that the density estimate using the new sample has one mode. I repeat the process B times to obtain h_1,...,h_B. The p-value of the test is

$$\hat{p} = \frac{\#\{h_j > h_0\}}{B}. \tag{1.11}$$

For a multimodal density, it takes a sufficiently large bandwidth, h_0, to smooth it such that it only has one mode, and thus will be more likely to reject unimodality. This test is done for each contest and the results are broken down by the various contest settings at the 10%, 5%, and 1% significance levels.

1.4 The Data

All of the player performance and DraftKings salary data used in this paper cover NFL seasons between 2016 and 2018. Data on DraftKings contest lineups and results only cover 2018.

Since the covariates in the player-vs-player approach represent the relative differences between players *m* and *n*, each of these variables are calculated by taking the difference between the observations. They test whether player experience, quarterback ability, offensive line ability, opposing defense ability, injury, and weather conditions are properly factored into DraftKings' pricing mechanism. Summary statistics for the covariates are in Table 1.3. The middle columns provide summary information for the observations in the sample, while the right-most columns provide the average difference (in absolute value terms) in the covariates for each position.

1.4.1 DraftKings Price Data

Historical DraftKings data, including prices and points scored, come from RotoGuru.[20] Prior to the contest becoming public, DraftKings sets player prices. Once published, these prices do not change, even if the player is ruled out for the game. Fantasy participants have 50,000 units of virtual currency to spend on their lineups, with player prices ranging from 2,500 to 10,100.

Table 1.1 provides total salary ranges by position. Each position has a designated minimum salary. The minimum QB salary was changed from 5,000 to 4,000 after 2016, but for the purposes of identifying a minimum QB level, 5,000 will be used. Even after reducing it, the minimum QB salary is significantly higher than that of any other position. This is because even the worst QB will have a relatively decent projectable floor. Meanwhile, backup RBs and WRs are much more likely to score zero or few points, making them less worthwhile pickups. In fact, the only reason a participant would select any minimum salary player is because of the 50,000 unit salary cap constraint. A substantial number of player

[20] http://rotoguru1.com/cgi-bin/fyday.pl?game=dk

matches are located at the minimum salary levels, as illustrated in Figure 1.3. To account for this, additional models are included that exclude combinations with minimum salary players.

Players score points based on their real-life offensive statistics on the field. Table 1.2 lists each of the categories that are counted when calculating a player's score.[21] QBs have unique passing categories, but they can also score points based on their rushing production. RBs mostly score based on their contributions to the running game, but they can also score points by catching passes from the quarterback. WRs and TEs mostly score points based on their catching production.[22]

1.4.2 Teammate and Opposing Player Ability

Football statistics are very teammate dependent. The best wide receiver in the league will only make catches if his quarterback is able to throw the ball to him, and the quarterback might only be able to do that if the offensive line is able to provide him enough time to get rid of the football. This model addresses potential teammate related factors on offense that may impact fantasy performance.

Player ability is measured using grades provided by Pro Football Focus.[23] These grades attempt to factor out teammate ability and only measure the performance of that player.[24] More traditional quarterback performance metrics—

[21] https://www.draftkings.com/help/rules/nfl

[22] As an example, in Week 3 of the 2016 NFL season, Todd Gurley put up the following stats: 85 rushing yards (8.5 pts), 2 rushing TDs (12 pts), 1 reception (1 pt), -5 receiving yards (-0.5 pts). Therefore, Gurley scored 21 points that week.

[23] Website subscription is required to access the data: https://www.profootballfocus.com/

[24] The exact formulas Pro Football Focus uses to calculate these grades are proprietary, but these measures of player performance are well accepted by sport pundits and publications.

passing yards and passing touchdowns—and offensive line metrics—pancakes and sacks allowed—do a poor job as measures of individual player performance and ability since they are strongly dependent on the performance of players around them.

First accounting for the effects of the quarterback's ability on WRs, RBs, and TEs, the variables QB Overall Grade and QB Passing Grade measure the relative performance of the quarterback for player m versus player n. QB Overall Grade is the difference in the relative overall QB grades by Pro Football Focus, while QB Passing Grade only measures the difference in the relative passing QB grades. The overall grade includes passing ability, as well as rushing ability, discipline, and other QB attributes. The passing grade only includes the QB's ability to pass the football. Models are estimated using each measure of QB ability separately. Given the nature of the interaction between the QB and the other positions, QB Passing Grade was used in the WR and TE models, and QB Overall Grade was used in the RB and FLEX models. Both measures are averages of their individual game grades for the given season weighted by snaps per game, and the projected starting quarterback for a given game is the player that actually started that week.[25]

The offensive line also impacts the QB, WRs, RBs, and TEs. The linemen give time for the QB to throw to his WRs and TEs, and they also provide block-

[25]In almost all cases, the team announces their starting quarterback ahead of time. Except due to a last-second injury, rarely do teams deviate from their starting quarterback once it is announced. Also, teams rarely switch QBs during a game unless there is an injury sustained or significantly poor performance.

ing for RB carries. Pro Football Focus provides two measures of offensive line performance: a Run Blocking grade and a Pass Blocking grade. The variables Run Blocking and Pass Blocking are again relative measures. The higher player m's grade, the better his offensive line compared to player n. Offensive lines are made up of five or six players, so I averaged their individual grades over a full season, weighted by their snaps per game. The overall line score was a weighted average of the projected starters' season grades, weighted by the total number of snaps they were on the field for during the season. Full mathematical notation of the different averaging derivations is available in Appendix A.1.

The opposing defense also has an obvious impact on an offensive player's performance. Pro Football Focus provides overall team defense grades, as well as grades for particular defensive skillsets, including the ability to rush the passer, defend the run, and defend the pass (pass coverage). The ability to rush the passer measures defensive line performance on passing plays, while the ability to defend the pass measures the ability of corners and safeties. Grades were averaged first on an individual player basis for the season, and then on the projected starters for the game. The differences in relative skillsets are measured by Run Defense, Pass Rush, and Coverage Defense.

1.4.3 Player Experience, Injury Status, and Weather Conditions

The model accounts for player experience with data coming from Pro-Football Reference.[26] The attribute, League Tenure, is the number of years since the player

[26] https://www.pro-football-reference.com/

was drafted. The variable measures the difference in the number of years the players have been in the league. The more recent the draft year for player *m*, the smaller League Tenure becomes. Players that are in the league longer have more relative experience. Rookies have zero League Tenure.

Control variables are included for the injury status of the players. DraftKings does not offer prices for players they know ahead of time will not be playing, but they do include salaries for players that are either probable (Prob) or questionable (Ques) on the injury report.[27] If a player is ruled out after prices are released but before matchups begin, that player will score zero points for the week. It is assumed that the basic DFS contestant is aware of general injury status and will choose not to select that player if he is not going to play due to injury. The variables Healthy vs Ques and Healthy vs Prob capture the impact on the probability of scoring more points when one player in the matchup is healthy and the other is either listed as day-to-day (less serious injury) or questionable (slightly more serious injury). These variables should also capture the average decrease in performance for players that do play as a result of the injury.

It is assumed that two players that are both healthy or both injured have no distinct advantage over the other.[28] The variable Healthy vs Ques takes on a value of one if player *m* has an injury status of healthy and player *n* has an injury status of questionable, negative one if player *m* is questionable while player *n* is healthy, and zero otherwise. Health vs Prob is similarly calculated for healthy

[27] Injury data come from The Football Database https://www.footballdb.com/transactions/injuries.html?yr=2018

[28] This assumption is necessary since it is difficult to know from the data how debilitating the injury is for either player.

versus probable.

The final set of covariates includes measures for weather conditions. It is more difficult to pass the ball when there is heavy snow and wind. Two weather variables, Wind Speed and Bad Weather, control for the differences in relative weather conditions. Wind Speed takes the absolute difference in mile per hour wind speeds in games played by player m and n, where it is expected that relative wind differences should negatively impact quarterbacks. Bad Weather takes a value of one if there was snow or heavy rain at the start of kickoff. All recent, historical NFL weather data is available on NFLweather.com.[29] For dome stadiums, Wind Speed and Bad Weather both take values of zero.

1.4.4 Contest and Usage Data

For the 2018 NFL season, I obtained data from over a thousand DraftKings contests. Each contest file includes data on participant usernames, players selected for each lineup, the percentage of lineups players were picked for, lineup point totals for each participant, and the payout format for the contest. I also have data on the number of participants for each contest, each contest's win condition, and each contest's entry fee.

After Week 4, I switched my data collection procedure, which allowed me to capture significantly more contests. Table 1.4 provides a breakdown of the number of contests in the sample by week and by various contest attributes. Average point totals vary since certain NFL weeks are higher scoring, leading to

[29] http://www.nflweather.com/en/

more fantasy points. The contests captured only make up a fraction of the total contests hosted by DraftKings.

Searching for evidence of chance versus skill, I analyze lineups based on whether they may come from 'skilled' players. To do so, I eliminate any contest lineup that either includes players who were ruled out for the game or is incomplete. Blank lineups are not representative of a player's skill level, and lineups that include players who are out would likely represent very low skill (or low knowledge) participants. When compiling population combinations for Week 1, those players are also removed so that the population lineups compare more closely with the contest lineups in the sample.

After checking to see if the pricing mechanism is efficient, I analyze whether DFS contestants adjust their selection habits based on the potential inefficiencies in the pricing mechanism. For example, if having a better QB makes it more likely for one RB to outperform another given equal salaries, when choosing between two equally priced RB DFS participants should be more likely to select the RB with the better QB. Usage data come from the contest data files and are averaged across contests by week for each player.

1.5 Player Pricing Results

Coefficient estimates impacting the probability of winning for each of the positional matchups for the full data set are in Table 1.5. Estimates for matchups excluding the minimum salary players are in Table 1.6. It should be noted that the sample sizes drop significantly between the two tables.

1.5.1 Player Experience, Weather, and Injury Results

In the full sample, League Tenure is negative and significant for QBs, WRs, and TEs. Since player tenure and age are highly correlated (players are drafted into the league around the same age), another way to interpret these results is that younger QBs, WRs, and TEs perform better than their older counterparts, at least among minimum level players. In the non-minimum sample, these results are nonexistant.

The RB experience coefficients are flipped compared to the other positions. Positive and statistically significant coefficients for League Tenure in both the full model and non-minimum model suggest that younger RBs are overvalued and older RBs are undervalued. With less career data available, younger RB (especially rookies) projections can be higher than their actual abilities, while older RBs have a more projectable outcome. There is also most likely survival bias here. RBs tend to have short lifespans in the NFL, so those minimum salaried ones that do make it past their first few years can probably provide added value when they play.[30]

In terms of the effects of weather on player performance, there are many takeaways for the full sample, but almost no evidence of mispricing in the non-minimum sample. The one consistent result is a positive and statistically signif-

[30]https://www.statista.com/statistics/240102/average-player-career-length-in-the-national-football-league/

icant coefficient on Wind Speed for quarterbacks. This is a peculiar result given that throwing the football with accuracy should be more difficult when it is exceptionally windy. The coefficients for Wind Speed and Bad Weather are negative and statistically significant for RBs in the full model, as one would expect them to perform worse when the weather is bad and it is windy. The same is true for wide receivers when it is windy, given that quarterbacks should have a more difficult time getting them the football. The coefficient for TEs is positive and statistically significant, so maybe quarterbacks substitute more difficult longer throws to WRs to less difficult, shorter throws to TEs when it is windy. Overall, the negative and statistically significant coefficients on Wind Speed in both tables for the flex position suggests increased wind speeds is generally bad for player performance.

Regarding injury for the full sample, results are inconsistent when comparing healthy players with those deemed questionable. For healthy versus probable players, results suggest that those labeled as probable are better picks. It is possible that without the injury designation, those probable players would have been priced higher, potentially greater than the minimum. Players labeled as probable are generally full participants and usually see no adverse affects from whatever injury gave them the designation. In the non-minimum sample, the only statistically significant coefficients come from RBs. Not surprisingly, healthy RBs perform better than questionable RBs, who, unlike probable RBs, are likely adversely affected by their injuries. This suggests avoiding RBs deemed questionable as their prices have likely not been reduced enough to compensate for their injuries. The same negative statistically significant coefficient on healthy versus probable suggests that probable RBs should not be penalized for their injuries as they are likely to be full participants and not effected by their injuries.

1.5.2 Team and Opponent Ability Results

Football is a team game, and as such, player production is directly impacted by the quality of teammates and opposing players. This subsection examines if the ability of the team's QB, offensive line, and opposing defense is properly priced into a player's DraftKings salary. Table 1.11 provides marginal effects for statistically significant results showing the change in the probability of player m outperforming player n if the variables change from having no difference to the average difference given in the sample.

First, there is evidence that the ability of the team's QB is not properly priced in DraftKings' pricing mechanism. In both the full and non-minimum RB models, the coefficients on QB Grade are positive and statistically significant. This would suggest that, all else equal, taking RBs on teams with quality QBs is a better bet. Better QBs are likely to have more opportunities in the red zone, which leaves for more opportunities to score touchdowns. In the minimum model, choosing the better QB increases the probability of selecting the correct player by 0.88%. That increases to 1.38% for the non-minimum model.

The coefficients flip between the two models for WRs. For the full sample, the coefficient is negative, suggesting that minimum-salary players with better QBs perform worse on average compared to players with worse QBs. This makes sense when considering the makeup of minimum-salary WRs. Receivers with good QBs are likely to not be at the minimum simply because of the potential for better performance. So the players likely at minimum salaries are those either with bad QBs but are decent or with good QBs but are awful. The model suggests the latter group does better on average. In the non-minimum sample, the coefficient is positive with an average effect of 1.16%, suggesting that overall quarterback ability is undervalued by the pricing mechanism.

Next, there is evidence that team blocking ability is mispriced for almost every position. While there is no statistically significant evidence in the full model, the non-minimum model shows that Pass Blocking is significantly overvalued for QBs. Choosing the QB with the better team pass blocking actually decreases the probability of that player outperforming an equally priced player by 7.75%. This means that QBs likely receive too large of a price boost for having an offensive line that does well blocking during passing plays. For RBs, Run and Pass Blocking are both negative and statistically significant in the full model, and Run Blocking is positive and statistically significant in the non-minimum model. This means Run Blocking is generally undervalued in RB pricing, but among minimum-salary players, it is better to go with better players with worse offensive lines than worse players with better offensive lines. The positive Run Blocking and negative Pass Blocking coefficients for TEs in the full model may suggest something about the way TEs are used to blocking depending on the quality of the offensive line, although the exact reasoning is unclear.

WR results for blocking are more peculiar. Pass Blocking is negative and statistically significant in the full model, but not statistically significant in the non-minimum model. Run Blocking is negative and statistically significant in the non-minimum model, but not the full model. The full model Pass Blocking story for WRs is similar to the Run Blocking story for RBs. For Run Blocking, if the team has a better offensive line, it will choose to run more, and thus there are less opportunities for receivers. This feature is not properly captured in DraftKings prices.

Finally, the most convincing evidence of inefficient pricing comes from the opposing defense results. The individual defensive components paint a clear picture of how opposing defense attributes are not efficiently factored into player prices. Run Defense coefficients are negative and statistically significant for QBs and WRs in both models. Being able to establish an effective run game makes it easier to pass the football, so the better the rush defense, the less effective the passing game will be, which negatively effects QBs and WRs. The same idea can be applied to the negative coefficient for Coverage Defense for RB. Limiting the effectiveness of the passing game has a negative impact on the running game.

That story also explains the negative Pass Rush and Coverage Defense coefficients for QBs and WRs in the full model. Improving either of those units makes it more difficult to pass the football, which suggests playing the matchups in the secondary when choosing a minimum salary WR or QB. However, the Coverage Defense coefficient is positive and statistically significant for WR in the non-minimum model. While playing the matchups are beneficial at the lower price level, it appears that Coverage Defense is underpenalized.

For RBs, the positive coefficient on Run Defense and the negative coefficient on Pass Rush in the non-minimum model indicate that DraftKings places too much focus on overall defensive ability when pricing RBs instead of emphasizing their ability to stop the run. Run Defense is underpenalized and Pass Rush is overpenalized.

The coefficients for TEs follow a slightly different narrative. The full sample has positive coefficients for Run Defense and Pass Rush, illustrating the role the TE plays when the team does not have an effective run game or when the QB is under pressure and needs a security net to target.[31] There were not enough observations to show evidence of mispricing for non-minimum TEs.

1.5.3 Splitting Results By Year

Up to this point, the results have been aggregated over three years of player matchups. Tables 1.7, 1.8, 1.9, and 1.10 split each of the models by year to see if mispricings are consistent or if they go away over time. Consistent mispricing suggests there are long-term strategies that can be applied to selecting lineups.

First, I will briefly discuss the results for the minimum level players. DraftKings is often updating its pricing strategy, which may result in varying estimates across years. For example, most of the statistically significant RB estimates in the full model vary across years including League Tenure, QB grade, Run and Pass Blocking, and Coverage Defense. The same holds true for League Tenure, Pass Rush, and Coverage Defense in the full TE model. Some of the estimates

[31] The TE is often called a QB's security blanket because they are generally available to make short completions when the QB has nobody else to target. Targeting the TE generally results in minor but positive yards.

do tend to zero. For example, Coverage Defense in the QB model, Run Defense in the WR model, and Run and Pass Blocking in the TE model all tend towards zero in 2018 after beginning statistically significant. The results that consistently return the same sign are the attributes where mispricing has and continues to exist and can be used strategically by participants looking for minimum level players. The full list of strategies to finding diamonds-in-the-rough is in Appendix A.2.

The results of greatest interest to DFS participants are the ones that pertain to the non-minimum players. In the aggregated sample for QBs, Pass Blocking and Run Defense were both negative and statistically significant. Both variable coefficients were negative when split out across the three seasons, but only statistically significant in one season. So while there is some evidence that those trends hold true across the three years, the small samples (120, 60, and 136 observations, respectively) make it difficult to definitively make those conclusions.

For RBs, in cases where there were weaker signs of statistical significance, the coefficients generally jump signs. This was the case for the QB Grade and Run Blocking Grade. There was some evidence that League Tenure was consistently positive, but 2018 was the only year the coefficient was statistically significant. Run Defense was consistently positive and statistically significant, and Coverage Defense was consistently negative and statistically significant. This supports the earlier conclusion that overall defense was valued too highly in DraftKings' pricing mechanism for RBs, while run-specific defense was not valued highly enough.

The WR results show evidence of adjustments being made by DraftKings. The Pass Rush coefficients jump around by year suggesting tinkering with the mechanism. Coverage Defense shows some evidence supporting the positive and statistically significant coefficient in the aggregate model, although only one season has a statistically significant result. Run Blocking and Run Defense both show signs of mispricing existing at one point, but Run Blocking has trended towards zero mispricing through the years, and Run Defense is still negative and statistically significant, but also trending towards zero. This means it may only be another season or two until DraftKings has Run Defense properly priced as it does now with Run Blocking.

To make clear, the probability model in the previous tables should not be used to make DraftKings wagers. For one, the model requires salaries of comparable players to equal each other. It is not clear what exactly the optimal prices should be. Second, there are almost surely additional characteristics that impact player scoring that are either unobservable or not included in this model due to data availability. Third, DraftKings is constantly changing their pricing mechanism, so the particular characteristics mispriced at this point in time may not be mispriced in future iterations of the mechanism.

1.5.4 Choosing Positions In The Flex Spot

Another strategic decision involves choosing which position to start in the FLEX spot. Table 1.12 provides the positional coefficient estimates from the FLEX model broken down by salary range. Since 3,000 is the floor salary for RBs and WRs, Table 1.12 does not include any minimum-salary players. The results show a shift in strategy depending on the amount of virtual currency the DFS participant plans to allocate to the FLEX position.

In the lowest range, 3,000–4,000, TEs appear to be the best bet in terms of improving probability of selecting a player that scores more points. WRs are also better than RBs in this range. This could be based on the opportunities available to a lower string WR versus a backup RB. WRs have more opportunities to make catches on the field, so the scoring potential is greater for the lower-level WRs. For TEs, there is not much differentiation between the production of mid-level and low-level TEs. Since their expectations are lower, there are starting calibar TEs assigned low salaries. So the player is more likely to score because there may be more starting TEs in this range compared to starters at other positions (and starting players will have much higher scoring potentials than backups).

In the 4,000–5,000 range, there is still slight evidence that WRs are more productive than RBs, although the difference between WRs and TEs has gone away. When extending to the 5,000–6,000 range, no position seems to have a distinct advantage over the other. From a positional standpoint, prices seem more efficient in this range.

Finally, in the 6,000 range, RBs are now statistically better than WR. The elite RBs produce the most points, and thus, at the higher salary ranges, picking a RB is better than picking a receiver. There were not enough TE observations in this range to produce statistically significant results, but just looking at the coefficients suggests elite TEs are worse bets compared to their equally priced RB counterparts. There is still no difference between WRs and TEs.

These results are consistent with what is seen in the data. Table 1.13 shows the average scoring breakdown by position for multiple salary ranges. In the lower ranges, TEs tend to score more points. In the middle, no position seems to have a distinct advantage. Then finally in the upper ranges, RBs seem to score the most.

1.5.5 Usage Analysis

Just because mispricing exists does not necessarily mean participants are taking advantage of the arbitrage opportunity. If the average participant does not take advantage of the mispricing, that provides more opportunities for skilled players to gain an advantage. Appendix A.3 provides results from an OLS model with the dependent variable being the difference in usage percentage for pairs of players. Table 1.14 takes the statistically significant results from the OLS table and shows in what direction (if at all) participants adjust their usage based on changes in the covariates included in the model.

I look to see how usage spreads change based on the characteristics we capture. In Table 1.14, the usage effects are captured for both the full sample and the non-minimum sample. The change in probabilities (P) from the previous 2018 models are included next to the usages (U) to see if they are moving in the same direction. The magnitudes cannot be directly compared since they operate under different units, but the table does show the relative size of the change based on the magnitude of the marginal effect (P) and the effect (U).

The majority of the coefficients in the usage OLS table did not return statistically significant. Where there are adjustments made by participants, the magnitudes of the effects are mostly negligible and do not match up with the mispricings. These results would suggest that participants are not recognizing DraftKings' mispricing.

This conclusion is even more noticeable in Table 1.15. The table summarizes the success DFS contestants have in picking similarly priced players for each position and breaks those percentages down based on the full model and the non-minimum model. If participants were choosing at random, one would expect the success rates to be around 50%. Some of the success rates are not statistically different than 50%, the full sample QB and RB results for example. That means the average participant is not using much skill in selecting these players. However, the non-minimum samples for both of those positions are statistically greater than 50%, suggesting that for the more relevant players, participants are incorporating some skill in selecting these players.

The results are flipped for WRs and TEs. The full sample success rates for both positions came back statistically different than 50%. However, for TEs, the success rate is under 50%, which means participants actually do worse than 50-50 at picking these players. The non-minimum WR success rate not being statistically different than 50% means the average participant is not particularly skilled at choosing non-minimum equally priced WRs.

So while there is mixed evidence, where in some cases, participants use skill and make adjustments, and in other cases, they do not make adjustments, there is clearly room for skilled players to improve their probability of winning. The last column in Table 1.15 shows how a participant would have fared in picking players had they fully taken advantage of the market mispricings discussed in this paper. Each sample is statistically greater than 50% and statistically greater than the participant success rate. These differences represent opportunities for skilled players to gain an advantage against other players and improve their win probability.

This section has shown that there are mispricings in DraftKings pricing mechanism, and that participants are not fully recognizing and adapting to them. This means there are systematic opportunities for skilled players to improve their win probability and potentially win more than lose. The next section will discuss whether skilled players actually exist in the DraftKings universe.

1.6 Contest Scoring Distribution Results

This section details the results for the stochastic dominance and modality tests. Combining the results from the two tests show that skilled players do exist in DFS, and they strategically participate in profitable contest settings.

1.6.1 Tests For Stochastic Dominance

Tables 1.16, 1.17, and 1.18 show results for various specifications of the stochastic dominance test. For each table, Panel A aggregates the lineups across contests into groupings based on the specified setting, and provides the percentage of weeks stochastic dominance is rejected between the groupings. A larger percentage indicates stochastic dominance of G(x) over F(x) (the group on the top of the table versus the group on the left side) is rejected more often. Conversely, a smaller percentage indicates stochastic dominance is not rejected more often. Stochastic dominance rejection is conducted at the 5% significance level. Panel B is similar except it tests each individual contest within the category against individual contests in the other category. Some of the contests have small sample sizes, so it is generally more difficult to reject than when using the aggregated results. Panel C provides two sets of sample sizes. Along the diagonal are the number of contests in the particular category. Off the diagonal are the number of unique sets of individual contest combinations between the two designated groups.

Table 1.16 examines stochastic dominance between types of contests. Tournament contests seem to reject stochastic dominance most frequently while double-up contests reject stochastic dominance least frequently. Triple-up contests reject stochastic dominance over double-up contests in 9 out of 16 weeks, while double-up contests reject stochastic dominance over triple-up contests in only 6 out of 16 weeks. While not definitive, this would suggest higher scoring takes place in double-up contests over triple-up contests. Both double-up and triple-up contests stochastically dominate quadruple-up and deca-up contests. These results are mostly supported by the individual contest results. The takeaway here is that contests with more total winners will have higher scores and possibly attract more skilled players.

Table 1.17 examines contests of different sizes. Tiny contests are defined as having less than 1,500 participants, small contests have between 1,500 and 10,000 participants, medium contests have between 10,000 and 100,000, and large contests have anything more than 100,000. As with contest type, there is a clear stochastic dominance pattern pertaining to contest size. The aggregated large contests reject stochastic dominance over tiny and small contests in all 16 weeks, and medium contests in 15 out of 16 weeks. Conversely, tiny and small contests fail to reject stochastic dominance in medium and large contests in 11 out of 16 weeks. There is some evidence that tiny lineups tend to outperform small lineups, but the evidence is somewhat weak with tiny rejecting stochastic dominance only half the time. Overall, it is clear that contests with fewer participants attracts higher scoring participants.

Table 1.18 examines contests with different entry fees. Three dollar contests reject stochastic dominance most frequently. One dollar contests fail to reject stochastic dominance in 12 out of 16 against three dollar contests, but otherwise reject stochastic dominance frequently against the other contest fees. On the other end, $25 contests fail to reject the most frequently. Overall, while the cheaper contests reject stochastic dominance more frequently, more expensive contests reject less. This means higher entry fee contests are accompanied by higher scoring.

Each of the contest categories show clear patterns as it pertains to higher scoring participants. When considering the chance versus skill question, it is helpful to consider what a contest score distribution might look like if lineups were selected at random. Table 1.19 takes a random subsample of the potential set of combinations in Week 1 that add to the 50,000 units and tests for stochastic dominance against the various groups. As the results show, stochastic dominance is completely rejected for the population over every category and failed to reject in every week for each category type over the population. I also tested stochastic dominance for each contest in Week 1 over the population. The smallest p-value among the contests was 0.974. The largest p-value among testing for stochastic dominance of the population over the various contest distributions was $3.22 * 10^{-56}$, virtually zero. This is clear evidence that the majority of participants are not selecting from random and have some baseline level of skill.

1.6.2 Tests For Modality

The modality test was done for each contest. The null for each test is a unimodal density for the contest lineup score density, and the alternative is a multimodal density. Table 1.20 provides the percentage of contests in various categories and significance levels that reject unimodality. Panel A analyzes the modality test by contest type, Panel B by number of participants, and Panel C by entry fee.

The results in Panel A show weak evidence of bimodal densities across contest types. Double-up contests have the highest rejection rates, and yet they are still below 25%. The results in Panel B show no linear relationship between contest size and the percentage of contests that reject modality. Tiny and large contests both have smaller rejection rates compared to small and medium contests. While small and medium contests reject unimodality at least 30% of the time at the 5% significance level, the main driver of those results come from one particular contest category, as will be discussed in the subsequent paragraph.

The clearest evidence of rejecting unimodality comes from $25 contests in Panel C. At the 10% significance level, unimodality is rejected in more than 53% of contests. At the 5% level, that drops to a still high 44%. As mentioned in the previous two paragraphs, double-up results rejected unimodality slightly more often than other contest types and small and medium lineups reject modality more often than the other two types. The last two columns in Panel C break down the fee tests conditional on those contests being double-ups. Double-up contests make up 65 of the 86 $25 contests in the sample, and reject unimodality in nearly 57% of contests. Of those 65 contests, 45 of them are classified as small. Even in the $1–$10 results, the percentage rejecting unimodality seems to go up with double-up contests.

The results in the previous subsection show that stochastic dominance was rejected less often for higher dollar contest amounts. Combining that result with the results from this section leads to the conclusion that more skilled participants enter higher stakes contests, especially double-up contests. Rejecting unimodality in a majority of these contest types suggests heterogeneity in player skill level in these contests. There are participants with more skill who drive up the contest averages, and the remaining participants have a normal, baseline skill level. There could be more than two types of participants, although that is not tested in this paper. Overall, these results suggest that there is heterogeneity in skill level, and that skilled players likely see their largest profit margins in higher stakes double-up contests.

1.7 Summary And Conclusions

Results in this paper provide clear evidence that DraftKings Fantasy Football is a game of skill, not chance. Although this paper does not test for this result in other DraftKings sport offerings, it is likely those sports and contests have similar mispricings, opportunities for skilled players to gain an advantage, and skilled players who win more often than lose. Therefore DFS should be covered by the exemption in the Federal Wire Act carved out by the Unlawful Internet Gambling Enforcement Act, making it legal in all states without competing legislation. The results for the logistic models in this paper provide evidence that certain aspects of DraftKings' pricing mechanism are not efficient relative to prices determined by open market forces. Analyzing usage data shows that these mispricings are not being accounted for by the average participant, providing skilled players with an opportunity to beat the market and improve their expected winnings, which rejects the efficient market hypothesis.

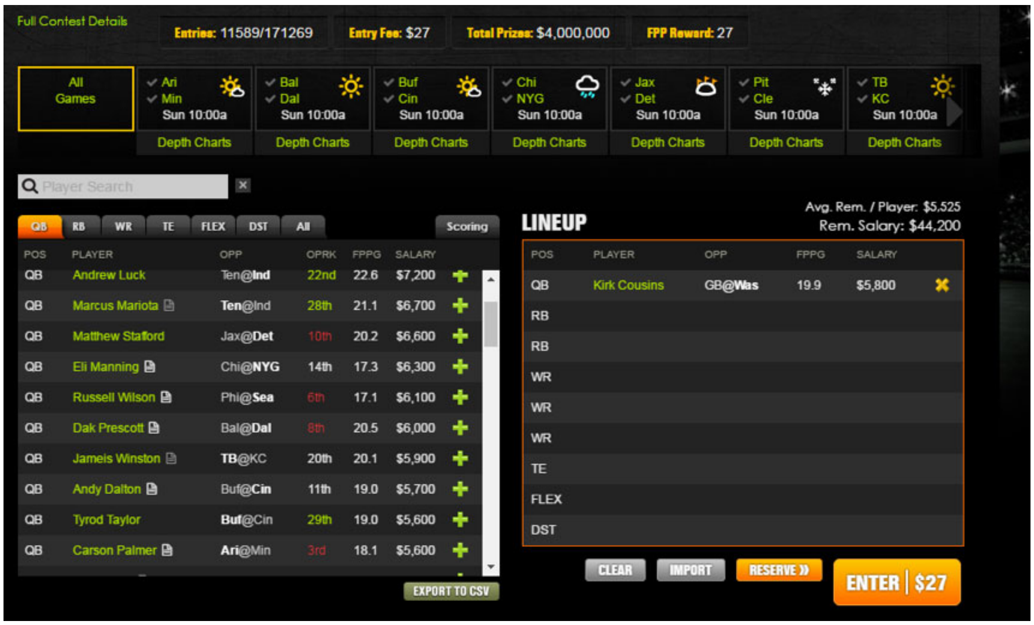
Non-parametric tests of stochastic dominance and modality show that the average DraftKings Fantasy Football participant has a baseline level of ability, supporting the results in Easton and Newell (2019), above what would be expected if participants were randomly selecting lineups. These tests also show heterogeneity in skill level, as skilled players exist and target particular types of contests that maximize their expected profits. Players with higher skill levels likely target some of the mentioned inefficiencies in DraftKings' pricing mechanism, along with incorporating other strategies not mentioned in this paper.

While results show heterogenity of skill level, they do not directly identify the actual skill levels of specific players. Overall contest skill level is inferred by the shape of the contest lineup densities. The results also do not specify to what extent prices should be adjusted to obtain an efficient mechanism, nor do they specify an exact percentage as to how the mispricing impacts a participant's probability of winning their contest. Rather, this paper only illustrates that skilled players exist, and there are arbitrage opportunities for them to improve their probability of winning. By doing so, the null hypothesis that DFS is a game of chance can be rejected in favor of the alternative that it is a game of skill. That is not to say there is not some chance involved. This paper does not directly measure how much chance exists, even when skill level is present and heterogeneous across players.

This paper shows that, even after allowing a reasonable baseline of skill, the null hypothesis can still be rejected. The act of giving players of different skill levels the opportunity to take advantage of inefficient pricing to improve their winning odds suggests that DFS is a game of skill, where players with more skill have positive expected earnings. This does not mean skilled players never lose. Rather, it shows that in the long-run skilled players should win more often than lose.

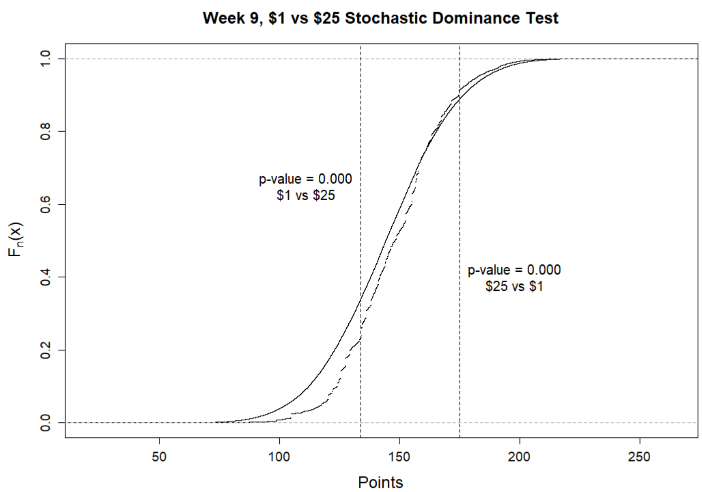
In the history of the NFL, only one team has ever gone undefeated. The better teams generally do not win all their games, but do win more often than lose. Yet most people would agree that NFL contests are games of skill; the teams with better skilled players and better coaches implementing superior strategy are going to win more often than not. As this paper has shown, the same holds true for DFS contests. There are participants with more skill and strategies they can incorporate to improve their probability of winning. Therefore, DFS should be considered a game of skill under federal law.

Figure 1.1: DFS contest layout



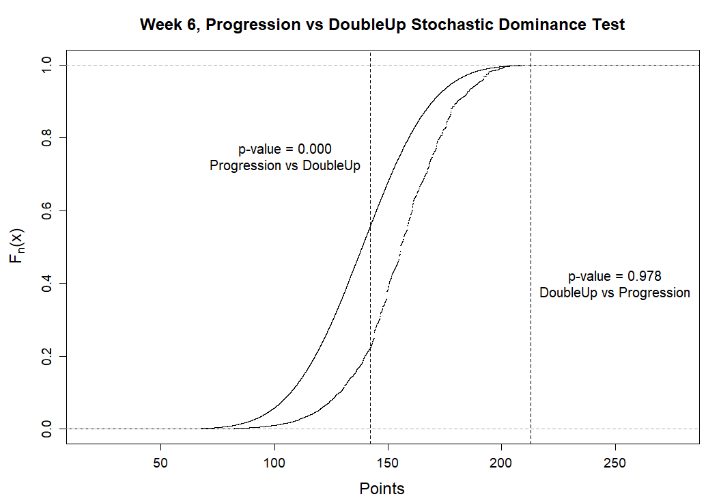
The figure is a layout of the screen DFS contestants see when selecting a lineup. The linuep consists of one QB, two RBs, three WRs, a TE, a FLEX, and a DST. Contestants select players by position, each player is given a salary, and each lineup is capped at spending $50,000 on salaries.

Figure 1.2a: Stochastic dominance rejection example



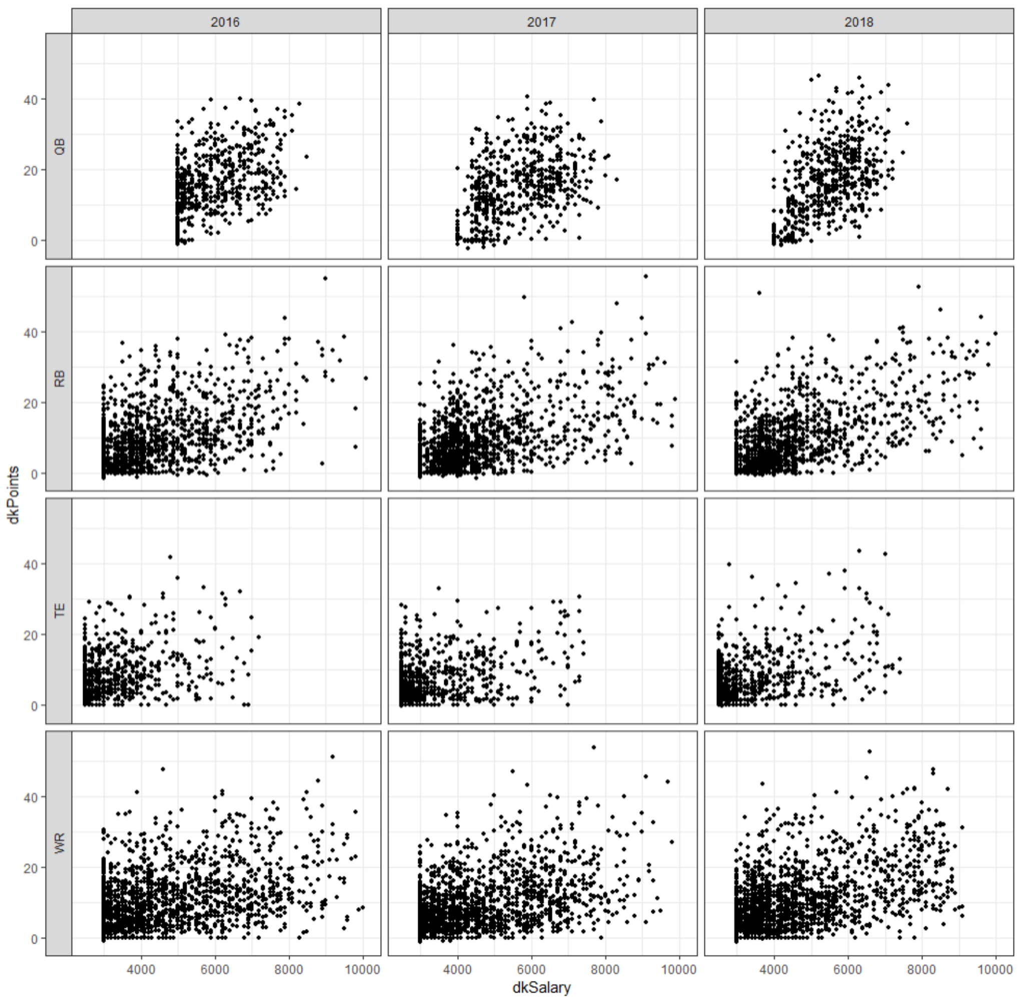
This is an example of a stochastic dominance test, examining $1 contest lineups and $25 contest lineups for Week 9. The test is conducted in both directions and the point total where the test statistic is calculated is represented by a vertical line with the p-value next to it. The more solid line represents the $1 contests while the other line represents the $25 contest.

Figure 1.2b: Stochastic dominance fail to reject example



This is another stochastic dominance test example, examining tournament contest lineups and double-up contest lineups for Week 6. The more solid line represents the tournament contests while the other line represents the double-up contest.

Figure 1.3: Player salary and points scored scatterplots



Above are scatterplots of player salaries and points scored by position and season. Notice that the majority of observations are clustered around the salary minimums. Also, notice that in most cases as the salary increases so to does the points scored.

Table 1.1: Salary Averages By Position

Position	Avg Salary (Min Excluded)	Min Salary	Max Salary
QB	5,593 (5,662/6,116)	4,000/5,000	8,500
RB	4,097 (4,728)	3,000	10,100
WR	4,147 (4,848)	3,000	10,000
TE	2,935 (3,635)	2,500	7,400
FLEX	3,821 (4,069)	2,500	10,100

Notes: Because there are a large number of players worth the positional minimums, averages are provided for the overall sample as well as for the non-minimum observations. The QB minimum salary changed after 2016 from 5,000 to 4,500, so minimum excluded averages are calculated using both cutoffs. Minimum salaries are set by DraftKings, but the maximum salary is the highest priced player for the position.

Table 1.2: DraftKings Scoring Settings

Scoring Categories			
Passing TD	+4 Pts	Rushing TD	+6 Pts
25 Passing Yards	+1 Pt	10 Rushing Yards	+1 Pt
300+ Yard Passing Game	+3 Pts	100+ Yard Rushing Game	+3 Pts
Interception	-1 Pt	Fumble Lost	-1 Pt
Receiving TD	+6 Pts	Special Teams TD	+6 Pts
10 Receiving Yards	+1 Pt	2 Pt Conversion	+2 Pts
100+ Receiving Yard Game	+3 Pts	Offensive Fumble TD Recovery	+6 Pts
Reception	+1 Pt		

Notes: This table contains the various DraftKings scoring categories for passing, rushing, receiving, and defense/special teams. DraftKings did not change their scoring settings during the sample.

Table 1.3: Covariate Calculations and Summary Stats

Type	Variable	Summary Data			
		Mean	SD	Min	Max
Experience	League Tenure	3.58	3.28	0	18
Quarterback	QB Overall Grade	66.63	7.03	25.30	83.58
	QB Passing Grade	65.62	7.05	25.60	82.87
Offensive Line	Run Blocking	61.56	3.35	53.58	69.51
	Pass Blocking	67.53	4.54	49.28	78.26
Opp Defense	Run Defense	63.33	1.67	57.12	69.05
	Pass Rush	63.08	2.46	55.77	72.08
	Coverage Defense	62.53	2.06	53.19	67.86
Weather	Wind Speed (mph)	3.40	3.31	0	18
	Bad Weather	0.02	0.14	0	1

Type	Variable	Mean Difference			
		QB	RB	WR	TE
Experience	League Tenure	4.78	2.63	2.68	3.04
Quarterback	QB Overall Grade		8.24		
	QB Passing Grade			7.76	7.18
Offensive Line	Run Blocking	3.03	3.03	3.16	3.05
	Pass Blocking	4.99	5.11	4.90	4.83
Opp Defense	Run Defense	1.55	1.80	1.80	1.80
	Pass Rush	2.55	2.79	2.54	2.65
	Coverage Defense	2.12	2.41	2.30	2.29
Weather	Wind Speed (mph)	3.40	3.32	3.23	3.28
	Bad Weather	0.04	0.04	0.04	0.04

Notes: Average absolute value differences are broken down by position. All player grades come from Pro-Football Reference. The summary information are averages for the players that make up the sample combinations.

Table 1.4: Summary statistics for collected DraftKings contests

Week	# Contests	Mean Score	Min Avg	Max Avg
1	40	157.75	153.22	203.66
2	35	137.77	132.95	138.97
3	37	137.58	126.21	151.21
4	31	160.85	154.61	202.75
5	161	134.12	131.30	146.84
6	164	140.39	137.46	163.41
7	162	133.53	121.98	136.53
8	159	138.99	134.80	147.35
9	159	145.26	142.73	154.83
10	162	130.82	124.96	138.47
11	159	129.04	121.04	141.02
12	123	152.50	151.28	167.54
13	162	138.46	135.71	154.88
14	158	137.69	133.87	146.93
15	166	115.73	113.83	127.57
16	95	146.94	143.66	170.00

Notes: The mean provides the average of all contests aggregated for the week. The minimums and maximums are the minimum contest averages and maximum contest averages among the collected contests.

Table 1.5: Aggregate Results: All Data, All Positions

	Dependent variable: Probability of Winning				
	(QB)	(RB)	(WR)	(TE)	(FLEX)
League Tenure	−0.044***	0.009***	−0.019***	−0.014***	−0.019***
	(0.009)	(0.003)	(0.002)	(0.002)	(0.001)
QB Overall Grade		0.005***			−0.003***
		(0.001)			(0.0004)
QB Passing Grade			−0.004***	-0.0002	
			(0.001)	(0.001)	
Run Blocking	−0.013	−0.022***	−0.002	0.021***	0.005***
	(0.016)	(0.003)	(0.002)	(0.002)	(0.001)
Pass Blocking	−0.011	−0.006***	−0.003***	−0.009***	−0.006***
	(0.009)	(0.002)	(0.001)	(0.001)	(0.001)
Run Defense	−0.101***	0.002	−0.016***	0.036***	0.009***
	(0.027)	(0.005)	(0.003)	(0.004)	(0.002)
Pass Rush	−0.030*	0.003	−0.030***	0.013***	−0.016***
	(0.016)	(0.003)	(0.002)	(0.002)	(0.001)
Coverage Defense	−0.064***	−0.015***	−0.006**	−0.006*	−0.003*
	(0.020)	(0.004)	(0.003)	(0.003)	(0.001)
RB vs WR					−0.219***
					(0.006)
WR vs TE					−1.053***
					(0.026)
RB vs TE					−1.269***
					(0.032)
Wind Speed	0.032***	−0.012***	−0.017***	0.012***	−0.005***
	(0.011)	(0.002)	(0.002)	(0.002)	(0.001)
Bad Weather	0.121	−0.166***	−0.025	0.103**	−0.063***
	(0.290)	(0.055)	(0.036)	(0.041)	(0.017)
Healthy vs Questionable	0.101	0.037	0.159***	−0.211***	0.026
	(0.318)	(0.044)	(0.031)	(0.033)	(0.016)
Healthy vs Probable	−0.503***	−0.171***	−0.002	−0.088***	−0.088***
	(0.123)	(0.039)	(0.021)	(0.023)	(0.011)
Constant	−0.163***	−0.567***	−0.455***	−0.620***	−0.526***
	(0.050)	(0.010)	(0.007)	(0.008)	(0.004)
Observations	1,741	40,324	80,164	69,732	318,939

*p<0.1; **p<0.05; ***p<0.01

Table 1.6: Aggregate Results: Non-Minimum Data, All Positions

	Dependent variable: Probability of Winning				
	(QB)	(RB)	(WR)	(TE)	(FLEX)
League Tenure	−0.019	0.022**	0.012	−0.041	0.004
	(0.019)	(0.009)	(0.008)	(0.042)	(0.003)
QB Overall Grade		0.007*			0.006***
		(0.004)			(0.0002)
QB Passing Grade			0.006*	0.031	
			(0.003)	(0.021)	
Run Blocking	0.043	0.017*	−0.023***	−0.097	−0.010**
	(0.037)	(0.010)	(0.008)	(0.067)	(0.004)
Pass Blocking	−0.070***	−0.00002	−0.003	0.057	−0.001
	(0.023)	(0.007)	(0.005)	(0.059)	(0.003)
Run Defense	−0.144**	0.063***	−0.059***	0.116	−0.021***
	(0.058)	(0.016)	(0.013)	(0.095)	(0.006)
Pass Rush	0.024	−0.056***	0.011***	−0.105	−0.016***
	(0.038)	(0.011)	(0.009)	(0.078)	(0.004)
Coverage Defense	0.011	−0.020	0.024**	0.053	0.002
	(0.049)	(0.013)	(0.010)	(0.077)	(0.005)
RB vs WR					−0.187***
					(0.023)
WR vs TE					−0.253***
					(0.042)
RB vs TE					−0.389***
					(0.049)
Wind Speed	0.059**	−0.003	−0.009	0.005	−0.009**
	(0.028)	(0.009)	(0.007)	(0.056)	(0.003)
Bad Weather	−0.409	−0.045	−0.242	−0.123	−0.168**
	(0.736)	(0.181)	(0.161)	(1.476)	(0.075)
Healthy vs Questionable	1.043	0.459***	0.067	0.639	0.222***
	(0.750)	(0.148)	(0.087)	(0.611)	(0.046)
Healthy vs Probable	−0.140	−0.171*	0.082	−0.692	−0.028
	(0.226)	(0.096)	(0.072)	(0.472)	(0.035)
Constant	−0.151	0.003	−0.064**	−0.098	−0.036**
	(0.119)	(0.037)	(0.029)	(0.207)	(0.014)
Observations	316	3,071	4,728	113	20,211

*p<0.1; **p<0.05; ***p<0.01

Table 1.7: Results By Year: All Data, QB and RB

	Dependent variable: Probability of Winning					
		QB			RB	
	(2016)	(2017)	(2018)	(2016)	(2017)	(2018)
League Tenure	-0.042^{***}	-0.038	-0.051^{**}	0.026^{***}	0.005	-0.033^{***}
	(0.011)	(0.026)	(0.018)	(0.005)	(0.007)	(0.006)
QB Overall Grade				0.009^{***}	-0.012^{***}	0.004^{**}
				(0.002)	(0.002)	(0.002)
Run Blocking	-0.017	-0.119^{***}	0.041	-0.066^{***}	0.011^{**}	0.032^{***}
	(0.022)	(0.040)	(0.035)	(0.005)	(0.005)	(0.007)
Pass Blocking	0.012	-0.004	-0.074^{***}	0.005^{*}	-0.026^{***}	0.002
	(0.013)	(0.024)	(0.022)	(0.003)	(0.004)	(0.004)
Run Defense	-0.118^{***}	-0.055	-0.149^{**}	0.040^{***}	-0.033^{***}	-0.040^{***}
	(0.037)	(0.068)	(0.058)	(0.008)	(0.009)	(0.010)
Pass Rush	-0.014	-0.097^{**}	0.006	-0.004	-0.011^{*}	-0.005
	(0.021)	(0.039)	(0.041)	(0.005)	(0.006)	(0.007)
Coverage Defense	-0.037	-0.073^{*}	0.004	-0.022^{***}	0.023^{***}	-0.004
	(0.029)	(0.043)	(0.055)	(0.006)	(0.007)	(0.008)
Wind Speed	0.033^{**}	-0.023	0.052^{**}	-0.024^{***}	0.005	0.002
	(0.013)	(0.045)	(0.026)	(0.003)	(0.007)	(0.004)
Bad Weather	0.335	-1.575^{**}	0.385	-0.266^{***}	-0.305^{***}	0.041
	(0.351)	(0.784)	(0.949)	(0.090)	(0.091)	(0.115)
Healthy vs Ques	0.101	0.220	14.245	0.117^{**}	-0.325^{***}	0.616^{***}
	(0.349)	(1.082)	(824.777)	(0.053)	(0.094)	(0.145)
Healthy vs Prob	-0.662^{***}	-0.629^{*}	-0.271	-0.151^{***}	-0.098	-0.298^{***}
	(0.160)	(0.358)	(0.264)	(0.056)	(0.072)	(0.085)
Constant	-0.170^{***}	-0.205	-0.120	-0.475^{***}	-0.693^{***}	-0.628^{***}
	(0.061)	(0.135)	(0.124)	(0.015)	(0.020)	(0.021)
Observations	1,175	260	306	18,890	10,970	10,464

*p<0.1; **p<0.05; ***p<0.01

Table 1.8: Results By Year: All Data, WR and TE

	Dependent variable: Probability of Winning					
	WR			TE		
	(2016)	(2017)	(2018)	(2016)	(2017)	(2018)
League Tenure	-0.031^{***}	-0.008^{**}	-0.020^{***}	-0.038^{***}	0.023^{***}	-0.021^{***}
	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
QB Passing Grade	-0.004^{***}	-0.002	-0.009^{***}	-0.08^{***}	-0.019^{***}	0.015^{***}
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
Run Blocking	0.028^{***}	-0.004	-0.029^{***}	0.100^{***}	0.027^{***}	-0.043^{***}
	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)
Pass Blocking	-0.016^{***}	-0.012^{***}	0.023^{***}	-0.039^{***}	-0.004	-0.0001
	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
Run Defense	-0.022^{***}	-0.006	0.005	0.038^{***}	0.029^{***}	0.026^{***}
	(0.006)	(0.006)	(0.007)	(0.007)	(0.006)	(0.007)
Pass Rush	-0.033^{***}	-0.018^{***}	-0.049^{***}	0.044^{***}	-0.039^{***}	0.018^{***}
	(0.004)	(0.004)	(0.005)	(0.005)	(0.004)	(0.005)
Coverage Defense	-0.019^{***}	-0.012^{***}	-0.024^{***}	-0.072^{***}	0.029^{***}	-0.015^{***}
	(0.005)	(0.004)	(0.006)	(0.006)	(0.005)	(0.006)
Wind Speed	-0.013^{***}	-0.031^{***}	-0.019^{***}	0.002	0.003	0.029^{***}
	(0.003)	(0.004)	(0.003)	(0.002)	(0.004)	(0.003)
Bad Weather	0.077	-0.208^{***}	0.062	0.125^{*}	-0.061	0.397^{***}
	(0.053)	(0.066)	(0.074)	(0.072)	(0.065)	(0.082)
Healthy vs Ques	0.447^{***}	-0.182^{***}	0.231^{***}	-0.107^{*}	-0.025	-0.694^{***}
	(0.047)	(0.050)	(0.088)	(0.058)	(0.057)	(0.063)
Healthy vs Prob	-0.114^{***}	0.235^{***}	-0.132^{***}	-0.278^{***}	0.009	0.064
	(0.034)	(0.036)	(0.043)	(0.039)	(0.041)	(0.041)
Constant	-0.403^{***}	-0.477^{***}	-0.532^{***}	-0.575^{***}	-0.659^{***}	-0.656^{***}
	(0.011)	(0.013)	(0.015)	(0.014)	(0.014)	(0.014)
Observations	34,408	26,274	19,482	23,348	23,809	22,575

$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 1.9: Results By Year: Non-Minimum Data, QB and RB

	Dependent variable: Probability of Winning					
	QB			RB		
	(2016)	(2017)	(2018)	(2016)	(2017)	(2018)
League Tenure	−0.026	−0.032	−0.028	0.012	0.016	0.026*
	(0.046)	(0.055)	(0.027)	(0.020)	(0.013)	(0.016)
QB Overall Grade				0.015*	−0.006	0.012*
				(0.009)	(0.007)	(0.007)
Run Blocking	−0.010	−0.015	0.062	0.008	0.047***	−0.019
	(0.094)	(0.098)	(0.058)	(0.025)	(0.014)	(0.020)
Pass Blocking	−0.054	−0.076	−0.075**	0.002	−0.019*	0.021*
	(0.045)	(0.068)	(0.036)	(0.014)	(0.010)	(0.013)
Run Defense	−0.159	−0.384**	−0.106	0.091**	0.058**	0.070**
	(0.117)	(0.163)	(0.092)	(0.039)	(0.026)	(0.028)
Pass Rush	0.059	−0.018	0.040	−0.047**	−0.072***	−0.044**
	(0.068)	(0.100)	(0.060)	(0.023)	(0.017)	(0.021)
Coverage Defense	0.004	−0.071	0.018	−0.032	0.001	−0.052**
	(0.092)	(0.124)	(0.085)	(0.035)	(0.019)	(0.026)
Wind Speed	0.021	−0.199*	0.125***	−0.003	0.001	−0.001
	(0.045)	(0.114)	(0.045)	(0.016)	(0.018)	(0.014)
Bad Weather	−0.269	−18.316	0.845	−0.503	−0.004	0.272
	(1.042)	(3,956.180)	(1.280)	(0.411)	(0.272)	(0.325)
Healthy vs Ques	0.051	18.136	14.493	0.433*	0.571**	0.410
	(0.944)	(2,524.379)	(1,029.108)	(0.255)	(0.269)	(0.255)
Healthy vs Prob	−0.088	−0.444	−0.200	0.124	−0.167	−0.343*
	(0.354)	(0.653)	(0.374)	(0.245)	(0.133)	(0.176)
Constant	−0.161	−0.095	−0.179	−0.055	0.007	0.029
	(0.197)	(0.323)	(0.193)	(0.076)	(0.055)	(0.066)
Observations	120	60	136	723	1,381	967

*p<0.1; **p<0.05; ***p<0.01

Table 1.10: Results By Year: Non-Minimum Data, WR and TE

	Dependent variable: Probability of Winning					
	WR			TE		
	(2016)	(2017)	(2018)	(2016)	(2017)	(2018)
League Tenure	0.050***	0.042***	−0.031***	0.057	−0.145	0.058
	(0.015)	(0.016)	(0.012)	(0.079)	(0.109)	(0.112)
QB Passing Grade	0.009	0.006	0.010**	0.054	0.021	0.025
	(0.007)	(0.006)	(0.004)	(0.049)	(0.080)	(0.035)
Run Blocking	−0.059***	−0.031**	−0.016	−0.322	−0.049	−0.014
	(0.022)	(0.014)	(0.014)	(0.211)	(0.147)	(0.203)
Pass Blocking	0.030**	−0.015	−0.006	0.185*	−0.089	0.167
	(0.012)	(0.009)	(0.009)	(0.106)	(0.099)	(0.108)
Run Defense	−0.135***	−0.045**	−0.037*	−0.124	0.169	−0.010
	(0.030)	(0.022)	(0.022)	(0.266)	(0.225)	(0.284)
Pass Rush	0.073***	0.018	−0.033**	0.029	−0.003	−0.639*
	(0.021)	(0.014)	(0.015)	(0.147)	(0.165)	(0.332)
Coverage Defense	0.032	0.049***	0.024	0.002	0.280	−0.007
	(0.026)	(0.017)	(0.017)	(0.184)	(0.206)	(0.245)
Wind Speed	0.012	0.005	−0.022**	−0.003	−0.282*	−0.032
	(0.014)	(0.017)	(0.009)	(0.089)	(0.162)	(0.173)
Bad Weather	−0.322	−0.119	−0.394	−17.093		15.789
	(0.455)	(0.254)	(0.241)	(2,399.545)		(2,399.545)
Healthy vs Ques	−0.035	0.141	0.017	1.115	−2.051	−1.102
	(0.157)	(0.150)	(0.151)	(1.140)	(1.872)	(1.680)
Healthy vs Prob	−0.042	0.309**	−0.037	−1.692	−1.420	−0.654
	(0.176)	(0.130)	(0.101)	(1.536)	(1.364)	(0.957)
Constant	−0.184***	0.010	−0.059	0.190	−0.434	−0.651
	(0.062)	(0.050)	(0.045)	(0.381)	(0.499)	(0.634)
Observations	1,115	1,631	1,982	40	35	38

*p<0.1; **p<0.05; ***p<0.01

Table 1.11: Marginal Effects For Statistically Significant Variables

Position	QB		RB		WR		TE
Sample	(F)	(NM)	(F)	(NM)	(F)	(NM)	(F)
QB Grade			0.88%	1.38%	-0.71%	1.16%	
Run Blocking			-1.53%	1.37%		-1.81%	1.48%
Pass Blocking		-7.75%	-0.65%		-0.40%		-1.03%
Run Defense	-3.68%	-5.59%		2.76%	-0.68%	-2.64%	1.49%
Pass Rush	-1.81%			-3.79%	-1.77%		0.76%
Coverage Defense	-3.21%		-0.80%		-0.34%	1.39%	-0.29%
Experience Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Weather Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Injury Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,741	316	40,324	3,071	80,164	4,728	69,732

Notes: Model results are given for the full (F) and non-minimum (NM) samples. Marginal effects are calculated comparing the effects of there being no difference in the variable versus the average difference in the variable. See Table 1.3 for the average differences for each position. Non-minimum marginal effects were not calculated for QBs because of the small sample.

Table 1.12: Choosing Flex Position By Salary Range

	Dependent variable: Probability of Winning			
	$s \in (3000, 4000]$	$s \in (4000, 5000]$	s $\in (5000, 6000]$	$s > 6000$
RB vs WR	-0.260^{***}	-0.087^{*}	0.009	0.365^{***}
	(0.028)	(0.049)	(0.085)	(0.133)
WR vs TE	-0.377^{***}	0.038	0.210	-0.107
	(0.051)	(0.096)	(0.176)	(0.247)
RB vs TE	-0.558^{***}	-0.259	0.186	0.265
	(0.060)	(0.217)	(0.203)	(0.328)
Observations	14,366	4,089	1,167	589
Note:			*p<0.1; **p<0.05; ***p<0.01	

Notes: Salary is represented by s. A positive coefficient means that a participant is more likely to win selecting the first position over the second position, conditional on the two having the same salary. The minimum salaries are 3,000 for RBs and WRs and 2,500 for TEs.

Table 1.13: Points Scored By Salary Range And Position

Salary	QB	RB	WR	TE
2,500–3,000		1.65	2.57	2.78
3,001–4,000	2.33	6.26	6.80	8.51
4,001–5,000	8.74	10.20	10.06	10.68
5,001–6,000	16.98	12.24	12.23	11.11
6,001–7,000	19.74	15.93	14.54	15.77
7,001–8,000	21.14	19.26	16.85	15.55
8,001–9,000	26.31	21.83	18.92	
9,001–10,100		25.83	20.38	

Notes: Scoring is broken down by salary range and position to examine which positions offer better returns for the flex spot given a predetermined level of investment.

Table 1.14: Comparing Probability And OLS Model Effects

	QB				RB				WR				TE	
	(F)		(NM)		(F)		(NM)		(F)		(NM)		(F)	
	P	U	P	U	P	U	P	U	P	U	P	U	P	U
QB Grade	NA	NA	NA	NA	$\uparrow$		$\Uparrow$		$\Downarrow$		$\Uparrow$		$\Uparrow$	
Run Blocking					$\Uparrow$				$\Downarrow$				$\Downarrow$	
Pass Blocking	$\boldsymbol{\Downarrow}$		$\boldsymbol{\Downarrow}$			$\uparrow$	$\Uparrow$		$\Uparrow$					
Run Defense	$\boldsymbol{\Downarrow}$	$\boldsymbol{\uparrow}$			$\Downarrow$		$\Uparrow$				$\Downarrow$		$\Uparrow$	
Pass Rush		$\downarrow$					$\Downarrow$		$\Downarrow$		$\Downarrow$		$\uparrow$	
Coverage Defense							$\Downarrow$		$\Downarrow$				$\downarrow$	

For Prob Model: ME $< 1\% =\uparrow$; ME $> 1\% =\Uparrow$; ME $> 5\% =\boldsymbol{\Uparrow}$
For Use Model: E $< 0.5\% =\uparrow$; E $> 0.5\% =\Uparrow$; E $> 1\% =\boldsymbol{\Uparrow}$
Same applies for negative effects

Notes: ME are marginal effects for the average difference and E is the linear effect on the difference in usage. Both effects are calculated by taking the difference between having a zero difference and the average difference. The single arrows denote a small effect, the double arrows denote a slightly stronger effect, and the bolded double arrows denote the strongest effects. The cutoffs for these are mostly arbitrary, but are meant to help the reader visualize the size of the effects.

Table 1.15: Usage Select Success Rate

Position	Min Included	Pairs	Participant Success Rate	Prob Model Success Rate
QB	Yes	235	53.62%	63.83%***
QB	No	124	58.87%**	63.71%***
RB	Yes	7,169	50.13%	63.86%***
RB	No	869	53.74%**	57.54%***
WR	Yes	15,604	51.03%**	62.45%***
WR	No	1,790	49.11%	54.30%***
TE	Yes	18,614	48.94%***	65.77%***

Notes: This table analyzes participants' abilities to pick the higher scoring player and comparing that to how someone would perform if fully and properly incorporating the various player attributes. Results are provided for both the full set of players and the set of players after removing minimum salary pairs. The last column calculates an expected probability of winning by obtaining fitted values using the coefficients from the logistic regressions for 2018 in Tables 1.7 through 1.10.

Table 1.16: Type Stochastic Dominance Test Results

		G 2x	3x	4x	10x	GPP
	2x		0.625	0.688	0.938	1.000
	3x	0.375		0.750	0.938	0.938
F	4x	0.125	0.125		0.563	0.813
	10x	0.375	0.313	0.188		0.875
	GPP	0.563	0.313	0.063	0.125	

Panel A: Aggregate Type Results

		G 2x	3x	4x	10x	GPP
	2x		0.135	0.342	0.474	0.537
	3x	0.117		0.198	0.391	0.472
F	4x	0.073	0.017		0.140	0.241
	10x	0.137	0.068	0.050		0.155
	GPP	0.169	0.091	0.068	0.082	

Panel B: Individual Type Results

		G 2x	3x	4x	10x	GPP
	2x	1,042				
	3x	15,115	192			
F	4x	5,795	1,042	75		
	10x	25,633	4,584	1,788	321	
	GPP	29,215	5,261	2,031	8,895	391

Panel C: Individual Type Sample Sizes

Notes: In Panel A, the numbers in the table represent the percentage of weeks during the 2018 NFL season that the aggregate contest type in G fails to reject stochastic dominance at the 5% level over the contest type in F. In Panel B, the numbers represent the percentage of times a contest in a particular week of type G fails to reject stochastic dominance at the 5% level over a contest of type F in the same week. In Panel C, the numbers along the diagonal represent the number of unique contests of each type during the 2018 NFL season, and the numbers off the diagonal represent the pairs of each combination of contest types.

Table 1.17: Size Stochastic Dominance Test Results

		G Tiny	Small	Medium	Large
F	Tiny		0.813	0.938	1.000
	Small	0.500		0.938	1.000
	Medium	0.313	0.313		0.938
	Large	0.313	0.313	0.313	

Panel A: Aggregate Size Results

		G Tiny	Small	Medium	Large
F	Tiny		0.298	0.420	0.659
	Small	0.290		0.677	0.874
	Medium	0.202	0.467		0.868
	Large	0.116	0.279	0.243	

Panel B: Individual Size Results

		G Tiny	Small	Medium	Large
F	Tiny	1,575			
	Small	22,219	196		
	Medium	14,882	1,902	158	
	Large	4,279	538	440	44

Panel C: Individual Size Sample Sizes

Notes: See comments under Table 1.16 for a description of the layout of this table. This table looks at stochastic dominance as it relates to contest size. Tiny contests are defined as having less than 1,500 participants, small contests have between 1,500 and 10,000, medium contests have between 10,000 and 100,000, and large contests have more than 100,000.

Table 1.18: Fee Stochastic Dominance Test Results

		G							
		$1	$2	$3	$5	$10	$12	$20	$25
F	$1		0.313	0.625	0.375	0.467	0.143	0.154	0.286
	$2	0.875		0.875	0.250	0.333	0.786	0.923	0.286
	$3	0.250	0.188		0.375	0.333	0.071	0.231	0.357
	$5	0.875	0.938	0.875		0.467	0.929	0.923	0.357
	$10	0.800	0.733	0.867	0.733		0.714	0.846	0.500
	$12	0.714	0.429	0.786	0.429	0.643		0.769	0.500
	$20	0.923	0.462	0.846	0.462	0.615	0.231		0.385
	$25	0.857	0.857	0.857	0.857	0.429	0.857	0.923	

Panel A: Aggregate Fee Results

		G							
		$1	$2	$3	$5	$10	$12	$20	$25
F	$1		0.193	0.209	0.192	0.163	0.448	0.628	0.193
	$2	0.155		0.169	0.152	0.116	0.383	0.538	0.105
	$3	0.208	0.216		0.222	0.187	0.467	0.630	0.218
	$5	0.256	0.248	0.241		0.177	0.478	0.672	0.217
	$10	0.255	0.268	0.257	0.242		0.542	0.719	0.206
	$12	0.236	0.169	0.282	0.198	0.204		0.731	0.302
	$20	0.151	0.106	0.213	0.157	0.158	0.231		0.293
	$25	0.341	0.343	0.341	0.357	0.301	0.718	0.813	

Panel B: Individual Fee Results

		G							
		$1	$2	$3	$5	$10	$12	$20	$25
F	$1	438							
	$2	11,744	307						
	$3	9,955	7,115	263					
	$5	13,139	9,412	7,984	343				
	$10	6,112	4,382	3,828	4,931	161			
	$12	784	543	478	607	299	24		
	$20	384	264	235	293	146	26	13	
	$25	2,663	1,875	1,635	2,095	1,030	149	75	72

Panel C: Individual Fee Sample Sizes

Notes: See comments under Table 1.16 for a description of the layout of this table. This table looks at stochastic dominance as it relates to contest entry fees.

Table 1.19: Week 1 Stochastic Dominance Test Versus Population

Type	vs Type	vs Pop	Size	vs Size	vs Pop	Fee	vs Fee	vs Pop
2x	0.000	1.000	Tiny	0.000	1.000	$1	0.000	1.000
3x	0.000	1.000	Small	0.000	1.000	$2	0.000	1.000
4x	0.000	1.000	Medium	0.000	1.000	$3	0.000	1.000
10x	0.000	1.000	Large	0.000	1.000	$5	0.000	1.000
GPP	0.000	1.000				$10	0.000	1.000
						$12	0.000	1.000
						$20	0.000	1.000
						$25	0.000	1.000

Notes: The table is split up by test category. The left column in each is the name of the category, the middle columns are the p-values for the population distribution versus the test type stochastic dominance test, and the right columns are the p-values for the type versus the population stochastic dominance test.

Table 1.20: Modality Tests

	Category	n	10%	5%	1%
Type	2x	1,015	0.250	0.188	0.124
	3x	181	0.160	0.077	0.022
	4x	73	0.014	0.014	0.000
	10x	309	0.074	0.036	0.000
	GPP	370	0.100	0.062	0.027

Panel A: Modality test by contest type

	Category	n	10%	5%	1%
Size	Tiny	1,554	0.124	0.080	0.034
	Small	194	0.438	0.330	0.253
	Medium	156	0.385	0.301	0.224
	Large	44	0.159	0.114	0.068

Panel B: Modality test by contest size

	Category	n	10%	5%	1%	DU	5%
Fee	$1	512	0.154	0.102	0.053	200	0.165
	$2	362	0.124	0.080	0.030	207	0.101
	$3	310	0.132	0.081	0.042	202	0.089
	$5	409	0.203	0.144	0.091	208	0.250
	$10	197	0.218	0.168	0.117	132	0.227
	$12	27	0.037	0.000	0.000	0	NA
	$20	14	0.143	0.143	0.071	1	0.000
	$25	86	0.535	0.442	0.326	65	0.569

Panel C: Modality test by contest entry fee

Notes: Each panel analyzes the percentage of contests that reject a unimodal distribution at each designated significance level for the various contest attributes. The last two columns of Panel C examine the number of double-up contests analyzed for each entry fee, and the percentage of double-up contests at each entry fee level that reject unimodality at the five percent level.

CHAPTER 2

DOES BASEBALL ARBITRATION PRIORITIZE SKILL SETS DIFFERENTLY THAN FREE AGENCY? A FRONTIER ESTIMATION APPROACH

2.1 Introduction

Since the abolishement of the reserve clause in the 1970s, Major League Baseball (MLB) player salaries as a percentage of total baseball revenues have skyrocketed.[1] The reserve clause gave teams full autonomy over the renewal of player contracts, permitting teams to set player wages. The establishment of free agency—an open market where teams bid against each other for players' services—was a huge win for the players and their union, the MLB Player's Association (MLBPA).

With the evolution of baseball analytical thinking and the increasing amount of available data, the way in which teams value players in free agency has changed over time. Over the past 20 years, team strategy has evolved, and so has the methodology used to measure player contributions. Stemming from the "Moneyball" revolution, teams' strategies are more statistically driven, including a shift from power focused production to on-base ability, and a shift away from tradi-

[1]While estimates vary depending on the data source and the benefits used to determine total player compensation, actual player compensation as a percentage of total revenues is somewhere near 50%. https://www.theringer.com/mlb/2018/2/21/17035624/mlb-revenue-sharing-owners-players-free-agency-rob-manfred

tional statistics (home runs, runs batted in, etc.) in favor of more advanced metrics.[2] That mindset shift is illustrated by free agent salaries during that time period, as shown by Hakes and Sauer (2007) and Brown, Link, and Rubin (2017), among others.

However, before players can file for free agency, they must accrue a certain amount of Major League service time, which was negotiated as part of the 1973 MLB Collective Bargaining Agreement (the union's contract with MLB, also known as the Basic Agreement).[3] During a player's first three years of service, he typically receives close to the Major League minimum salary since his team has the right to dictate compensation terms, and after six years of service, the player can enter free agency. If the player has between three and six years of Major League service, he can elect for final-offer salary arbitration (FOA).[4] During the arbitration process, the ball club and the player submit salary figures to an arbitration panel, and the three-person panel chooses (via majority rule) which is closer to the player's appropriate salary.

While the literature provides clear evidence that the free agent market evolves when presented with new ways to evaluate players, no paper has looked

[2]Moneyball was coined in Michael Lewis's famous book "Moneyball: The Art of Winning an Unfair Game". The book details how the Oakland Athletics used analytics to find advantages, despite having one of the lowest payrolls in the Major Leagues. Other teams began to mimic the Athletics, and this shift to a more analytically minded approach is called the "Moneyball" revolution.

[3]See http://www.mlbplayers.com/ViewArticle.dbml?DB˙OEM˙ID=34000ATCLID=211157624

[4]Players with less than three years of service time could also be eligible to file for arbitration if they rank within the top 22% of service time for players with between two and three years of service. These players are known as Super Two players. http://m.mlb.com/glossary/transactions/super-two

at whether the arbitration market evolves in a similar fashion. If it does, both markets should value players similarly. Otherwise, there will have been a divergence where the arbitration market continues to value attributes that are no longer valuable. The results from this paper support that idea.

Using the frontier estimation free disposal hull (FDH) estimator, this paper estimates unique arbitration markets for eligible players and identifies where they fit in those markets. Hadley and Ruggiero (2006) first introduce the approach, but do not account for separability concerns in their model. Separability is an issue when environmental variables, such as a player's position, service time, and contract signing date, impact the location or shape of the frontier. In order to implement a sensible second-stage model, separability must be satisfied. This paper builds a unique arbitration market for each player, addressing these separability concerns and allowing for the implementation of a second-stage model. The second-stage model captures the impact of different player attributes and environmental variables on a player's relative placement in his market.

The second-stage model provides evidence that power hitters are systematically overpaid in arbitration compared to a counterfactual system based on players' wins above replacement (WAR), while players who specialize in getting on base and drawing walks are underpaid. This is contrary to what has happened in the free agent market where teams have reduced the amount spent on power production and have increased the amount spent on on-base ability. There is no evidence that defense or speed is mispriced in arbitration.

Empirical results allow for discussion on resulting market outcomes. Players being undercompensated and overcompensated in the framrework of the arbitration market does not simply result in a redistribution of funds from undervalued to overvalued players. Teams have the ability to release arbitration players because their contracts are not fully guaranteed until the beginning of the MLB season.[5] Teams also have the option to non-tender a player prior to arbitration if the team believes the player would make more through the process than what he is worth on the open market.[6] As Lock and DeSerpa (1986) note, a player will be released or non-tendered if his minimum expected salary exceeds the team's perception of the player's value.[7] Meanwhile, players whose skills are undervalued have no recourse if they are underpaid. This has implications on arbitration negotiations (threat of being non-tendered sometimes forces players to take pay cuts) and potentially future contract extension negotiations.[8]

Section 2 provides an overview of the arbitration process. Section 3 gives background of the data and estimation techniques used in determining a player's relative arbitration market value. Section 4 provides results from the empirical

[5]A player who is tendered a contract, but released prior to a specified date in the middle of Spring Training, is entitled to 30 days pay. A player who is released after the previous date but before the start of the season is entitled to 45 days pay. A player released after the start of the season is owed his full contract. This all comes from the Basic Agreement.

[6]When a player's contract expires, prior to that player having at least six years of service time and being eligible to file for free agency, the team has to decide whether or not to tender the player a contract. If the team chooses to tender a contract, the team will auto renew the contract near the minimum salary (if the player is not arbitration eligible) or go through the arbitration process (if the player is arbitration eligible). If the team chooses to non-tender a player, that player instantly becomes a free agent eligible to sign with any team in baseball.

[7]The team's perception of the player's value is strongly based on the market's perception of that player's value. If the team values the player less than the market does, the team can choose to trade the player and enjoy the surplus.

[8]If a player accepts a lower salary in arbitration, that player may be more willing to except lesser valued forward contracts (contract extensions) that buy out years of free agency.

specifications. Section 5 summarizes the key conclusions and provides extensions for future research.

2.2 The Arbitration Process

Arbitration was negotiated into the 1973 Collective Bargaining Agreement as a way for player salary to increase without the team giving up exclusive rights to the player's services. Arbitration dictated that players with at least two full seasons of Major League service time could elect to allow a third-party panel to decide his salary. In 1975, with owners concerned about the potential negative effects to profits caused by the dawning of free agency, Marvin Miller, head of the player's union, negotiated a free agency system where players could only file for free agency after accumulating six years of service time. Therefore, today's arbitration system only covers players with between three and six years of service (with the exception of Super Two players).[9]

In its basic form, the arbitration system was designed to provide players with raises to their salaries as they reach closer to six years of service time. The style of arbitration incorporated by baseball is known as final-offer arbitration (FOA). In this setup, the player and the team submit salary figures to a panel of arbitrators, and the panel chooses either the player's or the team's offer. FOA differs from other forms of arbitration in that the arbitrators are unable to choose a number between the two proposals.

The FOA process generally results in an increased rate of settlements due

[9]See https://baseballhall.org/discover/short-stops/free-agency-still-fuels-baseball

to the exposed risk both sides face in relying on a hearing. The difference between offers are often worth hundreds of thousands of dollars, if not millions. For example, Jonathan Schoop of the Baltimore Orioles, going through his second season of arbitration prior to the 2018 MLB season, submitted a $9 million salary figure to the arbitration panel, while the team submitted a $7.5 million salary figure. Rather than taking what effectively would have been a $1.5 million gamble going to an arbitration hearing, both parties settled at an $8.5 million salary.

From 2001–2018 (the time period for this paper's sample), about 96% of all eligible cases resulted in agreed upon settlements.[10] Parties can either agree to one-year settlements (most common) or multi-year extensions. For players that go to a hearing, their case is decided by a panel of arbitrators. The panel is made up of three individuals picked from a pool of labor arbitrators, not necessarily baseball experts, jointly selected by the league and player's association. Arbitrators have prescribed criteria, as defined in the Basic Agreement, from which to evaluate players.

> *The criteria will be the quality of the Player*FL*s contribution to his Club during the past season (including but not limited to his overall performance, special qualities of leadership and public appeal), the length and consistency of his career contribution, the record of the Player*FL*s past compensation, comparative baseball salaries, the existence of any physical or mental defects on the part of the Player, and the recent performance record of the Club including but not limited to its League standing and attendance as an indication of public acceptance.*[11]

[10]This percentage includes both position players and pitchers. The percentages per group are approximately the same.

[11]See http://www.mlbplayers.com/pdf9/5450407.pdf

The most prominent tool arbitrators have to decide the value of a player's worth is comparable cases. In a typical hearing, the team will cite players who it deems to be of similar or better statue to the player and argue for the player to earn no more than those comparable contracts. Conversely, the player will cite comparable cases who he deems to be of similar or worse statue to him and argue for a salary greater than those players. It is common practice for the arbitration panel to consider a player's comparable market, the collection of potential comparable cases, and assign that player's true value based on where he fits within that market.

The arbitration panel can only pick one of the two offers as the player's salary and cannot pick a middle-ground number. That, combined with the fact that most cases end in settlements, leads one to question the magnitude of the role the arbitration panel plays in the process. Dworkin (1977) and Wittman (1986) model the decision making of the arbitration panel by assuming that the panel establishes a "true" arbitration market price for the player and then selects the offer which is closer to that true value. If that is in fact the case, the higher the arbitration market values a player, the more that player can expect to earn, either via a hearing (in expected value terms) or via a settlement. Hanany, Kilgour, and Gerchak (2007) show that, depending on the relative risk preferences of the two parties, there exists a set of settlement outcomes that dominate the expected utility from going to a hearing, and this set of values is determined based on the expected payoffs from going to a hearing. The greater the arbitration market values the player (team), the greater (lower) the magnitude of the set of preferable settlement offers. Thus, in effect, the arbitration market's value of players directly impacts every eligible case, whether or not they go to a hearing.

Previous research in baseball arbitration focuses on the impact of bargaining (Dworkin, 1977; Faurot and McAllister, 1992) and arbitration outcomes (Wittman, 1986). Gustafson and Hadley (1995), although somewhat dated, estimate the extent to which arbitration suppresses player contracts compared to what players can earn in free agency. The arbitration exchangeability hypothesis, as defined by Ashenfelter (1987), states that while each arbitrator has specific tendencies and evaluative methods, the expectation is that they conform to the same valuation in the long run. If they do not conform, or consistently favor one side over the other, they can be replaced in future cases. So the most sensible strategy for an arbitrator is to select a valuation system similar to what other arbitrators would select. Faurot and McAllister (1992) discuss the shortfall of the arbitrator exchangeability hypothesis in that it accounts for differences between arbitrators but not for any systematic bias in their evaluative criteria. This paper's main question of interest is whether that systematic bias exists and in what direction in the baseball arbitration market. More specifically, I am interested in whether the arbitration system is structured to handle new information and correct for systematic biases brought about by older evaluation criteria.

2.3 Data and Methodology

This paper extends the double frontier approach used by Hadley and Ruggiero (2006) to evaluate the MLB arbitration market. Free disposal hull (FDH) is a production frontier estimator developed by Deprins, Simar, and Tulkens (1984). FDH involves estimating production functions and the extent to which firms are efficiently turning their inputs into outputs. In the double frontier setting, technical efficiency is defined for both the player and the team.

For the player, technical efficiency is defined as the highest salary a player could expect to earn in arbitration given his abilities and performance to date.[12] Let $x \in \mathbb{R}_+^{\mathrm{p}}$ and $y \in \mathbb{R}_+^{\mathrm{q}}$ denote *p* player inputs and *q* player outputs, respectively. Using notation from Park, Simar, and Weiner (2000), and following standard production assumptions described in the efficiency literature, the estimated player frontier is

$$\hat{\Psi}_P(S_n) = \Big\{(x, y) \in \mathbb{R}_+^{\mathrm{p+q}} \mid y \leq \mathrm{Y_i}\,, x \geq \mathrm{X_i}\,, (\mathrm{X_i}, \mathrm{Y_i}) \in \mathrm{S_n}\Big\}. \qquad (2.1)$$

For the team, technical efficiency is the lowest possible salary it could expect to have to pay a player in arbitration given his abilities and performance to date. This frontier is given by

$$\hat{\Psi}_T(S_n) = \Big\{(x, y) \in \mathbb{R}_+^{\mathrm{p+q}} \mid y \geq \mathrm{Y_i}\,, x \leq \mathrm{X_i}\,, (\mathrm{X_i}, \mathrm{Y_i}) \in \mathrm{S_n}\Big\}. \qquad (2.2)$$

[12] Arbitration only considers past performance, not expectations of future performance. In the free agent market, teams pay players based on how they expect the players to perform. Arbitration rewards players based on past contributions.

The player has the upper frontier $\hat{\Psi}_P$ and the team has the lower frontier $\hat{\Psi}_T$ as illustrated in Figure 2.1.

2.3.1 Inputs and Outputs

The player and team frontiers are constructed using one input and one output. The input is the player's Wins Above Replacement (WAR). Wins above replacement is considered the best publicly available measure of the value of a player's total contributions to a team. The measure is comparable between teams since it controls for ballpark effects, comparable between leagues since it controls for league effects, and compare across seasons since it controls for year effects.[13] While individual teams may value player attributes differently, and the roster composition and strategy of particular teams make certain attributes more valuable in different settings, the internal team valuation of a player is not relevant in the arbitration decision.[14] Instead, the arbitration system produces its own valuation of player production, which may prioritize skills differently than WAR and differently than team's internal calculations. Any conclusions drawn in this paper regarding arbitration valuation mispricing are in comparison to a counterfactual in which arbitration salary is conditioned on a player's WAR. The empirical approach used in this paper is easily reproducible for any player production input.

The output is the player's salary. Salary data come from Wasserman's

[13]For example, if MLB makes a change to the baseball resulting in more total runs scored, WAR accounts for that change in run scoring.

[14]The arbitration panel will take park factors into account. For example, the Colorado Rockies are known to play in a ballpark that is strongly conducive to offensive performance. Arbitrators acknowledge that when they analyze player production.

baseball database.[15] Nominal yearly base salary is the primary metric for examining player arbitration compensation, but this paper also considers additional measures, including nominal salary raise, real yearly base salary, and real salary raise.[16] From speaking to those working in the industry, salary raise can actually be more important than base salary for certain cases, especially for second year (2 SAE) and third year (3 SAE) arbitration eligible players.

Table 2.1 summarizes the different salary figures for various groups of players in the sample. The first four rows break down average salary figures by position. Middle infielders (MI) earn the highest average salary in arbitration, and catchers (C) earn the lowest. Corner infielders (CI) and outfielders (OF) earn comparable salaries. The bottom three rows break down average salary figures by service year. Consistent with the findings from Gustafson and Hadley (1995), players with more years of service earn higher salaries in arbitration. First-time arbitration eligible players tend to earn higher raises, but those figures include their pre-arbitration salaries, which tend to hover around the Major League minimum salary. Excluding pre-arbitration salaries from the average raise for first time arbitration players, the three groups earn close to the same raise on average.

Players earn money in arbitration based on their performances and abilities in both the platform season (the season prior to arbitration) and throughout their careers. To quantify a player's ability and production, this paper uses WAR—a

[15]Wasserman Media Group is a sports marketing and talent management company based in Los Angeles. Salary data come from their baseball team sports division.

[16]Salary figures are adjusted for inflation using GDP deflator numbers from FRED (Federal Reserve Economic Data) with 2009 as the base year.

measurement that quantifies everything a player does to impact a game including the player's production with the bat, on the base paths, and in the field defensively. A player worth one WAR is by definition worth one win to his team over a potential replacement level player.[17] Players with equal WAR are said to have provided the same amount of wins to their teams, and thus, all else equal, should receive the same financial compensation in arbitration.

To test the robustness of WAR as the input measure, I also run model specifications that include plate appearances (PA) as the input instead. Plate appearances removes much of the context behind a player's production. On one hand, using PA as the input allows for a more agnostic view of production, with a player's efficiency estimate measuring how well players turn their opportunities into salary. One of the drawbacks of using WAR as the input is that the components used in the second stage impact the WAR value. For example, if a player increases his power production or on-base production, by definition he will have a higher WAR. The same is not true of PA. On the other hand, the number of plate appearances a player accumulates is not necessarily a representative measure of a player's production or his input to his team winning. There is also a strong relationship between the number of plate appearance a player receives and the magnitude of his various counting stats such as home runs (HR), runs batted in (RBI), hits (H), stolen bases (SB), etc. This makes disentangling a player's profile from his raw production difficult in the second-stage model. I use both WAR and

[17]See FanGraphs' stat glossary for a discussion on replacement level. https://www.fangraphs.com/library/misc/war/replacement-level/

PA to see if second-stage results are sensitive to the input specification.

The difficulty in using WAR, PA, or any other value-based stat is figuring out how much weight should be given to a player's platform value versus career value. While it is possible to include platform and career numbers as separate inputs, doing so comes at the cost of added complexity and limits the ability to clearly define a player's contract zone (it is easier to define a two-dimensional shape than a three or more dimensional shape).[18] Instead, this paper uses two different weighting schemes. The first utilizes the dimension reduction technique introduced by Wilson (2018). This approach takes advantage of collinearity by taking a weighted average based on an eigendecomposition. Table 2.2 provides distributional summary stats for the R_X values obtained from the dimension reduction approach, where R_X is the ratio of the largest eigenvalue to the sum of all eigenvalues. The R_X value relays the percentage of total information from the joint inputs captured by the weighting mechanism in the single-dimension reduced input. Each arbitration market goes through a dimension reduction procedure, providing unique R_X values. The average R_X value for the WAR input is 0.9077, meaning, on average, about 91% of the available information for the platform and career WAR values is captured by the dimension reduction technique. A first quartile R_X value of 0.8713 shows dimension reduction is capturing a sufficient amount of information in most cases. The average R_X value for platform and career PA is even greater at 0.9850.

[18]The contract zone is the feasible set of potential salaries a player could earn in arbitration. See Section 3.3 for more details.

Alternative to the dimension reduction approach, a composite WAR measure is calculated using a simple, multivariate, ordinary least squares regression to determine the relationship between a player's salary and his platform and career value, controlling for position and service time. The relative magnitudes of the coefficients are used to weight the importance of platform versus career.[19] Values of composite WAR are slightly different depending on the salary output measurement used. Distributions for the different composite WAR measures are shown in Figure 2.2. Most players in the sample, over 80%, have between 0 and 5 composite WAR, while about 15% of the sample have more than 5 composite WAR.

2.3.2 The Arbitration Market

When determining how much a player earns in arbitration, he is compared to other players in his "arbitration market". In a hearing, a team identifies comparable cases to the player and argues why the player should earn no more than them. Meanwhile, the player will identify comparable cases that make him look favorable and argue why he should earn more than them. The player's arbitration market is determined based on his service time and position. For example, the case of a 3 SAE catcher is generally only compared to other cases of 3 SAE catchers. Catchers are expected to provide less offense compared to other positions due to their responsibilities to the pitching staff; so to not penalize them, their production is compared to other catchers.[20] There are differing offensive skillsets at other

[19] A full derivation of composite WAR is available in the Appendix.

[20] A catcher has to help call the game for the pitcher, which requires an immense amount of additional preparation. So while most offensive players need to study the tendencies of the oppos-

positions as illustrated in Table 2.3, and those differences are generally captured in the arbitration process. Corner infielders and outfielders show more offensive ability, while middle infielders show more speed and defensive ability.

An arbitration market is filled with player comparables (comps) which help dictate the player's arbitration salary. Arbitration markets are generally filled with more recent comps, but when the market lacks depth, it is common to see team and player representatives use comps from over a decade ago. In determining which comps compose a player's market, for empirical estimation purposes, the following procedure is used:

[label=)]

1. All comps had to be of the same service year and similar position group.[21]
2. All comps had to sign their contract prior to the arbitration player signing his. Once a player signs his deal, his case can be immediately used as a comp in subsequent cases, even in the same year.
3. All comps had to have signed one-year settlement deals—arbitration cases that ended in either multi-year deals or arbitration hearing decisions are not included.
4. The initial arbitration market is composed of all comps with cases settled within the previous five seasons. If the initial group is composed of less than

ing pitcher, catchers need to do that plus study the tendencies of the opposing batters as well. In addition, catching is generally a much more physically demanding position, as the player spends half of the game crouched down behind the plate.

[21] Players are clustered together based on the similarity of their positions. First and third baseman (corner infielders - CI) are generally less defensively demanding positions, and thus typically are more power oriented. Middle infielders and center fielders (second baseman, shortstops, and center fielders - MI) are generally much more athletic and strong defensive players, but are not traditionally your best hitters. Left fielders and right fielders (Outfielders - OF) are grouped together. Catchers (C) have their own positional group, and designated hitters are usually included in the CI group.

30 comp cases, an additional previous year was pulled. This process continues until either the arbitration market consists of 30 cases or the previous 10 years of comps are pulled.

5. The player is only included in the sample if his market has more than 10 comp cases.

Because arbitration markets update the moment a player signs his contract every arbitration market is unique. Hadley and Ruggiero (2006) ignore the strategic nature of the available comparable cases prior to signing and have instead opted for a static setting where all cases are compared to each other.

Hadley and Ruggiero (2006) also did not consider the impact environmental variables had on the frontiers, failing to account for separability. Separability holds when environmental variables have no effect on the shape or location of the frontier. Given a set of *r* environmental variables, $z \in \mathbb{R}^r_+$, separability holds when

$$\Psi^z = \Psi, \ \forall \, z \,, \tag{2.3}$$

where $\Psi^z := \Big\{(x, y) \mid x \, \text{can produce} \, y \, \text{when} \, z = Z\Big\}$. With the addition of environmental variables, the data sample S_n is now $S_n = \{(X_i, Y_i, Z_i)\}_{i=1}^n$.

Simar and Wilson (2007) identify multiple issues that arise from ignoring separability concerns, the most pressing being that efficiency estimates lose operational meaning when environmental variables affect the frontier. Environmental variables can affect production in two ways. The first is that the environmental variables can influence the distribution of efficiency.[22] The second is that the en-

[22]As a practical example, suppose there are two otherwise identical accounting firms, with similar employee and capital inputs, working in two different types of office spaces. Also, suppose

vironmental variables can impact the shape and location of the frontier.[23] The first case motivates second-stage efficiency models and does not violate separability. The second case violates separability, rendering inference useless.

In this setting, environmental variables that most likely violate separability include: a player's position, a player's service time, and a player's previous salary.[24] Differences between positions and the reasons why those differences may affect arbitration salary were mentioned previously. Players with different service times, but the same ability, would face different frontiers simply because arbitrators only consider comps in the same service class. Player salaries are also partially conditioned on the amount of money the player made in the previous season. By building arbitration markets that condition on similar service time and similar positions, and by using raise as the output variable for second- and third-year players, the resulting frontiers are not affected by separability. Additional environmental variables used later on in the second-stage models are all assumed to not affect the frontier, but rather the distribution of the efficiency measures. See

the first firm works in the nicer offices with noise cancelling walls and on-site kitchens. That company would likely have the better efficiency score, meaning they are able to complete more tax returns. That being said, the quality of the office space does not necessarily preclude either office from completing the same amount or the technically efficient amount of returns. In this example, the quality of the office space is an environmental variable that impacts the distribution of efficiency.

[23]Continuing with the example from the previous footnote, suppose the two accounting firms complete different types of tax returns. Suppose one firm completes tax returns for individuals and the other completes tax returns for businesses. Because they are two very different types of returns, the technically efficient amount they would be able to complete are inherently different. Therefore, the type of tax returns the firms complete is an environmental variable that affects the shape or location of the frontier.

[24]Current tools make proving separability difficult. Testing separability is computationally possible, and will be tested in future versions of this paper, but for now intuitive economic reasoning is still the best way to identify these potential issues.

Section 3.5 for more details on this.

2.3.3 Relative Contract Position

The composite WAR and arbitration salaries for the player and each comp in the arbitration market make up the sample used to estimate the player and team frontiers. A player's contract zone is defined as the set of feasible salaries he can expect to earn in arbitration, given by the salary points along a vertical line at some input level connecting the player and team frontier. A player's relative distance between the two frontiers, or relative location along the contract zone, is called his relative contract position (RCP). Refer back to Figure 2.1 and suppose a player's arbitration salary and composite WAR puts him on the illustrated contract zone directly in the middle of the player and team frontiers. In that case, his RCP would be 0.5. As the salary approaches the team frontier, RCP approaches zero. As the salary approaches the player frontier, RCP approaches one. As introduced in Hadley and Ruggiero (2006), the equation for RCP is

$$RCP_i = \frac{SAL_i - CZL_i}{CZH_i - CZL_i}, \tag{2.4}$$

where SAL_i is defined as player i's final arbitration salary, CZH_i is the top of the contract zone along the player frontier for player i, and CZL_i is the bottom of the contract zone along the team frontier for player i. Calculating CZH_i and CZL_i require estimating efficiency relative to each frontier.

Estimating CZH_i is fairly straightforward using the Shephard (1970) output distance function, $\lambda(x, y \mid \Psi)$. The vector of efficiency estimates, $\hat{\lambda}$, from $\hat{\Psi}$ is estimated by

$$\hat{\lambda}(x, y \mid \hat{\Psi}_{FDH}) = \sup\Big\{\lambda \mid (x, \lambda y) \in \hat{\Psi}_{FDH}\Big\}. \tag{2.5}$$

This equation is estimated using the FEAR package in R (Wilson, 2008). By definition, each of the estimates, $\hat{\lambda}_i$, must be less than or equal to one, where a value of one indicates that the salary is along the estimated player frontier. After extracting $\hat{\lambda}_i$ for player of interest i, CZH_i can be estimated using

$$C\hat{Z}H_i = \frac{SAL_i}{\hat{\lambda}_i}. \tag{2.6}$$

As an example, if the estimated efficiency is 0.5 and the player's arbitration salary is \$1 million, the upper bound of the contract zone is \$2 million.

Calculating the bottom, team frontier is slightly trickier, especially since it would appear unnecessary in most practical economic applications to reduce outputs without also reducing inputs. In order to complete FDH estimation, the data are transformed along the x-y axis. Swapping the axes and utilizing the FDH estimator and the Shephrad (1970) input distance function, $\theta(x, y \mid \Psi)$, recovers a vector $\hat{\theta}$ in a similar way as was done to recover $\hat{\lambda}$. Figure 2.3 illustrates the transformation process. Note that by estimating in the input direction, relative efficiency is still being derived for the intended output. By definition of the Shephard input distance function, the values in $\hat{\theta}$ must be greater than or equal to one, where a value of one indicates that the salary is along the team frontier. After estimating $\hat{\theta}_i$ for the player i of interest, CZL_i is calculated by

$$C\hat{Z}L_i = \frac{SAL_i}{\hat{\theta}_i}. \tag{2.7}$$

Similar to the previous example, if the estimated efficiency equals two and the salary the player earned is $1 million, the CZL for that player is $500,000. Using the $2 million CZH from the previous example, the estimated contract zone would be [$0.5 million, $2 million] for that player and the estimated RCP for that player would be one-third.

With CZL and CHL estimated, RCP can be calculated using Equation 2.4. Figure 2.4 provides an example of the RCP estimation procedure using FDH. An RCP between zero and one means the player received a salary within the framework of the previous market. However, an RCP of zero or one indicates the player establishes a new part of the market not previously defined. In an ex-post analysis, this does not present any issues. However, this does limit the model's ability to predict arbitration salary ex-ante for future players that may exist outside the market.

Another concern is potential bias in the efficiency measures. Kneip, Simar, and Wilson (2015), establish for the FDH estimator that the bias term is said to be negligible if $p + q < 2$. Since this paper uses one input and one output for a given specification, and the underlying measure, RCP, is a construction of efficiency estimates from two different frontiers, the bias for the efficiency estimates is estimated using the subsampling bootstrap method in Kneip, Simar, and Wilson (2008).

2.3.4 Testing Convexity Assumption

The assumptions made about the shape of the frontier directly impact the resulting efficiency estimates. While estimators, such as data envelopment analysis (DEA), assume convexity in the production frontier, FDH makes no such assumption. This paper checks for convexity in the player and team frontiers using a test introduced by Kneip, Simar, and Wilson (2016).[25]

When the frontier is convex, both FDH and DEA will be consistent. But when the frontier is not convex, only FDH will be consistent. Korostelëv, Simar, and Tsybakov (1995) establish that the estimated frontier under DEA, assuming variable returns to scale, $\hat{\Psi}_{VRS,n}$, converges to the true frontier, Ψ, at rate $n^{\frac{2}{p+q+1}}$. The estimated frontier under FDH, $\hat{\Psi}_{FDH,n}$, converges to Ψ at rate $n^{\frac{1}{p+q}}$. Since the DEA-VRS estimator has a faster rate of convergence, failing to reject convexity may suggest better results using the DEA-VRS estimator.

Kneip et al. (2016) propose splitting the sample, runing FDH on one group and VRS on the other, and then comparing mean efficiencies using the following

[25] Theorem 4.1 in Kneip et al. (2015) establishes consistency and other properties for mean efficiency estimates needed for the test. The theorem relies on appropriate regularity conditions discussed in the paper.

test statistic,

$$\hat{\tau}_{n_{1,k},n_{2,k}} = \frac{(\hat{\mu}_{1,n_{1,k}} - \hat{\mu}_{2,n_{2,k}}) - (\hat{B}_{1,k,n_1} - \hat{B}_{2,k,n_2})}{\sqrt{\frac{\hat{\sigma}^2_{1,\gamma,n_1}}{n_{1,k}} + \frac{\hat{\sigma}^2_{2,\gamma,n_2}}{n_{2,k}}}} \longrightarrow N(0,1), \tag{2.8}$$

where k is the convergence rate for either FDH or VRS, $\hat{\sigma}^2_{i,\gamma,n_i}$ is the variance of the bootstrapped efficiency estimates for group $i \in \{1,2\}$ where group 1 is the FDH group and group 2 is the VRS group, γ is the efficiency estimator λ for the player frontier and θ for the team frontier, and n_i is the group sample size. The $\widehat{B}$ terms are the bias for both sample means. Kneip et al. (2016) suggests using the generalized jack knife bootstrapping method to estimate the biases for the sample means. Rejecting the null means rejecting convexity (DEA-VRS) in favor of non-convexity (FDH). This test is done for each arbitration market for both the team and player frontiers.

2.3.5 Second-Stage Model

The main purpose of this paper is to identify skillsets that may be overvalued or undervalued in the arbitration process. Two players, with otherwise identical win contributions and all else equal, should earn the same salary. This paper looks to determine whether players with different profiles (power hitter vs. an elite defender for example) are compensated differently in arbitration.

These player attributes are captured in various environmental variables with summary stats in Table 2.3. Measures similar to those used in Hakes and

Sauer (2007) are the key batting related skillsets. These include a player's ability to hit for average ("BAT"), ability to get on base ("EYE"), and ability to hit for power ("POWER").[26] A player's ability and production in the platform year is highly correlated with his career production. To account for this, a career variable and a growth variable are used for each attribute. The career variable captures a baseline of the player's ability. The growth variable captures the percentage difference between a player's career production (excluding the platform year) and the platform year. For example, from 2007–2010 Jacoby Ellsbury produced 1.390 Power. In 2011, Ellsbury produced 1.717 Power. So for his 2012 arbitration case, Ellsbury had 1.390 career power value and 0.23 power growth value.[27]

In addition to those offensive measures, environmental variables that capture defensive ability are also included. In the base model, whether a player has played multiple positions defensively during his career proxied for his defensive abilities. Just under half the players in the sample have played multiple positions. Catchers, not surprisingly, show the least positional flexibility, while outfielders show the most positional flexibility. More detailed defensive specifications and a variable capturing a player's speed are included to test the sensitivity of the base model. Baseball Reference provides data on the number of appearances a player makes at a particular position. The main variables used to dictate defensive performance are whether a player won a Gold Glove award (given each season to the player deemed to have the best defensive season at a particular position) and

[26] Derivations for these are in the Appendix.

[27] $Power\_g = \frac{Power\_p - Power\_c}{Power\_c}$

Total Zone (TZ) per 1,350 innings (an advanced defensive stat provided by Baseball Reference). Players' TZ scores are standardized based on their performances relative to the league and positional group and identified as "Great" if finishing in the top 20th percentile, "Good" if between the top 20th and 40th percentile, "Average" if between the top 40th and 60th percentile, "Poor" if between the top 60th and 80th percentile, and "Bad" if in the bottom 20th percentile. This paper looks at these attributes measured both during the platform year as well as accumulated over a full career.

The Basic Agreement mentions additional criteria in which to determine a player's salary in arbitration, including career consistency, team performance, and any physical defects. Career consistency is measured as the percentage of seasons in which a player accumulated within 80% of his best season WAR. A consistent player is defined as one that puts up similar WAR numbers each season. An example of an inconsistent player is a one-hit-wonder—a player who has one great season but is otherwise mediocre. Injury history accounts for the number of days the player spent on the disabled list (DL) in the platform season and his career (excluding the platform year). The average number of days a player spends on the DL during the platform year is just under 30 days, while the total career number is over 110 days. Finally, team win percentage is included as a measure of team performance.

The second-stage model is estimated using an approach similar to the one introduced in Simar and Wilson (2007), but is extended to estimate RCP in the double-frontier setting. They recommend a bootstrapped, truncated regression process in estimating the second-stage model, instead of the often-used censored normal (tobit) specification. The tobit model is often used because a number of efficiency estimates tend to equal one (or in this setting an RCP equalling one or zero), suggesting censoring may be taking place at full efficiency. But as Simar and Wilson (2007) note, the true underlying model does not have that mass point feature, and the mass point of full efficiency estimates is simply a consequence of the bias in the estimated efficiencies. Truncation occurs at the two RCP endpoints: zero and one.

As mentioned earlier, it is assumed that these second-stage variables satisfy separability requirements. The null hypothesis being tested in the second stage is that the environmental variables are properly priced into arbitration and thus have no effect on a player's pay outside of their contributions to winning. If separability does not hold for a given environmental variable, for example if power impacts the shape or location of the frontier, then by definition the null hypothesis is incorrect in favor of the alternative that the environmental variable does impact the frontier. Suppose power hitters face a stochastically dominant contract zone compared to contact hitters (and therefore the frontiers are located in different places), as illustrated in Figure 2.5. In that case, there may not be evidence of mispricing in the second stage because RCP estimates could be similar. Simply facing a stochasitcally dominant contract zone would indicate that power hitters are overcompensated in arbitration relative to on base hitters. Therefore, obtaining statistically significant results in the second-stage are sufficient to rejecting the null hypothesis, regardless of if separability holds. In absence of statistically significant results, testing for separability in the second-stage variables can provide a secondary source of evidence to test the alternative hypothesis. This paper leaves testing separability in these variables for future work and analyzes just the second-stage results.

2.4 Results

2.4.1 RCP Distribution

Table 2.4 provides the mean and standard deviation for an array of RCP specifications. Unsurprisingly the mean real RCP estimates are slightly lower than the nominal ones. The arbitration player is always going to be the most recent player in his market, so when adjusting salaries for inflation, he will get the biggest downward adjustment, moving him closer to the bottom frontier.

The bias correction reduces the RCP slightly for each specification. This makes sense considering the nature of the two frontiers. The lower bound of the team frontier is the minimum salary point since the team cannot pay lower than that amount. Meanwhile, there is no theoretical bound for the player frontier. This means there is more room for bias towards the upper frontier, leading to downward adjusting bias corrections.

In regards to the input weighting approach, the WAR RCP estimates were greater for the dimension reduction method in the salary specifications and greater for the composite method in the raise specifications. In each of the PA specifications, the composite RCP estimates are greater. This suggests that the frontier approach is sensitive to the inputs used and the mechanism used to weight them in the career versus the platform, which will be apparent in some of the second-stage model estimates.

Figure 2.6 displays a multi-dimensional density plot of the RCP distribution for different WAR levels for the bias adjusted composite weighting approach (the images are similar for the dimension reduction weighting specifications). There are more observations at the lower WAR levels, which is supported by the WAR density plot in Figure 2.2. Mass points at zero and one RCP are consistent across WAR levels.

2.4.2 Second-Stage Model Results

Base Model RCP Specifications

Table 2.5 provides results for the base models for various combinations of inputs, input weighting approaches, and real outputs. The career variables have larger magnitudes in the salary arguments than the raise arguments, which makes sense since the salary argument is going to rely more heavily on total career production. The growth variables show more evidence of statistical significance in the raise arguments.

Eye estimates are mostly negative, especially in the career WAR specifications. While the various specifications are negative, the composite input weighting scheme shows stronger effects of Eye on RCP than the dimension reduction technique. So while the magnitudes are dependent on the input-output specifications, these results provide somewhat strong evidence that Eye is undervalued in arbitration.

The Bat coefficients are less conclusive. The larger magnitudes in the PA specifications point to an emphasis on counting stats. A higher batting average is also associated with more hits (H) and runs batted in (RBI), and a player with more plate appearances and a higher batting average will have even more of those stats. While PA is non-decreasing with each additional at-bat, WAR can increase or decrease depending on the result of the appearance. Thus, WAR does a better job capturing the ability to get hits without capturing these counting stat effects. The signs also flip between positive or negative depending on the weighting scheme. Given the noise in the coefficients, it is difficult to come to any conclusions about Bat.

The most conclusive results are in the Power numbers. As was speculated, Power growth and career return positive and statistically significant in nearly all of the model specifications. The magnitudes are greater in the PA specification—more PAs means more HRs—but the steadily positive and statistically significant results indicate that hitters who hit for power have a greater RCP and are thus better compensated in arbitration relative to their win production.

Table 2.6 takes some of the coefficient estimates for Power and Eye from Table 2.5 and calculates the impact of increasing the career variable by one standard deviation and the growth variable by one percentage point. This is done using the average estimated 2017 contract zone length for the designated specification. Standard deviation values for Power and Eye are computed using player stats from

2001–2017. The table suggests that increasing career power by one standard deviation could cause a player to be overpaid in arbitration by at least $340,000 in salary or over $170,000 in raise. Increasing eye has an even greater negative effect. Increasing career eye by one standard deviation could cause a player to be underpaid by over $450,000 in salary or $260,000 in raise.[28] It is important to remember that the second-stage results hold the input level constant. Power cannot increase by one standard deviation with WAR remaining constant without decreasing some other attribute. So it is difficult to make conclusions about the impact of the mispricing.

In the base model, defensive ability is proxied by whether the player is versatile enough to play multiple positions. The effect is negative or close to zero for each of the specifications, but generally not statistically significant. In this initial model, there is not enough evidence to suggest defense is mispriced. A different model with more detailed defensive metrics is analyzed in Table 2.8.

Going back to Table 2.5, the additional environmental variables provide some interesting results. Career consistency is not statistically significant and is generally inconclusive with signs flipping depending on the specification. While the disabled list coefficients are inconclusive, the plate appearance variable in the WAR specifications is positive and statistically significant. In fact, the marginal effects on plate appearances are fairly substantial, most likely capturing the impact of counting stats, which, as mentioned previously, play a major role in arbitration

[28]These figures are in 2009 dollars. Converting them to 2017 dollars puts these values at $390,000 in salary and $190,000 in raise for power and $510,000 in salary and $294,000 in raise for eye.

outcomes. Having an additional 100 plate appearances means an increase in RCP between 0.245 and 0.419. Including these PA variables in the WAR model likely separates the counting stat nature of the variables that make up Eye, Bat, and Power, allowing for an analysis of the profiles illustrated by those variables rather than intertwining them with playing time effects. Finally, team win percentage is not statistically significant in any of the model specifications.

Model With Speed and Defense

Table 2.7 builds on Table 2.5 by including variables capturing a player's speed, including his ability to steal bases, advance on flyballs, and advance on the basepaths in other situations when possible. Because catchers and first basemen are generally slower players, the speed variable is interacted with an indicator variable that equals one if the player does not primarily play either of those two positions. This allows for the effects of speed to vary between the two sets of positions. The only time the coefficient is statistically significant is in the PA composite specification. Part of this is surely due to the increased opportunity in accumulated steals that is afforded by a player with more plate appearances. Therefore, these is no substantial evidence that speed players are mispriced in arbitration.

Table 2.8 introduces more detailed defensive variables. Gold Gloves, and player awards in general, are strongly relied upon by arbitrators. By analyzing defensive ability and Gold Glove status (Gold Glove winner or Gold Glove finalist), the model can evaluate whether defense is properly priced in arbitration and whether Gold Glove status is an effective proxy for defensive performance. Results are mixed. They differ between the WAR and PA specifications and between the dimension reduction and composite weighting schemes. The WAR coefficients are mostly negative, but that could be related to how WAR treats defensive ability compared to the TZ variable.[29] After including actual measures of defensive performance, multi-position eligibility is now negative and statistically significant across the majority of the specifications. This indicates that defensive versatility, often displayed from utility players, is an undervalued asset in arbitration. Overall, there is not much evidence to support that defensive ability is mispriced, but there is evidence to support that versatility is.[30]

2.4.3 Convexity Test

Table 2.9 shows the results for the FDH versus DEA-VRS convexity test. The number in the table is the percentage of arbitration markets where convexity (DEA-VRS) is rejected in favor of non-convexity (FDH). Rejection is defined in each market at the 90%, 95%, and 99% confidence levels. The top part of the table tests convexity using the dimension reduction weighting technique, while the bottom part of the table uses the composite weighting approach.

The obvious takeaway is that convexity is rejected much more often for the

[29]The career variables are being compared with the excluded "awful" variable which captures players who ranked in the bottom 20% for their position.

[30]Included in the model, not published in the table also statistically insignificant, were various interactions between Gold Glove status and defensive ability. In addition to Gold Glove status, various other defensive award status combinations were used. This includes number of Gold Gloves won in the previous three seasons (with prior three season performance interactions) and number of times finishing as a Gold Glove finalist (including interactions). None of these specifications produced statistically significant results on defense, partly due to a collinearity issue (in some cases there were zero observations in a given bucket).

player frontier than the team frontier. A potential theory for this has to do with the salary floor at the lower end of the frontier. A salary floor may lead to more observations clustered around the minimum (when they may otherwise be lower), leading to a convex shape (potentially exhibiting constant returns to scale). This theory is supported by the jump in rejection rate for the raise specification over the salary specification. There is no required minimum raise, so there is less likely to be a mass clustering of observations at the minimum points.[31]

One issue with this testing approach is that the various tests are not independent of one another. Many of these markets will share comparable cases, meaning the samples used to construct the frontiers are not independent of one another. Another issue with this testing approach is that there is no theory to suggest when it would be appropriate to conclusively reject convexity in favor of non-convexity over the full sample. Based on the results of this test, it would seem that convexity should not be assumed for the player frontier, but it may be an acceptable assumption for the team frontier. With that said, convexity is rejected in at least a sizeable percentage of team frontier cases, which would support not assuming convexity at all. FDH is consistent under both convexity and non-convexity, so it seems like the appropriate tool given the general uncertainty about the convexity assumption.

[31]Teams are not allowed to cut players' salaries by more than 20%. So while there is a minimum raise of sorts, that minimum is negative and not relevant here.

2.5 Conclusions

Baseball arbitration, while providing incentives for good player performance and getting contracts closer to true market value than what was possible under the Reserve Clause, suppresses overall contracts. But the arbitration market does more than suppress salaries; it also transfers wealth between different player profiles. It does so due to market divergence where the arbitration market prioritizes player skillsets differently than would the free agent market. This is prompted by a system that makes its decisions using player comparables, whose salaries were determined during a period that valued players differently than they are today.

Results show that power hitters are overcompensated in the arbitration market compared to similarly productive on-base (Eye) players. This trend is counter to what previous work has argued has taken place in the free agent market. Previous work show that the marginal returns to on-base skill are now better compensated in the open market in relation to its impact on winning (and similarly for power hitters). Conclusions based on defensive ability and speed are limited given the limited availability of quality variables used to measure them.[32]

Future work in this area will look at the evolution process of the free agent and arbitration market, specifically over the past 20 years. This paper does not analyze time series trends, so while results provide evidence of systematic pricing, they say nothing about the rate at which new information enters each market. In general, the Moneyball revolution has led to major improvements in how teams

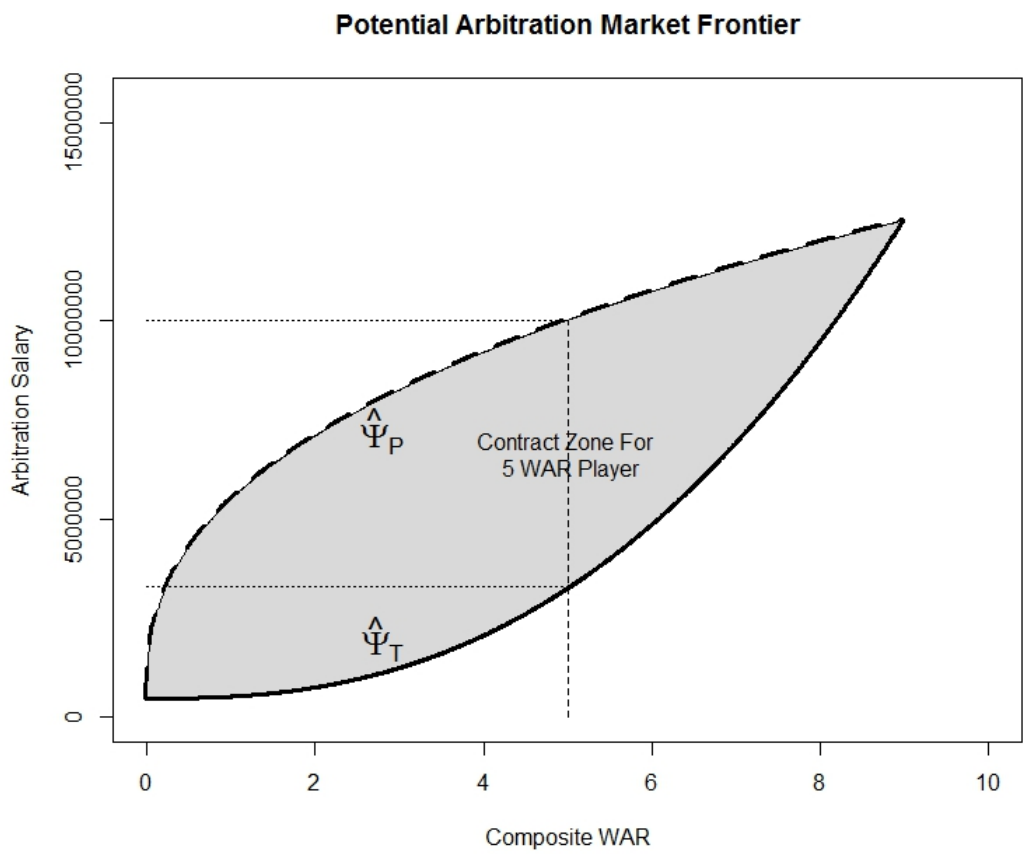
[32] Statcast data measuring defensive ability is fairly recent and only covers the past few seasons of data.

value defense and other previously lesser-valued skill sets. In a general economic context, this could provide empirical evidence to the downside of relying upon precedence in making current decisions.[33]

Another extension for future research is to analyze whether the current iteration of final-offer arbitration is ideal for the players. Because teams have the ability to release players before the arbitration process at no cost (non-tender prior to arbitration) or after arbitration for a percentage of the salary, this issue may be more than just a redistribution of wealth between players. If power players, who get overpaid in relation to players of other skill sets, can be released, then the only players that remain are those who are either underpaid relative to what arbitration should award, or underpaid relative to their market worth (or both). This would essentially be a redistribution of money from the players to teams.

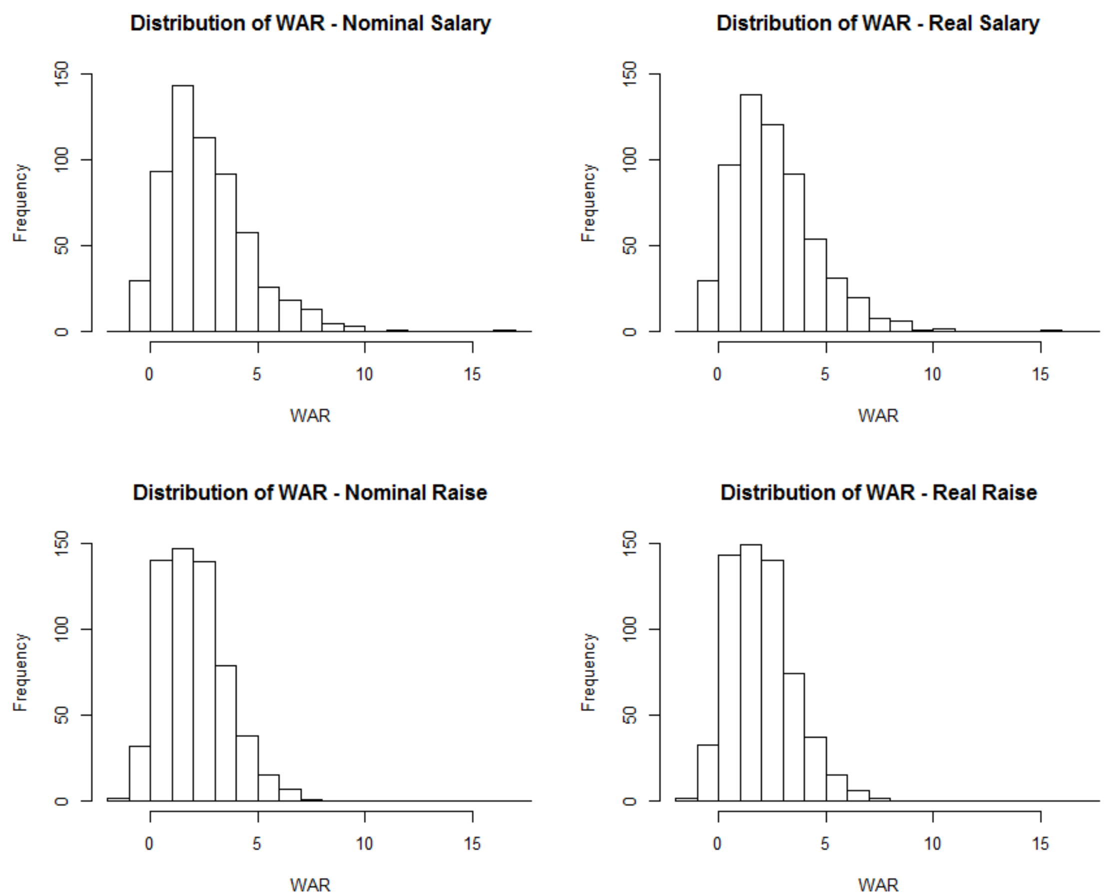
[33]The most obvious context would be in legal proceedings where public sentiment has changed but the law has not. Another obvious context would be in other labor market settings where better data allow for a more accurate measure of an employee's marginal revenue product, yet where salary decisions are being made based on antiquated criteria.

Figure 2.1: Generic arbitration market



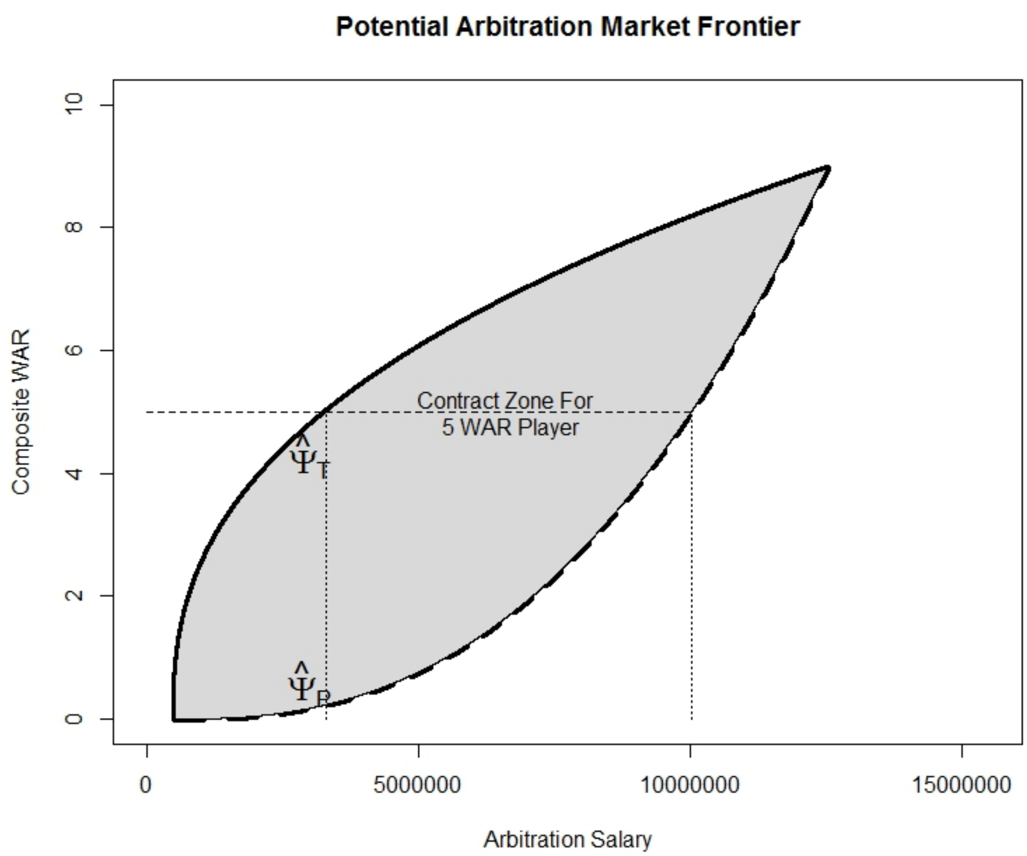
The top frontier is the player frontier and the bottom frontier is the team frontier. The vertical dotted line at 5 WAR represents a potential contract zone, with upper and lower bounds, for a player of that skill level.

Figure 2.2: WAR distribution by output type



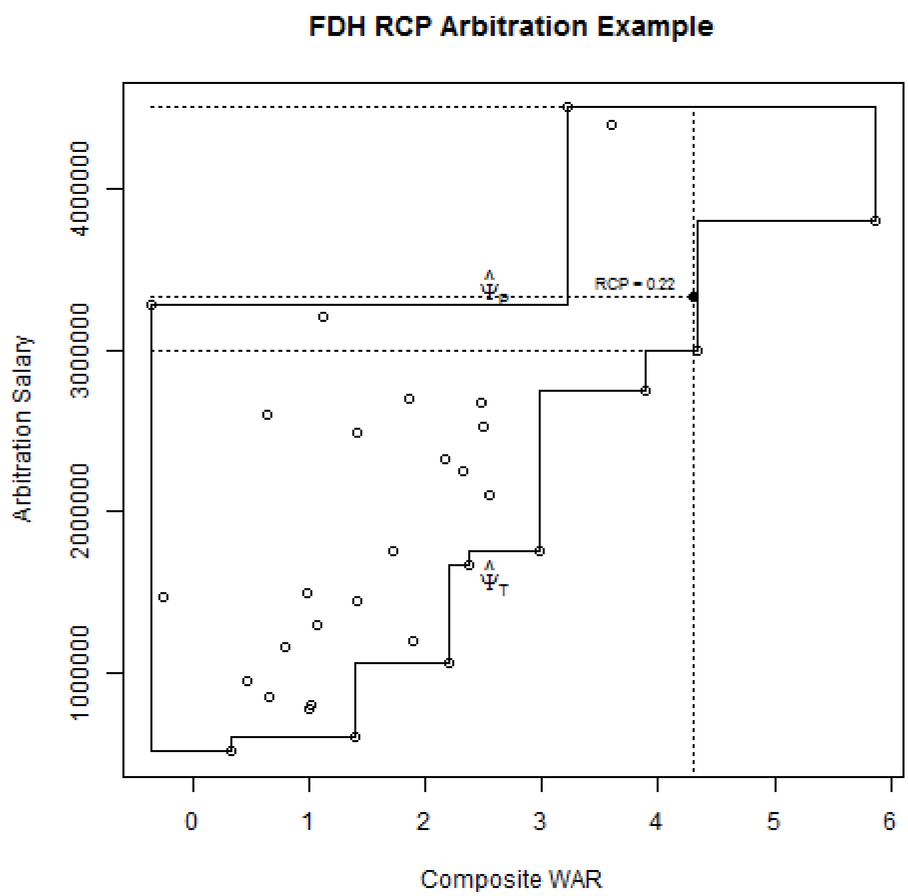
Displayed are distributions of the WAR input depending on the output type used to calculate it. The WAR value used comes from the composite measure.

Figure 2.3: Lower frontier rotation



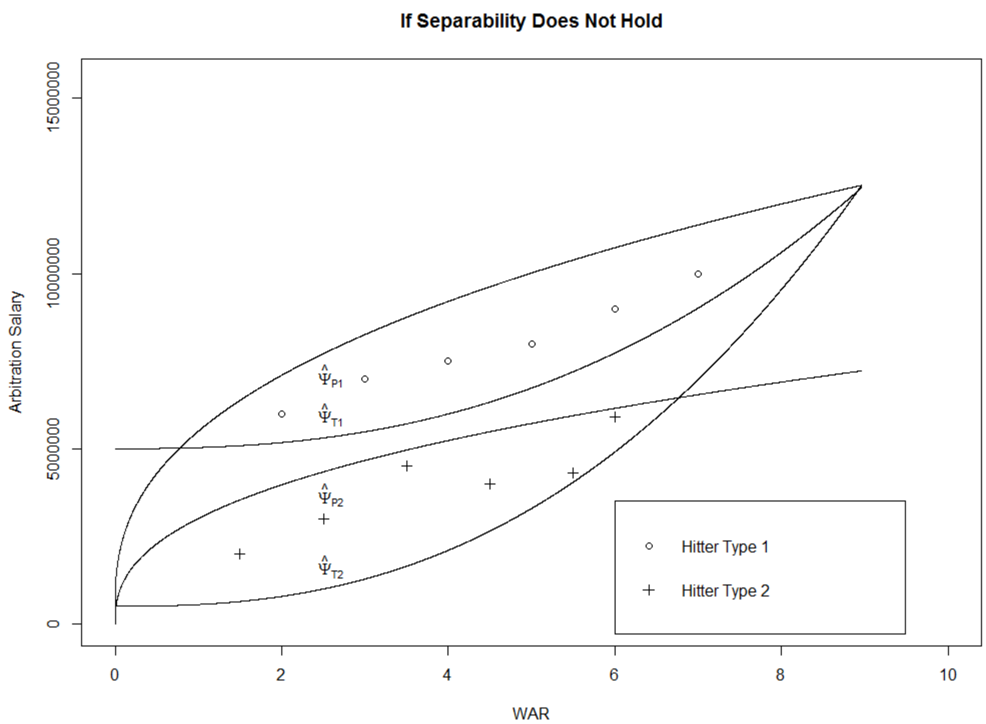
When rotating the axes, arbitration salary is now along the X-axis and WAR is along the Y-axis. The team frontier is now on top, and the player frontier is located on the bottom.

Figure 2.4: FDH double frontier example



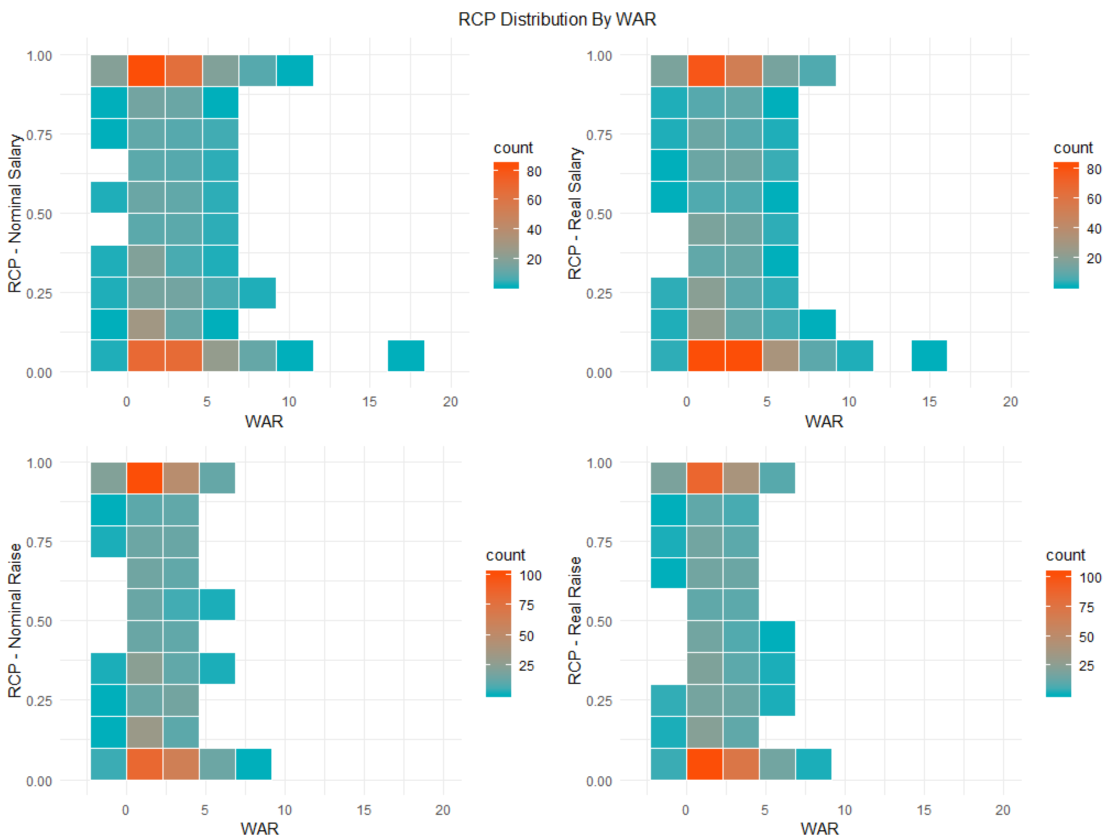
Above is an example of an estimated double frontier using the FDH estimator, illustrating how RCP is measured for a particular player. In the example above, the player of interest, filled in black, has an RCP of 0.22. That means he is closer to the team frontier (the bottom of the contract zone) than the player frontier (the top of the contract zone). The salary locations of the top and bottom of the contract zone are indicated by the top and bottom horizontal dotted lines.

Figure 2.5: If separability does not hold



This is what the actual frontiers could look like if separability did not hold for two different types of hitters. For example, if power hitters face a different frontier than on-base hitters, the RCP estimates from the first stage would be misleading and the results from the second stage model would be useless. However, this setup would support the initial hypothesis that certain attributes are mispriced in arbitration.

Figure 2.6: RCP distribution by WAR



This figure is a breakdown of the density of the different RCP estimates by WAR total for the bias adjusted composite weighting approach. The figure looks similar for the dimension reduction weighted specifications.

Table 2.1: Salary Summary Statistics.

	N	Nominal Salary ($)	Real Salary ($)	Nominal Raise ($)	Real Raise ($)
C	107	2,258,038	2,161,452	1,315,418	1,269,056
MI	265	3,877,244	3,801,798	2,099,801	2,065,967
CI	175	2,927,212	2,911,142	1,729,651	1,721,894
OF	191	2,969,175	2,930,704	1,666,164	1,658,315
1 SAE	365	1,957,572	1,946,003	1,957,572	1,946,003
2 SAE	210	3,314,453	3,267,516	1,340,023	1,321,680
3/4 SAE	163	5,229,463	5,100,231	1,772,337	1,733,311

Notes: Nominal and real salary and raise are broken down by service time and position. For the positions, C, MI, CI, and OF refer to catchers, middle infielders (second basemen, shortstops, and centerfielders), corner infielders (first basemen, third basemen, and designated hitters), and outfielders (left fielders and right fielders), respectively. The raise salary for first time arbitration eligible players is equal to their actual salary since their previous salary would have been the league minimum. All real variables are in terms of 2009 dollars.

Table 2.2: R_X Values, Information Captured By Dimension Reduction Technique

WAR (mean R_X: 0.9077)				
Min	1st Quartile	Median	3rd Quartile	Max
0.7343	0.8713	0.9059	0.9518	0.9878

PA (mean R_X: 0.9850)				
Min	1st Quartile	Median	3rd Quartile	Max
0.9649	0.9775	0.9834	0.9935	0.9981

Notes: R_X values capture the amount of information retained from the dimension reduction approach going from two inputs to one. Each arbitration market goes through a dimension reduction procedure, so each market has its own R_X value for WAR and PA. The numbers in the table are summary statistics of the distribution of the individual arbitration market R_X values.

Table 2.3: Second-Stage Variable Summary Statistics

	Summary Stats				Position Averages			
Variable	Mean	SD	Min	Max	C	MI	CI	OF
Eye_g	0.051	0.302	-1.000	3.052	0.103	0.032	0.074	0.029
Eye_c	0.085	0.025	0.027	0.168	0.086	0.080	0.083	0.092
Bat_g	0.019	0.141	-1.000	1.703	0.045	0.021	0.016	0.004
Bat_c	0.261	0.022	0.178	0.324	0.245	0.262	0.264	0.265
Power_g	0.020	0.113	-1.000	0.793	0.032	0.015	0.035	0.005
Power_c	1.565	0.172	1.158	2.035	1.548	1.481	1.629	1.630
Speed_p	0.655	3.334	-9.800	14.00	-1.500	2.458	-0.941	0.925
Speed_c	1.912	7.549	-19.80	34.90	-3.209	6.651	-2.938	3.171
MultiPos_c	0.477	0.500	0	1	0.065	0.457	0.429	0.780
TZ_p	1.034	13.58	-73.10	75.20	-0.285	1.830	1.621	0.131
TZ_c	0.890	10.36	-38.90	64.20	0.074	2.181	0.170	0.216
GG_w_p	0.038	0.191	0	1	0.009	0.057	0.040	0.026
GG_f_p	0.083	0.276	0	1	0.075	0.098	0.097	0.052
PA_p	449.9	165.7	0	744	318.4	481.9	492.2	440.4
PA_c	1213.4	640.4	64	6661	808.6	1261.9	1406.3	1196.1
DLDays_p	29.2	53.4	0	192	35.7	28.7	22.7	32.0
DLDays_c	111.1	135.7	0	661	124.3	109.4	99.0	117.3
Consistency_c	0.297	0.144	0	1	0.278	0.308	0.293	0.295
Win_Pct	0.509	0.065	0.315	0.716	0.511	0.504	0.509	0.514

Notes: Summary stats for variables used in the second-stage model. Growth variables are denoted by $\_g$ and career variables are denoted by $\_c$. The last four columns provide means by position.

Table 2.4: Mean RCP For Different Efficiency Specifications

Input:	Wins Above Replacement (WAR)			
Bias Corrected:	Yes		No	
Weighting:	Comp	Red	Comp	Red
Nominal Salary	0.5011	0.5059	0.5038	0.5092
	(0.4150)	(0.4062)	(0.4146)	(0.4056)
Real Salary	0.4512	0.4572	0.4537	0.4603
	(0.4162)	(0.4038)	(0.4160)	(0.4036)
Nominal Raise	0.5000	0.4782	0.5034	0.4836
	(0.4071)	(0.3832)	(0.4064)	(0.3821)
Real Raise	0.4529	0.4414	0.4562	0.4468
	(0.4058)	(0.3788)	(0.4055)	(0.3781)

Input:	Plate Appearances (PA)			
Bias Corrected:	Yes		No	
Weighting:	Comp	Red	Comp	Red
Nominal Salary	0.6249	0.5700	0.6264	0.5722
	(0.4116)	(0.4143)	(0.4110)	(0.4137)
Real Salary	0.5346	0.5136	0.5359	0.5157
	(0.4269)	(0.4186)	(0.4293)	(0.4183)
Nominal Raise	0.6108	0.5380	0.6129	0.5420
	(0.4098)	(0.3979)	(0.4089)	(0.3969)
Real Raise	0.5456	0.4963	0.5477	0.5005
	(0.4183)	(0.3984)	(0.4177)	(0.3977)

Notes: Standard deviation in parentheses; Comp is the composite weighted platform and career input specification; Red is the platform and career dimension reduction weighted input specification.

Table 2.5: Real Base Model With Attribute Growth Variables

Output	Real Salary				Real Raise			
Input	WAR		PA		WAR		PA	
Weighting	Comp	Red	Comp	Red	Comp	Red	Comp	Red
Model #	(1a)	(2a)	(3a)	(4a)	(5a)	(6a)	(7a)	(8a)
Eye_g	-0.599	-0.156	0.001	-0.377	-0.289**	-0.036	0.641	-0.089
	(0.405)	(0.145)	(0.149)	(0.570)	(0.146)	(0.107)	(1.683)	(0.181)
Eye_c	-13.01*	-5.797***	-2.018	-0.648*	-6.998***	-3.887***	38.21*	0.550
	(7.246)	(2.110)	(1.966)	(5.646)	(2.254)	(1.378)	(21.16)	(2.214)
Bat_g	-0.608	0.320	0.517	6.191	-0.731*	0.834***	13.84***	3.283***
	(0.807)	(0.364)	(0.418)	(5.584)	(0.402)	(0.295)	(5.260)	(1.084)
Bat_c	-4.296	-2.960	1.963	31.78	-4.046*	-1.950	85.16***	14.09***
	(4.895)	(2.267)	(2.278)	(28.71)	(2.302)	(1.679)	(19.77)	(4.874)
Power_g	0.246	0.192	0.944*	3.145	0.080	0.646**	13.83**	1.798**
	(0.774)	(0.393)	(0.501)	(3.058)	(0.373)	(0.307)	(5.738)	(0.728)
Power_c	0.984	0.641**	1.237***	3.074	0.521**	0.517***	10.71***	1.537***
	(0.664)	(0.263)	(0.389)	(2.819)	(0.251)	(0.185)	(3.481)	(0.537)
PA_p	0.004**	0.002***			0.003***	0.003***		
	(0.002)	(0.001)			(0.001)	(0.000)		
MultiPos_c	-0.258	-0.098	-0.065	-0.678	-0.142*	-0.031	-1.138	-0.058
	(0.193)	(0.075)	(0.081)	(0.644)	(0.076)	(0.053)	(0.914)	(0.096)
Consistency_c	0.084	0.013	-0.312	-0.498	-0.351	0.236	0.567	0.409
	(0.536)	(0.263)	(0.304)	(1.032)	(0.266)	(0.192)	(2.870)	(0.358)
DLdays_p	0.002	0.000	0.000	-0.001	0.000	0.000	-0.005	-0.002**
	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.008)	(0.001)
tmWinPct	-0.835	-0.299	-0.654	-4.459	0.434	0.041	5.258	-0.800
	(1.170)	(0.546)	(0.659)	(4.394)	(0.542)	(0.426)	(6.481)	(0.786)
Constant	-0.587	-0.353	-1.469	-10.25	-0.214	-0.959	-44.62***	-5.464***
	(1.614)	(0.768)	(0.941)	(9.748)	(0.749)	(0.601)	(9.035)	(1.997)
Observations	276	304	202	268	303	378	262	324

Notes: Standard errors in parentheses,
all numbers rounded to nearest thousandths

*p<0.1; **p<0.05; ***p<0.01

Table 2.6: Monetary Impact Of Mispricing Power and Eye

	Model #	(1c)	(2c)	(5c)	(6c)
2017 Avg Contract Zone:		\$2,200,000	\$2,533,000	\$1,566,000	\$2,157,000
Variable	Δ				
Power_c	1 SD (0.211)	\$457,000	\$343,000	\$172,000	\$235,000
Power_g	1% (0.01)	\$5,400	\$4,800	\$1,200	\$13,900
Eye_c	1 SD (0.031)	-\$887,000	-\$455,000	-\$340,000	-\$260,000
Eye_g	1% (0.01)	-\$13,200	-\$4,000	-\$4,500	-\$800

Notes: Approximates the impact of a 1 standard deviation (SD) change or a 1 percentage point change in various variables from Table 2.5. One SD calculations were done using MLB data from 2001–2017. The magnitude of changing RCP depends on the size of the contract zone. The above numbers are for the real salary and real raise WAR specifications and take the average contract zone length for the 2017 arbitration cases. Career numbers are rounded to the nearest thousand dollars and growth numbers are rounded to the nearest hundred dollars.

Table 2.7: Real Base Model With Speed and Attribute Growth Variables

Output	Real Salary				Real Raise			
Input	WAR		PA		WAR		PA	
Weighting	Comp	Red	Comp	Red	Comp	Red	Comp	Red
Model #	(1b)	(2b)	(3b)	(4b)	(5b)	(6b)	(7b)	(8b)
Eye_g	-0.591	-0.158	-0.010	-0.213	-0.314**	-0.039	0.619	0.122
	(0.368)	(0.155)	(0.144)	(0.415)	(0.148)	(0.110)	(1.435)	(0.173)
Eye_c	-11.37*	-5.930**	-2.443	-2.815	-6.340***	-3.509**	23.76	-0.396
	(6.020)	(2.309)	(1.934)	(5.144)	(2.206)	(1.421)	(41.09)	(2.122)
Bat_g	-0.584**	0.277	0.459	5.219	-0.826**	0.832***	8.570	3.137***
	(0.731)	(0.386)	(0.400)	(3.912)	(0.411)	(0.303)	(14.42)	(0.985)
Bat_c	-4.554	-3.134	2.309	27.89	-4.606*	-2.776	51.82	13.66***
	(1.583)	(2.472)	(2.227)	(20.90)	(2.374)	(1.788)	(86.76)	(4.496)
Power_g	0.090	0.193	0.883*	2.455	-0.096	0.568*	8.792	1.666**
	(0.697)	(0.419)	(0.477)	(2.078)	(0.372)	(0.313)	(14.84)	(0.661)
Power_c	0.848	0.662**	1.204***	2.757	0.291	0.406**	6.783	1.497***
	(0.578)	(0.291)	(0.373)	(2.108)	(0.249)	(0.194)	(11.35)	(0.502)
PA_p	0.004**	0.003***			0.003***	0.003***		
	(0.002)	(0.001)			(0.001)	(0.000)		
MultiPos_c	-0.306	-0.109	-0.046	-0.578	-0.119	-0.034	-0.603	-0.055
	(0.190)	(0.081)	(0.078)	(0.465)	(0.077)	(0.056)	(1.123)	(0.092)
Consistency_c	-0.094	-0.044	-0.311	-0.557	-0.372	0.235	0.256	0.344
	(0.497)	(0.283)	(0.295)	(0.874)	(0.271)	(0.200)	(1.787)	(0.335)
DLdays_p	0.002	0.000	0.000	-0.001	0.000	0.001	-0.002	-0.002**
	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
tmWinPct	-0.460	-0.359	-0.677	-4.159	0.419	-0.022	4.926	-0.919
	(1.057)	(0.588)	(0.635)	(3.426)	(0.551)	(0.440)	(9.099)	(0.758)
Speed˙p	0.027	0.023	-0.074*	-0.062	-0.030	0.023	0.146	-0.029
	(0.062)	(0.032)	(0.040)	(0.109)	(0.033)	(0.022)	(0.315)	(0.039)
X C/1B	-0.040	-0.027	0.079*	0.097	0.003	-0.044	-0.172	0.037
	(0.068)	(0.035)	(0.043)	(0.128)	(0.035)	(0.025)	(0.363)	(0.042)
Constant	-0.360	-0.297	-1.485	-8.699	0.220	-0.619	-28.29	-5.134***
	(1.496)	(0.835)	(0.919)	(6.959)	(0.761)	(0.619)	(47.88)	(1.790)
Observations	267	294	202	264	294	368	259	320

Notes: Standard errors in parentheses, all numbers rounded to nearest thousandths

*p<0.1; **p<0.05; ***p<0.01

Table 2.8: Real Base Model With Speed, Defense, and Attribute Growth Variables

Output	Real Salary				Real Raise			
Input	WAR		PA		WAR		PA	
Weighting	Comp	Red	Comp	Red	Comp	Red	Comp	Red
Model #	(1c)	(2c)	(3c)	(4c)	(5c)	(6c)	(7c)	(8c)
Eye_g	-0.468	-0.136	0.018	-0.149	-0.270**	-0.037	0.537	0.016
	(0.308)	(0.153)	(0.134)	(0.272)	(0.126)	(0.111)	(0.682)	(0.159)
Eye_c	-9.538**	-5.253**	-2.509	-0.565	-6.543***	-3.520**	12.54	-0.641
	(4.848)	(2.167)	(1.808)	(3.204)	(1.902)	(1.456)	(10.18)	(1.930)
Bat_g	-0.276	0.276	0.264	3.605**	-0.765**	0.819***	5.922*	2.551***
	(0.651)	(0.380)	(0.373)	(1.747)	(0.362)	(0.308)	(3.497)	(0.734)
Bat_c	-4.229	-3.193	1.214	18.20**	-4.373**	-3.098*	36.49*	11.28***
	(4.156)	(2.368)	(2.048)	(8.901)	(2.045)	(1.817)	(20.99)	(3.453)
Power_g	0.064	0.182	0.814*	1.600	-0.069	0.562*	5.316	1.503***
	(0.643)	(0.414)	(0.429)	(1.002)	(0.318)	(0.317)	(3.432)	(0.554)
Power_c	0.614	0.638**	1.050***	1.993**	0.242	0.387*	5.089*	1.321***
	(0.486)	(0.283)	(0.315)	(0.997)	(0.221)	(0.198)	(2.861)	(0.403)
PA_p	0.004**	0.002***			0.003***	0.003***		
	(0.002)	(0.001)			(0.001)	(0.000)		
MultiPos_c	-0.324*	-0.164*	-0.007	-0.431*	-0.130*	-0.057	-0.445	-0.048
	(0.177)	(0.084)	(0.074)	(0.237)	(0.068)	(0.057)	(0.406)	(0.084)
Consistency_c	-0.031	-0.009	-0.385	-0.513	-0.395	0.166	-0.075	0.358
	(0.464)	(0.280)	(0.274)	(0.580)	(0.240)	(0.201)	(1.119)	(0.311)
DLdays_p	0.002	0.000	0.000	-0.001	0.001	0.001	-0.002	-0.002**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)
tmWinPct	-0.580	-0.324	-0.812	-3.639*	0.407	0.006	2.935	-1.139
	(0.975)	(0.564)	(0.583)	(1.953)	(0.479)	(0.442)	(2.981)	(0.695)
Great_Defense_p	-0.510	-0.109	0.439*	1.220	-0.485**	-0.029	0.338	0.414
	(0.418)	(0.231)	(0.260)	(0.785)	(0.199)	(0.169)	(1.000)	(0.289)
Good_Defense_p	-0.667*	-0.169	0.196	0.965	-0.559***	-0.052	-0.803	0.358
	(0.402)	(0.195)	(0.200)	(0.663)	(0.176)	(0.138)	(0.926)	(0.259)
Average_Defense_p	-0.762*	-0.123	-0.023	1.118	-0.591***	-0.013	-0.473	0.309
	(0.452)	(0.225)	(0.210)	(0.721)	(0.196)	(0.159)	(0.970)	(0.271)
Poor_Defense_p	-0.684*	-0.155	0.130	1.017	-0.398**	-0.065	-0.615	0.319
	(0.396)	(0.187)	(0.188)	(0.662)	(0.159)	(0.132)	(0.853)	(0.246)
Great_Defense_c	-0.187	0.348	-0.234	0.991	-0.323	0.235	2.637	0.241
	(0.493)	(0.305)	(0.274)	(0.723)	(0.256)	(0.219)	(1.891)	(0.304)
Good_Defense_c	0.070	0.168	0.052	0.756	-0.002	0.118	3.283*	0.059
	(0.396)	(0.250)	(0.200)	(0.615)	(0.184)	(0.179)	(1.970)	(0.254)
Average_Defense_c	0.105	0.091	0.023	1.319*	0.011	0.090	3.281*	0.320
	(0.420)	(0.261)	(0.199)	(0.779)	(0.191)	(0.182)	(1.978)	(0.266)
Poor_Defense_c	0.114	0.325	0.013	0.742	-0.069	0.201	3.187*	0.193
	(0.389)	(0.249)	(0.188)	(0.596)	(0.180)	(0.172)	(1.911)	(0.247)
Gold_Glove_Winner_p	-0.186	-0.194	0.200	1.260	-0.079	-0.115		13.07
	(0.875)	(0.554)	(0.573)	(1.438)	(0.246)	(0.315)		(13.07)
Gold_Glove_Finalist_p	-0.402	-0.134	0.437*	0.491	0.085	-0.023		-0.053
	(0.538)	(0.276)	(0.231)	(0.568)	(0.242)	(0.211)		(0.294)
Constant	0.429	-0.367	-1.043	-7.190*	0.915	-0.604	-22.14*	-4.586***
	(1.411)	(0.843)	(0.840)	(3.715)	(0.693)	(0.646)	(11.97)	(1.462)
Observations	263	288	202	259	290	362	258	315

Note: Standard errors in parenthesis, all numbers rounded to nearest thousandths, all models include speed variables and interactions between Gold Glove status and platform ability. *p<0.1; **p<0.05; ***p<0.01

Table 2.9: Testing Convexity Assumption

Red		Player Frontier ($\hat{\Psi}_P$)			Team Frontier ($\hat{\Psi}_T$)		
Input	Output	10%	5%	1%	10%	5%	1%
WAR	Salary	.426	.347	.211	.347	.262	.176
PA	Salary	.651	.576	.421	.376	.302	.213
WAR	Raise	.329	.277	.167	.413	.343	.273
PA	Raise	.492	.397	.262	.440	.384	.316

Comp		Player Frontier ($\hat{\Psi}_P$)			Team Frontier ($\hat{\Psi}_T$)		
Input	Output	10%	5%	1%	10%	5%	1%
WAR	Salary	.490	.417	.250	.161	.118	.070
PA	Salary	.841	.824	.750	.138	.116	.085
WAR	Raise	.483	.374	.252	.215	.176	.105
PA	Raise	.824	.781	.686	.254	.211	.161

Notes: Mean efficiency is measured for each arbitration market using the jackknife sampling method. DEA-VRS, which assumes convexity, is rejected if mean efficiency for the FDH sample group is statistically different than mean efficiency for the DEA-VRS sample group. The table above shows the percentage of arbitration markets where convexity is rejected for each frontier using different inputs, outputs, and weighting schemes at different confidence intervals.

CHAPTER 3

MODELING THE COMPLETE BASEBALL ARBITRATION PROCESS

3.1 Introduction

The previous chapter discussed the role of comparable contracts in baseball arbitration and illustrated how the system undervalues and overvalues certain player types. This chapter takes a more holistic approach to analyzing the baseball arbitration process. When the Major League Baseball Player's Union (MLBPA) begins negotiation with Major League Baseball (MLB) owners for the new Collective Bargaining Agreement (CBA), arbitration will be one of many contentious negotiating battlegrounds. This paper extends prior theoretical research by constructing a model to analyze the elements of the arbitration process that are harmful to players. This model can be used in future research when analyzing arbitration outcomes.

The previous chapter discusses details on who is eligible to file for arbitration. The arbitration process consists of multiple elements, which I will summarize. First, teams have the opportunity to decide whether to engage in arbitration proceedings, with the ability to non-tender a player. If a player is non-tendered, he is released and becomes a free agent. If the player is tendered a contract, he and the team have the opportunity to negotiate a settlement offer that would preclude the parties from going to an arbitration hearing. In the absence of a settlement, the team and player submit salary figures to the arbitration panel in anticipation of going to a hearing. Prior to the hearing, the parties have one final opportunity to come to a settlement. If a settlement is not reached, the parties proceed to a final-offer arbitration (FOA) hearing. In an FOA hearing, the arbitration panel must select one of the two submission offers as the player's salary.

To date, no theoretical research has attempted to model the entire baseball arbitration process. So far, research has addressed the figure submission and second period bargaining stages. Wittman (1986) identifies the existence of an equilibrium using the Nash bargaining concept that results in a settlement. Faurot and McAllister (1992) also use the Nash bargaining concept and introduce a basic model to analyze the second bargaining period. Hanany, Kilgour, and Gerchak (2007) produce a more general model for bargaining that identifies mutually improving settlements by allowing for differing bargaining processes other than Nash. However, they do provide an example with the Nash bargaining solution and link to a theoretical framework the general observation that most cases end in settlements.

These papers do not analyze the first bargaining period, nor the team's non-tender decision. Teams incorporate various negotiation strategies that dictate their willingness to negotiate, and this willingness can differ in each of the two bargaining periods. The 'file-and-trial' strategy is one of the more popular approaches, where teams will not negotiate with players after salary offers have been submitted, which can impact bargaining outcomes. In this paper, I model the complete arbitration process.

These papers also fail to address the reality that arbitration contracts are not fully guaranteed. If a player is released during Spring Training but before 16 days prior to the start of the regular season, the player is only entitled to 30 days termination pay. If a player is released after that point, but prior to the start of the regular season, the player is entitled to 45 days termination pay. Only after the season starts does the contract become fully guaranteed. If a salary arbitration outcome is sufficiently bad for the team, it can choose to release the player, which I account for in my arbitration model.

I also add to the discussion of risk preferences in baseball arbitration. Wittman (1986) shows that in arbitration figure exchange strategy, the more risk averse party moves away from his preferred position to a less favorable figure, but one that increases his probability of winning the hearing. I add to the discussion by showing how the effects of risk preferences on player salary change under various circumstances.

Finally, I analyze a potential impact of the market mispricing discussed in the previous chapter. I provide an illustration of how players could be made collectively worse off by mispricing, and teams collectively better off. I do not identify the extent to which players are made worse off, rather I identify scenarios where they could be. Further research is necessary to quantify these findings based on market mispricing and other features of the arbitration process.

Section 2 presents the arbitration model using a backward induction approach. Section 3 introduces the simulation procedure used to tease out some of the comparative statics. Section 4 identifies key results from the simulations.

Section 5 provides summarizing and concluding remarks.

3.2 Arbitration Model

I model the arbitration process, hereby called the ‘game’, using a backward induction approach. The game consists of two active participants, the team and player. The game also consists of an arbitration panel that does not have a strategic objective other than to determine which salary submission figure is closest to their perception of the player’s true arbitration market value. Figure 3.1 depicts the game, with solid dots indicating decision nodes for at least one of the participants. First, the team decides whether to tender the player a contract. If it non-tenders the player, the player becomes a free agent and the game is over. If the team tenders the player a contract, the game proceeds to the first bargaining period. During this bargaining stage, both sides either come to a settlement or not. If a settlement is reached, the player earns the settlement amount. If a settlement is not reached, the participants proceed to the arbitration figure exchange stage. Simultaneously, the participants publicly submit salary figures to the arbitration panel. The exchange is followed by a second bargaining period. If a settlement is not reached during this period, the arbitration case goes to a hearing, where the arbitration panel decides which of the two exchange numbers to award the player for the upcoming season. After the arbitration panel makes a decision, the team has the ability to keep or release the player.

It is assumed that each participant looks to maximize its gains from the arbitration process. The participants are looking to maximize

$$U_i(x) = x^{r_i}, \tag{3.1}$$

where U_i is the utility function for participant $i \in \{t, p\}$, and t and p are the team and player, respectively. The salary the player earns or the surplus the team enjoys is represented by x. Player and team risk preferences are captured by r_i, with $r_i > 1$ indicating risk seeking preferences, $r_i < 1$ indicating risk aversion preferences, and $r_i = 1$ indicating risk neutral preferences. This is not the only feasible utility specification, but it has the benefit of allowing for various risk preferences.

Since the model is solved using backward induction, the remainder of this section will follow the backward path of the game, detailing the strategic decision making at each point.

3.2.1 Release Decision

If the game reaches this point, the team makes this final move. Based on the salary awarded to the player, S, the team decides whether to keep the player and pay S or release the player and pay dS, where d is the percentage of the contract guaranteed to the player after the arbitration decision. The team will decide to keep the player if $U_t(X_m - S) \geq U_t(-dS)$, where X_m is the player's free agent market value. This condition can be rewritten as $X_m - S \geq -dS$, or

$$X_m \geq S(1 - d). \tag{3.2}$$

A team will only release the player if the non-guaranteed portion of the player's contract is greater than the player's market value. The guaranteed part is a sunk cost and not relevant to the decision.

If the team chooses to keep the player, the player's payoff is S. If the team chooses to release the player, it is assumed that the player will sign with another club at his free agent market wage, and earn X_m.

3.2.2 Arbitration Decision

When an arbitration case goes to a hearing, an arbitration panel, consisting of three arbitrators, decides the player's salary. The panel is given two contract figures, a and b, and must pick from those two. In final-offer arbitration, the arbitrators are unable to select a middle-ground figure. The salary figures come from the salary figure exchange, where a is the player's submission and b is the team's submission.

The panel has to select the figure that they deem closest to their perception of the player's arbitration market value, X_A. The arbitrators will compare X_A with the midpoint of the two salary figure submissions,

$$M(a,b) = \frac{a+b}{2}. \tag{3.3}$$

If the midpoint is greater than X_A, the arbitration panel selects the team's submission b as the winner. If the midpoint is less than X_A, the arbitration panel selects the player's submission a as the winner.

The exact value of X_A is unknown by the participants. The following assumptions are made regarding the participants' beliefs on how X_A is generated. The player has a true arbitration market value, A, known to both participants. The arbitration panel selects a value X_A from a known function, f with cumulative distribution function, F. The median of f is A, so half the X_A draws are below A and half are above A. Assumption 3.2.2 is straightforward. By examining the results of previous arbitration cases, participants can form beliefs about the player's true arbitration market value, A. The function, f, in Assumption 3.2.2, acknowledges the idea that there is a certain level of subjectivity in setting X_A. Another assumption is needed when defining the behavior of the arbitration panel. The arbitrator exchangability hypothesis holds. There are no individual biases in the selection of X_A. The arbitration exchangeability hypothesis, as defined by Ashenfelter (1987), states that while each arbitrator has specific tendencies and evaluative methods, the expectation is that they conform to the same valuation in the long run. If they do not conform, or consistently favor one side over the other, they can be replaced in future cases. So the most sensible strategy for an arbitrator is to select a valuation system similar to what other arbitrators would select. Making this assumption allows f to be predictable across arbitration panels, regardless of the arbitrators selected. Rather, differences in X_A come from random one-off subjective differences in opinion that are not predictable ahead of time.

By extension, the cumulative distribution function, F, provides the probability of the team winning the arbitration case, given a midpoint level, M. We know that as M increases, either due to increases in a or b, the probability that it is greater than X_A also increases. Since the team wins if $M > X_A$, increasing M increases F.

The team's probability of winning the arbitration hearing can be represented using F, given by $F(a, b)$. Alternatively, the probability of the player winning the hearing is given by $1 - F$. When the team wins the hearing, the player is awarded b. When the player wins the hearing, he is awarded a. However, what both parties actually end up paying is dependent on the team's keep or cut decision in the next period. Using the condition in Equation 3.2, define the expected utility for the team of going to an arbitration hearing as

$$\begin{aligned} h_t(a,b) := F[I(X_m \geq b(1-d))U_t(X_m - b) + \\ I(X_m < b(1-d))U_t(-bd)] + (1-F)[I(X_m \geq a(1-d))U_t(X_m - a) \\ + I(X_m < a(1-d))U_t(-ad)], \end{aligned} \tag{3.4}$$

and the expected utility for the player from going to an arbitration hearing as

$$\begin{aligned} h_p(a,b) := F[I(X_m \geq b(1-d))U_p(b) + I(X_m < b(1-d))U_p(X_m)] + \\ (1-F)[I(X_m \geq a(1-d))U_p(a) + I(X_m < a(1-d))U_p(X_m)], \end{aligned} \tag{3.5}$$

where $I(\cdot)$ is the indicator function. The expected utilities, h_t and h_p, clearly depend on X_m, A, d, a, and b. The first three in that list are given for a particular arbitration case. The last two are selected strategically in the arbitration exchange period described in Section 3.2.4.

The arbitration panel plays an important role in this game. Although most cases do not reach this stage, the threat of reaching it drives how players are valued throughout the game. In theory, the greater the arbitration market values the player, the more that player should expect to earn, which improves the player's optimal bargaining position, and alters his salary figure exchange strategy. A greater arbitration market value also impacts the non-tender decision at the beginning of the game. If the player is expected to earn a higher salary through the arbitration process, the team might be more inclined to non-tender the player.

3.2.3 Second Bargaining Period

Prior to going to an arbitration hearing, the participants have an opportunity to negotiate a settlement. At this stage in the game, the two exchange numbers, a and b, are known, as are the expected utilities, h_t and h_p, from going to a hearing. Just as was done in Faurot and McAllister (1992), I use a Nash bargaining concept to derive an equilibrium solution at this stage in the game. Optimal negotiated salaries will generate utility gains for both participants over what would come from going to a hearing. From Faurot and McAllister (1992):

> *If bargaining satisfies the axioms of Nash (1950), the agreement that maximizes the product of the utility gains, relative to not negotiating an agreement, is the unique solution to the bargaining problem. Negotiating this agreement is Nash bargaining.*

Nash bargaining is solved by optimizing the gains from utility,

$$\bar{x}_2(h_t, h_p) =_y \{[U_t(X_m - y) - h_t][U_p(y) - h_p]\}, \tag{3.6}$$

where $\bar{x}_2$ is the solution to the Nash bargaining problem, and $(U_i() - h_i)$ are the gains from bargaining for $i \in \{t, p\}$.

The team will agree to the settlement offer if

$$U_t(\bar{x}_2) \geq h_t(a, b) \tag{3.7}$$

and earn a surplus of $X_m - \bar{x}_2$. The player will agree to the settlement offer if

$$U_p(\bar{x}_2) \geq h_p(a, b) \tag{3.8}$$

and earn a salary of $\bar{x}_2$. If either Equation 3.7 or 3.8 do not hold, a settlement will not occur and the game continues to the arbitration hearing.

Settlements do not always occur, even if they are Pareto efficient. In reality, there are breakdowns in negotiations that can cause the parties to go to a hearing. There may be hard feelings between the team and player, an unwillingness to negotiate, or really anything else that leads to a case not resulting in a settlement. To account for this, I introduce a variable k_2, which captures the probability of a breakdown in negotiations during this stage of the process. Both participants should have a good sense of the other's negotiating tactics, and thus should have a good sense of how likely a breakdown in negotiations will occur. If a breakdown occurs, the participants proceed to a trial, even if there would have been an amenable settlement offer.

3.2.4 Arbitration Salary Figure Exchange

During this stage of the game, participants simultaneously submit salary figures to the arbitration panel to be considered during the hearing. This subgame is solved using a Nash equilibrium solution concept. The team will submit b, such that it maximizes its expected utility. The player will submit a, such that it maximizes his expected utility.

The participants have to consider whether there will be a mutually beneficial settlement offer later in the game. Recall that $\bar{x}_2$ is a function of the two submission figures. If there will be a mutually improving settlement offer, the participants will be making their decisions with respect to producing their best possible settlement outcome. If there will not be a Pareto improving settlement offer, the participants will make their submissions looking to optimize h_t and h_p.

First, assume there will not be a settlement in the second period. The team will look to pick the value of b that maximizes h_t. Setting $\frac{\partial h_t}{\partial b} = 0$ and solving for b provides the team's best response function with respect to a, $b_{BR}(a)$. Meanwhile, the player will look to pick the value of a that maximizes h_p. Setting $\frac{\partial h_p}{\partial a} = 0$ and solving for a provides the team's best response function with respect to b, $a_{BR}(b)$. The point where $b_{BR}(a)$ and $a_{BR}(b)$ intersect, (a_l^*, b_l^*), represents the Nash equilibrium of the arbitration exchange subgame.

Next, assume there is, in fact, a mutually improving settlement offer during the second bargaining period. Now, the participants are looking to maximize their utility from a potential settlement. However, they also have to consider the possibility of negotiations breaking down, k_2, when picking submission offers. The more likely negotiations are to break down, the closer (a^*_{s2}, b^*_{s2}), the Nash equilibrium submissions when there is a preferred mutually improving settlement, will be to (a^*_l, b^*_l). When $k_2 = 1$, the two solutions will be equal. The team is now selecting b that maximizes

$$(1 - k_2)U_t(\bar{x}_2(a, b)) + k_2 h_t(a, b), \tag{3.9}$$

while the player is selecting a that maximizes

$$(1 - k_2)U_p(\bar{x}_2(a, b)) + k_2 h_p(a, b). \tag{3.10}$$

Solving for the best response functions, and finding where they intersect, produces a Nash equilibrium (a^*_{s2}, b^*_{s2}). Plugging those values into Equation 3.6 produces the Nash bargaining settlement value, $\bar{x}_2$.

If (a^*_l, b^*_l) exists, but (a^*_{s2}, b^*_{s2}) does not, (a^*_l, b^*_l) is chosen as the result of this game. If (a^*_{s2}, b^*_{s2}) exists, but (a^*_l, b^*_l) does not, (a^*_{s2}, b^*_{s2}) is chosen as the result. As long as non-tender is not the optimal strategy for the team, at least one of (a^*_l, b^*_l) or (a^*_{s2}, b^*_{s2}) will always exist. Conjecture 3.2.4 holds true in simulations discussed later.

Finally, if both equilibrium outcomes exist, the participants have to decide between them. The team will choose b^*_{s2} and the player will choose a^*_{s2} if

$$X_m - \bar{x}_2 \geq h_t(a^*_l, b^*_l) \tag{3.11}$$

and

$$\bar{x}_2 \geq h_p(a_l^*, b_l^*). \tag{3.12}$$

If either of these conditions do not hold, meaning if either participant (or both) prefers the submission strategy that results in a hearing, the team will choose b_l^* and the player will choose a_l^*. Define the team's final arbitration submission as b^*, and the player's final arbitration submission as a^*.

3.2.5 First Bargaining Period

This stage is similar to that of the second bargaining period. Knowing what a^* and b^* will be if the salary figure exchange takes place, the Nash bargaining solution, $\bar{x}_1$, can easily be calculated. First, define the team's expected utility from not settling in the first bargaining period as

$$H_t = \begin{cases} (1-k_2)U_t(\bar{x}_2(a^*, b^*)) + k_2 h_t(a^*, b^*) & \text{if Eq 3.11, 3.12 hold} \\ h_t(a^*, b^*) & \text{otherwise} \end{cases}, \tag{3.13}$$

and the player's expected utility from not settling in the first bargaining period as

$$H_t = \begin{cases} (1-k_2)U_p(\bar{x}_2(a^*, b^*)) + k_2 h_p(a^*, b^*) & \text{if Eq 3.11, 3.12 hold} \\ h_p(a^*, b^*) & \text{otherwise} \end{cases}. \tag{3.14}$$

The solution to the bargaining problem at this stage is

$$\bar{x}_1(H_t, H_p) =_y \{[U_t(X_m - y) - H_t(a^*, b^*)][U_p(y) - H_p(a^*, b^*)]\}. \tag{3.15}$$

The team will agree to the settlement offer, $\bar{x}_1$, if

$$U_t(\bar{x}_1) \geq H_t(a^*, b^*), \tag{3.16}$$

and earn a surplus of $X_m - \bar{x}_1$. The player will agree to the settlement offer if

$$U_p(\bar{x}_1) \geq H_p(a^*, b^*), \tag{3.17}$$

and earn a salary of $\bar{x}_1$. If the conditions in Equations 3.16 or 3.17 do not hold, a settlement will not occur and the game continues to the arbitration figure exchange. As with the second bargaining period, there is some chance negotiations will break down, captured by k_1.

3.2.6 Non-tender Decision

The first stage of the game, and last step in backward induction, is the team's decision on whether to tender the player a contract. That decision will depend on the team's expected utility if it tenders the player a contract. First, define the team's expected utility from tendering as

$$T = \begin{cases} (1-k_1)U_t(X_m - \bar{x}_1) + k_1 H_t(a^*, b^*) & \text{if Eq 3.16, 3.17 hold} \\ H_t(a^*, b^*) & \text{otherwise} \end{cases}. \tag{3.18}$$

Therefore, the team will decide to tender the player a contract if $T > 0$, meaning if the team expects to get positive surplus value.

If the team tenders the player a contract, the player goes through the arbitration process, and, on expectation, gets a utility of

$$P_U = \begin{cases} (1-k_1)U_p(\bar{x}_1) + k_1 H_p(a^*, b^*) & \text{if Eq 3.16, 3.17 hold} \\ H_p(a^*, b^*) & \text{otherwise} \end{cases}. \tag{3.19}$$

and a salary of

$$\begin{aligned} P_S = (1-k_1)\bar{x}_1 + k_1((1-k_2)\bar{x}_2 + k_2(F[b^* I(X_m \geq b^*(1-d)) + \\ I(X_m < b^*(1-d))X_m] + (1-F)[a^* I(X_m \geq a^*(1-d)) + \\ I(X_m < a^*(1-d))X_m])) \end{aligned} \tag{3.20}$$

if Equations 3.16 and 3.17 hold, and

$$\begin{aligned} P_S = (1-k_2)\bar{x}_2 + k_2(F[b^* I(X_m \geq b^*(1-d)) + \\ I(X_m < b^*(1-d))X_m] + (1-F)[a^* I(X_m \geq a^*(1-d)) + \\ I(X_m < a^*(1-d))X_m]) \end{aligned} \tag{3.21}$$

otherwise. If the team non-tenders the player, the player signs as a free agent for X_m.

3.3 Simulation Approach

To analyze different features of the arbitration process, I use the model discussed in the previous section to simulate player salary outcomes. I run 100,000 simulations each for various model specifications to calculate the player's expected salary, the frequency cases are settled, and the frequency players are non-tendered. I can also calculate optimal exchange figures.

For each simulation, the range of potential player arbitration salaries is from 10 to 20. I choose smaller, simpler numbers to simplify the analysis. When estimating effects for different players in future work, the range can be adjusted. Outcomes are simulated given different input parameters. The following variables are assigned values as part of the simulation: A, X_m, r_p, r_t, d, k_1, k_2, and σ. A is the player's true arbitration value, and in the simulation can take an integer value between 11 and 19, inclusive. X_m is the player's free agent market value, and in the simulation is assigned a value of 15, 20, or 40. The two risk variables for the player and team, r_p and r_t, respectively, can take on values of 0.5, 0.8, or 1. A risk value of one indicates risk neutrality, and a risk value less than one indicates risk averse preference. For the simulations, it is assumed that the team will never be more risk averse than the player.[1] The release percentage fee, d, takes on a value of 0, 0.17, or 1 in the simulation. A release percentage fee of one is the equivalent to a fully guaranteed contract, while a release fee of zero is the equivalent to a fully non-guaranteed contract. A release percentage fee of 0.17 corresponds to a release fee of about one-sixth of the salary.[2]

[1] The player's arbitration salary is likely to constitute a large percentage of the player's total income compared to the percentage of the team's total payroll. For example, if the player's salary increases by \$100,000, that will certainly increase the player's income by a much larger percentage than the increase in the team's payroll. So in theory, players have more to lose in the arbitration process than teams do.

[2] Depending on when the player is released, the team could be responsible for termination pay. In the case of releasing the player after the arbitration hearing, but before mid-way through Spring Training, the release fee is 30-days salary, which is one-sixth of a full MLB season. For the purposes of this paper, only the 30-day timeline is relevant for any arbitration decisions. If a

The probability that negotiations break down in either of the two negotiation periods, represented by k_1 and k_2, and the uncertainty in the spread of the arbitration panel's valuation, represented by σ, are discussed in more details throughout this section. Note that the purpose of this simulation exercise is to identify circumstances where players may be worse off. Further research is necessary to identify which scenarios are closest to that of a representative arbitration case, and therefore, how much worse off by the rules of the game players actually are.

3.3.1 Functional Form For Arbitrators Decision

In addition to Assumption 3.2.2, I incorporate an additional assumption when assuming a functional form for F. The density, f, has positive support over the range $[L, C]$, where L is the bottom support and C is the upper support. The upper support can be interpreted as a player's ceiling, the most a player can expect to earn in arbitration, while the lower support, the player's floor, is the minimum a player can expect to earn. These supports can also be used to characterize the player's comparable market. The upper support represents the most favorable comparable to the player, while the lower support represents the least favorable. Alternatively to Assumption 3.3.1, one could choose a functional form that does not have defined boundaries. The normal distribution, for example, would fit that criteria. I choose to incorporate bounded supports to better mimic the comparable

player gets released after that, it is likely that the arbitration salary was not the driving factor for the release.

based market setup.

Therefore, I assume a variation of the truncated normal distribution for F. First, define the normal truncated probability density function as

$$g(M) = \frac{\phi(\frac{M-A}{\sigma})}{\sigma(\Phi(\frac{C-A}{\sigma}) - (\Phi(\frac{L-A}{\sigma}))}, \tag{3.22}$$

where M is the midpoint as calculated in Equation 3.3, ϕ is the probability density function for the standard normal distribution, Φ is its cumulative distribution function of the standard normal distribution, A is the player's true arbitration market value, L and C are where left and right truncation occurs, respectively, and σ is the standard deviation of the function. In this application, σ can be interpreted as the participants' joint uncertainty in the spread of the arbitrator panel's valuation. Certain arbitrators may be more predictable than others, which would result in a lower σ. In the simulation, σ can take on values of 0.5, 1, 2, or 5.

Next, define $f(M)$ as the density function for the arbitrator's valuation of the player's arbitration value, X_A, given some midpoint, M, as

$$f(M) = \begin{cases} \frac{0.5g(M)}{\int_L^A g(M)dM} & M < X_A, \\ \frac{0.5g(M)}{\int_A^C g(M)dM} & M > X_A, \\ \frac{0.25g(M)}{\int_L^A g(M)dM \int_A^C g(M)dM} & M = X_A. \end{cases} \tag{3.23}$$

If the midpoint is closer to the upper boundary, more weight is given to the points to the right of A in order for $F(A) = 0.5$. The probability that the arbitrator selects the team's submission is given by $F(M)$, and is calculated using

$$F(M) = \int_L^M f(M)dM. \tag{3.24}$$

The probability that the arbitration panel selects the player's submission is $1 - F(M)$.

3.3.2 Estimating Best Response Functions

The functional complexity of $F(M)$, h_t, and h_p, necessitates the use of other estimation procedures to find $a_{BR}(b)$ and $b_{BR}(a)$ and solve for a^* and b^* in the arbitration salary figure exchange stage of the game. I estimate best response curves for both the player and team, and then estimate where the curves intersect to determine the Nash equilibrium of the game.

First, I restrict salaries to the range of feasible salary outcomes, salaries between 10 and 20, inclusive. For the team, for every potential submission by the player between 10 and 20 in intervals of 0.25, I determine the team's optimal submission in response. Next, I estimate a line for the best response function using a sixth degree polynomial, where the best responses represent the dependent variable, and the player submissions are the explanatory variable. Then, I repeat this process for the player, estimating a sixth degree polynomial for the best response function for the player given various team salary submissions. Finally, I find where the two best response functions intersect.

For example, suppose $X_m = 40$, $A = 15$, r_t and $r_p = 1$, $d = 0$, $\sigma = 1$, and k_1 and $k_2 = 1$. Figure 3.2a shows each participant's best responses for salary submissions by the other participant. Next, Figure 3.2b shows estimates of the best response functions using sixth degree polynomials, with the intersection of the two curves illustrated. In this example, $a^* = 16.267$ and $b^* = 13.733$.

This best response function estimation procedure does better when the best responses look like they come from a smooth function. Suppose instead that $X_m = 15$ instead of 40 like in the previous example. As can be seen in Figure 3.3a, when the team's submission is 15 or less, the player will always respond with a salary submission of 15. When the team's submission is greater than 15, the player's best response is to submit a salary of 20. Unlike in the previous example, these best responses clearly do not come from a smooth function. Figure 3.3b shows the estimated best response curves, and it should be immediately apparent that the estimated best response curve for the player does a poor job estimating the true best response curve. In this example, $a^* = 14.887$ and $b^* = 13.326$. While the estimates are going to be more widely off in this situation, there are clear takeaways. With the lower market value, the player's submission is much lower in this example than in the previous example. The player knows that if he is awarded a contract that is too large, the team will release him. Thus, there is no additional expected benefits from raising his offer past the free agent market value.

3.3.3 Potential Negotiation Breakdown

For any set of model parameters, I run 100,000 simulations to see how arbitration and salary outcomes vary. Before running the simulations, I identify whether the model determines that a settlement would occur in the first bargaining period or the second bargaining period if that stage is reached, whether the player would be non-tendered, and whether both parties would ultimately prefer to proceed to a trial. Then, I simulate the model, incorporating k_1 and k_2, the probability of a breakdown in negotiations occurring in the first and second bargaining periods, respectively. While a settlement may be optimal, a breakdown in negotiations could occur, leading to a less-than-optimal outcome.

Let $(k_1, k_2) = B_{1,2}$ be a pair of breakdown probabilities incorporated into the model parameters. Among model simulation parameters, $B_{1,2}$ can take the value of $(1, 1)$, $(0.9, 0.9)$, $(0.05, 0.95)$, $(0, 0)$, $(0.05, 1)$, or $(0.5, 0.5)$. When $B_{1,2} = (1, 1)$, negotiations will not occur, and the team can either non-tender the player or expect to go to trial. When $B_{1,2} = (0.9, 0.9)$, negotiations will almost surely not result in a settlement, but a settlement is possible. Having $B_{1,2} = (0.05, 0.95)$ or $B_{1,2} = (0.05, 1)$ mirrors the file-and-trial strategy, where teams refuse to negotiate with players after the exchange deadline, thus putting more pressure on a settlement to occur during the first settlement period. When $B_{1,2} = (0.5, 0.5)$ whether settlement talks break down is essentially a coin flip in each period. Finally, when $B_{1,2} = (0, 0)$, a settlement will always occur, if that is the optimal outcome.

Suppose $B_{1,2} = (0.5, 0.5)$ and the model predicts for a settlement to occur during the first bargaining period, or during the second bargaining period if not during the first. During the simulation, negotiations will break down during the first bargaining period with probability 0.5. If negotiations do not break down, a settlement occurs and the simulation is complete. If negotiations do break down, a settlement does not occur and the game continues. When the game reaches the second bargaining period, again, with probability 0.5 negotiations break down. If negotiations do not break down, a settlement is reached and the game is over. If negotiations do break down, the game proceeds to the arbitration hearing.

3.4 Simulation Results

Simulations provide results for three different arbitration scenarios. The first set of results assume that the player's free agent market value exceeds his true arbitration market value. The second set of results allow for the possibility that the arbitration market value exceeds the free agent market value. The third set of results introduces the effects of a release fee. Each model is run using a censored regression with censoring occurring at 10 and 20 for the player salary models, and 0 and 1 for the percentage of simulations ending in settlement models.

3.4.1 Base Model Results

Table 3.1 provides various model specifications and variable interactions when $X_m \in \{20, 40\}$. The dependent variable in each (a) column is player salary, and the dependent variable in each (b) column is the percentage of simulations that end in settlements. The models assume that the release fee is zero. The table illustrates the results for differing probabilities of negotiation breakdown, differing levels of risk preferences, and differing spreads in the expected beliefs of the arbitration panel's valuation of the player.

Models 1a and 3a show clear evidence that a higher breakdown probability in negotiations is better for a player's expected salary. Models 1b, 2b, and 3b show the anticipated result that increasing the probability of negotiation breakdown decreases settlements. The biggest effects on player salary occur when breakdown probability is higher in the second bargaining period. The first bargaining period precedes the arbitration figure exchange date, while the second bargaining period precedes the actual arbitration hearing. There is no additional opportunity to settle after the second bargaining period, so if one of the parties more strongly prefers a settlement outcome, it is likely to cost that party more to secure that outcome in the first bargaining period than the second. Given that the player has more to lose than the team as a percentage of total income/payroll costs, it would make sense that increasing the likelihood of breakdown in settlement talks would increase expect salary. The player has less of an opportunity to sell off risk and uncertainty and take a lower salary by settling.

While none of the breakdown coefficients listed in model 2a are statistically significant, the interactions between breakdown probability and risk preferences illustrate an interesting picture. Table 3.2 shows the results for the interaction terms from model 2a in Table 3.1. As with the results from models 1a and 3a, the coefficients are all positive, and are greater as the probability of breakdown in the second period increases. However, magnitudes are greatest, and statistically significant, when the team is risk neutral and the player is risk averse. When both parties are somewhat risk neutral, probability of breakdown has no statistical impact on player salary. However, when the team is risk neutral and the player is risk averse, reducing the ability for parties to settle is better for player salary. This is consistent with the story discussed in the previous paragraph.

These results would suggest that the 'file-and-trial' strategy incorporated by some teams may actually be detrimental, rather than beneficial, to their payoffs. A team that incorporates the strategy will not negotiate with players after the first bargaining period, instead choosing to submit figures and go directly to a trial. The thinking is that the threat of ceasing negotiations would pressure players in the earlier bargaining period to come to an agreement. This strategy is the equivalent of having the probability of breakdown in the second period equal to one. As the previous results suggested, removing bargaining in the second period actually increases players expected salary by preventing them from selling off risk and taking lower settlements. Therefore, the 'file-and-trial' strategy may actually be inefficient for teams.

The overall risk aversion story is supported in models 1a and 3a as well. When the team is risk neutral and the player is risk averse, expected salary goes down. Models 1b and 3b show that any sort of risk aversion on the side of players will increase the likelihood of a settlement. Risk aversion is even relevant when it relates to the uncertainty of the arbitration panel's valuation. A large uncertainty consistently negatively affects the player. This is consistent with Faurot and

McAllister (1992) who showed in their model that the difference between optimal salary proposals is proportional to the standard deviation of the distribution of the arbitrator's notion of a fair settlement. The more uncertainty in the arbitration panel's valuation, the more there is for a player to lose from going to an arbitration hearing since the range of potential valuations is greater. But it also appears that there is an optimal level of valuation uncertainty for the player. For each of the uncertainty variables, the coefficients are negative in models 1a, 2a, and 3a, and positive and statistically significant for models 1b and 3b. In models 2a and 3a, the coefficient magnitude is smallest (in absolute value terms) for a valuation uncertainty of 2. While there is a major difference between the coefficients in model 3a for valuation uncertainties of 0.5 and 5, the gap disappears when considering the interaction terms. The only statistically significant interactions in model 3a between r, the set of combinations of risk preferences for the team and player, and σ are $(1, 0.5)$ risk preferences interacted with 0.5 uncertainty (-1.073 coefficient with 0.184 standard error) and $(1, 0.8)$ risk preferences interacted with 0.5 uncertainty (-1.150 coefficient with 0.183 standard error). When the player is risk averse and the team is risk neutral, the player is made worse off when breakdown probability is low, and uncertainty in arbitration panel valuation is sufficiently different than some optimal uncertainty value.

Table 3.3 breaks down the player's average arbitration salary from the simulations, given the conditions in the models from Table 3.1, by risk of breakdown in negotiations, risk preferences, and arbitrator valuation uncertainty. In each box, the top left value sets $\sigma = 0.5$, the top right value sets $\sigma = 1$, the bottom left value sets $\sigma = 2$, and the bottom right value sets $\sigma = 5$. Consistent with the previous results, the highest player salaries are the ones in which the probability of a breakdown in negotiations equals one in the second bargaining period. Salary offers are closest to 15 when both parties are risk neutral. Players earn their lowest salaries when the team is risk neutral, the player is very risk averse, and there is little or no likelihood of a breakdown in negotiations.

3.4.2 Arbitration And Free Agent Market Value

Table 3.4 compares models when free agent market value is 15, versus when it equals either 20 or 40. Since the range of true arbitration market values and domain of arbitration panel valuations is between 10 and 20, there are instances where either of these values will be greater or equal to a free agent market value of 15. That is not the case for free agent market values of 20 and 40. Analyzing models under both scenarios illustrates the effects of having a market value less than the arbitration value.

The first set of results in Table 3.4 examines the impact of a player's true arbitration valuation on his expected salary, compared to a true arbitration value at 15. The right columns show the obvious result that as arbitration value increases, so does the player's salary. The left columns tell a slightly different story. Below the free agent market value of 15, the coefficients are negative and decrease as the arbitration value gets smaller. But above 15, the coefficients are stagnant. There is no positive benefit for the player to have an arbitration market value worth 16 versus 17 versus 18. The effects are the same. This is because in every simulation, the team chooses to non-tender the player if his true arbitration value is greater than his market value. So in each case, the player is non-tendered, becomes a free agent, and earns a salary equal to his free agent market wage.

Whether players are collectively better or worse off by market mispricing depends on the prevalence of mispricing in the market. For instance, suppose there are two players, Player A with an arbitration market value of 12, and Player B with an arbitration market value of 18, both with free agent market values of 15. They have the same free agent market values, but mispricing in the arbitration market results in different arbitration market valuations. Player A with the 12 arbitration valuation is 1.943 units worse off, while Player B with the 18 arbitration valuation is 1.201 units better off. In this scenario, players are collectively made worse off and teams collectively made better off by 0.742 units. This provides one example of how arbitration market mispricing may lead to a redistribution from players to teams.

The remainder of Table 3.4 examines differences in the effects of breakdown probability, risk preferences, and panel valuation uncertainty between the two free agent market value groups. The effects of breakdown probability and risk tolerance are clearly weaker. Since teams will non-tender players with market values less than true arbitration values, the only players proceeding through the arbitration process are those who are priced below market value. Since a player will be released if their arbitration salary is greater than his market value, players have incentives to reduce their submission figures, which, in kind, will incentivize teams to increase their submission offers, reducing the amount being risked by going to an arbitration hearing. Settlements will be less attractive since players' risk preferences will have less of an impact on their decision making.

3.4.3 Release Fees

Table 3.5 introduces release percentage fee into the model, and interacts it with different variables. A release fee of one represents a fully guaranteed contract, and a release fee of zero represents a non-guaranteed contract. Panel A shows the estimates for release fee percentage on player salary. Both variables are positive and statistically significant, indicating that, in general, higher release fees are good for players. A 100% release fee, fully-guaranteed contract, is even better than a 17% release fee, partially-guaranteed contract. With release fees, it becomes costlier for teams to release players, giving players slightly more bargaining power.

Panel B looks how interactions between release fee and true arbitration value impact player salary. All of the coefficients are negative, but the arbitration valuations greater than 15 have much stronger, and statistically significant, negative effects. Panel C analyzes interactions between release fee and free agent market value. For a market value of 15, the coefficient is positive and statistically significant, while the coefficient for a market value of 20 is statistically zero. Combining results from Panels B and C suggest that the release fee is most beneficial to players who are not likely to be non-tendered, but have an opportunity to earn more than the market value.

For example, suppose the player's arbitration value is 14 and his market value is 15. According to the model, the team will tender the player a contract. If a settlement is not reached and an arbitration hearing occurs, the player could potentially earn more than the market value of 15. If the release fee is zero, the contract is non-guaranteed, the team will choose to release the player if his earnings are greater than his market value. If the contract is partially or fully-guaranteed, the player could earn more than the market value, and the team may decide to keep the player if the cost to release him is greater than the cost to keep him. This opportunity to earn more than his market value will give the player more leverage in negotiations.

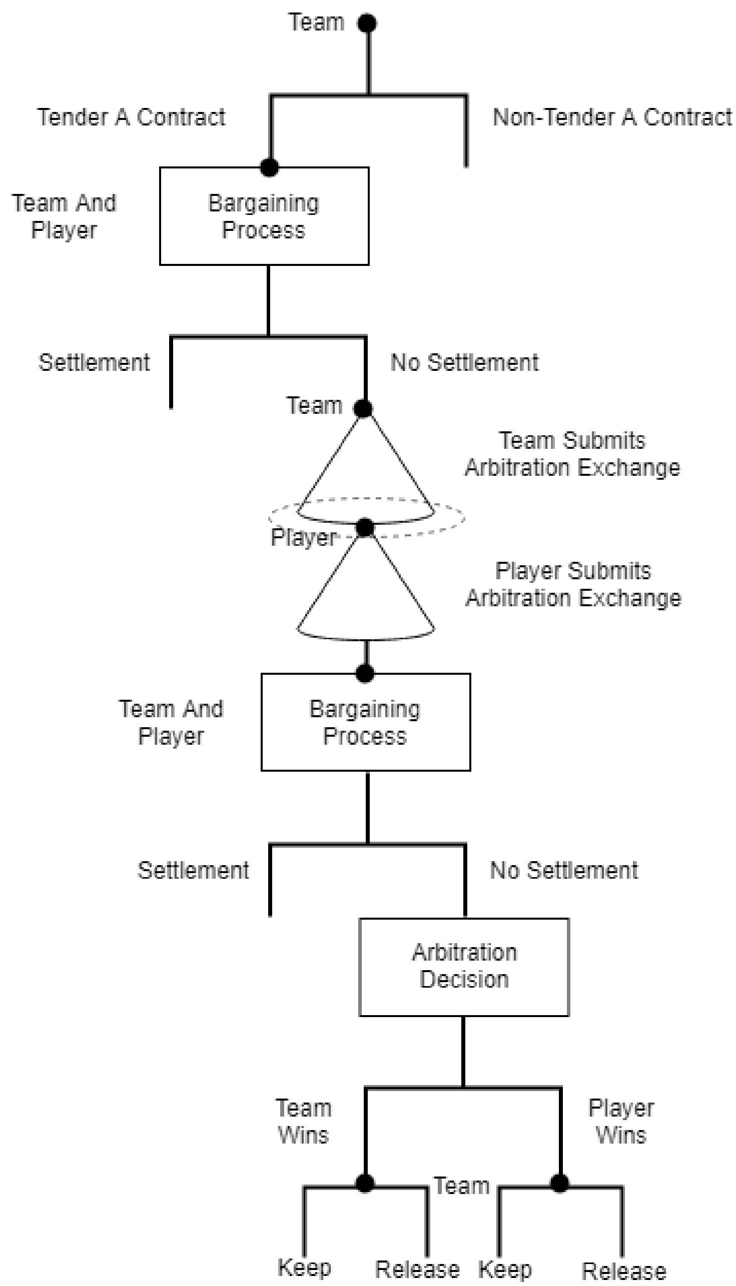
Panel D shows no interaction effects between risk aversion uncertainty and release fee percentage.

3.5 Conclusions

This paper supports many of the conclusions regarding risk aversion in previous work, while addressing aspects of the arbitration process that had not previously been addressed. Player salaries are unequivocally higher when players do not have the opportunity to negotiate a settlement. While individual players may prefer to trade away risk and accept lower salaries, the MLBPA has an incentive to maximize arbitration salaries. Once a player agrees to a lower settlement, that lower contract becomes a comparable case for future players, which serves to lower those future players' arbitration market values. Players collectively are better off when individuals behave neutrally in their risk preferences.

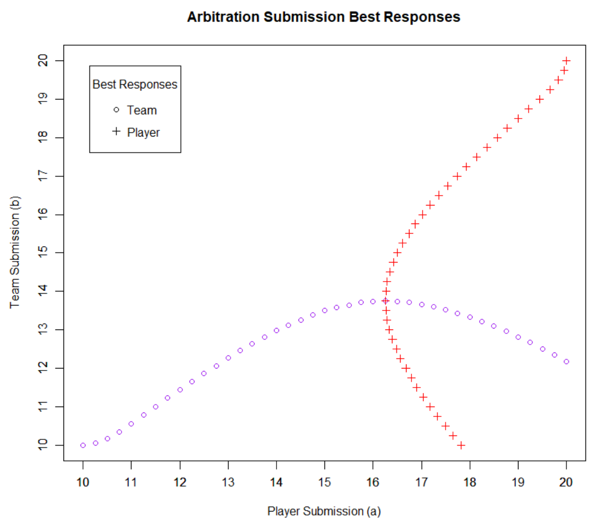
The model in this paper provides a theoretical framework for future work on baseball arbitration. One such topic is the impact of market mispricing on arbitration outcomes, numerically analyzing how market mispricing makes players collectively worse off. Future work may consider applying other bargaining solutions besides Nash, or may apply a more mathematical framework to finding the salary figure exchange team and player best response functions. Finally, future work may manipulate this model to accommodate other final-offer arbitration settings.

Figure 3.1: Arbitration model



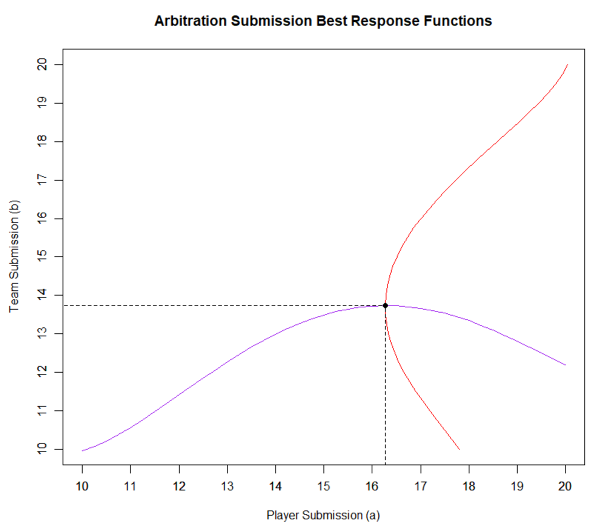
This is a visual representation of the arbitration game. Decision nodes, where either the team, player, or both, make a move in the game, are emphasized with black dots. Simultaneous moves is illustrated with a doted oval.

Figure 3.2a: Smooth best responses



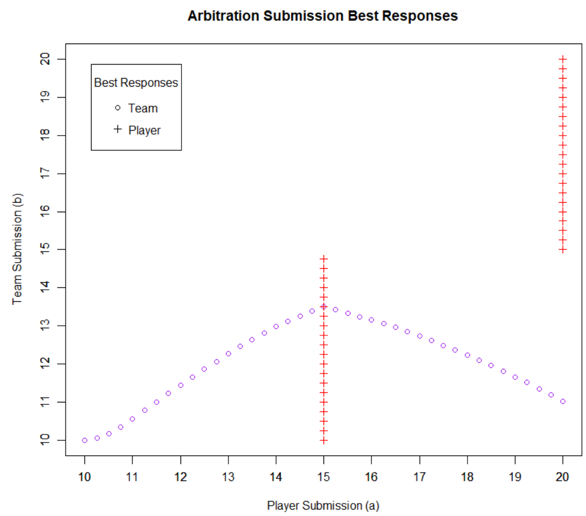
For a given salary submission by one participant, this shows the other participant's best response that maximizes expected utility.

Figure 3.2b: Smooth best response functions



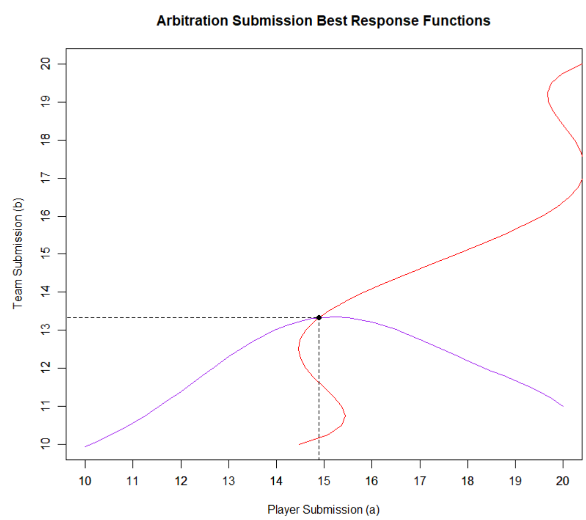
This uses the best responses in Figure 3.2a to estimate best response functions using sixth degree polynomials. The intersection of the two best response functions is the Nash equilibrium of the salary figure exchange subgame. In this example, the player's optimal submission is 16.267 and the team's optimal submission is 13.733.

Figure 3.3a: Best responses with market mispricing



For a given salary submission by one participant, this shows the other participant's best response that maximizes expected utility, except this time with $X_m = 15$.

Figure 3.3b: Best response functions with market mispricing



This uses the best responses in Figure 3.3a to estimate best response functions using sixth degree polynomials. In this example, the player's optimal submission is 14.887 and the team's optimal submission is 13.326. Note that the estimated best response functions are less accurate compared to those in Figure 3.2b.

Table 3.1: Base Model Results

Model #		(1a)	(1b)	(2a)	(2b)	(3a)	(3b)
Breakdown Probability ($B_{1,2}$) (1st Period, 2nd Period)	(0.05, 0)	0.043	-0.054^{**}	-0.0002	0.000	0.046	-0.054^{**}
		(0.080)	(0.022)	(0.179)	(0.037)	(0.077)	(0.022)
	(0.05, 0.5)	0.043	-0.292^{***}	-0.210	-0.106^{***}	0.045	-0.293^{***}
		(0.080)	(0.022)	(0.179)	(0.040)	(0.077)	(0.021)
	(0.05, 1)	0.421^{***}	-0.326^{***}	0.010	-1.097	0.422^{***}	-0.326^{***}
		(0.082)	(0.022)	(0.185)	(176.0)	(0.079)	(0.022)
	(0.5, 0.5)	0.216^{***}	-0.521^{***}	-0.001	-0.106^{***}	0.219^{***}	0.523^{***}
		(0.080)	(0.022)	(0.179)	(0.040)	(0.077)	(0.022)
	(0.5, 1)	0.423^{***}	-0.764^{***}	0.010	1.097	0.424^{***}	-0.764^{***}
		(0.082)	(0.023)	(0.185)	(176.0)	(0.079)	(0.022)
	(1, 1)	0.424^{***}	-2.440	0.010	-1.097	0.425^{***}	-2.369^{***}
		(0.082)	(97.13)	(0.185)	(176.0)	(0.079)	(57.85)
	(0, 0)	Excluded					
Risk Tolerance (r) (Team, Player)	(0.5, 0.5)	-0.085	1.84^{***}	-0.228	2.129	-0.215^{*}	1.293^{***}
		(0.069)	(0.021)	(0.178)	(162.1)	(0.129)	(0.042)
	(0.8, 0.5)	-0.103	1.195^{***}	-0.210	2.129	0.031	1.311^{***}
		(0.069)	(0.021)	(0.176)	(160.0)	(0.129)	(0.042)
	(1, 0.5)	-0.467^{***}	1.182^{***}	-0.923^{***}	2.129	-1.355^{***}	1.313^{***}
		(0.070)	(0.021)	(0.178)	(162.4)	(0.130)	(0.042)
	(1, 0.8)	-0.340^{***}	1.142^{***}	-0.691^{***}	1.186^{***}	-1.186^{***}	1.265^{***}
		(0.070)	(0.021)	(0.180)	(0.044)	(0.130)	(0.041)
	(1, 1)	Excluded					
Panel Valuation Uncertainty (σ)	0.5	-0.085	1.84^{***}	-0.228	2.129	-0.215^{*}	1.293^{***}
		(0.069)	(0.021)	(0.178)	(162.1)	(0.129)	(0.042)
	2	-0.103	1.195^{***}	-0.210	2.129	0.031	1.311^{***}
		(0.069)	(0.021)	(0.176)	(160.0)	(0.129)	(0.042)
	5	-0.467^{***}	1.182^{***}	-0.923^{***}	2.129	-1.355^{***}	1.313^{***}
		(0.070)	(0.021)	(0.178)	(162.4)	(0.130)	(0.042)
	1	Excluded					
Controls		A, X_m		A, X_m		A, X_m	
Interactions		None		$B_{1,2} * r$		$r * \sigma$,$A * X_m$	
Obs		2,328	2,328	2,328	2,328	2,328	2,328

*p<0.1; **p<0.05; ***p<0.01

Notes: The dependent variable in each column (a) is the player's salary. The dependent variable in each column (b) is the percentage of simulations that result in a settlement. In these models, $X_m \in \{20, 40\}$ and $d = 0$.

Table 3.2: $X_m \in \{20, 40\}$ Breakdown Risk Interactions

		r			
		(0.5, 0.5)	(0.8, 0.5)	(1, 0.5)	(1, 0.8)
$B_{1,2}$	(0.05, 0)	0.024	0.015	0.097	0.086
		(0.251)	(0.249)	(0.252)	(0.253)
	(0.05, 0.5)	0.021	0.015	0.099	0.087
		(0.251)	(0.249)	(0.252)	(0.253)
	(0.05, 1)	0.305	0.217	0.883***	0.665**
		(0.258)	(0.257)	(0.261)	(0.260)
	(0.5, 0.5)	0.149	0.107	0.470*	0.364
		(0.251)	(0.249)	(0.252)	(0.253)
	(0.5, 1)	0.288	0.220	0.900***	0.672***
		(0.258)	(0.257)	(0.261)	(0.261)
	(1, 1)	0.270	0.221	0.917***	0.678***
		(0.258)	(0.257)	(0.261)	(0.261)

Notes: This table provides the interaction effects between negotiation breakdown probability $(B_{1,2})$ and risk preferences (r) in model 2a in Table 3.1. For r, the team risk level is listed first, followed by the player risk level. For $B_{1,2}$, the probability of breakdown in the first negotiating period is listed first, followed by the probability of breakdown in the second negotiating period.

Table 3.3: Salary Breakdown

		r									
		0.5, 0.5		0.8, 0.5		1, 0.5		1, 0.8		1, 1	
$B_{1,2}$	0, 0	14.501	14.518	14.994	14.497	12.313	14.487	12.650	14.994	14.963	14.999
		15.101	14.998	15.002	14.683	14.957	14.748	14.983	14.561	14.998	14.998
	0.05, 0	14.526	14.543	14.994	14.522	12.445	14.512	12.765	14.993	14.963	14.998
		15.101	15.017	15.002	14.700	14.957	14.748	14.983	14.561	15.002	14.993
	0.05, 0.5	14.525	14.543	14.994	14.522	12.445	14.513	12.764	14.994	14.963	15.000
		15.099	15.007	15.002	14.700	14.958	14.751	14.983	14.562	14.998	14.995
	0.05, 1	15.213	15.014	15.210	14.999	15.015	14.988	15.018	14.760	15.020	14.763
		15.540	15.101	15.220	14.790	15.426	14.721	15.452	14.791	15.471	14.998
	0.5, 0.5	14.751	14.764	14.994	14.746	13.626	14.740	13.799	14.995	14.962	14.998
		15.074	15.093	15.002	14.842	14.967	14.781	14.987	14.572	14.998	14.996
	0.5, 1	15.212	15.010	15.210	15.000	15.016	14.993	15.018	14.763	15.020	14.763
		15.519	15.055	15.219	14.802	15.444	14.773	15.460	14.813	15.470	15.001
	1, 1	15.210	15.005	15.209	15.000	15.016	15.000	15.019	14.763	15.020	14.764
		15.497	15.006	15.221	14.811	15.462	14.827	15.470	14.829	15.471	15.000

Notes: This table compares average salaries using the assumptions from the models in Table 3.1, by probability of negotiations breaking down $(B_{1,2})$, risk preferences (r), and arbitration panel valuation uncertainty (σ). In each box, the upper left value assumes $\sigma = 0.5$, the upper right value assumes $\sigma = 1$, the lower left value assumes $\sigma = 2$, and the lower right value assumes $\sigma = 5$.

Table 3.4: Varying Market Value Model

Category	Variable	$X_m = 15$ Estimate	$X_m = 15$ Std. Error	$X_m \in \{20, 40\}$ Estimate	$X_m \in \{20, 40\}$ Std. Error
Arbitration Value (A)	11	-2.657^{***}	0.079	-3.602^{***}	0.092
	12	-1.943^{***}	0.078	-2.845^{***}	0.089
	13	-0.919^{***}	0.078	-1.831^{***}	0.094
	14	-0.032	0.078	-0.870^{***}	0.090
	16	1.201^{***}	0.078	0.644^{***}	0.092
	17	1.201^{***}	0.078	1.761^{***}	0.094
	18	1.201^{***}	0.078	2.800^{***}	0.089
	19	1.201^{***}	0.078	3.360^{***}	0.090
Breakdown Probability $(B_{1,2})$	(0.5, 0)	0.016	0.069	0.043	0.080
	(0.05, 0.5)	0.016	0.069	0.043	0.080
	(0.05, 1)	0.181^{***}	0.069	0.421^{***}	0.082
	(0.5, 0.5)	0.093	0.069	0.216^{***}	0.080
	(0.5, 1)	0.184^{***}	0.069	0.423^{***}	0.082
	(1, 1)	0.186^{***}	0.069	0.424^{***}	0.082
Risk Tolerance (r)	(0.5, 0.5)			-0.085	0.069
	(0.8, 0.5)			-0.103	0.069
	(1, 0.5)	-0.111^{**}	0.045	-0.467^{***}	0.070
	(1, 0.8)	-0.153^{***}	0.045	-0.340^{***}	0.070
Panel Uncertainty (σ)	0.5	-0.117^{**}	0.045	-0.416^{***}	0.060
	2	-0.211^{***}	0.052	0.180^{***}	0.060
	5	-0.111^{**}	0.052	0.118^{*}	0.064
Observations		753		2,328	

*p<0.1; **p<0.05; ***p<0.01

Notes: For r, the team risk level is listed first, followed by the player risk level. For $B_{1,2}$, the probability of breakdown in the first negotiating period is listed first, followed by the probability of breakdown in the second negotiating period. For $X_m = 15$, the team is assumed to be risk neutral. For both models, it is assumed that $d = 0$.

Table 3.5: Release Fee Interactions

Release Fee %	Estimate	Std. Error
0.17	0.174*	0.099
1.00	0.260***	0.099
0.00	Excluded	

Panel A: Release Fee on Player Salary

Release Fee	$d = 0.17$		$d = 1.00$	
	Estimate	Std. Error	Estimate	Std. Error
$A = 11$	−0.214*	0.119	−0.307***	0.119
$A = 12$	−0.177	0.116	−0.291**	0.116
$A = 13$	−0.189	0.121	−0.268**	0.121
$A = 14$	−0.178	0.118	−0.247**	0.118
$A = 15$	Excluded			
$A = 16$	−0.256**	0.120	−0.333***	0.120
$A = 17$	−0.245**	0.120	−0.338***	0.120
$A = 18$	−0.244**	0.116	−0.327***	0.116
$A = 19$	−0.247**	0.117	−0.314***	0.117

Panel B: Release Fee on Player Salary by Arbitration Value

Release Fee	$d = 0.17$		$d = 1.00$	
	Estimate	Std. Error	Estimate	Std. Error
$X_m = 15$	0.159**	0.069	0.185***	0.069
$X_m = 20$	−0.000	0.070	−0.000	0.070
$X_m = 40$	Excluded			

Panel C: Release Fee on Player Salary by Free Agent Market Value

Release Fee	$d = 0.17$		$d = 1.00$	
	Estimate	Std. Error	Estimate	Std. Error
$r_p = 0.5$	0.036	0.069	0.014	0.069
$r_p = 0.8$	0.019	0.069	0.005	0.069
$r_p = 1.0$	Excluded			

Panel D: Release Fee on Player Salary by Player Risk Preference

Notes: These panels analyze the effect of release fees on player arbitration salaries. Panel A examines the release fees themselves, while Panels B-D examine interaction effects between release fees and arbitration value, free agent market value, and risk preferences, respectively. This model assumes that the team is risk neutral.

REFERENCES

.

Ashenfelter, Orley (1987). “Arbitrator behavior”. In: *The American Economic Review* 77.2, pp. 342–346.

Barrett, Garry F and Stephen G Donald (2003). “Consistent tests for stochastic dominance”. In: *Econometrica* 71.1, pp. 71–104.

Billingsley, Patrick (1968). *Convergence of probability measures*. John Wiley & Sons.

Brown, Daniel T, Charles R Link, and Seth L Rubin (2017). “Moneyball After 10 Years: How Have Major League Baseball Salaries Adjusted?” In: *Journal of Sports Economics* 18.8, pp. 771–786.

Cabot, Anthony N and Keith C Miller (2011). *The Law of Gambling and Regulated Gaming: Cases and Materials*. Vol. 201. Carolina Academic Press, pp. 401–412.

Charnes, Abraham, William W Cooper, and E Rhodes (1978). “European J. of Oper”. In: *Res* 2, pp. 429–444.

MLB CBA. *Collective Agreement between Major League Baseball and the MLB Players Association*.

Deprins, D, L Simar, and H Tulkens (1984). *Measuring labor efficiency in post offices, The Performance of Public Enterprises: Concepts and Measurements, M. Marchand, P. Pestieau and H. Tulkens*.

Dworkin, James B (1977). “Final position arbitration and intertemporal compromise”. In: *Relations industrielles/Industrial Relations* 32.2, pp. 250–261.

Easton, Todd and Sarah Newell (2019). “Are daily fantasy sports gambling?” In: *Journal of Sports Analytics* 5.1, pp. 35–43.

Ehrman, Nathaniel J (2015). “Out of bounds: A legal analysis of pay-to-play daily fantasy sports”. In: *Sports Law. J.* 22, pp. 79–114.

Farrell, Michael James (1957). “The measurement of productive efficiency”. In: *Journal of the Royal Statistical Society: Series A (General)* 120.3, pp. 253–281.

Faurot, David J and Stephen McAllister (1992). “Salary arbitration and pre-arbitration negotiation in major league baseball”. In: *ILR Review* 45.4, pp. 697–710.

Gouker, Dustin (2017). *New Daily Fantasy Sports Industry Analysis: Uncertainty Still Reigns*.

Gustafson, Elizabeth and Lawrence Hadley (1995). “Arbitration and salary gaps in Major League Baseball”. In: *Quarterly Journal of Business and Economics*, pp. 32–46.
Hadley, Lawrence and John Ruggiero (2006). “Final-offer arbitration in major league baseball: A nonparametric analysis”. In: *Annals of Operations Research* 145.1, pp. 201–209.
Hakes, Jahn K and Raymond D Sauer (2007). “The moneyball anomaly and payroll efficiency: A further investigation”. In: *International Journal of Sport Finance* 2.4, p. 177.
Hanany, Eran, D Marc Kilgour, and Yigal Gerchak (2007). “Final-offer arbitration and risk aversion in bargaining”. In: *Management Science* 53.11, pp. 1785–1792.
Haruvy, Ernan, Dale O Stahl, and Paul W Wilson (2001). “Modeling and testing for heterogeneity in observed strategic behavior”. In: *Review of Economics and Statistics* 83.1, pp. 146–157.
Kneip, Alois, Byeong U Park, and Léopold Simar (1998). “A note on the convergence of nonparametric DEA estimators for production efficiency scores”. In: *Econometric theory* 14.6, pp. 783–793.
Kneip, Alois, Léopold Simar, and Paul W Wilson (2008). “Asymptotics and consistent bootstraps for DEA estimators in nonparametric frontier models”. In: *Econometric Theory* 24.6, pp. 1663–1697.
– (2015). “When bias kills the variance: Central limit theorems for DEA and FDH efficiency scores”. In: *Econometric Theory* 31.2, pp. 394–422.
– (2016). “Testing hypotheses in nonparametric models of production”. In: *Journal of Business & Economic Statistics* 34.3, pp. 435–456.
Korostelëv, Aleksandr Petrovič, Léopold Simar, and Alexandre B Tsybakov (1995). “Efficient estimation of monotone boundaries”. In: *The Annals of Statistics* 23.2, pp. 476–489.
Lindbergh, Ben (2018). *Baseball’s Economics Aren’t As Skewed As They Seem*.
Lock, Ethan and Allan DeSerpa (1986). “Salary Increases Under Major League Baseball’s System of Final Offer Salary Arbitration”. In: *Lab. Law.* 2, p. 801.
McFadden, Daniel (1989). “Testing for stochastic dominance”. In: *Studies in the Economics of Uncertainty*. Springer, pp. 113–134.
Meehan, Jeffrey C (2015). “The Predominate Goliath: Why Pay-to-Play Daily Fantasy Sports are Games of Skill Under the Dominant Factor Test”. In: *Marq. Sports L. Rev.* 26, pp. 5–36.
MLB Glossary: What is a Super Two?

Nash, John F (1950). "The bargaining problem". In: *Econometrica: Journal of the Econometric Society*, pp. 155–162.

Park, Byeong U, Léopold Simar, and Ch Weiner (2000). "The FDH estimator for productivity efficiency scores: asymptotic properties". In: *Econometric Theory* 16.6, pp. 855–877.

Rodenberg, Ryan (2017). *Daily fantasy sports state-by-state tracker*.

Sabermetrics Library: Replacement Level.

Shephard, Ronald W (1970). "Theory of cost and production functions Princeton University press". In: *Princeton, New Jersey*.

Silverman, Bernard W (1981). "Using kernel density estimates to investigate multimodality". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 43.1, pp. 97–99.

Simar, Leopold and Paul W Wilson (2007). "Estimation and inference in two-stage, semi-parametric models of production processes". In: *Journal of econometrics* 136.1, pp. 31–64.

Trippiedi, Michael (2014). "Daily fantasy sports leagues: Do you have the skill to win at these games of chance". In: *UNLV Gaming LJ* 5, p. 201.

Vardhman, R (2017). *The Unstoppable Growth of Fantasy Sports [Infographic] - Facts Stats*.

Which Daily Fantasy Sports Sites Allow Which States?

Wilson, Paul W (2008). "FEAR: A software package for frontier efficiency analysis with R". In: *Socio-economic planning sciences* 42.4, pp. 247–254.

– (2018). "Dimension reduction in nonparametric models of production". In: *European Journal of Operational Research* 267.1, pp. 349–367.

Wittman, Donald (1986). "Final-offer arbitration". In: *Management Science* 32.12, pp. 1551–1561.

Appendices

APPENDIX A

CHAPTER 1

A.1 Covariate Equations

The following equations were used to calculate the covariates, where t is the week of the season:

- QB Overall Grade: $\sum_{t=1}^{17} QBOverall_t \frac{TotalSnaps_t}{TotalSnaps}$, where $TotalSnaps_t$ are snaps in that game and $TotalSnaps$ are snaps over the full season
- QB Passing Grade: $\sum_{t=1}^{17} QBPassing_t \frac{PassingSnaps_t}{PassingSnaps}$, where $PassingSnaps_t$ are passing snaps in that game and $PassingSnaps$ are passing snaps over the full season
- Run Blocking: $\frac{1}{n} \sum_{i=1}^{n} \sum_{t=1}^{17} RunBlocking_{n,t} \frac{RunningSnaps_{n,t}}{RunningSnaps_n}$, where n is the number of players along the offensive line for that game (usually five)
- Pass Blocking: $\frac{1}{n} \sum_{i=1}^{n} \sum_{t=1}^{17} PassBlocking_{n,t} \frac{PassingSnaps_{n,t}}{PassingSnaps_n}$, where n is the number of players along the offensive line for that game (usually five)
- Defense Grade: $\sum_{i=1}^{n} \frac{Snaps_n}{Snaps} \sum_{t=1}^{17} OverallDefenseGrade_{n,t} \frac{Snaps_{n,t}}{Snaps_n}$, where n is the total numbers of starters on defense, $Snaps$ is the total number of snaps taken by the defense, and $Snaps_n$ is the total number of snaps by one particular player

- Run Defense: $\sum_{i=1}^{n} \frac{RunSnaps_n}{RunSnaps} \sum_{t=1}^{17} RunDefenseGrade_{n,t} \frac{RunSnaps_{n,t}}{RunSnaps_n}$

- Pass Rush: $\sum_{i=1}^{n} \frac{PassRushSnaps_n}{PassRushSnaps} \sum_{t=1}^{17} PassRushGrade_{n,t} \frac{PassRushSnaps_{n,t}}{PassRushSnaps_n}$

- Coverage Defense: $\sum_{i=1}^{n} \frac{CoverageSnaps_n}{CoverageSnaps} \sum_{t=1}^{17} CoverageGrade_{n,t} \frac{CoverageSnaps_{n,t}}{CoverageSnaps_n}$

The covariates are calculated for a given player or positional unit for a particular season.

A.2 Diamonds In The Rough

Results from Table 1.5 and Tables 1.7 and 1.8 support that the following strategies can be incorporated when selecting minimum-salary players in DraftKings contests:

- Quarterbacks:
    1. Select younger QBs.
    2. Select QBs facing better quality run defenses.
    3. Select QBs playing in very windy games.
- Running Backs:
    1. RBs listed as probable are really good bets to outperform healthy RBs.
- Wide Receivers:
    1. Select older WRs.
    2. Select WRs with lower quality QBs.
    3. Select WRs with lower quality pass blocking offensive linemen.
    4. Select WRs facing lower quality pass rushes and coverage defenses.
    5. Select WRs playing in less windy games.
- Tightends:
    1. Select TEs playing higher quality run defenses.

2. Select TEs playing in bad weather (snow/heavy rain).
3. Questionable TEs could prove valuable if they play.

A.3 Usage Regression Results

OLS Results For Usage Regressions By Position

	Dependent variable: Difference In Usage Percentage						
	(QB F)	(QB NM)	(RB F)	(RB NM)	(WR F)	(WR NM)	(TE F)
League Tenure	−0.020	0.007	−0.003	−0.013	0.004	0.037	−0.000
	(0.035)	(0.072)	(0.006)	(0.047)	(0.003)	(0.027)	(0.001)
QB Grade			−0.001	−0.009	−0.0002	−0.001	−0.000
			(0.002)	(0.020)	(0.001)	(0.010)	(0.000)
Run Blocking	0.015	−0.086	0.005	0.027	−0.001	−0.009	−0.002
	(0.069)	(0.155)	(0.006)	(0.059)	(0.003)	(0.032)	(0.001)
Pass Blocking	−0.052	−0.109	0.005*	0.061	0.001	0.009	0.001
	(0.043)	(0.093)	(0.003)	(0.038)	(0.002)	(0.020)	(0.001)
Run Defense	0.189*	0.251	−0.007	−0.050	0.002	0.049	0.001
	(0.114)	(0.242)	(0.009)	(0.088)	(0.005)	(0.049)	(0.002)
Pass Rush	−0.141*	−0.146	0.003	0.052	0.002	0.009	−0.003
	(0.082)	(0.159)	(0.006)	(0.062)	(0.004)	(0.035)	(0.002)
Coverage Defense	0.158	0.338	−0.0003	−0.044	−0.0004	0.015	0.0002
	(0.108)	(0.222)	(0.007)	(0.080)	(0.005)	(0.038)	(0.002)
Wind Speed	−0.006	0.076	−0.003	−0.018	−0.001	−0.012	0.001
	(0.049)	(0.112)	(0.004)	(0.042)	(0.002)	(0.022)	(0.001)
Bad Weather	−1.538	−1.327	0.172*	1.209	−0.048	−0.554	0.030
	(1.809)	(3.297)	(0.100)	(0.931)	(0.054)	(0.524)	(0.027)
Healthy vs Ques	1.643	2.164	−0.013	0.260	−0.108*	−0.325	−0.003
	(1.809)	(4.097)	(0.140)	(0.801)	(0.064)	(0.344)	(0.021)
Healthy vs Prob	0.427	0.492	0.154**	0.566	−0.049	−0.278	−0.022
	(0.533)	(0.978)	(0.074)	(0.510)	(0.032)	(0.227)	(0.014)
Constant	0.372	0.123	0.125***	1.000***	0.085***	0.246**	0.012**
	(0.243)	(0.507)	(0.019)	(0.197)	(0.011)	(0.103)	(0.005)
Observations	235	124	7,169	869	15,604	1,790	18,614
R^2	0.035	0.066	0.001	0.011	0.0004	0.004	0.001

*p<0.1; **p<0.05; ***p<0.01

A.4 Data Sources

- DraftKing Points and Salary Data: http://rotoguru1.com/cgi-bin/fyday.pl?game=dk
- Experience and Lineup Data: https://www.pro-football-reference.com/
- Injury Data: https://www.footballdb.com/transactions/injuries.html?yr=2018
- NFL Player Grades (subscription required): https://www.profootballfocus.com/
- NFL Weather Data: http://www.nflweather.com/en/

APPENDIX B

CHAPTER 2

B.1 Composite WAR Derivation

Composite WAR is calculated using weights created by a simple OLS regression

$$\begin{aligned} salary_{i,j} = \beta_{0,j} + \beta_{1,j} * pWAR_i + \beta_{2,j} * cWAR_i + \beta_{3,j} \\ * pos_i + \beta_{4,j} * SAE_i + \beta_{5,j} * pWAR_i * SAE_i \\ + \beta_{6,j} * cWAR_i * SAE_i, \end{aligned} \tag{B.1}$$

where i is a given player's arbitration case, j includes the different types of salary measures (real and nominal, base and raise), $pWAR$ and $cWAR$ are platform and career WAR values, pos is a player's position, and SAE is the player's year of arbitration service. After getting the coefficients from Equation B.1, composite WAR is calculated using the relative magnitude of the coefficients of the platform and career coefficients

$$\begin{aligned} WAR_{i,j} = \frac{\beta_{1,j} + \beta_{5,j} * SAE_i}{\beta_{1,j} + \beta_{2,j} + (\beta_{5,j} + \beta_{6,j}) * SAE_i} * pWAR_i \\ + \frac{\beta_{2,j} + \beta_{6,j} * SAE_i}{\beta_{1,j} + \beta_{2,j} + (\beta_{5,j} + \beta_{6,j}) * SAE_i} * cWAR_i. \end{aligned} \tag{B.2}$$

B.2 Offensive Environmental Variable Calculations

The following equations make up the calculations for the offensive skill environmental variables used in the second-stage model:

- Eye = $\frac{UBB+HBP}{PA}$, where UBB is the total number of unintentional walks (intentional walks are not counted) a player accumulates in a season (or career), HBP is the number of hit by pitches, and PA is the total number of plate appearances.

- Bat = $\frac{H}{AB}$, where H is the total number of hits a player accumulates in a season (or career), and AB is the number of at-bats (which removes walks, HBPs, and sacrifice hits from plate appearances). This stat is equivalent to the well-known batting average number seen on a player's baseball card.

- Power = $\frac{TB}{H}$, where TB is the total number of bases a player accumulates in a season (or career) via hits (a single results in one base, a double in two bases, a triple in three bases, and a home run in four bases). This stat captures how many bases a player averages whenever he gets a hits. A player with more power is going to accumulate more bases on average.