

Ain't Played Nobody: Building an Optimal Schedule to Secure an NCAA Tournament Berth

by

Kevin Floyd

B.S., Utah State University, 2016

Project Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

in the
Department of Statistics and Actuarial Science
Faculty of Science

© Kevin Floyd 2019
SIMON FRASER UNIVERSITY
Summer 2019

Copyright in this work rests with the author. Please ensure that any reproduction or re-use is done in accordance with the relevant national copyright legislation.

Approval

Name: Kevin Floyd

Degree: Master of Science

Title: Ain't Played Nobody: Building an Optimal
Schedule to Secure an NCAA Tournament Berth

Examining Committee: **Chair:** Dr. Jinko Graham
Chair

Dr. Thomas M. Loughin
Senior Supervisor
Professor

Dr. Tim B. Swartz
Supervisor
Professor

Dr. Harsha Perera
Internal Examiner
Lecturer

Date Defended: August 12, 2019

Abstract

American men’s college basketball teams compete annually for the National Collegiate Athletic Association (NCAA) national championship, determined through a 68-team, single-elimination tournament known as “March Madness”. Tournament participants either qualify automatically, through their conferences’ year-end tournaments, or are chosen by a selection committee based on various measures of regular season success. When selecting teams, the committee reportedly values a team’s quality of, and performance against, opponents outside of their conference. Since teams have some freedom in selecting nonconference games, we seek to develop an approach to optimizing this choice. Using historical data, we find the committee’s most valued criteria for selecting tournament teams. Additionally, we use prior seasons’ success and projected returning players to forecast every team’s strength for the upcoming season. Using the selection criteria and these projections, we develop a tool to help teams build the optimal nonconference schedule to increase their NCAA tournament selection probability.

Keywords: College basketball, March Madness, Scheduling, Sports analytics, Machine learning

Acknowledgements

I would first like to thank my supervisor, Dr. Tom Loughin. You gave me the opportunity of a lifetime by taking me as your student. Without your unwavering support, prudent guidance, and ample patience, none of this would have been possible. I will always have fond memories of our weekly meetings that were as entertaining as they were informative.

Thank you to Marie Loughin for being everything I could ever ask for from a boss. You gave me patience when I needed it, praise when I earned it, and admonishment when I deserved it. I am so grateful to you for bringing me back to the Workshop every semester and letting me feel like I made a difference.

I am thankful for the department faculty, whether I took a class from you, TA'd your undergrad class, or just saw you in the hallways now and then. Your wisdom and warmth make SFU an enviable place to study and work. Thank you to Charlene Bradbury, Kelly Jay, and Sadika Jungic for supporting me and solving my problems from my very first day until the last.

Thank you to the sports analytics community at SFU for creating such an exciting environment for me to learn in. Thank you to Dani Chu for your friendship and motivation; being celebrated for my work by you is like being praised for my painting by Picasso. I am so grateful for all my friends, in and out of the department, who enabled me to pursue my passions, supported me through my ups and downs, and helped me feel a sense of acceptance and belonging in Canada.

I would like to thank my committee members, Dr. Tim Swartz and Dr. Harsha Perera. Your thoughtful commentary and patience was vital to this project. Thank you to the staff of Club Ilia and the Study for fueling me and keeping me sane through the writing process.

Thank you to Dave Donlevy, Charles "Rock" Stephenson, Dr. Ernie Rimer, and the Utah Utes' men's basketball program for turning my passion for college basketball into a viable field of study and for inspiring this project. Thank you to Denis Beausoleil and the SFU men's basketball team for giving me more exposure to college basketball analytics and for filling me with school spirit.

Finally, I want to thank my family for their love, inspiration, and advice. Even from afar, I knew you were always there to support me and to see the best in me. I wouldn't be where I am today without your influence. I'm so proud to be part of the #BrothersFloyd.

Contents

Approval	ii
Abstract	iii
Acknowledgements	iv
Table of Contents	v
List of Tables	vii
List of Figures	viii
1 Introduction	1
2 Background	4
2.1 Overview of D-I college basketball	4
2.1.1 Conferences	4
2.1.2 Tournament Selection	5
2.1.3 Scheduling Games	8
2.2 Past work in predicting NCAA tournament selection	10
2.3 Advanced metrics and preseason predictions	12
2.3.1 Advanced metrics in college basketball	12
2.3.2 Predicting team strength in future seasons	13
2.4 Other relevant college basketball research	14
3 Data	16
3.1 Information needed	16
3.1.1 Phase 1	16
3.1.2 Phase 2	16
3.1.3 Phase 3	18
3.2 Data sources and acquisition	19
3.3 Data cleaning and augmentation	21
4 Methods	23

4.1	Phase 1	23
4.2	Phase 2	25
4.3	Phase 3	26
5	Results	29
5.1	Phase 1	29
5.2	Phase 2	32
5.3	Phase 3	33
5.3.1	2018 Bubble	33
5.3.2	2010 Virginia Tech	38
5.3.3	2020 Bubble	41
6	Discussion	46
7	Future Work	49
	Bibliography	51

List of Tables

Table 2.1	NCAA Quadrant Tiering System	8
Table 3.1	Explanatory Variables Used in Model 1	20
Table 3.2	Explanatory Variables Used in Model 2	20
Table 4.1	At-Large Classification Modeling Techniques	25
Table 4.2	Future Team Strength Prediction Modeling Techniques	26
Table 5.1	BMA Selected Variables	31
Table 5.2	Correct Classifications, Positional Errors of 2018 and 2019 At-Large Predictions	31
Table 5.3	Top 20 Teams in Predicted AdjEM for 2019-20	33
Table 5.4	Top 20 Teams in Predicted AdjEM for 2018-19	34
Table 5.5	2017-18 Utah Nonconference Schedule	35
Table 5.6	Optimal Games to Add to 2017-18 Utah Nonconference Schedule	36
Table 5.7	2017-18 Middle Tennessee Nonconference Schedule	37
Table 5.8	Optimal Games to Add to 2017-18 Middle Tennessee Nonconference Schedule	38
Table 5.9	Sub-285 Opponents on 2010 Virginia Tech Schedule	39
Table 5.10	Most, Least Optimal Games for 2019-20 Georgia, UConn, NMSU	42

List of Figures

Figure 5.1	Boxplots of Scaled Score-Based Misclassification and Positional Error Rates, Test Set	30
Figure 5.2	Model 2 Predictive Error	32
Figure 5.3	Distributions of Score and Position Changes for 2017-18 Utah by Game Location	37
Figure 5.4	2009-10 Virginia Tech At-Large Positional Changes by Replaced Opponent, Replacement's Tier	40
Figure 5.5	Predicted 2020 NCAA Tournament Bracket	41
Figure 5.6	Estimated Added Score by Predicted AdjEM	44
Figure 5.7	Change in Score and Ranking by Opponent's Conference Level	44
Figure 5.8	Estimated Added Position by Opponent's Predicted RPI Ranking	45

Chapter 1

Introduction

Each spring, the National Collegiate Athletic Association (NCAA), the primary governing body of American intercollegiate athletics, hosts postseason championship tournaments for its member institutions' men's basketball teams. Of each of the NCAA's three divisions (I, II, III; determined by an institution's willingness to give full, partial, or no scholarships to its student-athletes), Division I (D-I) has by far the most nationally popular tournament. The D-I champion is determined through a 68-team, single-elimination tournament, commonly known as "March Madness", that includes teams from all 32 D-I conferences (alliances of several similar universities). Even the revealing of the tournament field at the conclusion of the regular season, on what is known as "Selection Sunday", is a major television event.¹ Merely being selected for this tournament can provide a large financial windfall for an institution and its conference, both in financial payout from the NCAA and in national publicity for its school.² As such, tournament selection can initiate and sustain a university's athletic and even academic growth, so it is important for a team to take any available actions to improve its likelihood of being selected.

Teams can be admitted to the NCAA tournament in one of two ways.³ First, the more than 300 teams in D-I are aligned into conferences. Each of the conferences is allotted one automatic tournament berth, which is given to the winner of that conference's own single-elimination postseason tournament. For teams that do not win their conference tournament, a tournament selection committee comprised of university athletic directors and conference commissioners selects the "best" remaining teams to fill the at-large tournament spots. These teams are chosen based on a number of performance metrics that measure various aspects of a team's success during the regular season and their respective conference

¹<https://www.sportsmediawatch.com/2019/03/ncaa-tournament-selection-show-ratings/>

²<https://www.usatoday.com/story/sports/ncaab/2019/03/26/whats-an-ncaa-tournament-unit-worth-millions-heres-how/39256149/>

³<https://www.ncaa.com/news/basketball-men/article/2018-10-19/how-field-68-teams-picked-march-madness>

tournaments, and also the personal judgment of the committee's members. This selection process is subjective, and every season there is a lively debate among media, fans, and even athletic departments over the relative merits of teams that make or miss the tournament.⁴ Because of this subjectivity, it can be difficult for teams to know exactly what they need to do to earn a tournament berth. It would be beneficial for college basketball teams, as well as media and fans, to better understand the committee's selection process.

The process is complicated because of the immense size of D-I. Teams play only a very small fraction of the other D-I teams. Thus, a team's win-loss record measures a combination of the team's quality and the quality of its opponents, and therefore is insufficient to determine how good a team has shown itself to be. The quality of a team's opposition is important to the selection committee.⁵ For example, the ratings percentage index (RPI), a sorting tool traditionally relied on by the selection committee, combines a team's win-loss percentage with that of its opponents and its opponents' opponents. Considering that the quality of opponents that a given team plays can have a tremendous effect on its performance metrics, a team with tournament aspirations should attempt to play opponents that help improve those metrics.

Teams have limited control over the selection of their opponents. A team's regular season consists of three parts: nonconference, conference, and conference tournament. A team's conference games are scheduled before the season by the conference. Conference tournament games are determined by all conference teams' records in conference play. However, in the months before the next season starts, individual teams are free to negotiate with teams from other conferences to determine who and where to play most of their nonconference games. This can have an impact on eventual tournament selection; committee chair Bernard Muir specifically referenced a lack of high quality nonconference wins when explaining why Texas Christian was left out of the 2019 NCAA tournament.⁶ Thus, at-large eligible teams that schedule the right opponents can gain an advantage on Selection Sunday.

The goal of this project is to develop a statistical tool that can estimate the impact that any potential nonconference opponent will have on a given team's chances of being selected for that season's NCAA tournament. In order to achieve this goal, this project has three primary parts. First, we model the at-large selection process of the NCAA tournament selection committee. We want to find which of the several metrics thought to be used by the committee are measurably relevant in the final selection process and to what degree these metrics impact eventual selection, particularly for the metrics based on a team's opponents.

⁴<https://www.usatoday.com/story/sports/ncaab/2018/03/11/middle-tennessee-state-usc-among-biggest-ncaa-tournament-snubs-selection-sunday/414183002/>

⁵<https://www.ncaa.com/news/basketball-men/bracketiq/2018-02-15/ncaa-selection-committee-adjusts-team-sheets-emphasizing>

⁶<https://www.star-telegram.com/sports/college/big-12/texas-christian-university/article228068039.html>

Second, we develop a model to predict the value of an established metric of D-I teams' strength in the upcoming season, using information that should be available at the time when nonconference games are being scheduled. Lastly, we use the tournament selection model and predictions of future strength to find which potential nonconference opponents would best improve their chances of being selected to the tournament.

The thesis is organized as follows. In Section 2, we take an in-depth look at the structure of college basketball (including the history and importance of the NCAA tournament and the logistics of nonconference scheduling), past work in modelling the NCAA tournament selection committee, various college basketball metrics and work done in predicting a college basketball team's future strength, and previous work on regular season scheduling. Section 3 explores the data that we need to solve this problem and details the acquisition and cleaning of this data. Section 4 explains the methodological approaches we took to navigate the three parts of our problem. Section 5 reveals the results of the methodologies outlined in Section 4. In Section 6, we make conjectures about the results from Section 5 and discuss the limitations in our data and in the nonconference scheduling process. Section 7 looks forward to future work on this topic, including developing this methodology into a usable tool for college basketball coaches and administrators.

Chapter 2

Background

2.1 Overview of D-I college basketball

2.1.1 Conferences

As of 2019, the NCAA has 353 Division I member institutions.¹ These universities divide themselves into conferences, groups of universities that usually have similar locations, enrollments, private/public status, and/or athletic department budgets (Perline, Stoldt, and Vermillion, 2011). Conference membership has a number of benefits for a school which include, but are not limited to, filling the bulk of its regular season schedule with competitive contests, revenue sharing from media contracts and NCAA tournament payouts, and a path to the NCAA tournament through winning the conference tournament. Indeed, no D-I school has competed outside of a conference in basketball since the 2014-15 season.²

There is a large disparity in athletic budgets, fan support, and overall perception of prestige among the 32 D-I conferences. Media and fans often speak of conferences in terms of tiers, where the most powerful conferences are considered “high-major”, and all others are deemed “mid-major”.³ Considerable debate exists about what exactly constitutes a high-major conference, but the boundary tends to involve comparison of athletic budgets or ability of member schools to both reach and advance in the NCAA tournament. One such boundary deems the Atlantic Coast Conference (ACC), Big XII, Big Ten, Big East, Pac-12, and Southeastern Conference (SEC) as high-major, while the rest are mid-major.⁴ (This divide exists concretely in college football, where the five of these six conferences that play football comprise the autonomous “Power 5” conferences.) In addition to the high/mid-

¹<https://web3.ncaa.org/directory/memberList?type=12&division=I>

²<https://www.app.com/story/sports/college/2015/06/12/njit-sheds-independent-status-thanks-atlantic-sun/71133514/>

³http://www.espn.com/mens-college-basketball/notebook/_/page/midmajors140127/so-think-mid-major

⁴<https://johngasaway.com/2018/10/03/what-we-talk-about-when-we-talk-about-major-conferences/>

major divide, there is inequality among mid-major conferences themselves; four conferences (Mountain West, Conference-USA, Atlantic-10, Missouri Valley) in 2007 had average men's basketball budgets above \$2 million, over four times as much as the lowest spending conference (SWAC).⁵ We hypothesize that grouping all of these conferences together ignores the meaningful differences between mid-major conferences.

Perline, Stoldt, and Vermillion (2011) review the somewhat fluid process by which teams join and leave conferences. The inequality among conferences encourages strong teams in less powerful conferences to seek membership in more prestigious conferences and encourages conferences to add institutions that can improve its financial wellbeing and/or athletic prowess. Conferences, whose revenues are often driven by football, frequently change membership, and the fortunes of many conferences and teams have changed drastically as a result of realignment. Sometimes, schools move between conferences of similar status. But in many cases, conference realignment can function like a food chain; the most powerful conferences scoop up the most appealing schools from weaker conferences, leaving these conferences to try to replace the lost teams with new teams from smaller conferences.⁶

These realignments can come in the form of large events that dramatically shift the conference landscape or switches of just a school or two in any given offseason. While a myriad of conference realignments have occurred in the past 25 years, the most important change for our study came in 1996. That year, the high-major Southwest Conference (SWC) split, with half of its membership combining with another high-major, the Big 8, to form a new conference, the Big XII. While high-major conference membership has fluctuated since then, the SWC's demise was the last time a high-major conference dissolved completely. In the 23 years since this major realignment, only the Ivy League has not changed their membership at least once. As a result of these many conference realignments, conferences can gain and lose quality basketball programs as their membership changes, redefining their national prestige along the way.

2.1.2 Tournament Selection

In addition to the conference realignment during the above period, the NCAA tournament itself expanded to 68 teams in 2011. This was only the second such expansion since 1985, and the first that added extra at-large teams. Adding extra teams slightly changes the format of the single-elimination tournament; adding one team in 2001 necessitated a "play-in" game between the two lowest ranked automatic qualifiers, while the three added teams in 2011 led to the development of the "First Four" play-in games between two pairs of low-ranked automatic qualifiers and two pairs of low-ranked at-large teams.

⁵<https://web.archive.org/web/20100410203339/http://www.midmajority.com/redline.php>

⁶https://www.espn.com/mens-college-basketball/story/_/id/7916050/a-look-realignment-perspective-smaller-basketball-conferences-college-basketball

Today, the NCAA tournament is a lucrative, weeks-long television event. An estimated 100 million people watched the 2019 tournament, and the NCAA generated roughly \$900 million through TV, sponsorship, and ticket revenue.⁷ A portion of these revenues are redistributed to the conferences in units. A conference earns a unit for each game that a member school plays in the NCAA tournament, and earned units carry over for six years. Each unit was worth approximately \$280,300 in 2019, and these values increase over time.⁸ These revenues are annually redistributed to member schools by each conference. Thus, a single NCAA tournament bid earned in 2019 would earn, at minimum, about \$1.7 million for a conference over a six year period; this only increases with the number of games won. This can have a large effect on a conference's financial well-being. The 12-team Ohio Valley Conference (OVC), whose two teams in the 2019 tournament each played two games, earned about \$94,000 for each conference member in 2020 (assuming equal redistribution). Tournament redistributions accounted for roughly 13% of the OVC's average men's basketball budget of \$1.46 million in 2016. The OVC's 2019 tournament success will result in at least a 50% increase in payouts (over the OVC's six units in 2019) every year for the next six years, which could provide a substantial financial boost to every OVC member.⁹ ¹⁰ As this shows, every additional tournament entrant is important, especially for smaller conferences.

Teams can reach the NCAA tournament either by earning their conference's automatic berth or by being chosen for an at-large bid by the tournament's selection committee.¹¹ Conferences are free to award their automatic berth in any way they choose, but since the 2016-17 season, each conference has awarded their berth to the winner of the end-of-season conference tournament.¹² For teams that don't win their tournament, the 10-member selection committee, comprised of a rotating ensemble of university athletic directors and conference commissioners, picks teams to fill the available at-large spots. The number of at-large bids has ranged from 34 to 37 since 1997, based on tournament field size and the number of conferences with an automatic bid; since 2014, there have been 36 available at-large bids.

⁷<https://www.usatoday.com/story/opinion/2019/03/18/march-madness-ncaa-revenues-student-debt-wrong-priorities-column/3165524002/>

⁸<https://www.usatoday.com/story/sports/ncaab/2019/03/26/whats-an-ncaa-tournament-unit-worth-millions-heres-how/39256149/>

⁹<https://www.midmajormadness.com/2018/6/11/17441968/ncaa-revenue-expense-statistics-college-basketball-2016-gonzaga-mid-major>

¹⁰<https://ope.ed.gov/athletics/#/>

¹¹<https://www.ncaa.com/news/basketball-men/article/2018-10-19/how-field-68-teams-picked-march-madness>

¹²<https://www.ncaa.org/about/resources/media-center/mens-basketball-selections-101-selections>

The committee determines what it considers to be the “best” at-large teams by a variety of metrics.¹³ The committee’s primary sorting tool, until the 2018-19 season, was the Ratings Percentage Index (RPI). This metric, which sorts all D-I teams, is a weighted average of a team’s win-loss percentage (WL%, weighted 0.25), its opponents’ combined win-loss percentage in all games except those against that team (Opp WL%, weighted 0.5), and its opponents’ opponents’ combined win-loss percentage (OppOpp WL%, weighted 0.25). An additional adjustment was made in 2004-05 to weight a team’s home wins and road losses in its own WL % by 0.6, and home losses and road wins by 1.4; this was intended to account for the difficulty in winning road games.¹⁴ The RPI was meant to be used as a sorting tool, so many metrics supposedly used by the committee are derived from the RPI. These include strength of schedule (SOS, the Opp WL% and OppOpp WL% portions of RPI) for overall and nonconference games and records against teams in different ranges of RPI ranking. Officially, the selection committee is not supposed to look at a team’s conference affiliation when determining at-large selections.¹⁵

The 2017-18 and 2018-19 seasons brought new changes to the selection process. The first change was the introduction of the “quadrant” system for ranking quality of wins.¹⁶ The previous system of using record against teams ranked 1-50, 51-100, 101-200, and 201 and above was altered to account (on top of the RPI adjustment) for the difficulty of winning road games. The breakdown of the new “quadrants” is listed in Table 2.1. The following year, the NCAA announced that a new metric, called NET, would replace the RPI as the committee’s primary sorting tool.¹⁷ The NET “relies on game results, strength of schedule, game location, scoring margin, net offensive and defensive efficiency, and the quality of wins and losses.” However, the exact formula, which is derived using “machine learning techniques” developed by Google, is still unknown. In its inaugural season of 2019, the NET seemed to have been used similarly to the RPI; time will tell how its use will change in future seasons.¹⁸

¹³<https://www.ncaa.org/about/resources/media-center/mens-basketball-selections-101-selections>

¹⁴<http://www.collegerpi.com/rpifaq.html>

¹⁵<https://www.thesportsbank.net/the-bank/conference-call-with-gene-smith-the-chair-of-the-ncaa-division-i-mens-basketball-committee/>

¹⁶<https://www.ncaa.com/news/basketball-men/bracketiq/2018-02-15/ncaa-selection-committee-adjusts-team-sheets-emphasizing>

¹⁷<https://www.ncaa.com/news/basketball-men/article/2018-11-26/net-explained-ncaa-adopts-new-college-basketball-ranking>

¹⁸<https://www.sbnation.com/college-basketball/2019/3/18/18270907/ncaa-tournament-2019-bracket-march-madness-selection-net-rpi-analytics-kenpom>

Table 2.1: NCAA Quadrant Tiering System

Game Tier	Opponent's Ranking by Site		
	Home	Neutral	Road
Quadrant 1	1-30	1-50	1-75
Quadrant 2	31-75	51-100	76-135
Quadrant 3	76-160	101-200	136-240
Quadrant 4	161+	201+	241+

2.1.3 Scheduling Games

Per NCAA regulations as of 2019, teams can generally play a maximum of 31 games prior to conference and/or postseason tournaments.¹⁹ This is conditional on participating in a multi-team event (MTE) that accounts for three or four games (27 games + MTE). If a team does not participate in an MTE, they can play a maximum of 29 games. One additional game may be played if a team travels to Hawai'i. A maximum of four games can be against non-Division I competition (these games aren't accounted for by RPI), and one-third of a team's games must be at home.²⁰ ²¹ A team's regular season consists of its conference and nonconference games. Conference games, usually played from January to March, are scheduled by a team's conference such that a team will play an equal number of home and away contests and play each conference member at least once. In 2018-19, conferences played anywhere from 14 to 20 conference games, leaving the remainder as nonconference games.²²

College basketball teams are responsible for scheduling most of these nonconference games. Common exceptions include interconference challenges (e.g. the ACC-Big Ten Challenge) and MTE games, though teams are free to accept or decline invitations to these. A small number of out-of-conference rivalry games are annual mainstays of many teams' schedules. For the remainder of the available nonconference games, teams have more flexibility to select whom and where to play.

While these remaining games aren't explicitly or implicitly scheduled for a team, they are subject to the restrictions associated with planning events in general. Many scheduled games are part of multi-year deals where games against the same opponent are scheduled before one season but played across multiple seasons. After all previously set games are

¹⁹<https://www.ncaapublications.com/p-4547-2018-2019-ncaa-division-i-manual-august-version-available-august-2018.aspx>

²⁰<https://graphics.fansonly.com/confs/c-usa/graphics/ncaa-membership.pdf>

²¹<https://www.ncaa.org/about/who-we-are/membership/divisional-differences-and-history-multidivision-classification>

²²<https://www.espn.com/mens-college-basketball/standings>

scheduled, teams have a limited number of available game dates. Logistics of travel or academics (i.e. not traveling during finals week) may restrict who, where, and when they play. Another factor at play is the “guarantee game”, in which a large school pays money to a smaller school to play a road game at the large school. Large schools may be expected to play a certain number of games at home, and often use these guarantee games to fill these. Small schools may make more money playing a road game against a large school than they can by playing a home game, and thus rely on these guarantee game paychecks to bolster their athletic department budgets. And in some cases, teams just don’t want to play specific teams; mid-major coaches have claimed that it is difficult to schedule high-major opponents, particularly at home.²³ These wide-ranging logistical difficulties make college basketball scheduling a difficult process; Iowa State director of men’s basketball operations Micah Byars stated that “it’s amazing that any games get scheduled at all. So many factors have to work.”²⁴

While some scheduling work takes place during the previous season, much of the non-conference scheduling for the upcoming season happens at the start of the offseason (David Donlevy, personal communication, July 31, 2018). By the end of the NCAA tournament, teams have only a handful of games left to schedule. In the first months of the offseason, the coaching staff in charge of building the schedule will reach out to potential opponents, sometimes tapping into existing connections with other programs to fill these games. Some of these connections can be with other local schools, opponents from previous years, or former coaching stops. Sometimes, special favors are granted to these connections; in one instance during the 2018-19 season, Oklahoma, of the Big XII, played a road game at Texas-Rio Grande Valley, a WAC school that was the first head coaching job of Oklahoma coach Lon Kruger.²⁵ Scheduling behavior can also be driven by a coach’s opinion of his team; in 2017, the Atlantic-10’s St. Bonaventure, believing they had a tournament-caliber team, scheduled (and won) a game at Syracuse of the ACC.²⁶ Based on individual circumstances, teams may schedule to generate revenue, minimize travel, build up win totals, prepare themselves for conference play, and/or maximize their chances of making the NCAA tournament. Our paper is aimed towards the last of these goals.

²³<https://www.cbssports.com/college-basketball/news/mid-major-coaches-need-to-stop-complaining-and-start-adapting-to-college-basketball-realities/>

²⁴http://www.iowastatedaily.com/sports/it-s-amazing-that-any-games-get-scheduled-at-all/article_59a2bdd0-303d-11e8-9203-93cf62fccf30.html

²⁵<https://oklahoman.com/article/5604161/sooners-coach-lon-kruger-to-be-inducted-into-utrgv-hall-of-fame>

²⁶<https://www.cbssports.com/college-basketball/news/mid-major-coaches-need-to-stop-complaining-and-start-adapting-to-college-basketball-realities/>

2.2 Past work in predicting NCAA tournament selection

While selection of teams to the NCAA tournament is one of the primary objectives of this project, much of the existing literature on the NCAA tournament is related to the prediction of tournament game outcomes after selection. Picking March Madness “bracket pools” is wildly popular in the United States, so there is a great deal of interest in being able to optimally select the winners of each game. Others have presented ideas on how the tournament field should be better selected (Harville, 2003). However, some research on the selection process has been done that is relevant to this thesis. In addition, websites such as <https://crashingthedance.com/> put out related predictions annually.

Coleman and Lynch (2001) used probit regression to find a six-factor model for tournament selection. They used an initial 42 predictors (29 straight from the nitty gritty sheet, 13 derived from it) to predict at-large selection. The model was built on the six tournament fields from 1994 to 1999, using only at-large eligible teams whose RPI ranking ranged from 25 to 80 so as to eliminate obvious inclusions and exclusions (the lowest ranked invitee, as of 2000, ranked 75, and the highest ranked non-invitee ranked 29). They tested the model on the 2000 tournament field. The six factors found to be important in tournament prediction were RPI rank, conference RPI rank, wins against the RPI top 25, against the RPI 26-50, against the RPI 51-100, and wins above .500 in conference play. Their model misclassified 13 of the 202 (6.4%) at-large bids insample (including obvious inclusions) and 3 of 34 in the test set (8.8%). In later seasons (2001-08), the model misclassified 20 of 273 at-large bids (7.9%).²⁷

The research of Sanders (2007) finds evidence of bias in the RPI formula. These findings show that the RPI may be biased against teams from high-major conferences versus teams from low- or mid-major conferences, even if the high-major team is of equal or greater talent. However, their underlying assumption that every high-major team is at least as talented as every low- or mid-major team seems very restrictive and unlikely to be true based on past history.

Zimmer and Koethe (2008) find moderate evidence of a tournament seeding bias based on conference affiliation between members of high-major and non-high-major conferences. Five of the six coefficients in a linear regression model for seed position (one for each conference) show that teams from these high-major conferences tend to receive a lower seed, compared to mid- or low-major teams, than their eventual tournament performance would warrant. However, only two of these individual coefficients are significant even at the 20% level (one positive coefficient, one negative), so it is unclear how much conference affiliation affects tournament seeding. The researchers do not examine the effect of conference affiliation on at-large selection.

²⁷<https://www.unf.edu/~jcoleman/perform.htm>

Shapiro, et. al. (2009) apply both direct and indirect explanatory variables to predict at-large selection using logistic regression. The direct variables include several variables used by Coleman and Lynch (2001), as well as preseason ranking and conference finish in both tournament and regular season play. Indirect (non-performance-based) variables include conference affiliation, tournament appearance in the previous season and previous 10 seasons, and a school's time zone, county population, and county per capita income. Conference affiliation was divided into three tiers that labeled the same six conferences as chosen above as Major, nine conferences as Mid-major, and the rest as Small. This model was built on the 1999-2007 tournament fields, using only at-large eligible teams in the RPI top 100. RPI ranking, top 50 RPI wins, top 51-100 losses, conference affiliation, and conference regular season finish were found to be significant predictors of at-large selection at the .05 significance level. 27 teams from 1999-2007 were misclassified in-sample. The researchers didn't ensure that only the top 34 (each season in this range had 34 at-large bids) teams in each season were selected, so more or less than 34 teams were selected in several seasons, distorting the true number of missed at-large bids.

Coleman, DuMond, and Lynch (2010) built on their original 2001 work by examining potential committee biases with regards to conference affiliation and selection committee representation. 61 predictors (41 team performance based, 20 conference/committee based) were used in the analysis, which was divided into five model specifications: all team performance variables, all variables, variables selected using stepwise methods, all variables except RPI-based team performance measures, and all except Sagarin-based team performance measures (team strength measures discussed later). Evidence was found of selection bias related to both conference affiliation (benefiting major and mid-major teams versus low-major) and committee representation.

Paul and Wilson (2012) examined the committee biases found by Coleman, DuMond and Lynch (2010) in the presence of the RPI and Sagarin Predictor ratings. RPI does not take a team's margin of victory into account, whereas the Sagarin Predictor does. The researchers fit two probit regression models, one with a team's RPI rating included with committee representation and conference affiliation (Major, Mid-major, Small) considered, and one with Sagarin Predictor instead of RPI. These models were fit on the 2004-11 tournaments, with all at-large eligible teams studied. Significant effects of committee representation and high-major bias were found in the RPI model, but not in the Sagarin model. This gives evidence that margin of victory is important to the committee, and perhaps explains the findings of Sanders (2007) of a positive mid-major bias in the RPI. No misclassification rates were reported.

Leman et al. (2014) studied several interesting problems in the realm of at-large selection using Bayesian methods. First, they researched the importance of various performance metrics to the committee in selecting at-large participants. While each of Sagarin's Predictor (as above), ELO (no margin of victory), Overall (a combination of Predictor and ELO), and

LRMC (Logistic Regression Monte Carlo) explain variation in at-large selection, none of them are significant predictors in the presence of RPI. Next, the researchers examined the committee’s use of RPI. By its calculation (25% W-L, 75% SOS), the RPI values strength of schedule three times as much as win-loss percentage. However, by individually weighting W-L% and SOS, the ratio of SOS to W-L% in predicting selection to the 2010 and 2011 tournaments exceeds 4, and its 95% credible interval does not include 3. This provides evidence that the committee weights strength of schedule more than the RPI does, which makes sense if the committee uses the SOS-centric metrics on the nitty gritty sheet. Third, the researchers examine the effect of “marquee” status on a “bubble” team’s selection. They define a team as marquee if, in two of the previous three years, that team played at a similar level to a top three tournament seed (as North Carolina was 11 times from 1994-2011), and as “bubble” if that team played at a similar level to one of the worst teams selected to or best teams omitted from the NCAA tournament. Eleven of the 12 marquee bubble teams got at-large bids, while 61% of the 527 non-marquee bubble teams did so. According to the researchers, if in 2010, Virginia Tech (who missed the tournament) had been marquee, their probability of selection would have risen from 28% to 86%.

Coleman, DuMond, and Lynch (2016) extended their previous research to develop an eight-factor stepwise probit regression model. This model, which initially studied 59 predictors, was trained on the 2009-13 tournament fields and tested on 2014-16. The eight factors found by this model to be significant predictors are RPI ranking, losses below .500 in conference play, wins against the top 25 and 26-50, games above .500 against 26-50 and 51-100, road wins, and being in the Pac-10/Pac-12. This model correctly predicted all but one bid (of 179) in-sample and misclassified just four when examining each season via leave-one-year-out cross-validation. The model misclassified 4 of 72 bids when predicting the 2014 and 2015 fields. During the modeling process, the researchers also found when fitting for all the years 1999-2013, the first ten seasons (1999-2008) produced 2.5 times as many misclassifications per season as 2009-13, suggesting that the selection process has either changed or at least become more predictable in recent seasons. They also investigated the influence of 25 different performance metrics (most of which account for margin of victory), finding that none of the 25 performed as well in place of RPI and only eight were significantly associated with prediction when included alongside RPI in their model. This suggests that the most important metric used by the committee was still RPI.

2.3 Advanced metrics and preseason predictions

2.3.1 Advanced metrics in college basketball

As in many sports, independent researchers have developed a wealth of advanced metrics intended to better describe the quality of college basketball teams. These team metrics can be divided into two groups: prospective and retrospective.

Prospective metrics attempt to quantify how good a team is, which makes these metrics useful for predicting future game results. Both final scores and win probabilities can be computed using prospective metrics. Among these prospective metrics are Ken Pomeroy's Adjusted Efficiency Margin (AdjEM, the difference in an adjusted offensive and defensive scoring margin)²⁸, Sagarin's Predictor²⁹, and ESPN's Basketball Power Index (BPI)³⁰. While Las Vegas betting markets do not release any team rankings (for obvious proprietary reasons), their final betting lines are also remarkably accurate in predicting final scoring margins (SD of difference between betting line and actual scoring difference is roughly 11 points), and could be thought of as a kind of prospective metric on an individual game basis.³¹

Retrospective metrics attempt to quantify what a team has accomplished, and thus are often referenced when identifying potential NCAA tournament teams. These differ from prospective rankings in that game results (who a team has played and where) are far more important than scoring margin. Thus, a team with blowout victories and narrow defeats would be ranked higher by a prospective metric than a retrospective metric. Examples of retrospective metrics include Sagarin ELO (not available for 2018-19, but used in previous literature), ESPN's Strength of Record³², and the NCAA's RPI and NET rankings. Top 25 polls could also be thought of as a retrospective measure, albeit one that is neither analytical nor comprehensive.

2.3.2 Predicting team strength in future seasons

By its very nature, college basketball is difficult to predict from season to season. Players only have four years of playing eligibility, so the most seasoned players on a team are the ones least likely to return to their teams. High achieving underclassmen are increasingly likely to leave school to pursue opportunities in professional basketball.³³ Players are also transferring to different schools in increasing numbers, particularly immediately eligible graduate transfers.³⁴ Many coaches are also hired and fired in what has come to be known

²⁸<https://kenpom.com/>

²⁹<http://sagarin.com/sports/cbsend.htm>

³⁰<https://www.espn.com/mens-college-basketball/bpi>

³¹<https://harvardsportsanalysis.wordpress.com/2011/10/17/how-much-does-vegas-learn/>

³²https://www.espn.com/mens-college-basketball/bpi/_/view/resume

³³<https://twitter.com/NCAAREsearch/status/1148943206119346176>

³⁴<https://www.indystar.com/story/sports/college/2017/11/07/college-basketball-transfers-arent-new-but-players-and-coaches-cope-increase/686784001/>

as the offseason’s “coaching carousel”.³⁵ With the amount of turnover in both team rosters and coaching staffs, it is an inexact science to predict how good a team will be in the upcoming season.

This unpredictability does not prevent people from trying to prognosticate the next season. Following the end of the NCAA tournament, it has become common for college basketball media to release their own preseason rankings.³⁶ More analytical and comprehensive methods exist as well, the relative performances of which can be explored on the website of Dan Hanner, a former developer of predictions for Sports Illustrated.³⁷ These rankings may incorporate returning players, incoming freshmen and transfers, coaching changes, historical success, or other factors. However, most of these are released close to the start of the college basketball season in November. Since so much scheduling takes place in April and May, we need a metric by April 30. Bart Torvik releases publicly available team rankings throughout the offseason that are similar to the efficiency metrics used by Pomeroy.³⁸ Pomeroy himself makes proprietary preseason rankings in the spring that he makes available to college basketball coaching staffs for a fee.

2.4 Other relevant college basketball research

Harville (2003) proposes an alternative method of ranking college basketball teams than RPI. As a foundational piece of this method, he proposes a least squares regression model for the difference in game score, y_{ijk} , in the k th game between home team i and visiting team j . This model, using the same defined notation, is

$$y_{ijk} = x_{ijk}\lambda + \beta_i - \beta_j + e_{ijk}; i \neq j = 1, \dots, t; k = 1, \dots, r_{ij}$$

where β_i and β_j are the team effects of teams i and j , x_{ijk} is an indicator variable set to 1 if team i is at home and 0 if the game is at a neutral site, λ is the effect of home court advantage, t is the number of teams in Division I, and r_{ij} is the number of games with team j playing as the designated road team (including neutral site games) against team i . The residuals e_{ijk} are assumed to be approximately normally distributed with mean 0 and common, unknown variance σ^2 . This is warranted by the Central Limit Theorem because each team has a roughly equal number of offensive possessions in a basketball game, and

³⁵<https://www.cbssports.com/college-basketball/news/college-basketball-coaching-changes-the-last-big-job-has-closed-as-mike-anderson-agrees-to-coach-at-st-johns/>

³⁶<https://www.si.com/college-basketball/2019/04/09/ncaa-basketball-rankings-early-top-25-kentucky-duke-michigan-state>

³⁷<https://yetanotherbasketblog.blogspot.com/2018/06/comparing-2017-18-preseason-rankings.html>

³⁸<http://barttorvik.com/trankpre.php>

thus differences in score in a college basketball game, shown here as y_{ijk} , are the sums of the score differences in each possession. While the y_{ijk} 's cannot be normal because game scores are naturally discrete, Harville deems normality to be a reasonable approximation in this case.

In addition to their work mentioned in Section 2.2, Leman et al. (2014) also discuss the impact of scheduling on the eventual probability of tournament selection. As an example, they use the 2009-10 Virginia Tech (VT) Hokies, a team that was not selected to the NCAA tournament. The researchers hypothesize that this was due in part to five of VT's nonconference wins coming against teams ranked 285 or lower by RPI. They implement four simulations of season schedules. In each, they replace one to five of these low-ranking wins with losses against the same number of teams ranked in each of the RPI top 25, 26-50, 51-100, and 101-200. After each, they use their previously stated selection model to calculate the estimated probability that VT would have received an NCAA tournament bid, comparing the simulated seasons to VT's estimated probability of 0.31 given their actual game results. The researchers found that when replacing the low-ranked wins with losses against the top 25, selection probability increased with any number of replacements in every simulated season. This effect diminished as the replacements got worse; only one or two replacements against the next 25 improved selection probability each time, and replacements against 51-100 and 101-200 ranked teams decreased selection probability. As all the replacements were losses, these changes represent the worst-case scenarios for playing better opponents; any win would increase the probability of selection in that hypothetical season.

Frank Dehel of www.dribblehandoff.com found a positive association between difficulty of a nonconference schedule and success in conference play.³⁹ Looking at 2012-2018 data and controlling for preseason ranking in Pomeroy's Adjusted Efficiency Margin (AdjEM), he found a small but significant relationship between playing Tier A and B games (designations based on AdjEM ranking and game location; Tiers A and B are equivalent to a top 50 and top 51-100 game at a neutral site, respectively) and winning percentage in conference play; a 24% increase in Tier A and B games was associated with a 6% increase in conference WL %, about one win in an 18-game conference slate. No further numeric results were reported in the article.

³⁹<http://www.dribblehandoff.com/stories/the-6-impact-of-scheduling>

Chapter 3

Data

3.1 Information needed

Our analysis consists of three phases:

1. Predicting at-large tournament selection in a given season from the performance metrics supplied to the tournament selection committee and conference affiliation.
2. Predicting a team's strength in the upcoming season using metrics available as of April 30 in each year, including head coaching continuity, production of returning players, quality of incoming freshmen, and previous seasons' team strength.
3. Simulating the upcoming season using predicted team strengths from the model built in Phase 2, using this to predict the probabilities of selection using the Phase 1 model, simulating this season for a selected team, adding one game against each potential nonconference opponent each time, and calculating change in selection probability from initial simulation.

3.1.1 Phase 1

We need the participating teams, game type (regular season, conference tournament, post-season tournament), location (home, away, neutral), and game results for all Division I-only games to calculate the RPI and corresponding metrics. If possible, we would like this data to go back to the 1996-97 season, the last time a high-major conference was created or dissolved. In addition, we need to know the conference affiliation of each team to study the impact of conference affiliation on at-large selection.

3.1.2 Phase 2

As noted in Section 2.3.2, there are a few available methods to predict a team's strength in the upcoming season as required by Model 2. But since we need a metric that can be calculated using publicly available data knowable by April 30, we will develop our own

projection method. (Others who use our tool can substitute their own preferred projection if desired). In order to create such a model, we will need an adequate measure of team strength and a knowable list of predictors.

Of the several prospective metrics used to measure team strength, we believe that Pomeroy's AdjEM rating is the best for our needs because it is superior in three key attributes: historical availability, metric interpretability, and predictive accuracy. Team AdjEM ratings are available going back to 2002 (with a website subscription). AdjEM can be strictly interpreted as the number of points above an average basketball team per 100 possessions, which, when combined with teams' adjusted tempos (measuring the number of offensive possessions per games), can give predicted score differentials. AdjEM is also closer to the final point differential slightly more often than Sagarin's Predictor.¹

To project future team strength, we also need to know about key personnel changes that will occur in the offseason. The changes we use must be known before April 30.

1. Most head coaching changes are known by this cutoff date; in 2019, only 5 of 59 coaching departures took place after that date.²
2. The departures of seniors, who have exhausted their eligibility, are also known prior to this date.
3. Underclassmen who choose to declare for the NBA draft must do so by a given date; in 2019, this deadline was April 21.³ Prior to 2016, these underclassmen would have become immediately ineligible to return to college basketball. However, recent rule changes allow players to declare for the draft, later remove their name from the draft process, and still retain college eligibility; in 2019, the deadline to withdraw and still retain eligibility was May 29. As our cutoff date lies between these two points, we thus know by April 30 who has declared for the draft, but not who has decided to return to school. Thus, we must assume that all players who have entered the NBA draft process will not be returning to their team.
4. Departing and incoming transfers are difficult to study. As of 2019, players can announce their intent to transfer from their school by entering their name into a "transfer portal" managed by the NCAA. A large number of players will have announced their decision to transfer by April 30.⁴ However, their decision of which school to attend is

¹<https://www.sportsbettingdime.com/guides/strategy/kenpom-vs-sagarin/>

²<https://www.cbssports.com/college-basketball/news/college-basketball-coaching-changes-the-last-big-job-has-closed-as-mike-anderson-agrees-to-coach-at-st-johns/>

³https://www.espn.com/nba/story/_/id/26586762/nba-draft-233-early-entry-prospects-declare-2019

⁴<https://watchstadium.com/news/comprehensive-list-of-college-basketball-players-transferring-12-29-2018/>

often unknown by this point. Most transfers are not immediately eligible to play in the next season, so their new location is irrelevant. In addition to changes in personnel, it is very unclear how a player's production will be affected by transferring to different tiers of D-I. A highly productive low-major player may not be as productive when transferring to a high-major school, just as a bench player for a high-major team may or may not flourish in a smaller conference. The impact of transfers is certainly worth studying, but we feel that given the uncertainty in eligibility and final destination as of April 30, we cannot consistently incorporate incoming transfers into our model. We do have a sufficient knowledge by April 30 of who is leaving their team, and thus we can use these departures when looking at a team's roster.

To capture the relative effects on a team due to departing and returning players, we need a measure of each player's contribution to the team's quality. Returning players' production was measured by Win Shares (WS). This metric, calculated by Sports Reference in all seasons from 2002-19, estimates a player's offensive and defensive contributions to team wins.⁵ The total WS of all players returning for the upcoming season was scaled to a 30 game season to account for differing numbers of games played by Division I teams. Thus, for example, we can calculate the proportion of a team's WS from the past season that is projected to return in the upcoming season.

While freshman recruiting is a fluid process, most commitments from incoming freshmen are known by April 30 as well. Of the top 100 high school seniors in 2019, as ranked by RSCI (a composite of four different recruiting services⁶), only 19 were not committed to play for a school by April 30, while two decommitted from their original schools after this date.⁷ For those uncommitted freshmen, we can rely on the projections of recruiting experts as of April 30 to approximate their eventual landing spots; 247Sports.com correctly predicted the commitment of 14 of these 19, meaning that 93 of the top 100 were either committed to or predicted to commit to their chosen location by April 30. Thus, we can measure freshman talent entering into each team using those players' RSCI rankings from their final high school seasons.

3.1.3 Phase 3

To model the changes in selection probability by scheduling a given team in the upcoming season, we would need to know the upcoming season's schedule for all of D-I. We cannot know the full schedule because the very purpose of this tool is to help create this schedule. Thus, we need to find a suitable way to approximate the next season's schedule, which we

⁵<https://www.sports-reference.com/cbb/about/ws.html>

⁶<https://sites.google.com/site/rscihoops/home/2019-final>

⁷<https://247sports.com/college/basketball/recruiting/>

do by simply using the previous season’s schedule. A team’s conference game opponents and locations are either exactly the same from year to year (in conferences where every team plays each other home and away) or one combination in the set of all possible combinations. Furthermore, as stated in Section 2.1.3, teams tend to have the same or similar nonconference opponents from season to season. We therefore assume that the best simple approximation to a team’s nonconference schedule is that team’s previous nonconference schedule. Thus, we use the previous season’s schedule as a proxy for the upcoming season.

3.2 Data sources and acquisition

Several online sources exist to provide the necessary schedule information; we chose to collect data from Sports Reference LLC, found at <https://www.sports-reference.com/cbb/>. Their schedule data is almost entirely complete extending back to the 1992-93 season, stored in a table format consistent over different teams and seasons. Their data isn’t entirely complete; teams transitioning to D-I prior to 2010 were considered “non-major”, yet these teams were included in the NCAA’s RPI calculations. Data for games between two transitioning teams were found from ESPN.com or, failing there, individual school websites. The full list of explanatory variables and their abbreviations for this model is given in Table 3.1.

The necessary data for our projections of team strength were taken from multiple sources. Team rosters from 2002-2019 and Recruiting Services Composite Index (RSCI) rankings for high school seniors from the graduating classes 2001-2018 were gathered from Sports Reference. Our roster data gives us, among other information, the head coach, player biographical information (helpful for identifying players who transfer), year of eligibility, and WS. The full list of explanatory variables and their abbreviations for this model is given in Table 3.2.

The RSCI data gives us a player’s RSCI composite recruiting rank, year drafted into the NBA (if applicable), and final school attended (if any). For modeling, ranks are converted into scores such that the #1 ranked freshman receives 100 points, #2 receives 99, etc. Note that the final school may be different from the initial school the player was recruited to; in these cases, we had to manually find the original school. The total points of all freshmen committed and academically admitted to a school, along with the points of previous years’ freshmen who did not play due to redshirt status, injury, transfer, missionary service, or other extenuating circumstances, were summed for each school.

Team AdjEM and pace data, going back to 2001-02, was gathered from Pomeroy’s site, kenpom.com. Teams transitioning to D-I are not always included in these data; for teams in their first season or previously considered “non-major” by the Sports Reference data, we assume that their head coach was retained from the previous season (far fewer than half of D-I head coaches leave their school each offseason) and their returning WS is zero, as they have no returning major D-I production. D-I newcomers with no previous Pomeroy data

Table 3.1: Explanatory Variables Used in Model 1

Variable	Description
RPI	Numeric value of Ratings Percentage Index; higher is better
Rank	Ranking of RPI; takes integers 1 – # of D-I teams, lower is better
ConfLevel (High)	Indicator for member of power conference [†]
ConfLevel (Mid)	Indicator for member of non-power conference with $\tilde{b}_{ct} \geq 2.5$
ConfLevel (Low)	Indicator for member of non-power conference with $\tilde{b}_{ct} < 2.5$
SOS	Numeric value of Strength of Schedule
SOSRank	Ranking of Strength of Schedule
NCSOS	Numeric value of Strength of Schedule for nonconference games
NCSOSRank	Ranking of Nonconference Strength of Schedule
RawW	Number of wins in all D-I games
RawL	Number of losses in all D-I games
W1_50	Number of wins against teams ranked 1-50 by RPI
L1_50	Number of losses against teams ranked 1-50 by RPI
W51_100	Number of wins against teams ranked 51-100 by RPI
L51_100	Number of losses against teams ranked 51-100 by RPI
W101_200	Number of wins against teams ranked 101-200 by RPI
L101_200	Number of losses against teams ranked 101-200 by RPI
RoadW	Number of wins in road games
RoadL	Number of wins in home games
AdjWL	Winning percentage (with 2005 RPI adjustment)
RawWL	Winning percentage (unadjusted)
OppWL	Opponents' winning percentage (50% of RPI)
OppOppWL	Opponents' opponents' winning percentage (25% of RPI)
NCOppWL	Nonconference opponents' winning percentage
NCOppOppWL	Nonconference opponents' opponents' winning percentage

†: ACC, Big XII, Big East, Big Ten, Pac-10/12, or SEC

Table 3.2: Explanatory Variables Used in Model 2

Variable	Description
SameCoach	Indicator of the previous season's head coach returning
Prev.Yr.KP	AdjEM rating from the previous season
TwoYrs. Prev.KP	AdjEM rating from two seasons prior
Returning.WS.30G	Returning players' number of win shares per 30 games
Freshmen	Sum of converted RSCI rankings of incoming freshmen

were given the average ratings of new D-I schools in the season prior and two seasons prior to major status, according to Sports Reference, as their Prev.Yr.KP and TwoYrs.Prev.KP.

In the absence of 2020 roster data, information regarding 2019 recruits, head coaching changes, player transfers, and NBA Draft declarations was found on 247Sports⁸, CBS Sports⁹, Stadium¹⁰, and ESPN¹¹, respectively. These sites appeared to have the most complete available data that we could find, though other sources may be privately available to those within the coaching industry.

Given the sheer amount of data needed from Sports Reference, we built web scraping programs, largely from scratch, using the Python package BeautifulSoup (in Jupyter) to automate the data collection process for the schedule, roster, and RSCI datasets. These pulled the necessary data out of tables stored within separate pages for every school in each season and put them into .csv files, which were then manipulated using Microsoft Excel.

3.3 Data cleaning and augmentation

Considerable effort was needed to get our data into a consistent, usable format for data analysis. Cleaning the schedule data required a several-step process that included removing Top 25 designations from team opponents, generating a different dataset of teams, matching different versions of team names recorded on different data sources, and adding and removing teams as their D-I designation changes. This was done on a yearly basis every season from 1993 to 2019.

Roster data, while not cleaned separately for each season, was similarly arduous to get into a usable format. Far more inconsistencies exist in this data, with the alignment of columns in the scraping program's .csv output changing with every unintentional comma hidden in names, hometowns, previous schools, and more. Player heights (useful for player matching) were automatically converted to dates by Excel, and thus had to be reconverted. These kinds of alterations had to be done on an almost case-by-case basis at times. For anyone seeking to use this tool or repeat this work, cleaning player data would be perhaps the biggest barrier to usability.

Following the cleaning of the data, we used R to calculate the RPI and corresponding metrics, found in Table 3.1, for each season. Calculating RPI is an iterative process; each team's WL% (adjusted or unadjusted) must first be calculated before a team's Opp WL%

⁸<https://247sports.com/college/basketball/recruiting/>

⁹<https://www.cbssports.com/college-basketball/news/college-basketball-coaching-changes-the-last-big-job-has-closed-as-mike-anderson-agrees-to-coach-at-st-johns/>

¹⁰<https://watchstadium.com/news/comprehensive-list-of-college-basketball-players-transferring-12-29-2018/>

¹¹https://www.espn.com/nba/story/_/id/26586762/nba-draft-233-early-entry-prospects-declare-2019

can be calculated, and all teams' Opp WL% must be calculated before finding Opp Opp WL%. RPI for seasons prior to 2005 used the old version of the RPI, while all seasons since then were calculated using the adjusted version. Both the rating (a decimal between 0 and 1, where higher is better) and ranking (integer-valued from 1 to the total number of D-I teams in a season, where lower is better) of RPI and SOS measures were used. Minor differences in calculated RPI versus the committee's RPI may exist, most often due to differences in home/neutral designation between the NCAA and Sports Reference.

The tiering of mid and low-major conferences was done using the number of historical berths to the NCAA tournament and National Invitational Tournament (NIT, a secondary tournament that includes the presumed best teams not invited to the NCAA tournament and any regular season conference champions who did not win their conference tournament). The true difference between conferences isn't just the difference in quality of their best teams, but also the quality of their average teams. Conferences that consistently earn multiple bids to postseason tournaments may be viewed differently than conferences that may usually send only automatic qualifiers to those tournaments. We calculate a weighted average, \tilde{b}_{ct} , of the number of combined bids from conference c to these tournaments using the formula, $\tilde{b}_{c,t} = .4(b_{c,t-1}) + .3(b_{c,t-2}) + .2(b_{c,t-3}) + .1(b_{c,t-4}); t \in [1997, 2020]$, where $b_{c,t}$ is the number of combined NCAA and NIT berths for a conference in any year t , c is the conference, and t is the season in question. These weights give greater value to the number of bids earned in more recent seasons. A mid-major conference was deemed to have a weighted average of at least 2.5 combined bids, while low-major conferences have a weighted average below 2.5. Data going back to the 1992-93 season were used to calculate these weighted averages.

Chapter 4

Methods

To obtain a list of potential nonconference opponents that increase a given team's likelihood of reaching the NCAA tournament, we need to create three different models. Our first model, Model 1, predicts which teams in a given season are most likely to receive an at-large bid given the outcome of a season. The model is based on a variety of variables thought to be considered important by the NCAA tournament selection committee, which include various performance measures and conference affiliation. Next, Model 2 estimates a measure of the quality of a given team in the next season, given the outcome of the previous season and the anticipated changes that the team will undergo. The model is based on measures of a team's quality in previous seasons, returning players and coaches, and incoming freshman recruits. Note that Model 2 is entirely separate from Model 1. It uses completely different data (roster and efficiency vs. schedule and conference) and predicts a different outcome (how talented a team will be vs. how well it has played). These two separate models come together in the final step to form Model 3. In Model 3, we first simulate game outcomes for an upcoming season, using that season's team-quality predictions from Model 2 to compute the expected results from every game. From these, we then calculate the performance measures used in Model 1 for each team, thus producing its predicted tournament qualifications. These measures are then put into Model 1 to obtain each team's ranking for determining the expected NCAA tournament at-large recipients. These results will serve as a baseline for each team, assuming that they play the same schedule as the previous season. Next, for a given team, we add one home, away, and neutral game against every potential nonconference opponent, recalculating the performance metrics and tournament probabilities after each added game. We then compare each new season, with its one added game, to the baseline season and find the resulting change in probability of selection and at-large ranking.

4.1 Phase 1

The primary objective of our first model is to predict team rankings for at-large tournament selection. In our study, this is done using a binary response variable, *Tourney*, which takes

on either “At-Large” for a selected team or “Missed” for an unselected team. Secondly, it is worthwhile to learn something about which factors the committee considers the most important. Additionally, we need our model to be set up so that simulated season output from Model 3 can be fed back into it.

Prior to classification, we used Bayesian model averaging (BMA) to perform variable screening. BMA operates under the assumption that out of all possible models, there is one true model. We fit a logistic regression model to each of the most likely possible combinations of explanatory variables, as found by a leaps and bounds algorithm. For each model, we computed a posterior probability that it was the true model given the data. The sum of posterior probabilities across all models in which a particular variable appears estimates the posterior probability that the variable is included in the final model, or $P(\textit{inclusion})$. Our full model has 24 predictor variables, but we believe that many of them may be useless in predicting at-large selection. So as a first step of our modeling process, we screened out all variables that had a $P(\textit{inclusion})$ of at least 0.25. We chose 0.25 instead of the more common 0.5 because our purpose is to eliminate completely useless variables, and variables with a $P(\textit{inclusion})$ between 0.25 and 0.5 may still have a limited impact.

The variables chosen by BMA were next fed into a battery of classification models with the goal of assigning each team a score that is monotonically related to the certainty of its tournament selection. We need these scores, rather than individual predictions of success or failure, because the tournament has a number of at-large berths, m_t , that must be filled no matter what quality of teams are available. As the value of m_t has changed over time with new conferences and tournament expansion, we find that simple binary classification using a rule of selecting teams whose success probability is at least 0.5 would often produce a number of at-large teams not equal to m_t . Rather, a scoring metric was derived from each model, and the top m_t teams for each year were labeled as successes. Table 4.1 has a list of classification techniques used and the derived score metric used. Details of each model are available in (Hastie, Tibshirani, Friedman, 2009).

We used team schedules and results from 1997-2017 to build our tournament selection model. Only at-large eligible teams (no automatic qualifiers or postseason ineligible teams) with an RPI ranking of 100 or better were used to fit the model ($n = 1618$). These observations, stratified by season, were randomly split ten times into training and test sets (15 seasons in the training set, six in the test set). Within each split, we used each technique to fit a model to the training data and used that model to predict the invitees for the years in the test data. Model types that require tuning of hyperparameters were tuned using leave-one-year-out crossvalidation. Models were judged based on performance in two criteria for each split. The primary criterion is a model’s score-based misclassification rate, meaning the percentage of observations where the actual at-large selection differs from the model’s classification of at-large selection based on an observations ranking in a score

Table 4.1: At-Large Classification Modeling Techniques

Technique	Score Metric	Hyperparameters	Reference [†]
Lin. Discrim. Analysis	Posterior probability	N/A	4.3
Quad. Discrim. Analysis	Posterior probability	N/A	4.3
Logistic Regression	Fitted probability	N/A	4.4
LASSO Log. Reg. (1)	Fitted probability	Shrinkage, $\lambda = \lambda_{minCV}$	4.4
LASSO Log. Reg. (2)	Fitted probability	$\lambda = \lambda_{minCV+1SE}$	4.4
Gen. Add. Mod. (GAM)	Fitted probability	Splines on all num. vars	6.5, 9.1
Support Vector Machines	Posterior probability	Cost, γ	12.1-12.3
Random Forests	Proportion of votes	# of vars, max. nodes	15
Neural Networks	Fitted probability	Size, decay	11.3-11.5

†: Section where found in (Hastie, Tibshirani, Friedman, 2009).

metric compared to m_t . The secondary criterion is positional error, defined as a misclassified observation’s number of score-based ranking places away from a correct classification. This can be defined as $poserr_i = \{pos_i - m_t; pos_i > m_t, pos_i = m_t + 1 - pos_i, pos_i \leq m_t$, where pos_i is the score-based ranking of team i . The best model was then used to predict the 2018 and 2019 tournament fields.

4.2 Phase 2

The primary objective of our second model is to predict a team’s quality in the upcoming season, which we estimate using Pomeroy’s AdjEM rating. In our study, our response variable is the upcoming season’s AdjEM rating, KP.Rating, a continuous variable which takes on positive and negative values. Unlike with Model 1, we aren’t as concerned with identifying influential features in the predictions from Model 2; we just want the best prediction available. Predictions from this model are primarily used to aid our simulations of upcoming seasons. As we only have five predictor variables, variable selection is not really needed here. The added variance from estimating unimportant parameters would seem to be minimal.

The five hypothesized explanatory variables for team strength were put into a series of regression models, the full list of which can be found in Table 4.2. Details of each model are available in (Hastie, Tibshirani, Friedman, 2009). We used team personnel data and AdjEM ratings from 2004-2019 to build our team strength prediction model. All Division I teams (including reclassifying teams playing a regular D-I schedule) were used to fit the model ($n = 5439$). These observations were randomly split ten times into training (75%) and test (25%) sets. Unlike with the tournament selection model selection, we do not need to stratify our observations by year because each team’s quality is assumed to be independent of other teams in a given season, whereas eventual tournament selection is somewhat dependent on the season (because of the m_t cutoff) and the performances of other teams in that season.

Table 4.2: Future Team Strength Prediction Modeling Techniques

Technique	Hyperparameters	Reference [†]
Linear Regression	N/A	2.3, 3.1, 3.2
LASSO Lin. Reg. (1)	Shrinkage, $\lambda = \lambda_{minCV}$	2.8.1, 3.4-3.9, 7.10
LASSO Lin. Reg. (2)	$\lambda = \lambda_{minCV+1SE}$	2.8.1, 3.4-3.9, 7.10
Relaxed LASSO	$\lambda = \lambda_{minCV}, \phi = \phi_{minCV}$	2.8.1, 3.4-3.9, 7.10
GAM	Splines on all vars	9.1
Projection Pursuit Reg.	Terms = [1, 2, 3, 3 from 6]	11.2
MARS	Degree, penalty	9.4
Random Forests	# of vars, node size	15
Neural Networks	Size, decay	11.3-11.5
Gradient Boosting Machines	Interaction depth, shrinkage	10, 16

†: Section where found in (Hastie, Tibshirani, Friedman, 2009).

Within each split, we used each technique to fit a model to the training data and used that model to predict the test data. Model types that require tuning of hyperparameters were tuned using ten bootstrap repetitions. Models were judged based on each split’s mean squared prediction error. The best model was then used to predict the strength of teams in the upcoming 2019-2020 season.

4.3 Phase 3

Finally, the predictive models for tournament selection and upcoming year’s AdjEM were combined to find optimal nonconference opponents. We used the previous season’s schedule as an approximation of next year’s schedule, with conference tournament games excluded because of the widely varying tournament formats between conferences and extreme unpredictability of conference tournament matchups. In every game, each team’s predicted AdjEM, as estimated by the final Model 2, was combined with their adjusted tempos (approximated by the previous season’s adjusted tempo) to calculate an expected score differential. Our model is similar to Harville (2003) as explained in Section 2.4, where our team effects β_i and β_j are the predicted AdjEM ratings of teams i and j , multiplied by the average of the teams’ tempos divided by 100 possessions, our home court advantage λ is 3.75 points¹, and x_{ijk} is an indicator equal to 1 if team i is at home, -1 if they are away, and 0 if the game is at a neutral site. We assume each score differential, y_{ijk} , to be approximately normally distributed with mean \hat{y}_{ijk} , the estimated score differential, and standard deviation 11.1779, the standard deviation of the difference between Las Vegas casinos’ closing point spreads and the actual final score in the 2016-17 season.² We use

¹<https://kenpom.com/> (reference behind paywall)

²<https://www.sportsbookreviewsonline.com/scoresoddsarchives/ncaabasketball/ncaabasketballoddsarchives.htm>

these Vegas point spreads to calculate our standard deviation under the assumption that a predicted point spread is an unbiased predictor of a given game’s score differential. This holds in most cases, although we acknowledge that some point spreads (typically involving highly popular teams) may be biased in a way that financially benefits the casino. Using this approximate distribution, we then calculate $P(y_{ijk} > 0)$, the expected probability of team i winning game number k against team j . These expectations estimate the results of a team’s games and overall season (i.e. a win probability of .6 means a team gets .6 wins and .4 losses for that game). Performance metrics are calculated from these and selection score metrics are estimated for this “baseline” simulated season by Model 1, the final model for predicting at-large selection. Then, for a given team, one game against each potential nonconference opponent at each of home, neutral, and away locations is added to the baseline season. Performance metrics and selection probabilities are calculated again for each new season with its one added game.

We examined example teams in three different seasons using this tool. First, we look at two bubble teams from the 2017-18 season: Utah, from the high-major Pac-12, and Middle Tennessee, of the low-major Conference-USA. Both teams missed the NCAA tournament and were invited instead to the NIT, receiving #2 and #3 seeds, respectively. Neither team was among the first four teams left out of the NCAA tournament, indicated by receiving a #1 seed in the NIT.³ Nevertheless, both teams had very good seasons that perhaps could have become NCAA tournament-worthy seasons given more optimal nonconference scheduling. We use our model on these two teams to see what games, if any, they could have scheduled to help them get into the NCAA tournament. We then explore the highest ranking games for both teams to see similarities between them that could motivate better scheduling strategies.

Prior to the 2017-18 season, three teams switched conferences: Wichita State from the MVC to AAC, Valparaiso from the Horizon to MVC, and IUPUI from the Summit to the Horizon. These realignments necessitated some minor adjustments of these conferences’ 2017 conference schedules to better approximate 2018.

In our second case, we revisit the 2009-10 VT team studied by Leman et al. (2014). Those researchers studied the effect of replacing wins against low-ranked opponents with losses against highly-ranked opponents. We study the same effects using our at-large selection model. The changes due to dropping each of VT’s five low-ranking opponents and those from adding a loss against each potential nonconference opponent were calculated. In this case, we used the actual 2009-10 season results as the basis for our simulations, rather than predicted outcomes based on the previous season’s schedule, so as to produce results comparable to the Leman et al. (2014) paper.

³<https://www.nytimes.com/2018/03/12/sports/ncaa-snubs-tournament.html>

Lastly, we take a look at the upcoming 2019-20 season. We select one high-, mid-, and low-major team projected by our models to be on the NCAA tournament bubble. We then investigate the effect of adding a nonconference game to explore which teams could be added to their current nonconference schedule to achieve a potentially better chance of tournament selection. The highest ranked teams on this list would be considered the top teams that each of these college basketball coaches should try to schedule for next season.

Chapter 5

Results

5.1 Phase 1

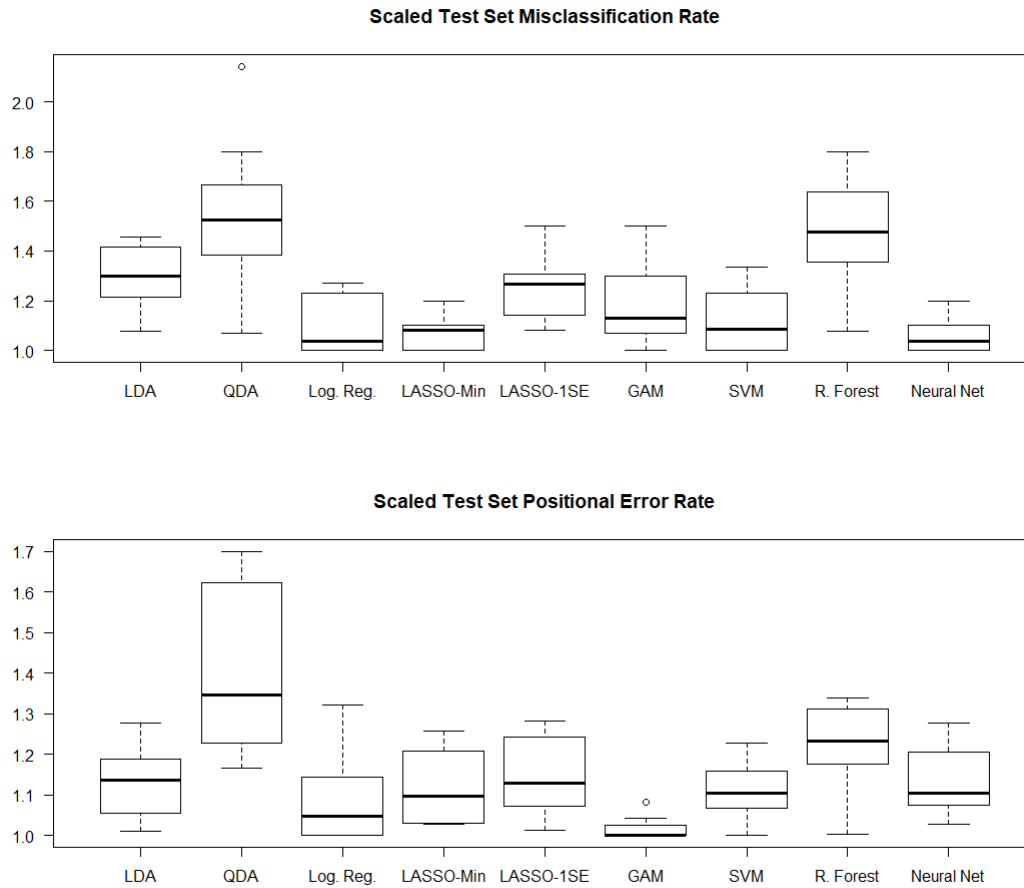
In our final model, BMA selects eight variables as having at least a 25% posterior probability of being included in the true at-large selection model, given our data. These variables, their posterior probabilities, and their parameter estimates from LASSO are found in Table 5.1.

Over the ten training and test splits, the model built using neural networks had the lowest average test set misclassification rate. This model had the lowest test set misclassification rate in five of ten loops. It had a mediocre positional error rate, meaning that the model's misclassifications were farther away from the cut line. Another model that had a very low test set misclassification rate was logistic regression with LASSO, using the minimum-CV value of the λ hyperparameter. While it was only the best in three of ten loops, this model was a very consistent predictor of at-large selection, and it had a slightly better positional error rate than neural networks. Boxplots comparing the models' misclassification and positional error rates, scaled to the lowest value of each loop, are found in Figure 5.1, respectively. Thus, as we proceed to the full dataset, we will use neural networks and LASSO logistic regression, as well as a simple logistic regression as a comparison, to model the 1997-2017 tournament fields and predict 2018 and 2019.

Tuning neural nets over size values of 1, 3, and 5 and decay values of 0.001, 0.01, 0.1, and 0.5, leave-one-year-out-CV finds 1 and 0.1 to be the optimal size and decay values for the final neural networks model. We fit 20 neural nets at each combination and selected the one with the best in-sample error, then selected the combination with the best CV error. For LASSO logistic regression, the value of λ with the lowest leave-one-year-out crossvalidation error was found to be 0.000112. Both models had a misclassification rate of 5.44% in-sample from 1997-2017.

When examining the LASSO logistic regression coefficients, we notice that RPI in both its ranking and decimal forms is highly important, but its LASSO coefficients are both negative. However, it is extremely difficult to interpret these coefficients separately; even after BMA variable selection, there is still a lot of correlation between variables. Even

Figure 5.1: Boxplots of Scaled Score-Based Misclassification and Positional Error Rates, Test Set



Note: observations are scaled to the minimum value in each test split (i.e. minimum for split #1 is 1.0, 2.0 means twice the minimum value, etc.)

Table 5.1: BMA Selected Variables

Variable	Rank	ConfLevel	RawWL	W1_50	W51_100	RoadW	RPI	SOSRank
$P(\beta \neq 0)$	100.0	100.0	100.0	100.0	94.5	49.9	100.0	89.9
Estimate	-0.098	1.458	44.33	0.781	0.240	0.213	-93.59	-0.050

Table 5.2: Correct Classifications, Positional Errors of 2018 and 2019 At-Large Predictions

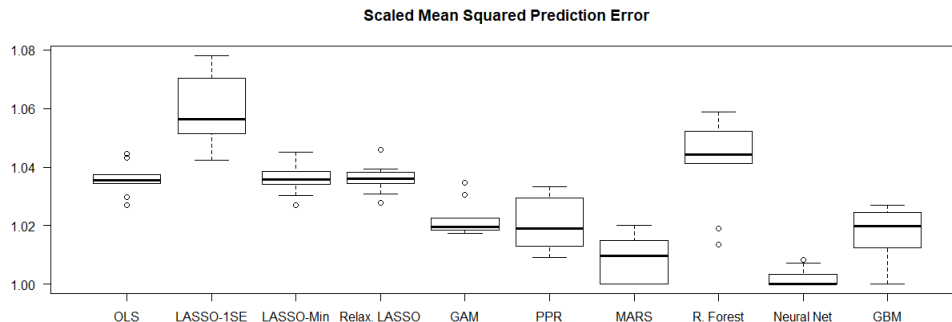
Year	2018		2019				
Metrics Used	T50	Quad	RPI,T50	RPI,Quad	NET,T50	NET,Quad	Mean
Neural Nets	34	32	34	33	33	34	33.33
	4.75	4.625	8.75	4.167	7.333	5.00	5.77
LASSO Reg.	34	32	34	33	33	34	33.33
	5.00	4.375	7.50	4.167	7.667	5.25	5.66
Logistic Reg.	32	32	32	33	32	32	32.17
	4.875	4.125	6.375	4.333	7.375	4.375	5.24

attempting to explain a change in RPI rating and ranking at different rank values would also involve the corresponding change in SOS (affecting SOSRank) or RawWL. This highlights one of the biggest difficulties in interpreting logistic regressions for tournament selection based on nitty gritty variables; almost everything is connected somehow.

One question of interest was to see how the tournament selection committee has utilized the changes to its selection criteria mentioned in 2.1.2. We used our models to predict the 2018 tournament field using both the old system of tiering wins (RPI top 1-50, 51-100, etc.) and the new quadrant system, and we predicted the 2019 field using all four combinations of tiering systems and rankings (RPI, NET). For combinations using the new metrics, we simply replaced the old metrics with the new ones in the fitted model, since NET and RPI rankings and the top 50 and quadrant systems use the same measurement scales. As such, we kept the same parameter values to ascertain if the committee used these new metrics differently than they used the old ones. In all six cases, neural networks and LASSO had the same number of correct classifications (out of 36), and each correctly classified at least as many selections as logistic regression. Though logistic regression had more misclassifications, those mistakes were smaller, on average, than the other two models. In 2018, the models correctly classified fewer at-large teams using the new quadrant system than the old top-50 system. However, the 2019 prediction that used the new NET ranking and quadrant system performed the best of the four combinations. Model 1’s correct classification of 34 of 36 at-large teams in 2019 would have put it in a tie for 98th place out of 195 tournament fields predicted by media and amateur “bracketologists”.¹ The full results are found in Table 5.2.

¹<http://bracketmatrix.com/>

Figure 5.2: Model 2 Predictive Error



Note: observations are scaled to the minimum value in each test split

5.2 Phase 2

Over these ten training and test loops, neural networks once again proved to be the best technique to predict estimated team strength, this time based on mean squared prediction error. This model had the lowest prediction error in five of ten loops and second lowest in another four loops. As Figure 5.2 shows, neural nets built a much better model than the other techniques used. We use neural networks to model the full 2004-2019 dataset and predict the best teams in 2020.

Though neural networks was again the best predictor, its tuning parameters took on very different values from the Model 1 fitting. We tuned over size values of 6, 10, and 14, and decay values of 1, 2, and 3. We once again fit 20 neural nets at each combination and selected the one with the best in-bootstrap error, then selected the combination with the best mean out-of-bag prediction error for the ten bootstraps. The best combination of size and decay for the full dataset was a size of 14 and decay of 3. The in-sample mean squared error of this model, when fitting on the full dataset, was 35.8899 points per 100 possessions, and the mean squared prediction error for the ten training/test splits was 35.9775.

This neural networks model predicted the AdjEM ratings of all 353 Division I teams for the 2019-20 season. The top 20 teams in predicted AdjEM are listed in Table 5.3, alongside those April 23 predictions of Andy Katz, a NCAA.com college basketball writer, as found in his “Power 36” column.² Sixteen of the top 20 teams in predicted AdjEM are found in Katz’s prediction of college basketball’s best 36 teams. However, only time will tell whether our final Model 2 will have given fairly reasonable results. The 2018-19 predictions, as seen

²<https://www.ncaa.com/news/basketball-men/article/2019-04-09/early-college-basketball-rankings-michigan-state-leads-andy>

Table 5.3: Top 20 Teams in Predicted AdjEM for 2019-20

School	Conference	Pred AdjEM	Pred Rank	Power 36 Rank
Michigan State*	Big Ten	26.20	1	1
Duke*	ACC	23.37	2	8
Michigan	Big Ten	22.64	3	7
Kentucky*	SEC	22.51	4	2
Villanova*	Big East	20.03	5	27
Purdue	Big Ten	19.83	6	34
St. Mary's (CA)*	WCC	19.59	7	NR
Virginia	ACC	19.22	8	6
North Carolina St.	ACC	19.01	9	NR
Providence	Big East	18.61	10	NR
Texas Tech*	Big 12	18.17	11	10
Houston*	American	17.69	12	30
Gonzaga	WCC	17.15	13	18
Wisconsin	Big Ten	16.82	14	Hon. Men.
New Mexico St.*	WAC	16.79	15	NR
Arizona*	Pac-12	16.39	16	26
Maryland	Big Ten	16.31	17	22
Auburn	SEC	15.87	18	4
Baylor	Big 12	15.78	19	29
Va. Commonwealth*	Atlantic-10	15.37	20	16

* denotes the highest ranking team in its conference by predicted AdjEM.

in Table 5.4, lined up very closely with Katz's earliest Power 36 rankings (6/4/18³) and was more accurate in predicting teams' final AdjEM ranking (average error of 11.8 places vs. 12.6 among predicted top 20 teams).

5.3 Phase 3

5.3.1 2018 Bubble

To establish a baseline for comparison, we first simulated the 2017-18 season, using the 2016-17 season as an approximation of the 2017-18 games and our AdjEM predictions from Model 2 to generate the expected number of wins for each team in each game. Then, we calculated the explanatory variables from Model 1 and used them to predict the at-large berths for the 2018 tournament. Teams that won their conference tournaments in 2018 were assumed to have done so in these simulations, so they were removed from the pool of at-large candidates.

³<https://www.ncaa.com/news/basketball-men/article/2018-06-04/andy-katzs-power-36-look-ahead-2018-19-season-post-nba-draft>

Table 5.4: Top 20 Teams in Predicted AdjEM for 2018-19

School	Conf	Pred Rtg	Final Rtg	Pred Rk	Power 36	Final Rk
Gonzaga	WCC	24.87	32.85	1	1	2
Virginia	ACC	24.30	34.22	2	3	1
Michigan St.	Big Ten	23.11	30.81	3	15	3
Duke	ACC	22.56	30.62	4	6	4
No. Carolina	ACC	22.49	27.69	5	8	7
Tennessee	SEC	21.72	26.24	6	4	10
Kentucky	SEC	21.64	27.57	7	7	8
Cincinnati	American	21.56	17.50	8	35	29
Villanova	Big East	21.36	17.33	9	10	30
Nevada	MWC	20.44	18.18	10	5	27
Kansas	Big 12	20.04	21.57	11	2	17
Mississippi St.	SEC	19.85	20.04	12	24	21
Florida	SEC	19.66	18.30	13	32	26
Michigan	Big Ten	19.35	28.32	14	17	6
West Virginia	Big 12	19.14	7.15	15	21	96
Syracuse	ACC	18.69	15.13	16	13	39
Purdue	Big Ten	18.01	26.81	17	20	9
Florida State	ACC	17.89	22.39	18	18	14
Clemson	ACC	17.59	16.14	19	19	36
Kansas St.	Big 12	17.56	20.06	20	9	20

However, our combination of Models 1 and 2 did not result in an accurate prediction of the 2018 tournament field. Fourteen of the 36 at-large berths were misclassified. The average positional error of these misclassifications was 18.36 places, including the massive misclassification of North Carolina State, an eventual at-large selection who was projected by our model to have the 104th best tournament profile going into the season. Though this tournament projection using information known as of April 30, 2017 was not very accurate in predicting the 2018 tournament field, it was superior to other simple potential prognostication methods, such as using the previous year’s field (20 misclassifications), or the previous year’s AdjEM ratings (18 misclassifications). Our projection was worse than the “way too early” bracket prediction of CBS’s Jerry Palm, who misclassified 12 at-large selections in July 2017.⁴

In our prediction of this simulated season’s at-large selection, the Utah Utes ranked 42nd (among eligible teams) in the selection scoring metric, six places away from the at-large cutoff $m_{2018} = 36$, and 54th in predicted AdjEM rating. In the real 2017-18 season, the Utes largely met these projections, finishing 42nd in scoring metric and 58th in AdjEM,

⁴<https://www.cbssports.com/college-basketball/news/too-early-for-bracketology-dont-be-ridiculous-and-enjoy-end-of-july-madness/>

Table 5.5: 2017-18 Utah Nonconference Schedule

Opponent	Location	Result	Score Change	Position Change
Prairie View A&M*	Home	W, 83-62	0.0048	+1
Mississippi Valley St.	Home	W, 91-51	0.0020	+1
Missouri	Home	W, 77-59	0.0087	+2
Mississippi*	Neutral	W, 83-74	0.0032	-1
UNLV*	Neutral	L, 58-85	0.0020	0
Eastern Washington*	Home	W, 85-69	0.0062	+2
Hawaii	Home	W, 80-60	0.0048	+2
Butler	Away	L, 69-81	0.0030	+1
Utah State [^]	Home	W, 77-67	0.0049	+2
Brigham Young	Away	L, 65-77	0.0077	+2
Northwestern St.	Home	W, 84-62	0.0009	+1

* MGM Grand Main Event game

[^] Game played in Salt Lake City, UT, but not at Utah's home arena. S-R labels this a home game, but Utah website labels it a neutral game

before receiving a #2 seed in the NIT and advancing to the championship game. Still, we can look at Utah's 2017-18 nonconference schedule to find the games they theoretically could have played that would best improve their chances of selection and to see how their actual schedule compares to those optimal games.

Utah played 29 regular season games in 2017-18, including 11 nonconference games. Per NCAA rules, they could have played up to two more games, due to their participation in the MGM Grand Main Event tournament. Utah won all of their nonconference home games, but went 1-3 in road and neutral games. The Utes' full nonconference schedule is shown in Table 5.5.

Given Utah's baseline score metric of 0.9759 and ranking of 42, we then simulate additional seasons using the following procedure:

1. Add one game against an opponent at a neutral site, calculating each team's win probability as described in Section 4.3.
2. Recalculate performance metrics, calculate selection score metric and associated position
3. Compute difference in score metric and position between new season and baseline season.

We repeat this process for home and away games against each opponent for all schools outside of Utah's conference, the Pac-12. The value of repeating each of the games against Utah's actual opponents is indicated in the last two columns of Table 5.5. The top fifteen possible games (ranked first by change in position, then change in score) are listed in Table 5.6.

Table 5.6: Optimal Games to Add to 2017-18 Utah Nonconference Schedule

Opponent	Conference	Location	Score Change	Position Change
Bucknell	Patriot	Home	0.0219	+6
Middle Tennessee	C-USA	Home	0.0215	+6
Vermont	Am. East	Home	0.0213	+6
St. Bonaventure	Atlantic-10	Home	0.0205	+6
UT-Arlington	Sun Belt	Home	0.0201	+5
Princeton	Ivy	Home	0.0199	+5
Bucknell	Patriot	Neutral	0.0186	+5
SMU	American	Home	0.0186	+5
Maryland	Big Ten	Home	0.0180	+5
Vermont	Am. East	Neutral	0.0179	+5
Minnesota	Big Ten	Home	0.0178	+5
Houston	American	Home	0.0180	+4
Gonzaga	WCC	Home	0.0178	+4
St. Mary's (CA)	WCC	Home	0.0176	+4
Cincinnati	American	Home	0.0174	+4

Note: +6 and above needed to meet $m_{2018} = 36$ threshold

Four potential games, all at home against low or mid-major schools, could have given Utah the estimated increase in score ranking needed to reach the 2018 NCAA tournament. Seven more games could have increased the Utes' ranking by an estimated five places, and another seven had an associated increase of four places. Among the existing games on their nonconference schedule, repeating the games against Missouri, Eastern Washington, Hawaii, Utah State, and BYU were associated with an increase of two. We cannot interpret these as the effect of adding these games to the original schedule, but it gives us a good sense of which games were the most beneficial to schedule. Note the decrease in ranking from playing Mississippi despite a positive change in score; because RPI is so heavily based on opponents' performance, adding a game can affect other teams' score metrics in unpredictable ways.

Figure 5.3 shows the changes in score and position after adding neutral, home, and away games against each opponent. Adding a home game seems to be more beneficial to Utah than adding an away game or even a neutral game, particularly in terms of the score metric. No potentially added home game could have lowered Utah's estimated ranking. There even seems to be a trend in the skewness of the distributions for the three locations. Home games are positively skewed, potentially indicating that scheduling home games may be a low-risk, high-reward scheduling strategy, while road games are negatively skewed. These trends ought to be confirmed by looking at other teams.

The second most optimal game Utah could have played was against Middle Tennessee (MT), the other 2018 example we will examine. The 2017-18 Blue Raiders were coming

Figure 5.3: Distributions of Score and Position Changes for 2017-18 Utah by Game Location

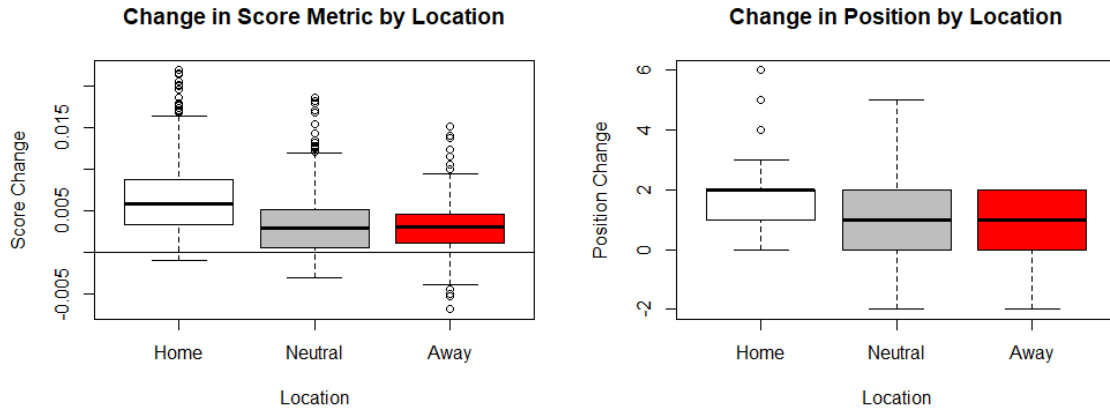


Table 5.7: 2017-18 Middle Tennessee Nonconference Schedule

Opponent	Location	Result	Score Change	Position Change
Trevecca Nazarene	Home	W, 104-52	Non D-I	Non D-I
Murray State	Away	W, 72-67	0.0069	0
Belmont	Home	L, 63-69	0.1379	0
Tennessee State	Away	W, 75-65	-0.0085	0
Fla. Gulf Coast	Home	W, 85-72	0.0655	0
Fla. Gulf Coast	Away	W, 81-76	-0.0146	0
Vanderbilt	Away	W, 66-63	-0.0296	0
Mississippi	Home	W, 77-58	0.0597	0
Auburn	Neutral	L, 70-76	-0.0660	-1
Princeton*	Neutral	W, 69-67	0.0902	0
USC*	Neutral	L, 84-89	0.0320	0
Miami (FL)*	Neutral	L, 81-84	0.0297	0

* Diamond Head Classic

off of two straight NCAA tournament appearances in which they won their first round game. However, their tournament chances were heavily damaged when they lost in the C-USA conference tournament, and they eventually settled for a #3 seed in the NIT. In our baseline season, MT ranked #54 in selection position, with a score metric of 0.5100, and #60 in predicted AdjEM. These predictions fell well short of MT's actual results, where they finished #37 in selection score ranking (the first team left out according to our model) and #45 in final AdjEM rating.

MT, a member of the low-major C-USA, played 11 nonconference games against D-I opponents. Unlike high-major Utah, MT played only three home games, instead playing four games each at neutral and away sites. While they finished 24-7, MT only beat one NCAA tournament team (Murray State) in five tries. The full MT nonconference schedule is found in Table 5.7.

Table 5.8: Optimal Games to Add to 2017-18 Middle Tennessee Nonconference Schedule

Opponent	Conference	Location	Score Change	Position Change
UT-Arlington	Sun Belt	Home	0.2390	+2
Bucknell	Patriot	Home	0.2369	+2
St. Bonaventure	Atlantic-10	Home	0.2302	+2
Vermont	Am. East	Home	0.2275	+2
Princeton	Ivy	Home	0.2211	+2
Utah	Pac-12	Home	0.2106	+1
Oregon	Pac-12	Home	0.2028	+1
Houston	American	Home	0.2011	+1
Iowa	Big Ten	Home	0.1973	+1
Michigan	Big Ten	Home	0.1951	+1

Note: +18 and above needed to meet $m_{2018} = 36$ threshold

MT is in a much more difficult situation when trying to formulate their schedule. From the perspective of college basketball structure, it is harder for low-major teams to schedule lots of home games, which, as we now have seen in two examples, appear to be much more helpful than neutral and road games. From a selection model perspective, MT has a low score metric, due to having a low predicted AdjEM (thus fewer expected wins) and a lack of quality opponents on their C-USA schedule. As a result, even the most beneficial added games do little to help MT’s tournament prospects. The ten most beneficial games for MT are found in Table 5.8.

MT is too far separated from the teams in front of them in the baseline season to be able to secure themselves an at-large bid, even with other predicted bubble teams like Utah, Oregon, Houston, and Iowa (along with eventual bubble team St. Bonaventure) on this list of top games. The most optimal games would only improve their standing by an estimated two positions. MT even played Princeton in the nonconference (albeit in a neutral site MTE, where they had no choice of opponent), which is the fifth most optimal home opponent. Utah is the sixth most optimal opponent, but even this highlights the difficulty that MT faces; the Utes would gain an estimated two places from playing at MT, while the hosts would only gain one. Though they eventually outperformed our preseason predictions, there is nothing that MT could have done from a scheduling standpoint, that could have substantially improved their chances of an at-large selection.

5.3.2 2010 Virginia Tech

We revisit the case of the 2009-10 Virginia Tech Hokies. They played five nonconference games against teams ranked 285 or below in RPI, the details of which can be found in Table 5.9. Leman et al. (2014) study the impact of removing anywhere from one to five of these games and replacing them with losses against teams in the RPI top 25, 26-50, 51-100, and

Table 5.9: Sub-285 Opponents on 2010 Virginia Tech Schedule

Original Game			Drop Win vs. Opponent		Best Replacement	
Opponent	Conf	RPI	Score Chg	Rank Chg	Score Chg	Rank Chg
Char. So.	B. South	285	0.0012	0	0.0002	0
Longwood	Ind.	286	0.0010	0	0.0012	0
UMBC	Am. East	333	0.0016	+1	0.0027	+2
NCCU	Ind.	347	0.0007	+1	0.0026	+3
VMI	B. South	310	0.0019	+1	0.0018	0

Best replacement game was Kansas (Home)

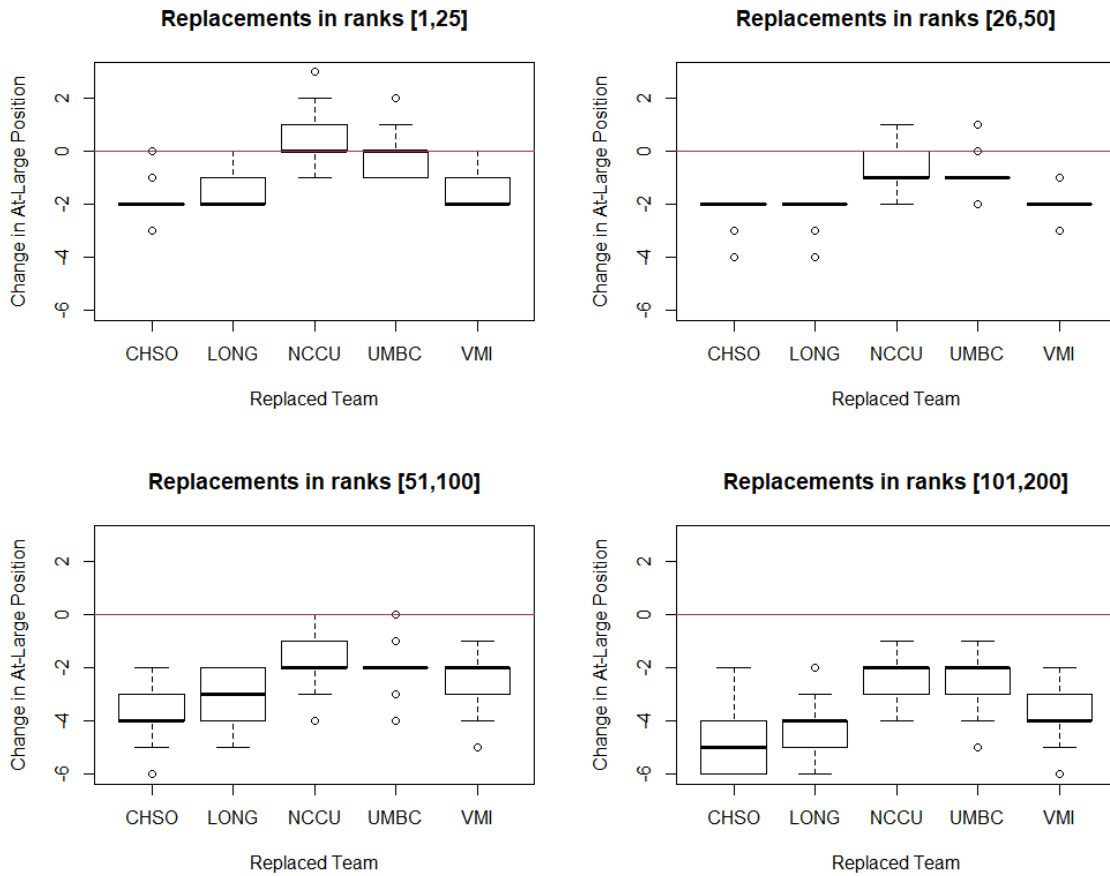
101-200. However, they did not report the impact of replacing each of these low-ranked games on an individual game basis. In our case study, we used the actual 2009-10 season results as our baseline season for the simulations. For each of the five low-ranked opponents, we removed the win against that team and calculated the change in score and score ranking associated with adding a loss to each potential nonconference opponent at all three locations (or adding no game at all).

According to our model, VT was actually predicted to make the NCAA tournament in 2010 as the 34th and final at-large team in the field (one of only two Model 1 misclassifications in 2010). Despite our model having a more optimistic viewpoint of VT than the 2010 committee, we can still make comparisons between our model's initial prediction and predictions based on replacement games, since we are looking to see if there is any improvement associated with replacing a low-ranked win with a high-ranked loss. These changes corresponding to each replacement game can be interpreted as the estimated change in at-large position associated with replacing a win against a particular team with a loss against the replacement team.

In all five cases, removing VT's game against a low-ranked opponent and replacing it with no D-I game at all increased their score metric, suggesting that playing these games at all was detrimental to VT's tournament hopes. Yet in two cases, Charleston Southern and Longwood, these incremental improvements in score were not large enough to lead to a higher ranked tournament profile. We found a home loss to Kansas to be the optimal replacement loss for VT in all five cases. However, not even this best replacement improved tournament standing for three of the five cases. Only one replacement game for Charleston Southern and five for Longwood, the two teams above 300 in RPI ranking, even managed to keep VT's estimated ranking the same. In almost all cases, erasing a sure win and replacing it with a sure loss against even an excellent team does not help VT, which differs from the conclusion of Leman et al.

As Figure 5.4 shows, replacing a win against a low-ranked team with a loss against a team in the RPI top 25 is, on average, better for VT's tournament prospects than replacements in lower tiers. This mostly aligns with the findings of the previous researchers. However,

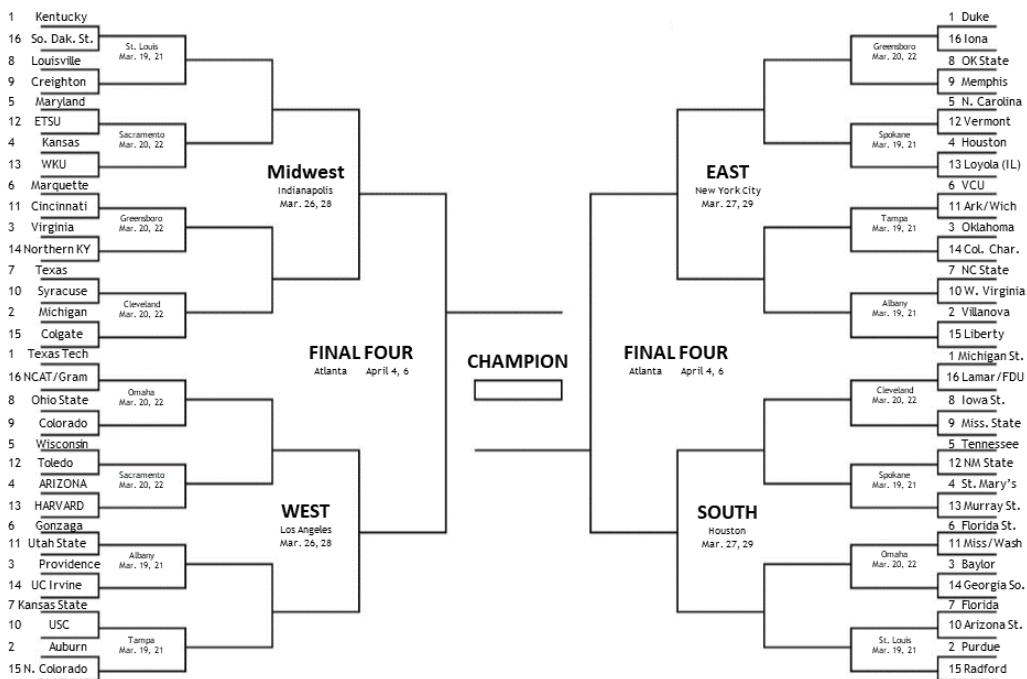
Figure 5.4: 2009-10 Virginia Tech At-Large Positional Changes by Replaced Opponent, Replacement's Tier



Note: Position changes reflect replacing win against Replaced Team with loss against team in a given RPI tier

this is not to say that even highly ranked replacements are of much benefit. For only two teams, North Carolina Central and UMBC, were there any replacement losses that were more beneficial than a) beating the low-ranked team or b) not even playing the game. For VMI, only the case where their game is left unplayed is associated with a positive ranking change for VT. This differs from the conclusion of Leman et al., whose model predicted a clear improvement in at-large selection probability for VT when replacing a low-ranked win with a single replacement in the [1,25] and [26,50] range. Again, both of these predictions are the worst case scenario for any replacement game, so any win in such a replacement game would likely improve the estimated chances of at-large selection.

Figure 5.5: Predicted 2020 NCAA Tournament Bracket



5.3.3 2020 Bubble

Finally, we look at the upcoming 2019-20 season using this tool. Using the same methods as in Section 5.3.1, we are able to predict the at-large recipients for the 2020 NCAA tournament. Unlike 2018, we do not know who will win each conference tournament. As we study these predictions, we assume for simplicity's sake that the best team, by predicted AdjEM, will win each conference tournament. We do not remove these teams from the dataset entirely, as we will more closely examine such teams. Based on our simple assumption, we find our expected cutoff $\hat{m}_{2020} = 47$, which includes the 36 available at-large berths and 11 conference winners ranked above the 36th most qualified at-large contender. Our 2020 "bracketology" prediction, based on our conference tournament assumption and the NCAA selection committee's bracketing guidelines, is found in Figure 5.5.

We look more closely at three teams close to the cutoff:

- Georgia (SEC), ranked #53 by score and #63 by predicted AdjEM
- Connecticut (American), ranked #48 by score and #55 by predicted AdjEM
- New Mexico State (WAC), ranked #58 by score and #15 by predicted AdjEM

The best ten and worst five games that each team can play, according to added score and score ranking, are found in Table 5.10.

Table 5.10: Most, Least Optimal Games for 2019-20 Georgia, UConn, NMSU

Georgia			Connecticut			New Mexico State		
Opponent	Site	Rk Chg	Opponent	Site	Rk Chg	Opponent	Site	Rk Chg
Toledo	H	+7	Syracuse	H	+6	Murray St.	H	+11
Murray St.	H	+7	NMSU	H	+5	VCU	H	+11
R. Island	H	+7	C. Mich.	H	+5	Gonzaga	H	+11
Belmont	H	+7	R. Island	H	+5	Vermont	H	+11
Furman	H	+7	Murray St.	H	+5	Belmont	H	+11
Utah St.	H	+7	Belmont	H	+5	Houston	H	+10
NMSU	H	+6	Toledo	H	+5	Wofford	H	+10
Memphis	H	+6	Davidson	H	+5	Duke	H	+9
Davidson	H	+6	Vermont	H	+5	St. Mary's	H	+9
ETSU	H	+6	Washington	H	+5	Michigan	H	+9
Indiana	A	-1	MVSU	N	-1	Idaho St.	H	-6
Inc. Word	N	-1	Nebraska	A	-1	Idaho	N	-6
MVSU	N	-1	Chicago St.	H	-2	Idaho St.	N	-6
Chicago St.	N	-1	Nebraska	N	-2	UMES	H	-7
UMES	N	-1	Chicago St.	N	-2	UMES	N	-7

Georgia needs +6 or better, UConn needs +1 or better, NMSU needs +11 or better

Though the Georgia Bulldogs find themselves outside the NCAA tournament field according to our preseason projections, they are not too far removed from an at-large berth. Sixty different opponent-location combinations could push Georgia over the projected cut-off. In fact, one such game, a home game against projected Sun Belt champion Georgia Southern (0.0064, +6), is on Georgia's 2019-20 schedule (released 7/9/2019), with another two games on the road against optimal home opponents Memphis and Arizona State.⁵ Again, these are interpreted as the estimated effect of adding a game to Georgia's approximated (2019) schedule, but this list gives a rough estimate of which teams can be the most beneficial. It seems that Georgia coach Tom Crean may already be utilizing some good scheduling strategies.

According to our preseason projections, the Connecticut (UConn) Huskies are the first team out of the tournament field. Thus, UConn would seem to be in a position where one or two smart scheduling choices could put them safely into the field. Our tool says that only 354 of the 1023 possible nonconference games would not put UConn into the tournament field. There are few games that the Huskies could schedule that would hurt their estimated chances of reaching the NCAA tournament. However, they happen to have already scheduled one of these, a neutral site game against Indiana (-0.0009, -1).⁶

⁵<https://georgiadogs.com/news/2019/7/9/mens-basketball-bulldogs-non-conference-schedule-set.aspx>

⁶<https://uconnhuskies.com/news/2019/7/29/mens-basketball-announces-2019-20-non-conference-schedule.aspx>

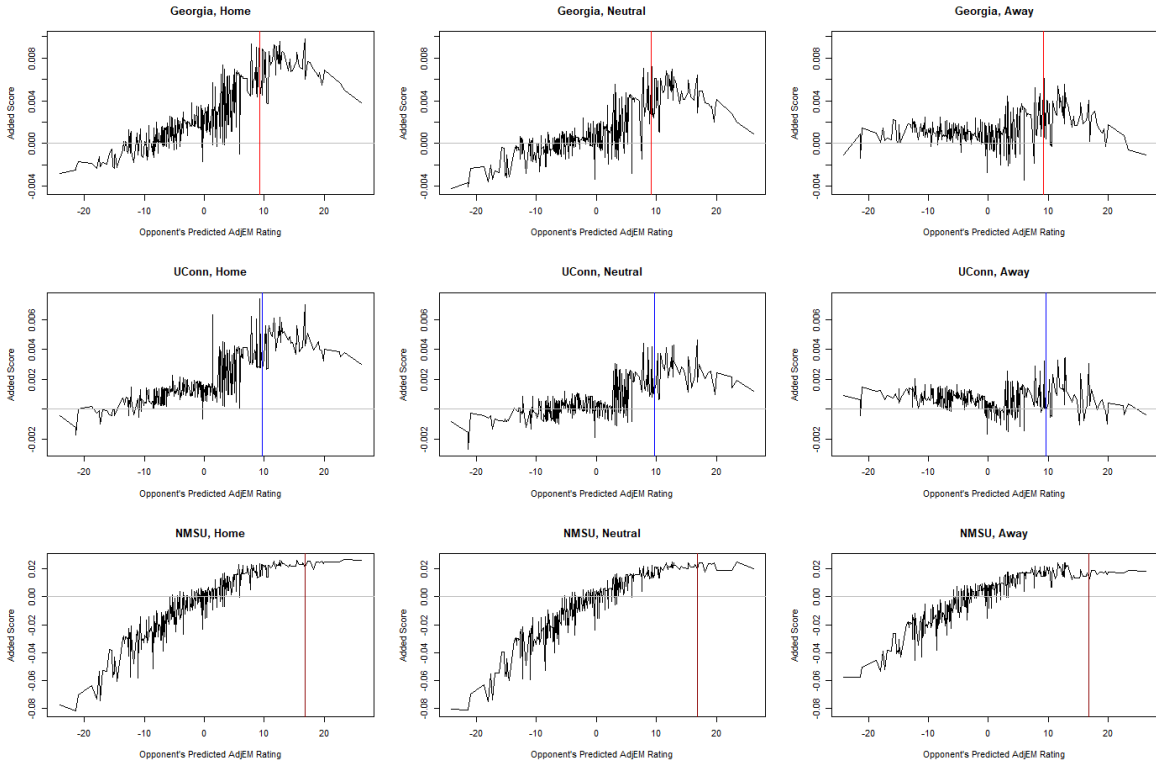
The New Mexico State (NMSU) Aggies are in a more difficult position than the previous two teams, due mostly to their membership in the low-major WAC. Their estimated ranking as the 15th best team in the country according to Model 2 but only the 58th best tournament profile shows that their weak conference schedule is very detrimental to their tournament hopes. However, this leaves NMSU in a position where their nonconference scheduling strategy has an outsized effect on their eventual selection potential. NMSU, as a top 15 team, has the potential to beat just about any team in the country, which raises the ceiling on the benefit of scheduling even dominant opponents. Thus, good scheduling choices will greatly improve their at-large selection chances, with five individual games even putting them over the projected cutoff. Bad scheduling choices, on the other hand, could be devastating; 120 possible games are associated with a decrease of five or more ranking spots. NMSU's range of estimated positional changes is 18, which is far higher than any of the other four case studies. NMSU finds itself in a position where it needs to schedule its nonconference games optimally to give themselves a chance at an at-large berth.

Schedulers may be interested in knowing more about the relationship between an opponent's talent, a given team's talent, and the improvement in tournament chances. According to the plots in Figure 5.6, there seems to be a positive relationship between the projected talent of an opponent and estimated increase in the scoring metric. However, it may not be a monotonic relationship; for UConn and Georgia, the highest points of added score occur against teams predicted to be a little bit better than them, only for those gains to diminish as opponents become tougher. This trend is most present in potential home and neutral games but diminishes in road games. This does not seem to be the case for New Mexico State, which is a much better team by predicted AdjEM and thus is more likely to beat a potential opponent. In terms of talent, the most beneficial opponents to play are pretty talented, but not so much more talented that they are difficult to defeat.

Perhaps our most interesting findings concern the estimated effect of scheduling teams from different tiers of conferences. Figure 5.7 shows the distribution of both score changes and the resulting ranking changes for each case study. From these, we see a noticeable effect of conference level on score change for all three case study teams, as well as an effect on position change for NMSU and Georgia. NMSU seems to more affected by an opponent's conference than UConn or Georgia. On average, it is more beneficial for a team to schedule a high or mid-major opponent than a low-major team. However, these boxplots paint an incomplete picture of the effect of a potential opponent's conference on at-large selection.

Accounting for the predicted RPI ranking of a potential opponent, by comparison, leads to an entirely different conclusion. As Figure 5.8 shows, high-major opponents tend to be less beneficial than low- or mid-major opponents with a similar RPI ranking. This is an interesting case of Simpson's Paradox. Teams with good RPI rankings are more beneficial to play than teams with poor rankings. High-major teams tend to have much higher rankings, on average, than low- or mid-majors. Thus, the average high-major team is more beneficial

Figure 5.6: Estimated Added Score by Predicted AdjEM



Note: Colored vertical line indicates given team's predicted AdjEM rating

Figure 5.7: Change in Score and Ranking by Opponent's Conference Level

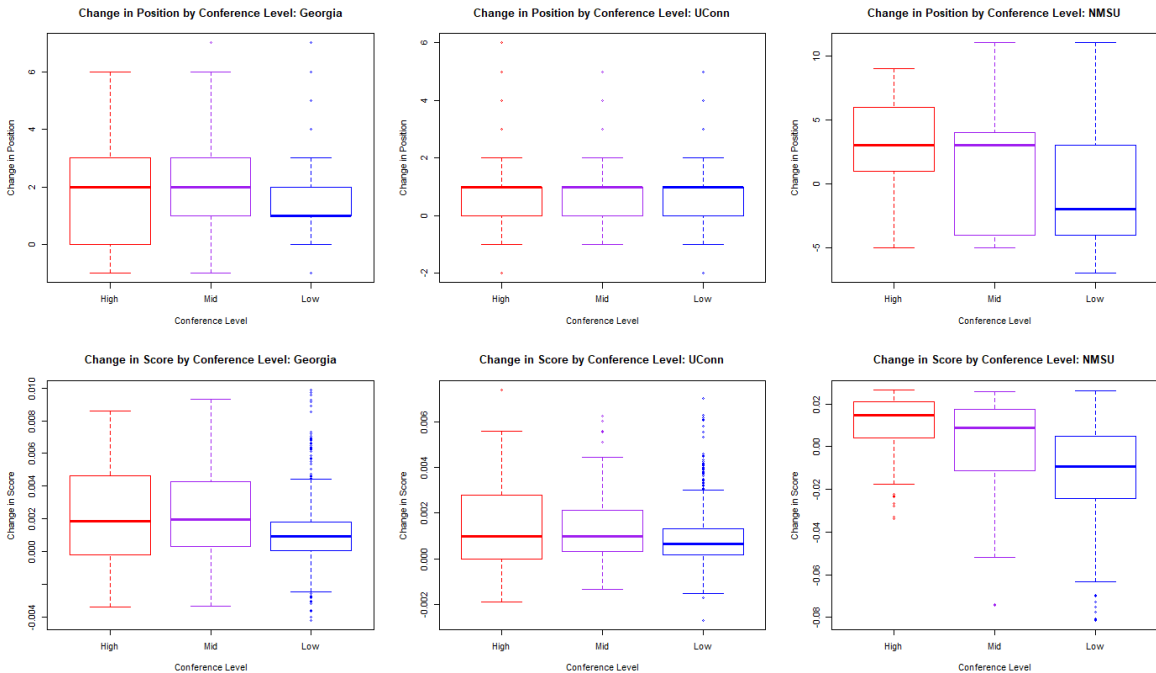
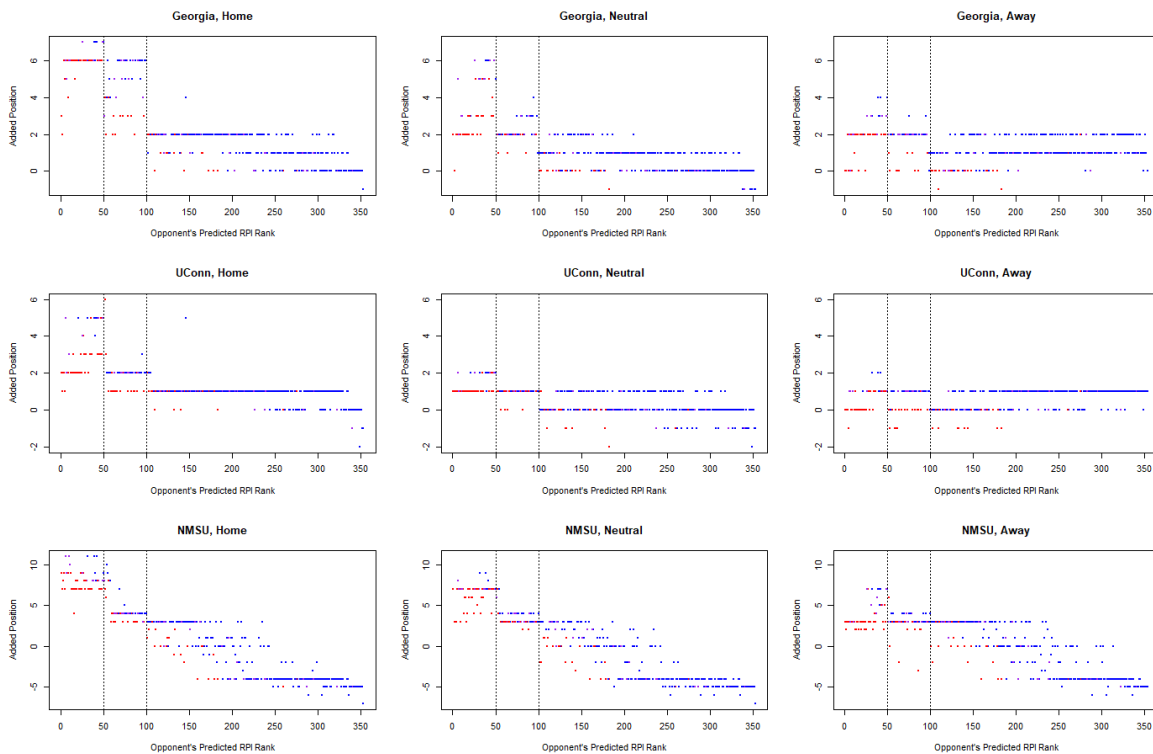


Figure 5.8: Estimated Added Position by Opponent's Predicted RPI Ranking



Red: High-Major, Purple: Mid-Major, Blue: Low-Major
 Vertical lines at Top 50 and 100 cutoffs

to play than the average low- or mid-major. But among teams with good RPI rankings, low- and mid-major teams are preferable to high-majors.

Another interesting phenomenon is shown by Figure 5.8. For potential opponents in the top 50, as an opponent's RPI ranking increases, the change in position associated with playing them tends to rise as well. However, this benefit discontinuously drops at around the 50th ranked team, rises again until about 100, then drops again. The increases come from the decreasing difficulty of the opponent (lower ranked teams tend to be easier to beat), and the drops at 50 and 100 are due to the large benefit of beating a team in the RPI top 50 and the smaller benefit of beating a team ranked 51-100. Though the NCAA has adopted the new quadrant system, we would expect a similar trend to exist, but with the discontinuities corresponding to the cutoffs for Quadrant 1 and Quadrant 2 wins. For opponents outside the top 100, there is an inconsistent change in position depending on the team in question and location. For Georgia and UConn, there is almost no change in added at-large position with worsening RPI ranking until a drop for opponents in the RPI bottom 25. This does not hold for NMSU; their potential opponents show a very consistent decrease in added at-large position among teams ranked outside the RPI top 100 at all locations.

Chapter 6

Discussion

Teams from low and mid-major conferences are at a distinct disadvantage when it comes to tournament selection. The importance of the variables in the logistic regression show that there are explicit (Conference Level) and implicit (Q1, Q2 wins, which are much easier to obtain in high major conference play) biases against smaller schools. The large effect of conference affiliation is especially apparent in the 2020 projections, where the projected 15th best team, NMSU, is well below the estimated cutoff line because of the schedule they play. However, the presence of WL% and road wins as important predictors give some hope to these teams, as these are categories where small conference bubble teams tend to have an advantage over high-major bubble teams.

Though the 2018-19 predictions of team strength largely aligned with media projections, the 2019-20 predictions differ much more from the media. Some of this difference may be due to influence from the 2018-19 season; two predicted top 20 teams missing from Katz's rankings, NC State and Providence, missed the 2019 tournament, and the other two, St. Mary's and New Mexico St., won their conference's automatic bid, but otherwise were unlikely to reach the tournament. On the other hand, the prediction of NC State as the ninth best team highlights a model shortcoming; NC State had two top 100 freshmen who were predicted, as of April 30, to commit to them, but they instead chose to play for North Carolina (a "marquee" team notably absent from these rankings). The uncertainty in freshman recruiting and the recent spike in both player transfers and underclassmen declaring for the NBA draft only to return to school has made it more difficult to predict the 2020 season, and likely beyond.

Using our process, we recommend these guidelines to college basketball coaching staffs as they build their schedule for the upcoming season:

1. Play nonconference games at home whenever possible
2. Playing quality low- and mid-major opponents is more beneficial than playing high-major teams
3. Play teams at the lower ends of Quadrants 1 and 2

We find from Model 3 that teams that can schedule quality opponents at home are at a much greater advantage come Selection Sunday than those who are forced to play these teams on the road. Much of this is due to the inherent advantage that comes with playing home games, accounted for in our model by the 3.75 point home court advantage. Even though RPI gives less weight to home wins than road wins, the greatly increased probability of winning at home overcomes the disadvantage in the RPI calculation. This is not necessarily ground-breaking news; coaches in large and small conferences alike have been using that knowledge to manipulate the RPI to their advantage for years.

The paradoxical effect of an opponent's conference affiliation on at-large selection is another important finding of this project. The average high-major team is more beneficial to play than the average low- or mid-major, but highly ranked high-majors are less beneficial than similarly ranked low- or mid-majors. This paradox may be present because high-majors of a certain RPI ranking might be more talented, according to our predicted AdjEM, than similarly ranked teams from smaller conferences, and thus provide fewer expected wins. These findings align with the work of Sanders (2007) and Paul and Wilson (2012), who found evidence of a bias in favor of small conference schools in the RPI formula.

Teams at the lower ends of Quadrants 1 and 2 are easier to beat than teams at the higher ends, and thus playing them provides more expected Quadrant 1 and Quadrant 2 wins. However, it is a gamble to play these teams; if a team plays a neutral site game versus an opponent with a predicted RPI (or NET) ranking of 99, but that opponent finishes ranked 102, the benefit of that game is highly diminished even though the opponent only slightly underperformed. But for a team that needs a good nonconference schedule to obtain an at-large bid (like low-major NMSU), scheduling a low-ranked Quadrant 1 game is a gamble worth taking.

One shortcoming of our process lies in the approximation of next season's schedule using the previous season's schedule. While it is difficult to work around this, as the primary purpose of this tool is to build the next year's schedule, better approximations may exist. As it stands, the true effect of scheduling an opponent may be distorted by the differences between the two schedules. For example, a team that played a schedule with many quality teams in one season but schedules easier games the next season may appear more beneficial to opponents than they really are.

The results from Model 3 may not have actionable in real-world college basketball scheduling. For example, our model says that a home game against Duke would be the eighth most optimal game that NMSU could schedule. It is highly unlikely that Duke, a perennial powerhouse with several national championships in its storied basketball history, will travel roughly 1,500 miles to Las Cruces, NM to play a WAC school. Coaches who use this tool, especially at smaller schools, must balance the optimality of opponents with the reality that most of these opponents are not very interested in playing them.

Finally, the selection committee's adoption of the quadrant system in 2017-18 and the NET ranking metric in 2018-19 represent a major change in the tournament selection process. This raises two issues. First, we have a very small sample of seasons where the committee has used these measures, and thus we have a very incomplete picture of how the selection committee will use them. So far, our at-large selection model, when using both old and new metrics, has done a pretty good job of predicting at-large bids in the 2018 and 2019 tournaments. Caution should still be taken when using this prediction model, as the committee may adjust their selection criteria as they continue to use the new metrics. Second, we do not know the new NET formula, and thus we cannot use our model to optimize a team's NET ranking or associated metrics. RPI and top 50 tiering correctly classified as many at-large selections in 2019 as NET and quadrant tiering, so our model may yet be very useful to college basketball coaching staffs. But without the NET formula, our process should be considered more of a template to study optimal nonconference scheduling.

Chapter 7

Future Work

As mentioned above, the tournament selection process has undergone big changes in the last two years. We do not have a very good understanding of how the selection committee will choose to implement the quadrant system and NET, nor how its eventual implementation will compare to the old RPI and top-50 tiering system. Unlike the RPI, the exact NET formula remains a secret. Future work could identify the impact of these changes on committee behavior, and finding the NET formula (or a good approximation) would help teams know what criteria they are ranked on.

Some of the biggest errors in the predicted AdjEM ratings came from teams with many transfers. Their contributions in their first season aren't accounted for in this model. Future work could identify how to measure the impact of transfers, particularly those moving between different tiers of Division I. However, many transfer decisions aren't made before the end of April, so these player movements may not be applicable for our purposes anyway.

An interesting takeaway that may be worth further investigation is the influence of geography on these rankings; two of NMSU's top ten games and three of the bottom five are against teams in the Pacific or Mountain time zones. Georgia's best and worst scheduling choices have a similar trend; five of the top ten and two of the bottom five games are in states with SEC schools. Playing regional opponents seems to have a compounding effect (for good or bad), as that regional opponent may have games against other conference members, so the RPI's SOS components are affected not only for the team in question but for several other teams on their schedule. However, further study is needed to why these changes seem to occur and to find if this trend is anything more than anecdotal.

This project gives predicted changes in score and position, but we do not have any procedure to find the variability in these predictions. Models 1 and 2 give associated error rates, but Model 3 as yet does not. Further work could develop a way to assess variability in predicted changes from Model 3, perhaps using deviations from the predicted AdjEM of a team or its opponents in a predicted season or simulating many predicted seasons for a single team by adding noise to the expected score differential in each game.

While this work may be enlightening to followers of college basketball, it is most useful in the hands of a college basketball coaching staff. As such, we would like to develop this work into a user-friendly program that coaching staffs could use as a guide in scheduling their opponents. Developing this into an R Shiny program with the capability of adding or subtracting high school recruits and returning players from the team strength projection data would make this far more useful to college basketball teams. Adding capabilities to account for upcoming nonconference games that are known before the offseason could make a more accurate approximation of next season's schedule.

Bibliography

- [1] Coleman, B. J., DuMond, J. M., & Lynch, A. K., 2010. Evidence of bias in NCAA tournament selection and seeding. *Managerial and Decision Economics*. 31(7), 431–452.
- [2] Coleman, B. J., Dumond, J. M., & Lynch, A. K. 2016. An easily implemented and accurate model for predicting NCAA tournament at-large bids. *Journal of Sports Analytics*, 2(2), 121– 132.
- [3] Coleman B. J., Lynch, A. K. 2001. Identifying the NCAA tournament ‘dance card’. *Interfaces* 31(3): 76–86.
- [4] Harville DA. 2003. The selection or seeding of college basketball or football teams for postseason competition. *Journal of the American Statistical Association* 98(461): 17–27.
- [5] Hastie, Trevor,, Robert Tibshirani, and J. H Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. New York: Springer. Ebook.
- [6] Leman, S. C., House, L., Szarka, J., & Nelson, H., 2014. Life on the bubble: Who’s in and who’s out of march madness? *Journal of Quantitative Analysis in Sports*. 10(3), 315–328.
- [7] Paul, R. J., & Wilson, M., 2012. Political correctness, selection bias, and the NCAA basketball tournament. *Journal of Sports Economics*. 16(2), 201–213.
- [8] Sanders S. 2007. A cheap ticket to the dance: systematic bias in college basketball’s ratings percentage index. *Economics Bulletin* 4(34): 1–7.
- [9] Shapiro, S. L., Drayer, J., Dwyer, B.,& Morse, A. L., 2009. Punching a ticket to the big dance: A critical analysis of at-large selection into the NCAA division I men’s basketball tournament. *Journal of Issues in Intercollegiate Athletics*. 2, 46–63.
- [10] Zimmer T, Kuethe T. 2008. Major conference bias and the NCAA men’s basketball tournament. *Economics Bulletin* 12(17): 1–6.