Real Case Analysis
Analysis of the “Big 4” Metrics in A10 College Basketball Games

Authors: Kelly Ellis, Scott Hoyland, Joseph Patrick, David Weaver

Table of Contents

Summary

Schools with a basketball program that makes it to the NCAA Tournament earn revenue from inclusion and gain exposure to a national audience which can lead to an increase in applications from students. Four factors have been identified as key for basketball success: rebounding percentage, free throw rate, effective field goal percentage, and turnover percentage. In this paper, we investigate what basketball metrics lead to an improvement in these four factors for the winning team over the losing team within George Mason University's (GMU) conference, the Atlantic 10. Models were created based on the LASSO regression, F tests, correlation analysis, and decision trees. Recommendations for improving GMU's likelihood of participating in the NCAA tournament based on our findings are provided and include exploiting home court advantage and recruiting players with better three-point shooting and skills that lead to more assists.

Introduction

A successful basketball program can bring prestige, interest, and money to colleges and universities nationwide. Each spring, the top-ranked 68 Division I teams in the National Collegiate Athletic Association (NCAA) compete for the National Championship in the NCAA tournament. Participating in this competition, popularly known as "March Madness", can bring increased recognition to basketball programs, and in turn, the schools. In 2006, the last time George Mason University (GMU) was featured in the "Final Four" teams of the championship, the tournament had more viewers than the 2006 Super Bowl or National Basketball Association (NBA) finals (Zullo, 2006). This increased attention is not only related to increased revenue, but also increased applications from students; thus, it is of utmost importance for universities with Division 1 basketball programs to make it into the NCAA tournament, especially teams from lower, lesser-known conferences.

GMU is a part of the Atlantic 10 (A10) basketball conference, which is made up of 14 teams that are primarily located in states on the Eastern Seaboard, as well as some in the Midwest. The A10 is a mid-major conference that receives one automatic bid to the NCAA tournament and typically sends 1-3 teams to the NCAA tournament each year. For instance, for the 2021 NCAA tournament the A10 sent two teams - St. Bonaventure and Virginia Commonwealth University (VCU). This is in contrast to teams from major conferences like the Big 10 or Southeastern Conference (SEC), which can send upwards of 8 to 9 teams to the tournament in a year. This difference, based on conference, places much higher importance on conference tournaments for mid-majors because the only teams that may go to the tournament in a season are conference tournament winners, which get automatic bids.

Four factors have been identified as the most important in winning a college basketball game (Oliver, 2011). These factors are rebounding percentage, free throw rate, turnover percentage, and effective field goal percentage, and are commonly referred to as the "big four". Determining the game variables that are associated with these four factors may lead to insights for improving GMU's conference game performances and increase the likelihood of receiving an automatic or at-large bid to the NCAA tournament. *Our primary research objective is to determine the variables that are significantly associated with rebounding percentage, free throw rate, turnover percentage, and effective field goal percentage for A10 basketball games.*

This report consists of a data description section, detailing the data used in analysis. The methodology section describes the statistical methods utilized. The analysis section is divided into five subsections, one for the four factors combined, and one for each of the four factors separately. A conclusion section ends the report and features recommendations for improving the regular season performance of GMU's men's basketball team.

Data Description

The data collected for this analysis is game-by-game data for teams in the Atlantic 10 (A10) conference during regular season games from 2015 to 2020. The data is obtained from the March Men's Basketball Machine Learning Mania 2021 competition from Kaggle.com, a machine learning competition website. Each row includes losing and winning team statistics for each game between two A10 schools. The original data includes box scores from each game for many common basketball metrics. From this data, we calculated advanced statistics for each game that were not included in the original data set.

The metadata included in the original data for each game are the season the game was played, the number of overtimes (if any) in the game, the location of the game (whether the winner was at home, away, or on a neutral court), whether the game was a tournament game, an indicator of the two teams that played in the game, and an indicator of which team won. The game metrics for each of the two teams are separated out by winner and loser and include score, field goals made and attempted, three point field goals made and attempted, free throws attempted and made, offensive rebounds, defensive rebounds, assists, turnovers, steals, blocked shots, and personal fouls. From these basic statistics, advanced statistics are calculated including the number of possessions in the game, a metric for the pace of the game, field goal percentage, three point field goal percentage, free throw percentage, effective field goal percentage, turnover percentage, offensive and defensive rebounding percentages, free throw rate, offensive rating, defensive rating, free throw attempt rate, three point attempt rate, and true shooting percentage.

The final data is split into two datasets, one with the winning and losing team statistics and the other with the differences between the losing and winning team statistics. This is important because this allows us to understand what factors affect overall percentages, such as overall effective field goal percentages, and what factors affect which teams win a certain statistical battle, such as the difference in effective field goal percentage between winning and losing teams.

Methodology

Prior to analyzing each of the four factors, it is important to understand how outcomes in games are related to the "big four" factors; furthermore, it is important to understand the effect size of each of the four factors. To do so, we created a multiple linear regression model with the difference in score as the outcome variable and the differences between the winning and losing teams metrics for each of the "big four" factors.

Many previous studies' main objective is to predict the winner of the basketball game where the outcome variable is a binary variable - win or lose (Zimmerman, 2016). In this case, there are two observations for each game (winner and loser), which creates a dependency across observations. Additionally, the win or lose framework does not easily allow for game specific variables, such as pace, because the variables are common to both teams. Calculating the difference in game metrics between the winning and losing team allows us to have one observation per game, mitigating concerns over dependency across observations. Additionally, we can assess whether a game specific variable, such as pace, increases the difference in the winner's margin on the "big four" factors. Our approach also allows for a meaningful interpretation of the intercept, namely, whether the factor is associated with winning after the inclusion of explanatory variables. Finally, modeling wins as a function of the "big four" factors would likely add very little to the literature as it is widely agreed these factors affect winning. Our methodology allows us to look at issues that have been previously unexamined: how differences in the factors predict differences in the score between the winner and loser and what increases or decreases the winner's margin with regard to each of the factors.

A separate analysis is performed for each "big four" factor, where we take a broad-to-narrow approach in each analysis. Within each analysis a broad descriptive analysis is performed on each variable to understand the univariate distributions of each factor, as well as a multivariate analysis to understand relationships between each variable.

Following the descriptive analysis, various techniques are used to build models for linear regression. The usage of linear regression follows from the calculation of the differences in each of the game metrics creating continuous variables. The techniques for building the regression models come from analyzing the correlation of the “big four” factors with other game metrics, the LASSO regression, and creating a decision tree. The use of correlation helps to show which feature variables have a relationship with the “big four” factors. The LASSO regression uses shrinkage, encouraging a model with fewer parameters, by reducing the influence of noninformative variables. The LASSO regression allows the coefficient of a variable to shrink to zero, effectively removing that variable from the model. Finally, a decision tree is useful for visualization and predicting outcomes in regression. Following this model exploration step, a final model was created for each factor. We then use the final models to perform inference, and develop interpretations and recommendations for GMU’s men’s basketball team.

Analysis of the Big Four Factors

To support our focus on the “big four” measures and to provide context for this study, an analysis of the difference in the score for the winning team and the score for the losing team was conducted. The difference in score was regressed on winner minus loser difference in effective field goal percentage, offensive rebounding percentage, turnover percentage, and free throw rate. All four variables have expected direction of effects. When winning teams have relatively high effective field goal percentages, rebounding percentages, and free throw rates, the score difference is widened, whereas when winning teams have relatively high turnovers the score difference is narrowed. For every additional percentage point difference in effective field goal percentage and rebounding percentage, the mean difference in score increases by 0.89 points and 0.31 points respectively. A one percentage point increase in the difference in turnover percentage lowers the mean difference in score by 0.82 points. Finally, a 0.01 increase in the difference in the free throw rate increases the score by 0.14 points. All “big four” variables are highly significant (p-values < 0.001). Additionally, these four variables explain about 86 percent of the variation in difference in score, suggesting they are key elements to a winning-game strategy. We will now investigate game variables associated with each of the four factors individually.

Rebounding

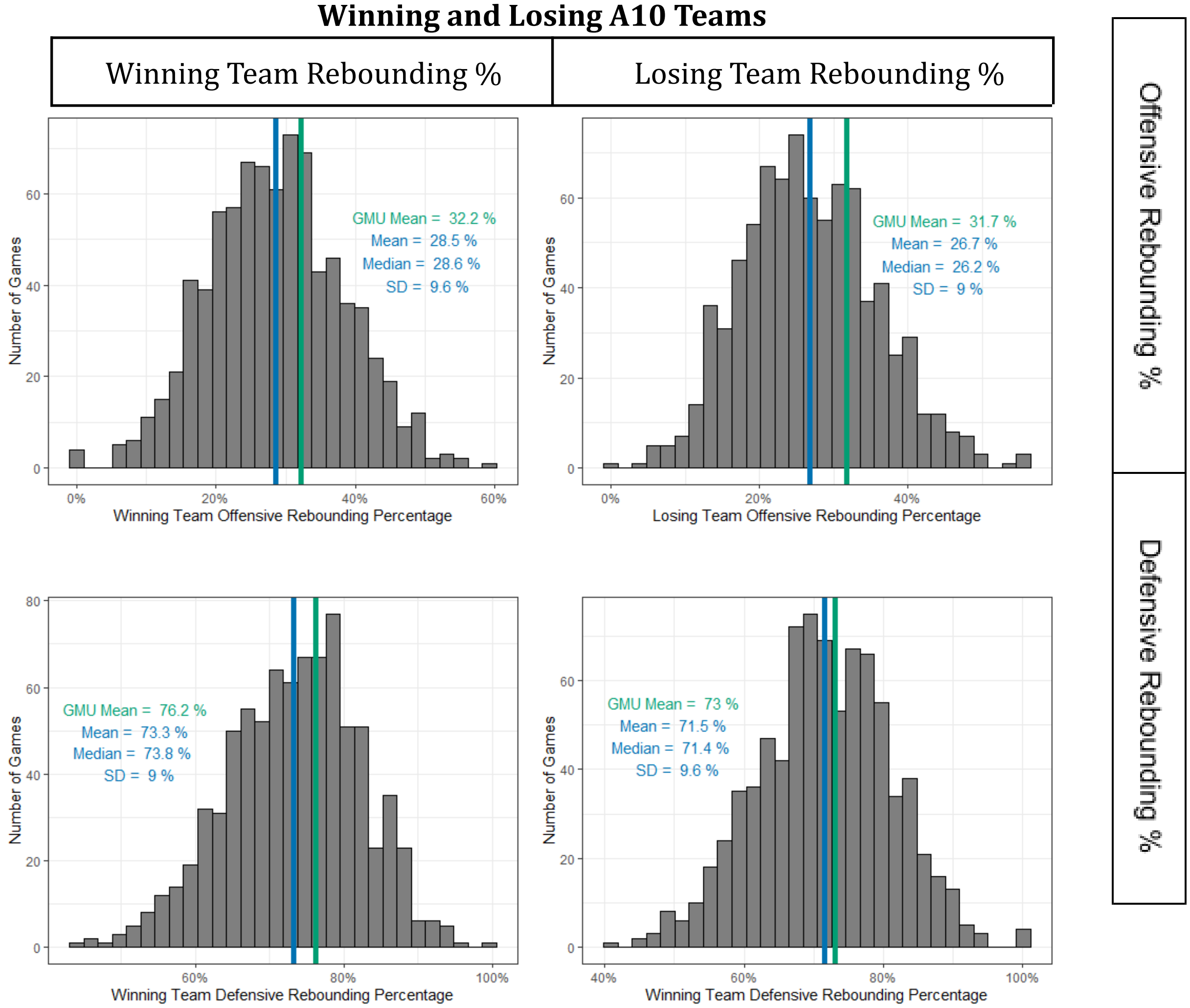
Rebounding percentages are broken into two separate categories for both teams in a game: winning team offensive rebounding percentage, losing team offensive rebounding percentage, winning team defensive rebounding percentage, and losing team defensive rebounding percentage[1]. Rebounding percentages are a better, and more frequently used,

[1] See appendix for offensive and defensive rebounding percentages

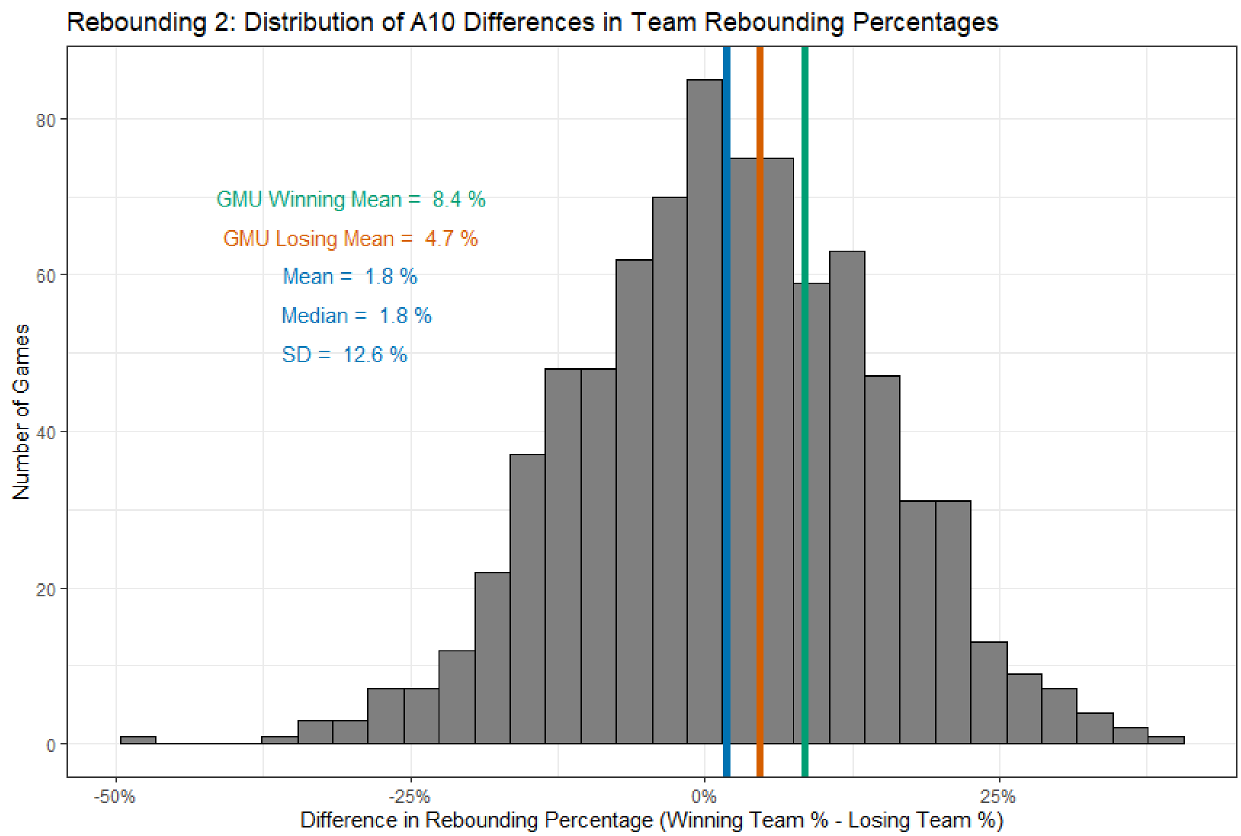
metric for analysis of rebounding because rebounding percentages account for the number of rebounds that a team was vying for during the course of a game. This is important because if we analyze rebounding based on counts of rebounds, the data will be biased towards games with higher pace and more shots taken.

Rebounding figures 1 and 2 below show the distributions of winning and losing teams offensive and defensive rebounding percentages of all A10 basketball games from 2015-2020. Offensive and defensive rebounding characteristics are substantially different due to offensive rebounding being much harder to secure than a defensive rebound, and thus, having much lower rates. In these distributions we see that offensive rebounding averages are around 28.5% (+-9.6%) for winning teams and 26.7% (+-9.0%) for losing teams. For defensive rebounding we see that winning teams have an average of 73.3%(+-9.0%) and losing teams an average of 71.5%(+-9.6%). These data provide preliminary evidence of the predictive power of rebounding percentages on the outcomes of basketball games. This is also seen in the following graph of the differences between rebounding percentages, where the winning team, on average, has a rebounding percentage 1.8% greater than the losing team. Within each of these graphs the mean, median and standard deviation are provided; furthermore, a vertical line is drawn for the average and GMU’s averages for each respective rebounding category. We see that GMU is already a very strong rebounding team with rebounding percentages well above average for both winning and losing games. This is also seen in raw rebounding numbers, where GMU ranks in the top two in the A10 for average offensive and defensive rebounds in both winning and losing games.

Rebounding 1: Distributions of Offensive and Defensive Rebounding Percentages for Winning and Losing A10 Teams

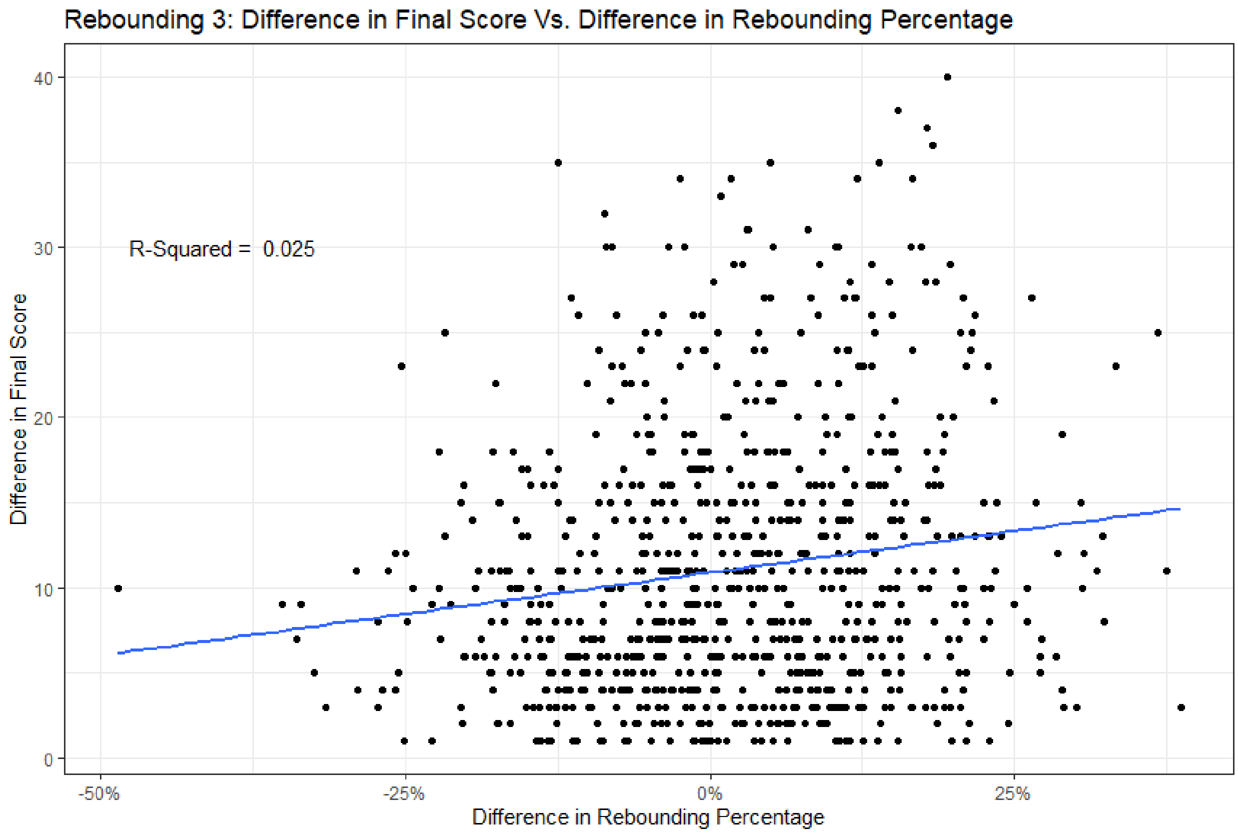


Shown: Distribution of offensive and defensive rebounding percentage for winning and losing teams from all A10 conference games from 2015-2020



Shown: Distribution of the difference in rebounding percentage from all A10 conference games from 2015-2020

Along with understanding the univariate distributions of each of the rebounding variables, we are interested in understanding the relationships between each of the rebounding variables and other variables, to include: end-game scores, effective field goal percentage, pace, and location of the game. The figure below shows the relationship between the difference in the final score and difference in rebounding percentages for winning and losing teams. We find a slight positive correlation between the two variables, however, it is a very weak association (R-squared =0.025). This is in stark contrast to what we see with other variables like effective field goal percentage, which, unsurprisingly, is highly correlated to differences in final score.



Shown: Scatterplot of the difference of final scores with the difference in rebounding percentage from all A10 conference games from 2015-2020

We also evaluated correlations between the difference in rebounding percentage and all the other continuous explanatory variables. The strongest correlation found being the difference in effective field goal percentage with a correlation of -0.27. This means that as the winning team increases the difference in effective field goal percentage, the difference in the rebounding battle goes down. This is an interesting finding in that both of these are part of the "big four" factors, but a teams' success in one may be hurting their success in the other. A possible interpretation of this is that teams that win rebounding battles in games may tend to be teams that are taller and more physical. Such teams are typically weaker in finessed skills such as shooting, ball handling, and passing. Since effective field goal percentage is so predictive of difference in scores, this relationship may be indicative that teams in the A10 should spend less energy on winning the rebounding battle or recruiting players who are talented rebounders, and spend more energy on recruiting talented shooters.

In the modeling phase three separate models for rebounding are developed due to the different stories seen when looking at overall offensive rebounding percentage, overall defensive percentage, and the differences in rebounding percentages. In modeling the differences in rebounding percentage the coefficients with the largest effect sizes are effective field goal percentage, which has a significant negative effect, and turnover percentage, which has a significant positive effect. In modeling overall winning defensive rebounding percentage the largest effect sizes coming from the winning teams effective field goal percentage, with a significant negative effect, and the winning team turnover percentage, with a significant positive effect. In modeling overall winning offensive rebounding percentages the largest effect sizes coming from, again, the winning teams effective field goal percentage, with a significant negative effect and the winning teams turnover percentage, with a significant positive effect.

From each of the models the most significant variables that affect defensive, offensive, and the difference in rebounding percentages between winning and losing teams is effective field goal percentage and turnover percentage. Specifically, we find a negative effect for effective field goal percentage, which coincides with the initial findings discussed at the beginning of this section. The intriguing finding from the model that was not found in the descriptive analysis is the significant, and substantial, positive effect for turnover percentage. This effect means that as a winning team has more turnovers they seem to get more rebounds; conversely, as a winning team has more turnovers they tend to win the rebounding battle by a wider margin. This may be due to the fact that if a team turns the ball over more the other team will get more shot attempts, thus making more defensive rebounding attempts, which will benefit the other teams' rebounding percentage.

Overall, we find a compelling story about a potential tradeoff that occurs between a more finessed team that values precise shooting and less turnovers, versus a team that values larger, more physical, and better rebounding players. An ideal team would likely be physical enough to obtain rebounds and finessed enough to be effective shooters, however, such a combination is not easy to come by for lower conferences such as the A10. From these findings, and findings about the importance of effective field goal percentage on winning basketball games, these results may indicate that GMU should pursue more finessed players, and reduce playing more physical and less finessed players.

Free Throws

In this section of the analysis, factors that are associated with the free throw rate and two related measures are examined. The free throw rate is defined as the ratio of free throws made to the number of field goal attempts. It gives weight to a team with a high free throw percentage because the numerator is free throws made as opposed to free throw attempts. The denominator, field goal attempts, can be thought of as the number of scoring opportunities a team has during the game. If many of those scoring opportunities result in free throws, then the expected value of points could be higher because free throw points can be in addition to a field goal and because free throw percentages tend to be high.

Although not official "big four" measures, it is also worthwhile to consider two related dependent variables: the total number of free throws made during the game and the free throw percentage, which reflects the number of free throws made as a percentage of free throw attempts. Free throws made is more directly associated with winning and free throw percentage can provide insights for the purposes of recruiting players with certain skills and for the role of practice before games.

The dependent variables are the winning team's values minus the losing team's values. For example, the difference between the winning team's free throw rate and the losing team's free throw rate is a dependent variable; the difference in free throws made and free throw percentage are similarly constructed for the other two dependent variables.

A key explanatory variable is the court location of the winner (home court, neutral court, or away court). Court location is a key variable for free throw measures because home teams may get favoritism from the referees in terms of fouls called. Additionally, familiarity with the court and behavior of fans may influence the percentage of free throws made by the home team. If either of these factors were true, this would suggest a strategy during home games of trying to draw fouls.

In the basic model, the other explanatory variables include pace of the game, time point in the season (measured as days since the season started), whether the game is a conference tournament, whether the game went to overtime, and the season. Pace squared and point-in-season squared are also included allowing for nonlinear relationships between explanatory and dependent variables. In the expanded model, the other three "big four" factors (winner's effective field goal percentage minus the loser's, winner's turnover ratio minus the loser's, and winner's offensive rebounding percentage minus the loser's) are also brought in as explanatory variables.

This section is divided into three parts. The first part, *Exploratory Analysis*, provides some descriptive statistics on dependent and explanatory variables and establishes the relationship between the free throw rate and the other three factors of the “big four”. The second part, *Basic and Expanded Models*, examines the basic and expanded models. The third and final part synthesizes findings and discusses the role of court location on free throw outcomes.

Exploratory Analysis

Appendix Table FT-1 contains means for the analysis variables in this section. On average the free throw rate of winning teams is about 0.28 (+/- 0.12) as compared to about 0.21 (+/- 0.10) for losing teams, suggesting higher free throw rates help improve a team's chance of winning. Additionally, total free throws made during a game average about 15.2 (+/- 5.9) for winners and 11.7 (+/- 5.2) for losers. The free throw percentage for winning teams is about 72% (+/- 12) and about 68% (+/- 13) for losing teams.

Some general information provides context for this section's findings. Looking at the court location of winners, 54.9% are home games, 8.7% are neutral, and only 36.3% are away. This suggests home courts are associated with winning. About 8% of games are conference tournament games and slightly more than 6% of games go to overtime.

Results from regressing the difference in free throw rates between winners and losers on differences for the other three "big four" factors provides insight into the relationship of the "big four" factors. When the winning team has a low effective field goal percentage relative to the losing team, the winning team also tends to have a high free throw rate. This likely reflects the role of fouls: if the winning team is fouled frequently during shooting, it would be expected that the effective field goal percentage would be low but the free throw rate would be high. When the winning team has a high turnover ratio relative to the losing team, the free throw rate of the winning team tends to be higher. This could occur if the winning team is pursuing an aggressive offensive strategy and turns the ball over frequently but also draws a number of fouls. Finally, when the winning team has a relatively high offensive rebounding percentage, the free throw rate of the winning team tends to be lower. All of these three "big four" factors have statistically significant effects on the difference in free throw rates between winners and losers.

Basic and Expanded Model

Regression results for the basic model are presented in Appendix Table FT-2 (column 2). The intercept, which reflects whether winning teams have higher or lower free throw rates when other variables are zero, is positive but statistically insignificant. Playing on a home court amplifies the difference between the winner and loser free throw rates. On average, winners have free throw rates that exceed those of losers by about 0.07. If the winner is playing at home as opposed to away, the difference rises by about 0.05, which reflects a sizable association between court location and free throw rates. The p-value is less than 0.0001, indicating a statistically significant effect. Neutral courts also amplify the difference in free throw rates between winners and losers, relative to playing away games. That effect (about 0.17) is quite large relative to the mean value of the difference in free throw rates between winners and losers (0.07) and statistically significant at the 5 percent level.

Explanatory variables other than court location do not offer clear patterns with differences in free throw rates. Pace, point in season, quadratic terms, tournament games, overtime games, and year of season are statistically insignificant at the 5 percent level. An F-test of the null hypothesis that all non-court-location variables are zero is not rejected at the 5 percent level.

The expanded model is presented in Appendix Table FT-2 (column 3). This model adds the other three factors of the "big four" to the regression. These variables have expected effects with high field goal percentage (fewer fouls) associated with lower free three ratios and more turnovers (aggressive play and more fouls) associated with higher free throw rates. However, including these variables does not change the findings, in general, from the basic model. Court location predicts free throw rates whereas many other explanatory variables are statistically insignificant.

In the basic model, the explanatory variables explain about 3% of the total variation in the difference in free throw rates between winners and losers, adjusted for the number of explanatory variables in the model. This does not invalidate the findings with regard to court location, but rather suggests the difference in free throw rates between winners and losers may be affected by many variables not available in our data. With regard to the expanded model, the explanatory variables explain about 17% percent of the total variation in the difference in free throw rates between winners and losers, adjusted for the number of explanatory variables in the model.

Regressions are also conducted using the difference (winners minus losers) in free throw percentage and in free throws made as dependent variables. These results help decompose the court location effects presented earlier. Home court does not increase the free throw percentage (in basic or expanded models) but does increase the number of free throws made (in basic and expanded models). This is consistent with a view that home teams get favorable treatment by the referees (more fouls are called in their favor) but do not have better free throw shooting as a percentage, which might have been suspected given the team's familiarity with the environment and the fans attempts to distract opponents.

Synthesis of Findings

Court location is found to be the important predictor in analyzing free throw variables. Home court could be important for two different reasons: referees may show favorable treatment to the home team and call more fouls in their favor, or home teams may have a higher free throw percentage. The latter may happen because the home team has practiced numerous free throws on that court (familiarity with the environment) and because home fans try to distract opponents but not the home team during free throws. The evidence presented does not support this possible latter effect. Rather, the advantage of home court is tied to the number of times the home team makes free throws. This finding is suggestive of referees showing a bias in their calls in favor of home teams.

There is some descriptive evidence in support of this claim. The average number of fouls called against the home team is 16.8, while the number of fouls called against the away team is 18.5. However, the free throw percentage of home teams is fairly close to that of away teams (70.4 versus 69.6). Home teams get to the free throw line more often, however, and thus have a higher number of free throws made (15.1 versus 11.8). This suggests a style of play at home should be different than away. Specifically, at home games, teams should try to draw fouls and at away games be careful to minimize the number of fouls.

Effective Field Goal Percentage

The third "big four" factor under consideration is effective field goal percentage. Effective field goal percentage (EFG%) is a statistic that provides more weight to three-point field goals than two-point field goals. This is an important factor because in a typical basketball game, the majority of points will be scored on the field (as opposed to through free throws).

If a team has a high EFG%, and its opposing team has a low EFG%, the team with the high percentage is performing well both offensively and defensively. Significant differences between the mean EFG% of the winning teams and the mean EFG% of the losing teams are found using Welch's t test (p-value <0.0001). A scatterplot of the difference in EFG% and the difference in final score of the game in Figure EFG-1 displays that a greater difference in EFG% between the winning and losing team is associated with a greater difference in the final score. To investigate the EFG% for GMU and how it compares to the other members of the A10, further exploratory data analysis is performed.

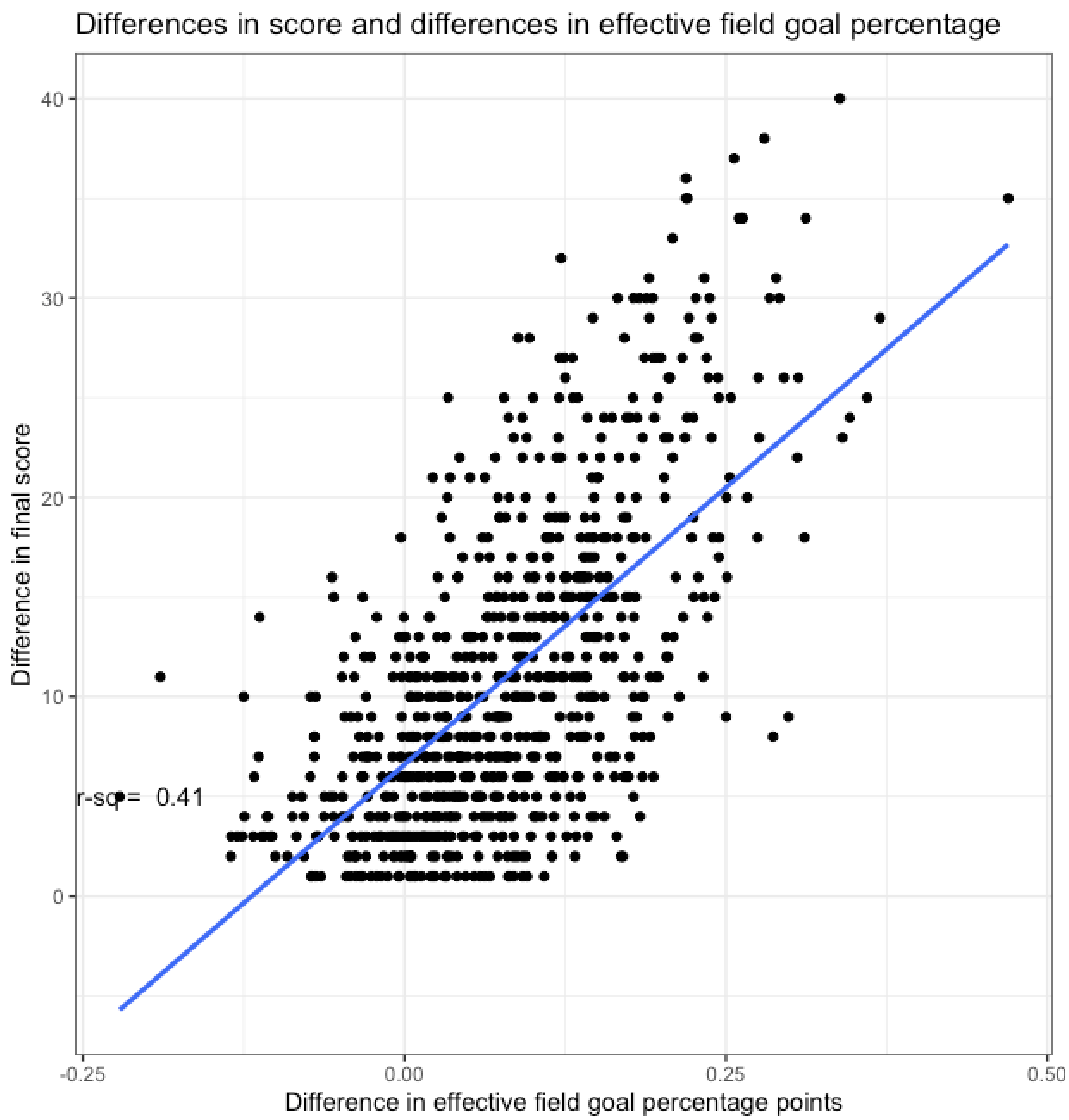


Figure EFG- 1: Scatterplot for the relationship between the difference in effective field goal percentage and the difference in final score in A10 regular season games, 2015-2020. The sloped line represents the estimated mean difference in final score based on the difference in effective field goal percentage from a linear regression. Pearson's r-squared = 0.41.

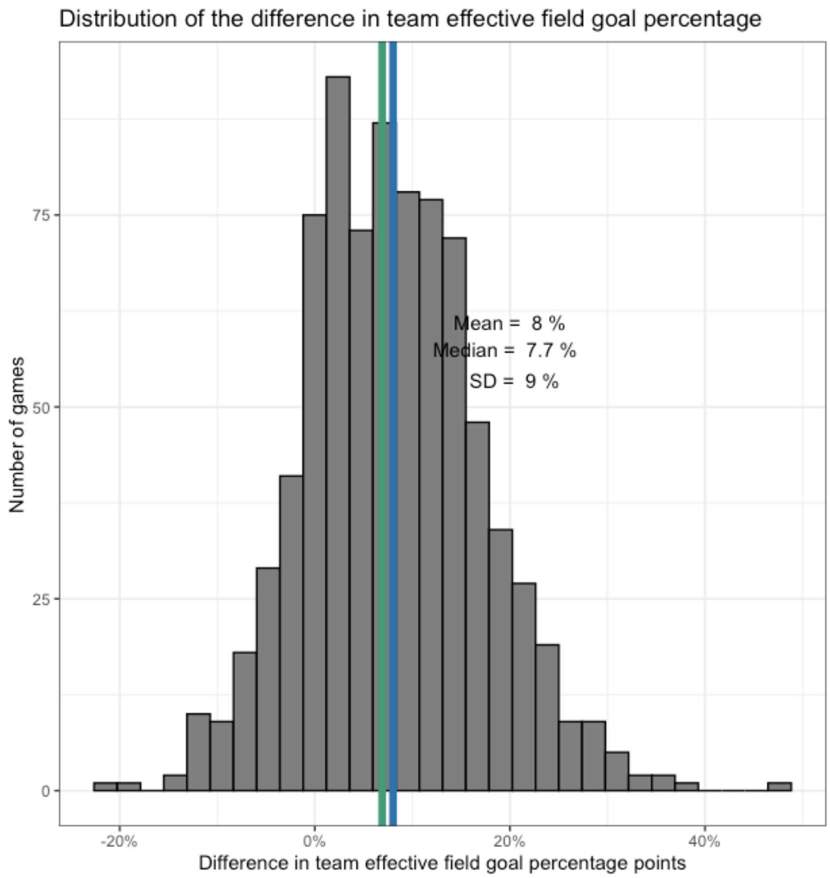


Figure EFG- 2: The distribution of the difference in effective field goal percentage across the A10 regular season games, 2015-2020. GMU's mean difference in effective field goal percentage is in green, the mean difference in effective field goal percentage for the entire A10 conference is in blue. The y-axis features the number of games within the A10 regular season. Mean, median, and standard deviation for the A10 conference are displayed.

Figure EFG- 2 displays the distribution of the difference between winning team and losing team EFG% across the A10. GMU's mean difference in EFG% is depicted in green, and the overall mean difference in EFG% is depicted in blue. The mean difference in EFG% between winning and losing teams among the A10 regular season games is 8% (+/- 9%). The mean difference in EFG% for games that GMU participated in is 6.9% (+/- 8.3%). The mean EFG% for games that GMU won is 5.2% (+/- 6.8%), whereas the mean winning team EFG% across the A10 is 5.4% (+/- 8%). The mean difference in field goal percentage for games that GMU participated in is also lower than that of the overall A10 - GMU's mean difference in field goal percentage is 5.5% (+/- 7.1%), and for the A10 the mean difference is 6.8% (+/- 7.9%). In contrast, games that GMU participated in featured a higher mean difference in 3-pointer field goal percentage (8.8% +/- 13.8%) than the mean difference in 3-pointer field goal percentage across the A10 (7.6% +/- 13.8%). For games that GMU was the winner, the mean three-pointer percentage was 3.9% (+/- 11.3%), only slightly higher than the mean three-pointer percentage for winning teams across the A10, 3.8% (+/- 11.1%).

The correlation plot in Figure EFG- 3 displays the magnitude and direction of the correlation between the difference in EFG% and other game variables. Stronger color intensity and larger size of circles denotes a stronger correlation, whether positive or negative. The difference in offensive rebounds is negatively correlated with the difference in EFG%. The difference in defensive rebounds and the difference in assists are positively correlated with the difference in EFG%. Weakly negative correlations for the difference in EFG% are found for the difference in free throws made and the difference in free throws attempted. Weakly positive correlations for the difference in EFG% are found for the difference in turnovers, and the difference in turnover percentage. These correlations are a useful starting point for investigating associations between EFG% and other games variables.

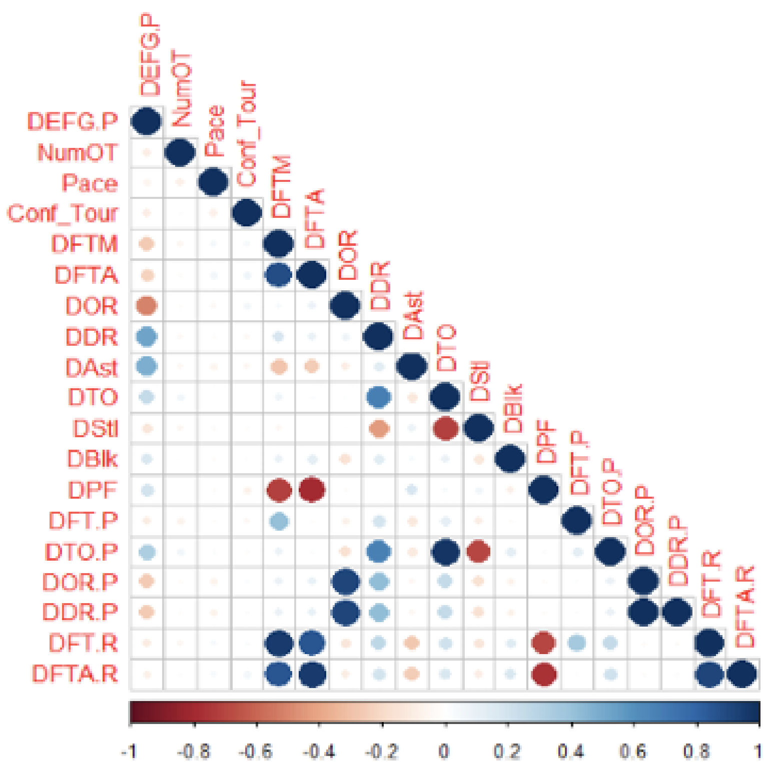


Figure EFG- 3: Correlation plot to display positively and negatively correlated game variables. Color intensity and size of circles denote strength of correlation, with deeper hues and larger circles representing stronger correlations. Negative correlations are depicted in red tones, and positive correlations are depicted in blue tones.

A linear model is used for the mean difference in EFG% and positively and negatively correlated variables. The mean difference in EFG%, the response variable, is significantly associated with the difference in free throws made, the difference in free throws attempts, the difference in offensive rebounds, the difference in defensive rebounds, the difference in assists, the difference in turnovers, and the difference in defensive rebounding percentage. The estimated mean difference in EFG% decreases with increases in the difference in free throws made, the difference in offensive rebounds, the difference in turnovers, and the difference in defensive rebound percentage. The estimated mean difference in EFG% increases with increases in the difference in defensive rebounds, difference in free throw attempts, and the difference in assists. The largest effect size is for the difference in defensive rebound percentage. The adjusted R-squared value for the model is 0.8051, meaning 80.51% of variability in the difference in EFG% is explained by the model, adjusted for the number of explanatory variables in the model. Table 1-EFG in the appendix displays the coefficients and associated p-values for this linear model.

To model the explanatory variables that are associated with differences in EFG%, a linear regression model is used. The mean difference in EFG%, the response variable, across the A10 is significantly associated with nine variables: number of overtimes, pace, difference in free throws made, difference in free throw attempts, difference in offensive rebounds, difference in defensive rebounds, difference in offensive rebound percentage, and the difference in assists. This model has an adjusted R-squared value of 0.8133, meaning 81.33% of the variability in EFG% is explained by the model. Table 2-EFG in the appendix displays the coefficients for this linear model. The estimated mean difference in EFG% decreases with increases in pace, difference in free throws made, difference in offensive rebounds, difference in turnovers, difference in offensive rebound percentage, and number of overtimes. The estimated mean difference in EFG% increases with increases in the difference in free throws attempts, the difference in defensive rebounds, and the difference in assists. The largest effect size is for the difference in offensive rebound percentage.

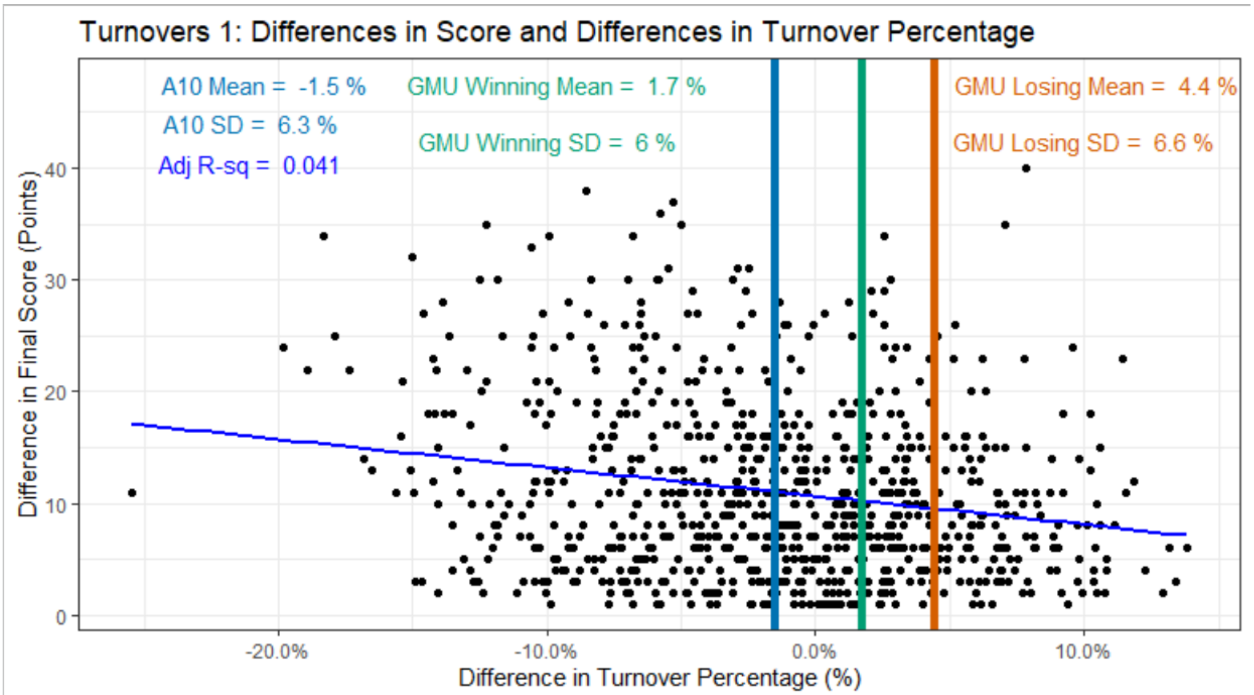
A Decision Tree (Figure EFG-4 in the appendix) is used for further visualization and investigation of the difference in EFG% among winning and losing teams. Decision Trees are useful for pattern detection and predicting outcomes for regression models. The predictor variable for the primary split for difference in EFG% is the difference in defensive rebounds. More than half of the games in the dataset (443 out of 823) feature a small difference in offensive rebounds, and a small difference in EFG%. These 443 games are further split by difference in offensive rebounds, difference in assists, and difference in free throws made. Games with the smallest difference in EFG% in the data set are associated with small differences in defensive rebounds, offensive rebounds, assists, and free throws made. For games that feature a larger difference in defensive rebounds (380 out of 823 games), a larger difference in EFG% is displayed. These games are further divided by difference in offensive rebounds, difference in assists, and difference in free throws made. The largest difference in EFG% is associated with larger differences in assists, defensive rebounds, and offensive rebounds.

Maintaining a relatively high EFG% is important for winning basketball games, and winning them by a high margin. Focusing on completing three-pointer shots over two-pointer shots is associated with an increase in the team EFG%, as three-pointer shots weigh more heavily in the calculation of EFG%. We find that a large difference in EFG% is coupled with a high number of assists, and a large difference in rebounds (both defensive and offensive). Focusing on assists during practice, and recruiting players who excel at passing and creating opportunities for teammates to score, as well as maintaining impressive rebounding percentages, may improve GMU's EFG% and lead to more wins within the A10 conference and tournaments. As discussed previously, finding players with both finesse and physicality will improve the EFG% of the GMU men's basketball team.

Turnover Percentage

In this section, we consider the "big four" factor turnover percentage. A turnover occurs when the opponent steals the ball, the team in possession commits a personal foul, goes out of bounds, commits a violation, or the team in possession loses a jump ball. The denominator of the turnover percentage formula is a proxy for the number of possessions in a game. The data collected allows the turnover percentage to be calculated for both winning teams and losing teams. In the process of determining how to improve the turnover percentage, it is important to understand what metrics help explain the number of turnovers by each team in the game. The variables used are the difference between the winning team's statistics and the losing team's statistics. For example, the dependent variable, *Difference in Turnover Percentage*, is calculated as the winning team's turnover percentage minus the losing team's turnover percentage.

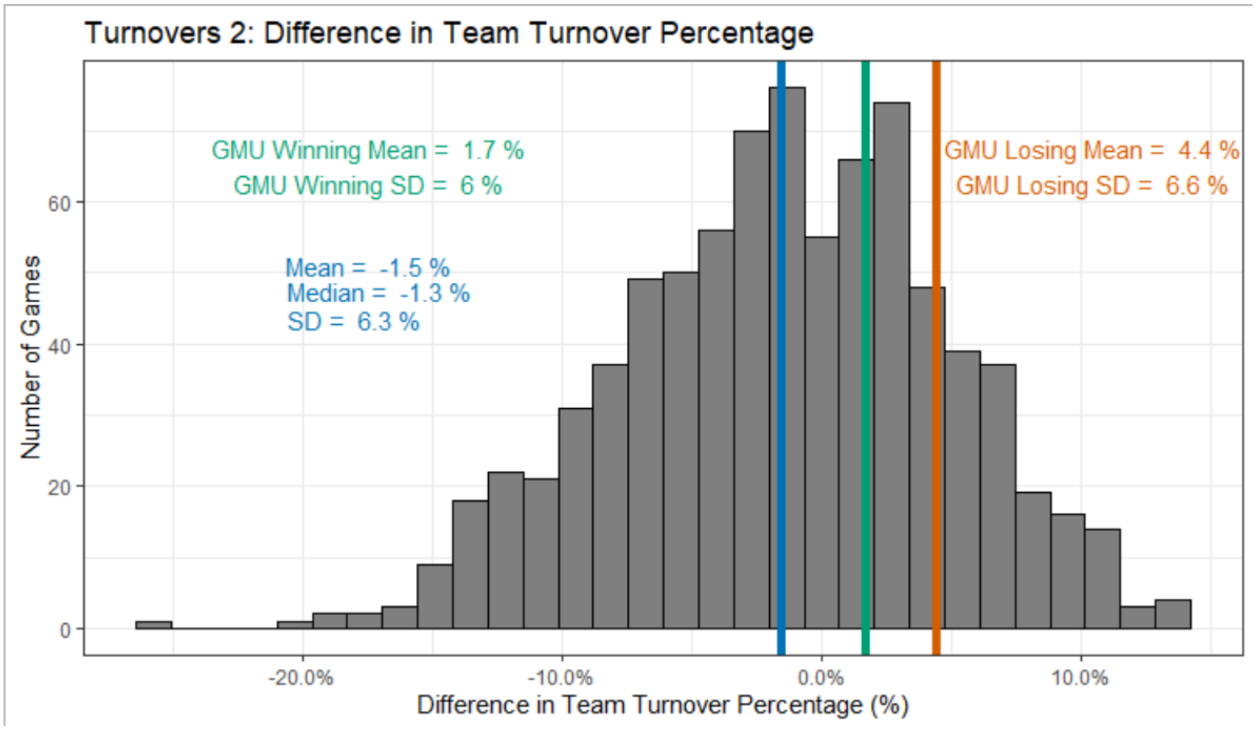
The way to win a basketball game is to score more points than the opponent, and every possession that ends in a turnover does not end in points. From this, it is reasonable to assume that, on average, teams that have a lower percentage of possessions ending in turnovers than their opponent (the difference in turnover percentage) will score more points. The figure Turnovers 1 shows the relationships between turnover percentage and the difference in score between the winning team and the losing team.



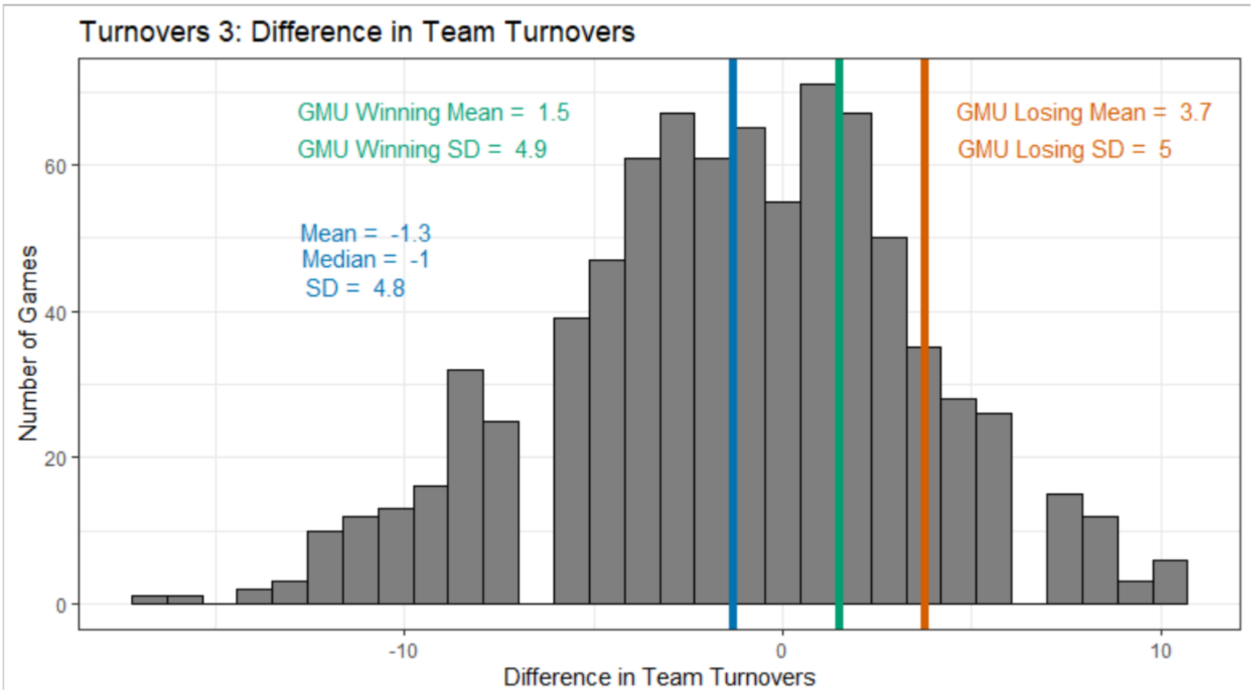
Turnovers 1: Scatter plot of the difference in final score and the difference in turnover percentage for all A10 games from 2015-2020. The leftmost vertical line is the average turnover percentage difference between the winning team and the losing team for all A10 games. The middle vertical line is the GMU Men's Basketball team's difference in turnover percentage when the team wins and the rightmost line is the team's difference in turnover percentage when they lose. The sloped line represents the estimated mean difference in score based on the difference in turnover rate from a linear regression.

The figure Turnovers 1 validates the earlier assumption for basketball games between A10 opponents. The adjusted R-squared value of 0.04 for the sloped line shows that the difference in turnover percentage and the difference in turnovers only explains about 4% of the variability in the difference in scores. This low R-Squared value suggests that winning the turnover percentage battle has little effect by itself on the difference in final scores. While having a lower turnover percentage creates more scoring opportunities for one team compared to its opponent, the ability to capitalize on those extra possessions is also important. While figure Turnovers 1 shows data points that are potential outliers (specifically the bottom left and the top right), the exclusion of those points do not drastically affect the slope of the regression line. Therefore, these points are included (here and in further analysis) to represent all A10 men's basketball games from 2015-2020.

While turnovers alone do not determine the winner of the basketball game, the winner, on average, has a lower turnover percentage and lower turnovers than its opponent. The figures Turnovers 2 and Turnovers 3 show the distribution of the difference in turnover percentage and the distribution of the difference in turnovers between the winning and losing teams in the A10.



Turnovers 2: Histogram of the difference in team turnover percentage for all A10 games between 2015-2020. The leftmost vertical line is the average turnover percentage difference between the winning team and the losing team for all A10 games. The middle vertical line is the GMU men's basketball team's difference in turnover percentage when the team wins and the rightmost line is the team's difference in turnover percentage when they lose.

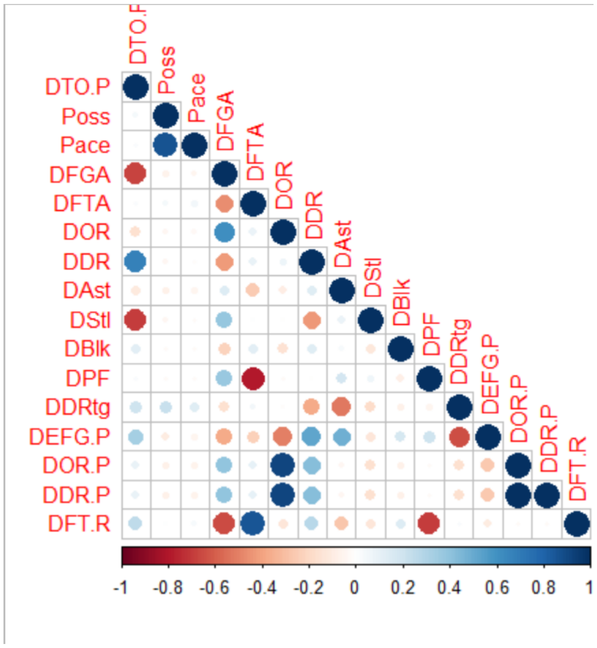


Turnovers 3: Histogram of the difference in team turnovers for all A10 games from 2015-2020. The leftmost vertical line is the average turnover difference between the winning team and the losing team for all A10 games. The middle vertical line is the GMU men's basketball team's difference in turnovers when the team wins and the rightmost line is the team's difference in turnovers when they lose.

Looking at figures Turnovers 2 and Turnovers 3, it is easy to tell that GMU often loses the turnover battle. Amongst all A10 schools over the 2015-2020 seasons, GMU has the 7th highest turnover percentage when they win (15.3%), the 2nd highest turnover percentage when they lose (18.0%), the 5th highest number of turnovers when they win (12.1), and the highest number of turnovers when they lose (14.2). For greater success in the A10, GMU should look for ways to minimize its own turnover percentage while maximizing the turnover percentage of its opponent.

One way to increase an opponent's turnovers is to create more steals, but there may be more areas within the basketball game that influences turnover percentage. The figure Turnovers 4 uses correlation to visualize what aspects of the basketball game GMU could focus on to improve its difference in turnover percentage.

Turnovers 4: Correlation of Difference in Turnover Percentage



Turnovers 4: A correlation plot of basketball game metrics for all A10 basketball games from 2015-2020. Larger circles represent larger absolute correlations.

The difference in field goal attempts (correlation =-0.68) and the difference in steals (correlation = -0.70) are strongly negatively correlated with the difference in turnover percentage while the difference in defensive rebounds (0.67) is strongly positively correlated with the difference in turnover percentage. One explanation for the negative correlation between the difference in turnover percentage and the difference in field goal attempts is that any possession that leads to a field goal attempt does not end in a turnover. Two explanations for the positive correlation between the dependent variables and the difference in defensive rebounds are that the other team attempts a field goal instead of a turnover and the rebounding team obtains another possession and another opportunity to lose a possession from a turnover.

Using correlation as an indicator of possible influential metrics, a linear model is created based on the features with an absolute correlation to the difference in turnover percentage greater than 0.5. Based on an F-test comparing linear models that include and do not include the game location, the game location is determined to be influential. The linear model is set up as the difference in turnover percentage modeled by game location, difference in field goal attempts, difference in defensive rebounds, and the difference in steals. The model predicts that the mean difference in turnover percentage between the winning and losing team decreases when the winner is playing at home or at a neutral site compared to playing an away game and decreases as the difference in field goal attempts or the difference in steals increase. The model suggests that as the difference in defensive rebounds increases, the mean difference in turnover percentage increases. The estimated coefficients in the model and corresponding p-values are given in Table TO-1 in the appendix. The assumptions of the model were verified using a residual versus fitted values plot and a normal QQ-plot. With an adjusted R-squared of 0.78, this linear model explains 78% of all variability in the difference in turnover percentage adjusted for the number of explanatory variables in the model.

The linear model created from the correlation plot explains a fair amount of the variability in the difference in turnover percentage, however, the correlation plot only shows the correlation of one independent variable with the difference in turnover percentage. Different linear combinations of variables may be influential in a way that does not show up on the correlation plot. Using LASSO regression is one way to find these influential linear combinations. The LASSO regression for the difference in turnover percentage identifies 18 influential game metrics. With such a large number of predictors, there is a risk that this model is overfitting the data. Some of the game metrics may not be providing enough predictive power to justify the increased complexity. Using F-tests to determine if each variable is significantly influential (and removing those that are not significantly influential), the model reduces to 9 variables. This new model estimates the difference in turnover percentage based on the location of the game, the difference in free throw attempts, the difference in offensive rebounds, the difference in defensive rebounds, the difference in steals, the difference in personal fouls, and the difference in true shooting percentage. This model suggests that the mean difference in turnover percentage decreases when the winning team is at home or a neutral site compared to at an away game, decreases as the difference in field goal attempts, the difference in free throw attempts, and the difference in steals increases. The mean difference in turnover percentage is estimated to increase as the difference in offensive rebounds, the difference in defensive rebounds, the difference in defensive rating, and the true shooting percentage increase. The influence of the difference in personal fouls, while positive (as expected), is nearly zero. The variable coefficients and p-values are available in Table TO-2 in the Appendix. The assumptions of the model were verified using a residual versus fitted values plot and a normal QQ-plot. With an adjusted R-squared of 0.93, this model explains nearly 93% of all the variability in the difference in turnover percentage between winning and losing teams adjusted for the number of explanatory variables in the model. Compared to the model created using the correlation plot, this model explains nearly 15% more variability in the difference in total turnover percentage adjusted for the number of explanatory variables in the model.

One aspect of the difference in turnover percentage is the difference in turnovers. Similar to the previous model, the LASSO regression can be run on the difference in turnovers based on all the available game metrics. The LASSO method finds 15 influential variables, which reduces to 10 variables after model reduction using F-tests. This model suggests that the mean difference in turnovers between the winning and losing teams decreases when the winning team is at home or a neutral site, decreases as the pace of the game, the difference in field goal attempts, and the difference in steals increases. The mean predicted difference in turnovers increases as the difference in three point field goals, the difference in defensive rebounds, the difference in offensive rebounds, the difference in personal fouls, the difference in three point field goal percentage, and the difference in defensive ratings increases. The assumptions of the model were verified using a residual versus fitted values plot and a normal QQ-plot. The estimated coefficients and p-values can be found in Table TO-3 in the Appendix. With an Adjusted R-Squared value of 0.87, nearly 87% of all variability in the difference in turnovers is explained by the model adjusted for the number of explanatory variables in the model.

While many aspects of basketball are interconnected, it is worth analyzing what factors influence the difference in turnover percentage and the difference in turnovers outside of the other "big four" factors. To test these relationships, the LASSO regression was run on a reduced dataset that did not include metrics related to field goals, free throws, or rebounds. The LASSO method identifies 7 influential variables and F-tests reduced this value to 4. This model estimates the difference in turnover percentage based on the location of the game, the difference in steals, the difference in blocks, and the difference in defensive rating between the winning and losing team. The estimated mean difference in turnover percentage decreases when the winning team plays at home or at a neutral site and decreases as the difference in steals increases. The estimated mean difference in turnover percentage increases as the difference in blocks increases. The influence of the difference in defensive rating, while positive, is nearly zero. The coefficients for each variable and the corresponding p-values can be found in Table TO-4 in the Appendix. The assumptions of the model were verified using a residual versus fitted values plot and a normal QQ-plot. With an R-squared of 0.51, 51% of the variability in the difference in turnover percentage is explained by the model, adjusted for the number of explanatory variables in the model.

The following are possible explanations for the directions of the coefficients in the model. A home game is estimated to decrease the mean difference in turnover percentage compared to an away game, because the home team may receive a psychological boost from the support of the fans, the players may be more comfortable at home, rather than traveling, the home team is familiar with the particulars of the home court/arena, and its possible referees give home teams the benefit of the doubt on calls. At a neutral site game, both teams are required to travel and neither team is likely to receive the typical fanbase support of a home game and neither team has familiarity with the arena. The difference in field goal attempts and the difference in free throw attempts are negatively associated with the difference in turnover percentage because, each field goal attempt or free throw attempt by the winning team is one less possession that ends in a turnover while one less field goal attempt or free throw attempt by the losing team is one more possession that ends in a turnover. The increase in the mean difference in turnover percentage created by an increase in the difference in blocks may be included in the model without other "big four" factors as an influential variable because blocks may be a proxy for field goal attempts. As field goal attempts increase by the losing team, the winning team has more opportunities to block shots. Possessions that end in a block are possessions that do not end in a turnover. The difference in assists is estimated to decrease the mean difference in turnover percentage, because a winning team with more assists is more likely to have better ball security than the losing team. With better ball security, the winning team is likely to have a lower turnover percentage. Obviously, an increase in the difference in steals will decrease the turnover percentage, because each steal by the winning team is a turnover by the losing team. An increase in defensive rebounds is estimated to increase the mean difference in turnover percentage, because each defensive rebound by the winning team is the result of a field goal attempt by the losing team. The increase in the estimated mean difference in turnover percentage as defensive rating increases follows from the formula of defensive rating which is based on the opponent's score per possession where a lower defensive rating is better. A lower defensive rating means opponents scored less per possession. Some of those possessions may have been turnovers. Similarly, the increase in the mean difference in turnover percentage estimated by an increase in true shooting percentage can be explained from its formula. The true shooting percentage uses field goal attempts and free throw attempts as a proxy for possessions. As field goal attempts and free throw attempts increase, the number of possessions with turnovers decreases. Therefore, the true shooting percentage is likely to decrease as the turnover percentage decreases.

Conclusion

Maintaining a strong performance throughout the regular season is essential for teams to receive an invitation to compete in the Division 1 Men's Basketball Tournament. Competing in this tournament, "March Madness," may positively impact the reputation of GMU's basketball program. Linear regression is utilized to investigate the game variables most highly associated with the identified "big four" factors for basketball success.

GMU has high offensive and defensive rebounding percentages in games they lose and games they win. The game variables most highly associated with rebounding percentages are effective field goal percentage and turnover percentage. The rebounding percentages, and difference in rebounding percentages, improve for the winning team as their effective field goal percentage worsens in comparison to the opposing team. Similarly, the rebounding percentage for the winning team improves as the winning team turns the ball over more. These findings seem to differentiate between two different types of teams in the A10 - larger and more physical teams versus smaller and more finessed teams. The findings indicate that the smaller, more finessed teams lead to winning games, even while performing badly in rebounding. Therefore, we recommend GMU focus more on recruiting and playing team members who have good shooting and ball handling skills, while reducing focus on larger players who do not have as refined finesse skills.

Examination of free throw data in A10 games indicates a home court advantage. The difference in free throw rates between the winning and losing team in a game is amplified if the game takes place on the winner's home court. This result is not due to a higher free throw percentage by home teams, but rather due to getting to the free throw line more often at home. This home court advantage could be due to preferential treatment of the home team by referees (calling more fouls against their opponents). This suggests GMU may want to use a style of play at home that draws fouls and one that minimizes fouls at away games. However, while free throw rate is predictive of winning score, it is less important than other "big four" factors. Thus, GMU may wish to apply a free-throw-based style of play selectively rather than as a general strategy. For example, a strategy to draw fouls at home may be a good overall game strategy if GMU suspects a particular opposing team would be susceptible to making mistakes resulting in fouls. That is, if a particular opposing team is known to make mistakes such as being out of position or not getting back to set defense, GMU may wish to run a style of play at home that draws fouls. Against other opponents, GMU may wish to focus on other "big four" factors.

Maintaining a higher effective field goal percentage than the opponents' is associated with winning basketball games. This factor is negatively associated with offensive rebounds and the difference in turnover percentage, and positively associated with the difference in assists and defensive rebounds. We recommend GMU focus on improving their three-pointer percentage, since three pointers more heavily impact EFG% than two-pointers. Upholding GMU's impressive rebounding percentage, and spending more attention on increasing assists and effective shooting will likely increase their overall EFG% and success in the A10 conference.

Through all of the models for the difference in turnover percentage, it is evident that the GMU men's basketball team should increase steals and field goal attempts are important to decreasing the difference in turnover percentage. While increasing the difference in defensive rebounds is associated with an increase in the difference in turnover percentage, the solution is not to perform worse at defensive rebounds, but to ensure that the opponent has fewer field goal attempts. Similarly, increasing the difference in offensive rebounding should not be abandoned in the attempt to lower the difference in the turnover percentage. Instead, GMU should focus on ball control and finding an opportunity for a field goal attempt following an offensive rebound to ensure the possession does not lead to a turnover. Finally, the GMU men's basketball team should focus on exploiting the advantages gained from games at home and at neutral sites. Topics for future research in using the difference in turnover percentage to win basketball games can be done on which metrics increase the difference in points off turnovers, which aspects of basketball games played at home lead A10 teams to have a better turnover percentage, and how these results can be expanded to improve GMU's chance of winning non-conference games.

Appendix

Acronyms list

Acronym	Full name
EFG%	Effective field goal percentage
DEFG%	Difference in effective field goal percentage
Poss	Possessions
DFGA	Difference in field goal attempts
DDRtg	Difference in defensive rating
NumOT	Number of overtimes
Conf_Tour	Conference tournament
DFTM	Difference in free throws made
DFTA	Difference in free throws attempted
DOR	Difference in offensive rebounds
DDR	Difference in defensive rebounds
DAst	Difference in assists
DTO	Difference in turnovers
DStl	Difference in steals
DBlk	Difference in blocks
DPF	Difference in personal fouls
DFT.P	Difference in free throw percentage
DTO.P	Difference in turnover percentage
DOR.P	Difference in offensive rebound percentage
DDR.P	Difference in defensive rebound percentage
DFT.R	Difference in free throw rate
DFTA.R	Difference in free throw attempt rate
DTS.P	Difference in true shooting percentage

Formulas

Defensive rating: 100*opponent points scored/ possessions

Defensive rebounding percentage: defensive rebounds / defensive rebounds + opponent offensive rebounds

Effective field goal percentage: field goals made + (0.5 * three-pointers made)/ field goals attempted

Field goal percentage: field goals made/field goals attempted

Free throw rate: free throws made/ field goals attempted

Pace: 40 * (possessions/(0.2*(5*(40+ number of overtimes * 5))))

Possessions: 0.5* (field goal attempts + 0.475 * free throw attempts - offensive rebounds +turnovers) + 0.5 * (opposition field goal attempts + 0.475 * opposite free throw attempts - opposition offensive rebounds + opposite turnovers)

Offensive rating: 100*points scored/ possessions

Offensive rebounding percentage: offensive rebounds / offensive rebounds + opponent defensive rebounds

True shooting percentage: team score/(2*field goal attempts + 0.475*free throw attempts)

Turnover percentage: turnovers/(field goals attempted + (0.44*free throws attempted) + turnovers)

Rebounding Model Results

Table RB-1: Defensive rebounding percentage

Variable	Coefficient	P-value
Intercept	0.8675	<.001
Winning Team Location - Home	0.0053	0.4382
Winning Team Location - Neutral	0.0008	0.9474
Winning Team Effective Field Goal Percentage	-0.2718	<.001
Winning Team Turnover Percentage	0.3196	<.001
Winning Team Assists	0.0020	0.0300
Winning Team Free Throw Rate	0.0330	0.4249
Winning Team Blocks	-0.0076	<.001
Winning Team Steals	-0.0026	0.0795
Winning Team Personal Fouls	-0.0045	<.001
Losing Team Effective Field Goal Percentage	0.2224	<.001
Losing Team Turnover Percentage	-0.2247	0.0126
Losing Team Free Throw Rate	0.1192	<.001
Losing Team Blocks	-0.0054	<.001

Losing Team Personal Fouls	-0.0007	0.5708

Table RB-2: Offensive rebounding percentage

Variable	Coefficient	P-value
Intercept	0.3053	<.001
Winning Team Location - Home	0.0009	0.9016
Winning Team Location - Neutral	0.0143	0.2537
Winning Team Effective Field Goal Percentage	-0.4342	<.001
Winning Team Turnover Percentage	0.4381	<.001
Winning Team Assists	0.0024	0.0173
Winning Team Free Throw Rate	-0.0741	0.0982
Winning Team Blocks	0.0029	0.0487
Winning Team Steals	0.0003	0.8727
Winning Team Personal Fouls	0.0002	0.8977
Losing Team Effective Field Goal Percentage	0.2463	<.001
Losing Team Turnover Percentage	-0.1555	0.1105
Losing Team Free Throw Rate	0.0524	0.2702
Losing Team Blocks	0.0027	0.1067
Losing Team Personal Fouls	0.0005	0.6784

Table RB-3: Difference in rebounding percentage

Variable	Coefficient	P-value
Intercept	0.2244	<.001
Winning Team Location - Home	0.0141	0.1116
Winning Team Location - Away	0.0088	0.5679
Pace	-0.0024	0.0027
Difference In Effective Field Goal Percentage	-0.4894	<.001
Difference in Free Throw Percentage	0.0080	0.7396
Difference in Turnover Percentage	0.5123	<.001
Difference in Free Throw Rate	-0.1332	<.001
Difference in 3-Point Attempt Rate	-0.2258	<.001

Free Throw Model Results

Table FT-1: Mean values for selected variables in free throw analysis

Variable	Mean
Winner free throw rate	0.2802
Winner free throws made	15.2041
Winner free throw percentage	0.7227
Loser free throw rate	0.2100
Loser free throws made	11.7254
Loser free throw percentage	0.6781
Difference in free throw rate	0.07026
Difference in free throws made	3.4787
Difference in free throw percentage	0.04462
Home court (winner)	0.5492
Neutral court	0.0875
Away court (winner)	0.3633
Game went to overtime	0.0632
Tournament game	0.0814
Pace	67.2365
Pace squared	4546.2400
Days since start of the season	36.7218
Days squared	1804.2040
Difference in effective field goal percentage	0.0802
Difference in turnover ratio	-0.0155

Difference in offensive rebounding percentage	0.0177
Difference in score	11.0632

Table FT-2: Difference in free throw rate regressed on explanatory variables in Basic and Expanded Model

	Basic Model, adjusted R^2=0.0298		Expanded Model, adjusted R^2=0.1714	
Variable	**Coefficient**	**P-value**	**Coefficient**	**P-value**
Intercept	0.1390	0.8290	0.2148	0.7185
Home court	0.0465	<0.0001	0.0611	<0.0001
Neutral court	0.1664	0.0107	0.1820	0.0025
Pace	-0.0037	0.8462	-0.0032	0.8551
Pace squared	0.00004	0.7936	0.00003	0.8464
Days	0.0002	0.8814	-0.0002	0.8605
Days squared	-0.000009	0.5524	-0.000004	0.7695
Tournament	-0.0913	0.1885	-0.1122	0.0808
Season 2015	0.0050	0.7832	-0.0003	0.9834
Season 2016	0.0039	0.8267	-0.0003	0.9866
Season 2017	-0.0044	0.8024	-0.0096	0.5589
Season 2018	-0.0346	0.0514	-0.0348	0.0344
Season 2019	-0.0156	0.3807	-0.0211	0.1997
Overtime	-0.0264	0.2062	-0.0621	0.0016

Difference in effective field goal %	-		-0.4581	<0.0001
Difference in turnover ratio	-		0.9200	<0.0001
Difference in offensive rebounding percentage	-		-0.1608	<0.0001

Effective Field Goal Percentage Model Results

Table EFG-1: Difference in effective field goal percentage regressed on positively or negatively correlated explanatory variables. The adjusted R-squared value is 0.8051.

Variable	Coefficient	P-value
Intercept	0.0280	<0.0001
Difference in free throws made	-0.0077	<0.0001
Difference in free throw attempts	0.0032	<0.0001
Difference in offensive rebounds	-0.0066	<0.0001
Difference in defensive rebounds	0.0126	<0.0001
Difference in assists	0.0042	<0.0001
Difference in turnovers	-0.0035	<0.0001
Difference in defensive rebounding percentage	-0.1140	0.0415

Table EFG- 2: Difference in effective field goal percentage regressed on nine explanatory variables. The adjusted R-squared value is 0.8133.

Variable	Coefficient	P-value

Intercept	0.0915	<0.0001
Number of overtimes	-0.0214	<0.0001
Pace	-0.0009	0.0011
Difference in free throws made	-0.0079	<0.0001
Difference in free throw attempts	0.0033	<0.0001
Difference in offensive rebounds	-0.0064	<0.0001
Difference in defensive rebounds	0.0125	<0.0001
Difference in turnovers	-0.0033	<0.0001
Difference in offensive rebounding percentage	-0.1244	0.0249
Difference in assists	0.0041	<0.0001

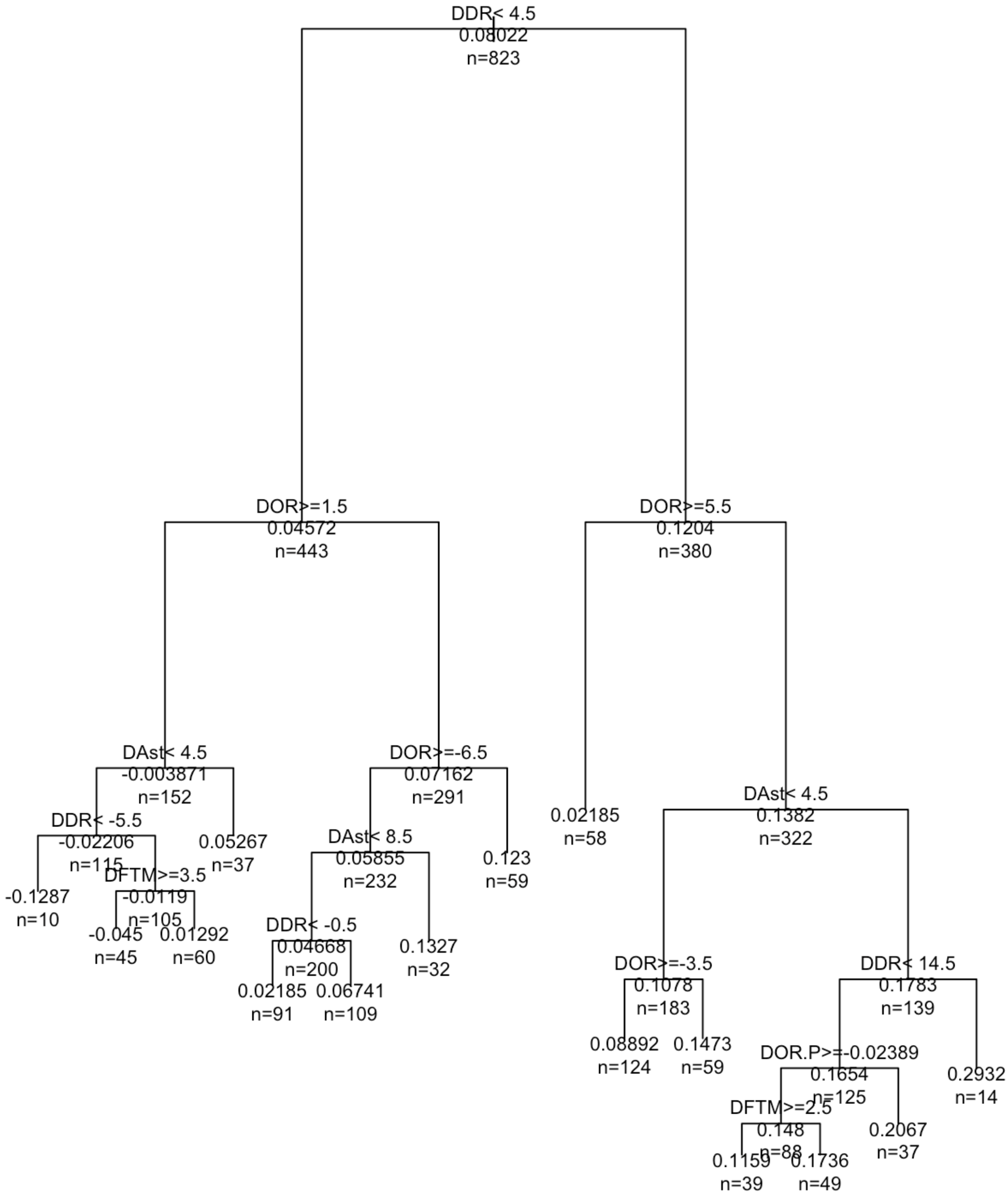


Figure EFG-4: Decision tree for the difference in EFG%. Splits occur for the difference in defensive rebounds, difference in offensive rebounds, difference in assists, and difference in free throws made.

Turnover Analysis Model results

Table TO-1: Linear regression of the difference in turnover percentage formulated based on correlation analysis

Variable	Coefficient	P-value
Intercept	-0.015	< 0.001
Winner Home Game	-0.017	< 0.001
Winner Neutral Game	-0.014	< 0.001
Difference in Field Goal Attempts	-0.003	< 0.001
Difference in Assists	-0.001	< 0.001
Difference in Steals	-0.007	< 0.001
Difference in Defensive Rebounds	0.004	< 0.001

Table TO-2: Linear regression of the difference in turnover percentage formulated based on the LASSO regression

Variable	Coefficient	P-value
Intercept	-0.002	0.237
Winner Home Game	-0.005	< 0.001
Winner Neutral Game	-0.003	0.164
Difference in Field Goal Attempts	-0.008	< 0.001
Difference in Free Throw Attempts	-0.003	< 0.001
Difference in Steals	-0.003	< 0.001
Difference in Offensive Rebounds	0.008	< 0.001
Difference in Defensive Rebounds	0.001	0.005
Difference in Defensive Rating	0.002	< 0.001
Difference in True Shooting Percent	0.289	< 0.001
Difference in Personal Fouls	< 0.001	0.030

Table TO-3: Linear Regression on the difference in turnovers formulated based on the LASSO regression

Variable	Coefficient	P-value
Intercept	3.16	< 0.001
Winner Home Game	-0.445	0.001
Winner Neutral Game	-0.385	0.101
Game Pace	-0.05	< 0.001
Difference in Field Goal Attempts	-0.254	< 0.001
Difference in Steals	-0.346	< 0.001
Difference in Three Point Field Goal Attempts	0.068	< 0.001
Difference in Defensive Rebounds	0.435	< 0.001
Difference in Offensive Rebounds	0.245	< 0.001
Difference in Personal Fouls	0.184	< 0.001
Difference in Three Point Field Goal Percentage	4.38	< 0.001
Difference in Defensive Rating	0.136	< 0.001

Table TO-4: Linear regression on the difference in turnover percentage based on the LASSO regression where the explanatory variables are restricted to those metrics not used in other "Big Four" Factors

Variable	Coefficient	P-value
Intercept	0.012	< 0.001
Winner Home Game	-0.019	< 0.001
Winner Neutral Game	-0.019	0.001
Difference in Steals	-0.012	< 0.001
Difference in Blocks	0.001	0.006
Difference in Defensive Rebounds	<0.001	0.004

References

Oliver, D. (2011). *Basketball on paper: rules and tools for performance analysis*. Potomac Books, Inc.

Zimmermann, A. (2016). *Basketball predictions in the NCAAB and NBA: Similarities and differences.* Statistical Analysis and Data Mining: The ASA Data Science Journal, 9(5)

Zullo, A. (2006). *March to madness: the ultimate fan's trivia book*. Running Press.