NBA Personal Project

Brady Biehn

April 2025

1 Background

Throughout the early 2000s the National Basketball Association (NBA) implemented various rule changes that limited defenses to promote offense in the game. The league hoped that an increase in points would combat declining ratings following Michael Jordan's second retirement in 1998. Defenses had become increasingly effective over the years, culminating in the statistically best defensive year league-wide in the 2004 NBA season. The two most notable rule changes implemented by the league were the defensive three-second rule (the defender cannot stand within the paint for longer than 3 seconds) and the banning of hand checking (the defender uses their hands to impede the movement of the opponent's ball handler) following the 2001 and 2004, respectively. The game continued to evolve when by the mid-2010s, head coaches Steve Kerr and Mike D'Antoni revolutionized the NBA landscape with their teams, the Golden State Warriors and Houston Rockets, respectively. They pioneered a modern offensive strategy centered on spacing and the three-point shot. Unlike traditional methods that emphasized inside play before moving the ball to the perimeter, these coaches made the three-pointer the primary scoring weapon in their offenses. The purpose of this paper is to use linear regression to quantify how these developments have changed the scoring and defense in the NBA.

2 Data

The data used for this analysis come from an NBA statistics repository compiled by GitHub user Brescou (github.com/Brescou). The data set contains information on the advanced, standard, and defensive stats of a team for 27 NBA regular seasons. It spanned from the 1996-1997 to the 2022-23 seasons.

This analysis aims to explore the impact of increased three-point shooting and changes in defensive rules on scoring, defense, and overall team success. To achieve this, a multiple linear regression approach is employed, using key performance metrics, such as total points and defensive rating, as response variables to evaluate how team performance has evolved over seasons. These metrics were averaged for each team over a season, with each team's season serving as an observational unit. The remaining explanatory variables used in the analysis are all numerical, except for the season (factor). All variables used in the analysis and modeling are shown and defined below.

- Defensive Rating (Amount of points a defense allows per 100 possessions) (DefRating)
- Total Points Scored on Offense (Pts)
- Three-Pointers Attempted (Fg3A)
- Two-Pointers Attempted (Fg2A)
- Free-Throws Attempted (FtA)
- Opposing Team's Field Goals Attempted (OppFgA)
- Opposing Team's Three-Pointers Attempted (OppFg3A)
- Opposing Team's Two-Pointers Attempted (OppFg2A)
- Opposing Team's Free-Throws Attempted(OppFtA)
- Season (factor) with the 2003-04 season serving as the baseline)
- Team name

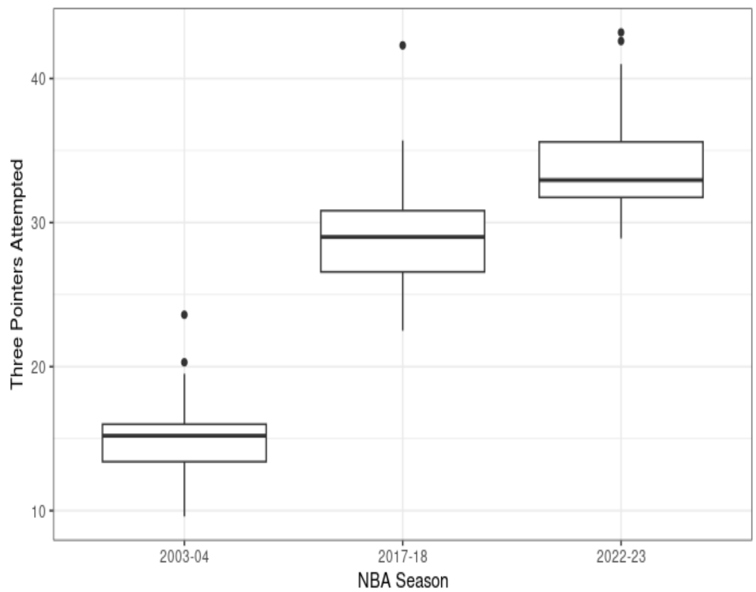
Before presenting the actual models, it is important to first examine how season trends influence response variables individually. This analysis will provide valuable context for understanding broader patterns and insights. The summary statistics for both variables by season are shown in Table 1. These are used when discussing the distributions for the individual response variables.

Table 1: Five-Number Summary

Variable	Min	Q1	Median	Q3	Max
FG3A					
2003-04	9.60	13.40	15.20	16.00	23.60
2017-18	22.50	26.58	29.00	30.83	42.30
2022-23	28.90	31.75	32.95	35.60	43.20
Defensive Rating					
2003-04	93.10	100.10	101.40	103.50	108.80
2017-18	103.00	105.95	108.05	109.93	111.70
2022-23	109.90	112.73	113.50	115.85	119.60

The distribution of three-point shot attempts by season is shown in Figure 1. For this distribution, the 2022-23 season appears to be right-skewed, the 2017-18 season symmetric, and the 2003-04 season is left-skewed. The 2003-04 season has a median of 15.20 shot attempts per game, and the 2017-18 and 2022-23 seasons have much greater medians of 29.00 and 32.95 shot attempts per game, respectively. This shows that three-point shooting has increased over time, since

Figure 1: Distribution of Three Pointers Attempted by Season

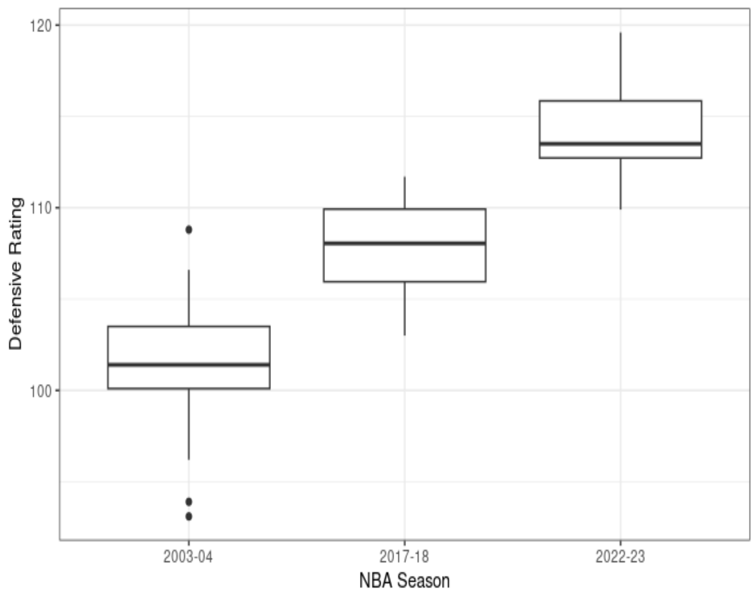


the median number of attempts has doubled over 20 seasons. Across all these seasons, the 2003-04 season has the smallest interquartile range (IQR) of 2.60, the 2022-23 season was in the middle with an IQR of 3.85, and the 2017-18 season has the largest IQR of 4.25 shot attempts per game. There is not much difference in the spread between seasons. This result is important because it indicates that the increase in the median number of shots is league-wide. If only a select number of teams had increased their amount of three-point shot attempts per game while the rest of the league did not, a much greater IQR value would be expected. Each season includes at least one team as an outlier for high-volume three-point shooting. The most notable outlier was the 2017-18 Houston Rockets, who shot a staggering 42.3 threes per game. This volume of three-point attempts has become increasingly prevalent league-wide as the median increases (four teams shot that amount of threes per game in 2022-23). However, at the time, no one was close to the Houston Rockets, who truly pioneered the three-point revolution. The 2017-18 Rockets would eventually lose in Game 7 of the Western Conference Finals when they missed an NBA playoff-record 27 STRAIGHT threes, which supports the old saying "live by the three, die by the three."

The distribution of defensive rating by season is shown in Figure 2.

For the distribution of defensive rating by season, the 2003-04 and 2017-18 seasons appear relatively symmetric, while the 2022-23 season appears to be right-slewed. The lowest median defensive rating is 101.40 points per 100 possessions for the 2003-04 season, the highest median is 113.50 points per 100

Figure 2: Distribution of Defensive Rating by Season



possessions for the 2022-23 season, and the 2017-18 season falls between them with a median of 108.05 points per 100 possessions. This demonstrates that the defensive rule changes in the early 2000s were effective in promoting scoring, as evidenced by the steady increase in median defensive ratings in the subsequent seasons. The IQRs were quite similar across the three seasons, with the 2017-18 season having the highest IQR of 3.98 points per 100 possessions and the 2022-23 season having the lowest at 3.125 points per 100 possessions. This fact indicates that defenses are deteriorating across the league as a whole because if only a few teams had declining defenses, then a much greater spread would be expected. There are at least three apparent outliers for the 2003-04 season and no outliers for the 2017-18 or the 2022-23 NBA seasons. These outliers include the defending NBA champion San Antonio Spurs with a defensive rating of 93.1, the eventual NBA Champion Detroit Pistons with a rating of 93.9, and the league-worst Orlando Magic with a defensive rating of 108.8. It should be noted that the abysmal defensive rating of 108.8 for the 2003-04 Orlando Magic would have been the best in the league for the 2022-23 season, further illustrating that defenses have significantly worsened league-wide in recent years.

3 Methods

Instead of using all 27 seasons in the dataset, this analysis focuses on the 2003-04, 2017-18, and 2022-23 NBA seasons. The 2003-04 season is the last season before the last of the two major defensive rule changes were enacted. Therefore, this season serves as the baseline for this analysis. The 2017-18 season was included because it captures the rise of three-point-centric offenses, as seen with Steve Kerr's Warriors and Mike D'Antoni's Rockets. Lastly, the 2022-23 season is the most recent in the data, offering insight into the current style of play. Two multiple linear models were fit to predict the amount of points scored by a team and the defensive rating of a team. These models aim to see how the evolution of the play style and rules has influenced these factors.

The population regression model for the Points Scored Model is defined in Equation 1:

$$PointsScored = \beta_0 + \beta_1(FG3A) + \beta_2(FG2A) + \beta_3(FTA) + \epsilon \tag{1}$$

This model uses the number of three-point attempts (FG3A), two-point attempts (FG2A), and free-throw attempts (FTA) as explanatory variables to predict the scored points.nts. The goal is to assess the relative impact of different types of shots on scoring efficiency and to determine whether the modern three-point dominant offensive approach is statistically justified or if alternative strategies may be more effective.

The population model for the defensive model is defined in Equation 2:

$$\begin{aligned}\text{DefensiveRating} = \beta_0 &+ \beta_1(\text{OppFG3A}) + \beta_2(\text{OppFG2A}) + \beta_3(\text{OppFTA}) \\ &+ \beta_4(D_{\text{Season}}) + \beta_5(\text{OppFG3A} \times \text{Season}) \\ &+ \beta_6(\text{OppFG2A} \times \text{Season}) + \beta_7(\text{OppFTA} \times \text{Season}) + \epsilon\end{aligned} \tag{2}$$

This model examines how the opponent's shot selection has impacted the defensive rating over time. In addition, it includes interaction terms between these variables and season dummies (D_{Season}) to capture how these relationships evolve over time. Dummy variables were created for the selected seasons (2017-18 and 2022-22), and the 2003-04 season served as the baseline for comparison.

4 Discussion

The resulting estimated regression line for the Points Scored Model is shown below in Equation 3.

$$\widehat{PointsScored} = -16.08 + 1.63FG3A + 1.03FG2A + 0.76FTA \tag{3}$$

All four assumptions for linear are reasonably met and multicollinearity appeared to not be an issue the interpretation of the fitted values can proceed (see Appendix A for further details).

The first test conducted was an F test to determine whether the explanatory variables were overall effective in predicting the average points scored per game by a team. With a p-value of ¡0.0001 (F = 256.7, df = 3 and 85), there is very strong evidence to suggest that at least one of the average numbers of attempts of free throw (either free throw,ee throw, two-point, or three-point) by a team contributes to predicting the number of points a team scores per game, after adjusting for simultaneous changes in the other variables. In addition, partial F tests were conducted to test the individual significance of each variable in predicting total points. For three-point shot attempts, the resulting p-value was less than 0.0001 (F = 357.24, df = 1 and 85); for two-point shot attempts, the p-value was less than 0.0001 (F = 68.87, df = 1 and 85); and for free throw attempts, the p-value was less than 0.0001 (F = 26.918, df = 1 and 85). Therefore, all the averages of these three types of shots are statistically significant in predicting the average number of points a team scores.

Next, the focus of the analysis shifts to measure the individual effect that all three type shot attempts had on predicting the average number of points a team scores.

The slope coefficient for three-point attempts was 1.48. This means that for every additional three attempted, the team is estimated to score an additional 1.48 points per game (95% confidence interval: 1.47 to 1.81 points per game), on average. The slope coefficient for two-point attempts was 1.03, which means that a team is estimated, on average, to score 1.03 more points per game with every two-point shot (95% confidence interval: 0.78 to 1.28 points per game). Lastly, the coefficient for free-throw attempts was 0.76. This slope dictates that the estimated average points per game increased by 0.76 points (95% confidence interval: 0.47 to 1.05 points per game) per game for every additional free-throw attempted per game. For each of these slope coefficients, the effect is after adjusting for simultaneous changes in the other two explanatory variables. Due to three-point shot attempts have the largest estimated slope coefficient and confidence interval, this model suggests that three-point shot attempts have the greatest effect on scoring compared to two-point and free-throw attempts. As a result, this model indicates that prioritizing three-point shot attempts would be optimal for scoring points. This may also help explain the increasing emphasis on three-point shooting in the modern game.

Next, let us examine the Defensive Model. The resulting regression line for each season is shown in Table 2. Additionally, the parameter estimates, standard error, t-statistics, and p-values are shown in Appendix B. The four assumptions of linear regression were reasonably met and multicolinearity was not a concern (see Appendix C for further details). The first test conducted on the model was an overall F-test. This test indicated that there was very strong evidence to suggest that at least one of the explanatory variables (NBA season, number of one-point, two-point, or three-point shot attempts allowed) is useful in predicting a team's defensive rating (p-value ¡ 0.0001, F = 46.35, df = 11 and 77). Next, the attention turns to what individual variables are useful in predicting defensive rating. To this partial F-test are conducted, due to the principle of marginality, the tests are initially conducted on the higher-order interaction terms to test their effectiveness.

The first interaction term tested was the one between free-throw shooting and season.

With a p-value of 0.30 (F = 0.89, df = 2 and 77), the interaction term between OppFtA and season was not found to be statistically significant. This means that the effect that OppFtA had on defensive rating did not depend on the season. Consequently, this interaction term was removed from subsequent models. However, the interaction terms between OppFg3A (p-value = 0.00042, F = 5.87, df = 2 and 77) and OppFg2A (p-value = 0.019, F = 4.18, df = 2 and 77) by season were both found to be statistically significant. This indicates that the effect of both the average three-point and two-point shots that the opposing team took on defensive rating depended on the season.

Following these partial F-tests, the final model contains all main effects and interaction terms expect the free-throw and season interaction term. The attention now turns to see the individual effect all of the explanatory variables had on defensive rating.

Table 2: Estimated Regression Line by Season

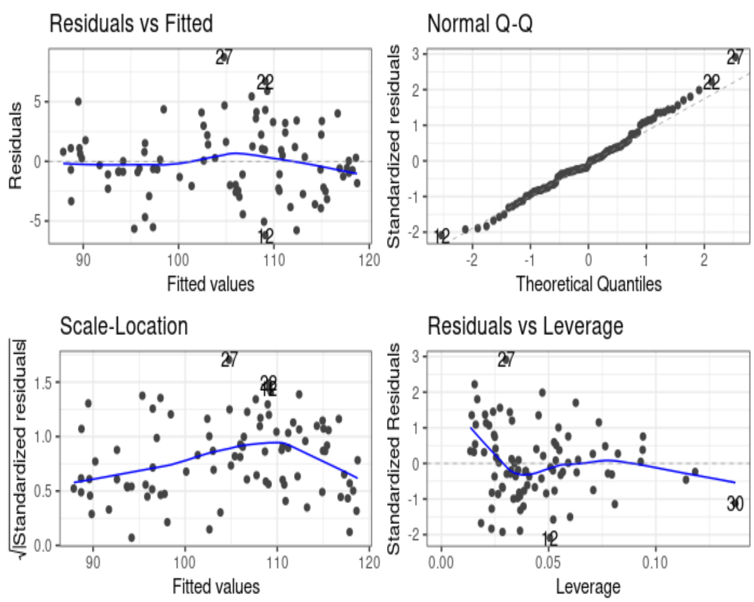
Season	Estimated Regression Line
2003-04	$\widehat{DefensiveRating} = 19.61 + 1.44(OppFG3A) + 0.75(OppFG2A) + 0.49(OppFTA)$
2017-18	$\widehat{DefensiveRating} = 68.24 + 0.78(OppFG3A) + 0.12(OppFG2A) + 0.49(OppFTA)$
2022-23	$\widehat{DefensiveRating} = 86.60 + 0.13(OppFG3A) + 0.21(OppFG2A) + 0.49(OppFTA)$

For the 2003-04 NBA season, the estimated average intercept was the lowest across all three of the estimated season with a estimated average defensive rating 19.61 compared to the 2017-18 seasons estimated intercept out of 68.24 and the 2022-23 season's estimated intercept of 86.60. The estimated average defensive rating increases by 48.64 (95% confidence interval: 9.58 to 87.70) for the 2017-18 season, and the same metric increased by 66.99 (95% confidence interval: 28.037 to 105.94) points per 100 possessions. These values are heavily extrapolated, but since 2003-04 has the lowest defensive rating and 2022-23 season has the highest value this suggest that defensive rating have grown worse over time. The estimated average increase in defensive rating is 1.44 (95% confidence interval: 0.91 to 1.96), 0.78, and 0.13 points allowed per 100 possessions for every three-point shot by the opposing team for the 2003-04, 2017-18, and 2022-23 NBA seasons, respectively. There was a decrease of 0.66 (95% confidence interval: -1.39 to 0.078) and 1.30 (95% confidence interval: -2.02 to -0.57) points allowed per 100 possessions on the estimated average effect that every additional three-point shoot allowed on defense had the defensive rating. These values suggests that the value of three-point shoots have decreased over the years. However, for two of the three models three point shooting still had the greatest impact on defensive rating. On reason for this is because when the teams turned to three-point shooting the floor became where more spaced out than before. Therefore, it a league consisting of the best basketball players on Earth, this additional spacing allowed for shooters to get more open shot. It also made much harder for defenders to cover the court. For these reasons, it allowed offenses to become much more efficient. This may be one of the reasons why defensive rating has declined in recent years. It may be that defensive have not gotten worse of the years, but that offenses have significantly better and the decline in defensive rating is a corollary of this newfound offensive efficiency. For every additional attempt at two points on a shot allowed, the estimated defensive rating of a team increases by 0.75 (95% confidence interval: 0.40 to 1.089), 0.12, 0.21 points allowed per 100 for the 2003-04, 2017-18 and 2022-23 seasons, respectively. Lastly, for every free-throw attempted allowed the average estimated defensive rating increases by 0.49 (95% confidence interval: 0.22 to 0.77) points per 100 possessions across all three selected seasons. One final interesting thing to note is that the effect of three-point shooting on defensive rating is smaller that the effect that two-point shots allowed had in the 2022-23 season. This fact suggests that three-point shooting is becoming less valuable. There are two likely factors causing this outcome. The first is that the long mid-range shot is becoming abandoned league-wide. This shot is by far the least efficient shot for a team to shoot, because all the player needs to do is take one step back to take a three-point, and as the Points Scored Model suggested this type of shot is the most effective. Therefore, these highly invaluable long mid-range are abandoned, hence only the most valuable two-points are being taken by teams, such as layups and dunks which increases the value in two-point shots. Secondly, three-point shooting is becoming over saturated. In the 2003-04 season, only the premier three-point shooters on teams were permitted to take a high volume of threes. Compared to the current game, only a selected number of particularly horrible shooters are not permitted to take a high volume of threes, while the rest of the time is permitted (the latter group consists of the vast majority of the league). Therefore, the three-point shot is becoming over saturated with average-to-bad three-point shooters causing to appear less valuable.

5 Appendix A: Points Model Assumption Checking

The diagnostic plots for the Points Model are shown in Figure A1. For the Points-Scored Model, the diagnostic plots are shown in Figure A1. The linearity assumption is reasonably met because the blue line follows the y = 0 line almost exactly in the residuals versus fitted plot. The normality assumption is reasonably met because the points follow the dashed line extremely well in the Q-Q plot. The equal variance assumption is reasonably met because there are no clusters or groupings in the residuals versus fitted plot. Lastly, there are potential concerns about the assumption of independence. However, because season averages were used across all games and the teams were separated by at least five seasons, this assumption is considered reasonably met. Therefore, accurate estimates are expected for this model because all four assumptions of linear regression are reasonably met.

Figure A1: Diagnostic Plots for the Points Scored Model



To ensure that the predicted variables are interpretable, multicollinearity was tested. If multicollinearity is present that means most of the variability in the explanatory variable is explained by some other variable(s) hence the values become less interpretable. To test for multicollinearity, the variance inflation factor (VIF) was calculated. The higher a VIF the more problematic it becomes to get an accurate interpretation of that respective explanatory variable. The general threshold for VIFs is that values over 10 indicate significant multicollinearity and values under 5 are acceptable. The resulting VIFs for the Points

Scored Model were 5.66 for the three-point shots, 5.53 for two-point shots, and 1.06 for free-throw shots. Despite the Fg2A and Fg3A variables being greater than 5, since there are close enough to 5, these variables will remain and be interpreted within the model.

6 Appendix B: Defensive Model Parameter Estimates

Table B1 presents the parameter estimates of the Defensive Model, with the season, the three-point, two-point, and free-throw attempts of the opposing team serving as explanatory variables. In addition to these variables, the estimates for the interaction terms for both opposing team's two-point and three-point shot attempts by season is also included.

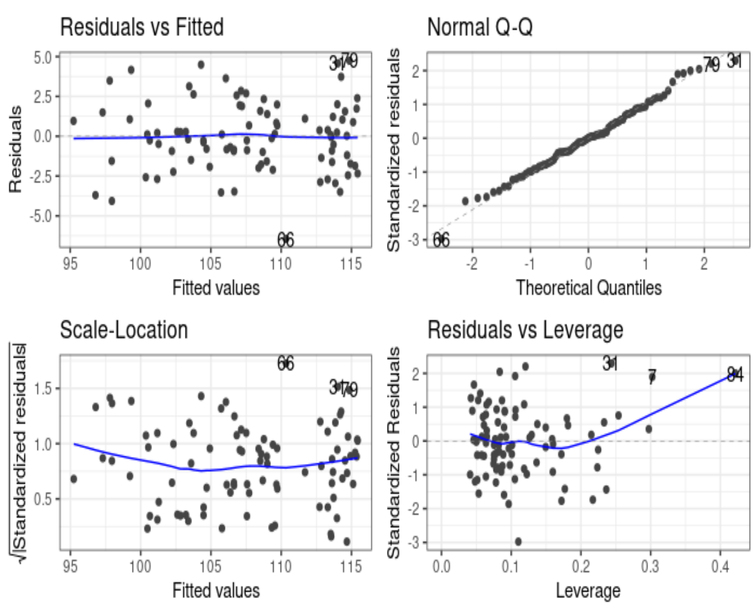
Table B1: Defensive Model Parameter Estimates

	Estimate	Standard Error	t-Statistic	p-value
Intercept	19.61	13.73	1.43	0.16
OppFg3A	1.44	0.26	5.43	$<$0.0001
OppFg2A	0.75	0.17	4.33	$<$0.0001
OppFtA	0.49	0.14	3.58	$<$0.001
Season				
2017-18	48.64	19.62	2.48	0.015
2022-23	66.99	19.57	3.42	$<$0.001
OppFg3A * Season				
2017-18	-0.66	0.37	-1.78	0.079
2022-23	-1.30	0.37	-3.56	$<$0.001
OppFg2A * Season				
2017-18	-0.63	0.24	-2.63	0.010
2022-23	-0.54	0.24	-2.27	0.026

7 Appendix C: Defense Model Assumption Checking

The diagnostic plot for the Defense Model is shown in Figure B1. The first assumption tested is normality. In the Q-Q plot, the points for the observed residuals closely follow the dashed line for the theoretical residuals extremely well. Furthermore, there are a large number of data points, with n = 89. Due to both of these factors, the normality assumption is reasonably met. Next, the linearity assumption is tested. The blue line in the Residuals versus Fitted graph is almost horizontal at y = 0, which is what is desired. Therefore, this condition is reasonably satisfied. After that, the equal variance was tested. In the residuals versus fitted graph, the variance was relatively equal across. There does not appear to be anything overtly concerning, so this condition is reasonably met. Lastly, there remains the same concern regarding the independence assumption. However, for the same reasons listed in Appendix A, it is assumed that the assumption is reasonably met. Therefore, all four assumptions are reasonably met for the defense model, thus accurate estimates are expected for the Defense Model.

Figure C1: Diagnostic Plots For the Defensive Model



The Defense Model was also tested for multicollinearity. However, since high-order terms were present in the model (the interactions terms), the generalized variance inflation factor (GVIF) was calculated instead ($GVIF^(1/(2*Df))$ more specifically). For GVIFs the threshold is 10 for acceptable and interpretable results (instead of 5 for VIFs). The resulting GVIF values were 3.65 for OppFg3A, 3.091 for OppFg2A, 1.21 for OppFtA, and 1.024 for seasons. All of these values are way below the threshold. Therefore, multicollinearity does not appear to be a problem in this model.