Writeup

Sunil Butler

May 1, 2020

In recent years, many have noticed a trend in basketball, particularly in the Nation Basketball Association (NBA) toward higher frequency perimeter shooting. This trend is generally credited both to the popularization of modern analytics applied to team strategies and to the success of the Golden State Warriors and Houston Rockets. This paper explores the possibility that this shift has been detrimental to the offensive performance of at least a singificant portion, if not most players in the NBA.

We use extensive compiled statistics from the 2018-2019 NBA regular season. We limit our scope to this season for two reasons: first, compiling and cleaning datasets of this size is time consuming for what is essentially a preliminary investigation. Second, if any models are created in the course of this study, having separate data sets from other seasons to test it against for validation purposes will be useful.

Investigation of Direct Relationship Between 3-Points Shooting Practices and Offensive Rating

Our data is compiled from NBAStuffer and BasketballReference. The data is gathered from various spreadsheets for the 2018-2019 regular season and combind by player. We omit several players with either multiple missing values or clear data entry errors; for example, Jordan Sibert's 2018-2019 offensive rating is listed as 300 and his scoring average is listed at zero. Foruntately, of the 5-15 players removed in each data set, none played for more than 5 minutes per game, making the loss of information minimal.

We start by looking at the relationship between players' three point attempts and their offensive ratings. An offensive rating is a widely used statistic measuring the point produced by a player per 100 possessions. This statistic can sometimes punish the best scorer on a team who is often forced to take low-percentage shots on unsuccessful plays, and can sometimes disproportionately reward post players or catch-and-shoot players who receive more open looks. Nonetheless, when analyzing large numbers of players, offensive rating is a good general measure of offensive performance.

Some modern analytics have suggested that players who shoot more 3-pointers often have better offensive performance. However, investigation suggests that this may be because players with high shooting percentages take a lot of perimeter shots, thus making them outliers. We can investigate this by controlling for 3-point percentage.

```r
#Three point attempts standardized by game
TPA_Min <- 48*Data_Main$`3PA`/(Data_Main$MPG*Data_Main$GP)

#Three point percentage
TPP <- Data_Main$`3P%`*100

fit_1 <- lm(Data_Main$ORTG ~ TPA_Min + TPP)

summary(fit_1)$coef
```

```
##                Estimate Std. Error   t value      Pr(>|t|)
## (Intercept) 105.3446426  1.3394876 78.645477 6.637918e-316
## TPA_Min      -1.2505743  0.1726875 -7.241833  1.380507e-12
## TPP           0.3358945  0.0425040  7.902657  1.334743e-14
```

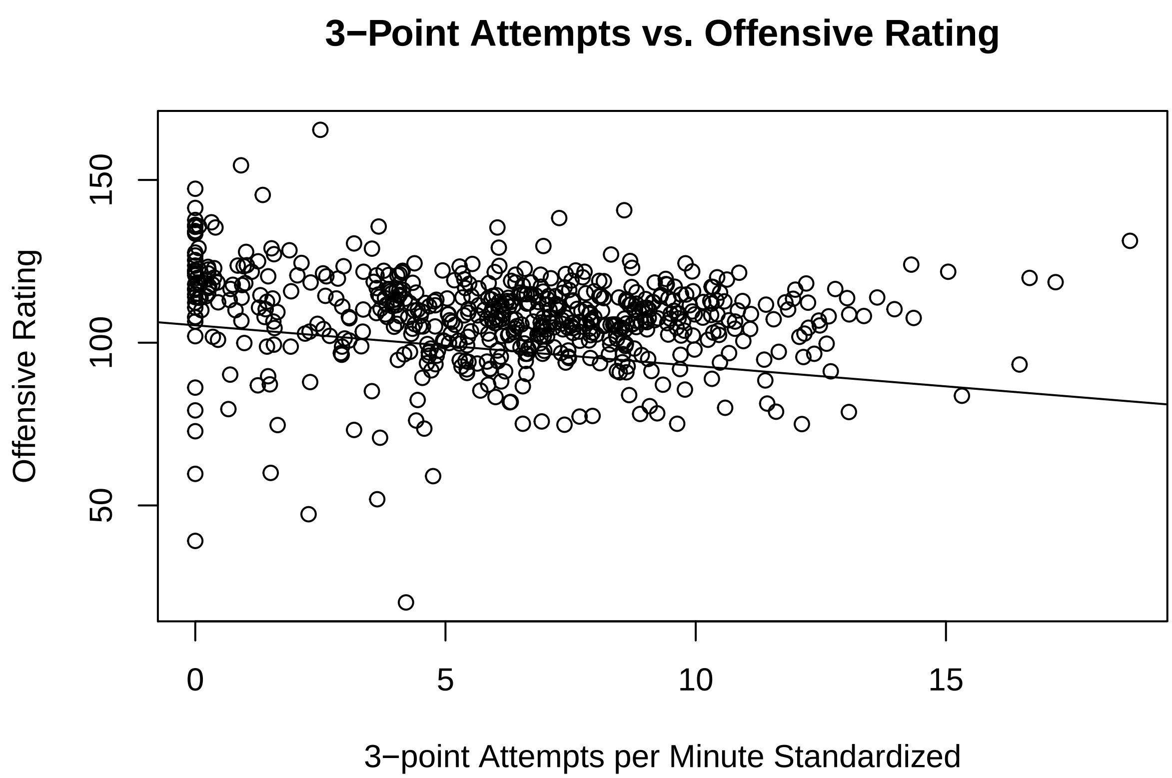
In this regression output, the variable TPA is rate of player 3-point attempts standardized by the length of a game for interpretability. It can be thought of as the number of three point shots a players would take, if that player played for the entire game. The variable TPP is the player's 3-point percentage.

The negative slope of TPA suggests that the higher the rate of 3-point attempts by a player, the lower the player's offensive rating, on average. Specifically, the slope coefficient implies that, if 3-point percentage is held constant, adding one 3-point attempts per 48 minutes would reduce that player's offensive rating by 1.33. The relationship is statistically significant at even the most stringent significance levels (we'll use $\alpha = 0.01$). Meanwhile, 3-point shooting percentage (TPP) shares a positive relationship with offensive rating with similarly clear statistical significance.

A visual display of the cross-section of the data containing 3-point attempts and offensive rating shows that while there is a lot of variance in the data, there is a clear general downward trends.

```
plot(TPA_Min,Data_Main$ORTG, xlab = '3-point Attempts per Minute Standardized',
     ylab = 'Offensive Rating', main = '3-Point Attempts vs. Offensive Rating')

abline(a = summary(fit_1)$coef[1], b = summary(fit_1)$coef[2])
```



We also look at the relationship between our two independent variables.

```
summary(lm(TPA_Min ~ TPP))$coef
```

```
##              Estimate  Std. Error   t value     Pr(>|t|)
## (Intercept) 2.7659141 0.293899951  9.411074 1.053487e-19
## TPP         0.1123273 0.008947281 12.554351 3.052090e-32
```

So players with higher 3-point percentages tended to take more 3-point shots. The realtionship is again statistically significant at the $\alpha = 0.01$ level. This could raise concerns of multicolinearity in our earlier model, especially given that the two variables share a relatively high correlation of 0.46. However, in this case the symptoms of multicolinearity would be an artificially positive relationship between three point attempts and offensive rating. We see the opposite in this case, implying that multicolinearity is not the underlying cause of our results.

While the slope coefficient for our second model may seem small, keep in mind the scaling of the variables; a nine percent increase in 3-point percentage encouraging a player to take about 1 additional shot every 48 minutes is a reasonable order of magnitude.

Investigation of Significance Accounting for Player Minutes

Before further investiagation, there is a major concern worth addressing: our investigation focuses on individual players because it seek to look at the results of player behavior in a general sense. In doing this, it doesn't take into account the impact of each player on the game, but rather weights each player equally. This could raise questions regarding how important this study really is to the outcome of individual games. We repeat the fitting of our above model with each player weighted by the number of minutes played. We could also weight each player by winshares, but this would put us in a sticky situation since our other independent variables are factored into already.

```
summary(lm(Data_Main$ORTG ~ TPA_Min + TPP, weights = Data_Main$MPG))$coef
```

```
##                Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 108.6039333 1.27624665 85.096351 0.000000e+00
## TPA_Min      -1.0554243 0.15015301 -7.028992 5.739964e-12
## TPP           0.2404276 0.04296551  5.595827 3.357777e-08
```

Our results are still significant and at a similar level. Our conclusion is that the relationship between three point attempts and offensive rating is singificant even when accounting for minutes played, and thus that our investigation is justified even from the perspective of someone who is mainly interested in impact on the final score.

Investigation of Changes in Relationship Based on Player Position

The increase in 3-point shooting in the NBA is most apparent when looking at 'big men', or centers and center/forward hybrids. While guards and conventional centers are undeniably shooting more than they used to, big men, who in the past could finish their career without a single 3-point attempt, are now expected to develop their permieter shooting skills. In this section we investigate how the relationship between 3-point attempts and offensive rating varies depending on position.

Some positions are prone to higher offensive rating than others; post-heavy centers and catch-and-shoot guards, for example, tend to have higher shooting efficiency due to the types of shots that become avilable to them and hence typically have a higher offensive rating. Therefore, we want to account for not only the interaction of position and 3-point attempts, but also the effect of position by itself.

```
summary(lm(Data_Main$ORTG ~ TPP + TPA_Min*Data_Main$POS))$coef
```

```
##                          Estimate  Std. Error   t value      Pr(>|t|)
## (Intercept)           116.0742286  1.87491538 61.909049 5.336814e-258
## TPP                     0.4546051  0.03972259 11.444496  1.721685e-27
## TPA_Min                -3.6935151  0.60927816 -6.062116  2.418740e-09
## Data_Main$POSC-F       14.5535824  5.71000761  2.548785  1.106590e-02
## Data_Main$POSF        -15.6989265  2.66507646 -5.890610  6.517428e-09
## Data_Main$POSF-C       -7.5666932  3.49399678 -2.165627  3.074587e-02
## Data_Main$POSF-G      -25.5254123 13.18506744 -1.935933  5.336059e-02
```

```
## Data_Main$POSG           -29.3966365  2.85135224 -10.309718  5.308181e-23
## Data_Main$POSG-F         -22.5012494  4.93859803  -4.556202  6.351549e-06
## TPA_Min:Data_Main$POSC-F -11.8957991  2.54664969  -4.671156  3.724627e-06
## TPA_Min:Data_Main$POSF     2.6788852  0.65531586   4.087930  4.965562e-05
## TPA_Min:Data_Main$POSF-C   2.0785723  0.76384986   2.721179  6.699926e-03
## TPA_Min:Data_Main$POSF-G   3.1941818  1.74162199   1.834027  6.716126e-02
## TPA_Min:Data_Main$POSG     4.1353142  0.65699201   6.294314  6.087303e-10
## TPA_Min:Data_Main$POSG-F   3.1072798  0.81560927   3.809765  1.539171e-04
```

The effect on offensive rating for each additional 3-point attempt can be calculated by adding the slope term and the interaction term. We still see an overall negative effect of three-point attempts on offensive rating for every position except guard. However, this effect iso small that calculating the standard error on each sum independently yields results that are not statistically significant except for at the center and center/forward positions. Thus, we can attribute the majority if not all of the negative relationship between three-point attempts and offensive rating to the 'big men', or those playing center and center/forward hybrids.

The intuitively obvious explanation is that guards and forwards tend to be better shooters, thus they would benefit from taking more perimeter shots; however, keep in mind that all our models have accounted for three-point percentage. So even among players with similar three-point percentages, big men's offensive ratings tend to be lower when they take more 3-point shots.

It helps to consider the nature of a players offensive rating. What is boils down to is 'how often does the player score when he tries to score?' With this is mind, we explore several possible explanations:

1. Missed three point shots by centers and forwards yield fewer second chances. Since the center and/or forward is outside of the three-point line, he is less likely to be available to rebound the shot. His post presence will be missed more than that of a guard; therefore, his missed shots are less likely to return to him for a second attempt.
2. Three point attempts ending in a foul are less likely to pan out. Centers and forwards tend to have a lower free throw percentage. Therefore, centers who attempt three pointers may be more likely to be fouled, likely missing the shot, and are less likely the recover those points at the free throw line.
3. Three point attempts are in general less efficient than post play. Since centers and forwards who take frequent alley-oops and layups tend to have high offensive ratings, it could be that their three point attempts are not specifically disadvatageous, but simply not as beneficial as their regular style of play.

Investigation of Possibility 1

The first and most basic step is to determine whether centers/forwards who shoot more get fewer rebounds.

```r
#Players who played centers or center/forwards
CF <- which(Data_Main_Test$POS == 'C' | Data_Main_Test$POS == 'F-C')
#Offensive rebounds standardized by game
ORB_GM <- 48*Data_Main_Test$ORB[CF]/(Data_Main_Test$GP[CF]*Data_Main_Test$MPG[CF])
#Three point attempts standardized by game for centers and center/forwards
TPA_Min_Test <- 48*Data_Main_Test$`3PA`[CF]/(Data_Main_Test$MPG[CF]*Data_Main_Test$GP[CF])

summary(lm(ORB_GM ~ TPA_Min_Test))
```

```
##
## Call:
## lm(formula = ORB_GM ~ TPA_Min_Test)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -5.5255 -0.9252 -0.2143  0.8202  6.4745
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    5.5255     0.2530  21.841  < 2e-16 ***
## TPA_Min_Test  -0.3365     0.0556  -6.052 6.33e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.641 on 70 degrees of freedom
## Multiple R-squared:  0.3435, Adjusted R-squared:  0.3341
## F-statistic: 36.63 on 1 and 70 DF,  p-value: 6.334e-08
```

We see a substantial and statistically significant negative relationship between 3-point attempts and offensive rebounds in big men. While this might seem inuitively obvious, the most interesting component is the r^2 value. The relatively high value $r^2 = 0.34$ shows that a substantial portion of the variation in offensive rebounds of a big man can be explained if the know the player's rate of 3-point attempts. In fact, the r^2 implies a correlation of about -0.59 (this can be confirmed by calculation), which denotes a strong inverse correlation between 3-point attempts and offensive rebounds.

While this does suggest that big men who take more three point shots miss out on rebound chances, it doesnt necessarily suggest that this is the cause of the offensive rating relationship. First, we don't know how many of those offensive rebounds would have found their way back to the big men who shot them, so the lack of rebouding power may not have directly detracted from the player's rating. Second, offensive rebounds themselves are factored into a player's offensive rating. This raises a unique situation where a variable we had hoped was a confounding variable may be in a fact a causative variable.

Due to the complexity of its formula, deducing the exact relationship between offensive rebounds and offensive rating would require a multi-dimensional differential equation that would give us little intuitive insight. Instead we can investigate this through looking at the correlation between offensive rebounding and offensive rating, and the correlation between three-points attempts and offensive rating.

```r
cor(ORB_GM,Data_Main_Test$ORTG[CF],use = "pairwise.complete.obs")
```

```
## [1] 0.09062158
```

```r
cor(TPA_Min_Test,Data_Main_Test$ORTG[CF],use = "pairwise.complete.obs")
```

```
## [1] -0.3999343
```

We can estimate the secondhand effect of offensive rebounds on offensive rating through three point attempts. Keep in mind that this calculation can in no way suggest causation, but with proper logical justification, causation can be suggested for future analysis.

```r
cor(TPA_Min_Test,Data_Main_Test$ORTG[CF],use = "pairwise.complete.obs") * cor(ORB_GM, TPA_Min_Test)
```

```
## [1] 0.2344049
```

So the secondhand correlation between offensive rebounds and offensive rating through three point attempts is higher than the direct correlation between offensive rebounds and three points attempts. This provides compelling evidence for possibility 1, as it suggests that big men with more three-point attempts may grab fewer offensive rebound to the extent that it noticeably affects offensive rating.

THus, possibility 1 can be attributed some validity.

Investigation of Possibility 2

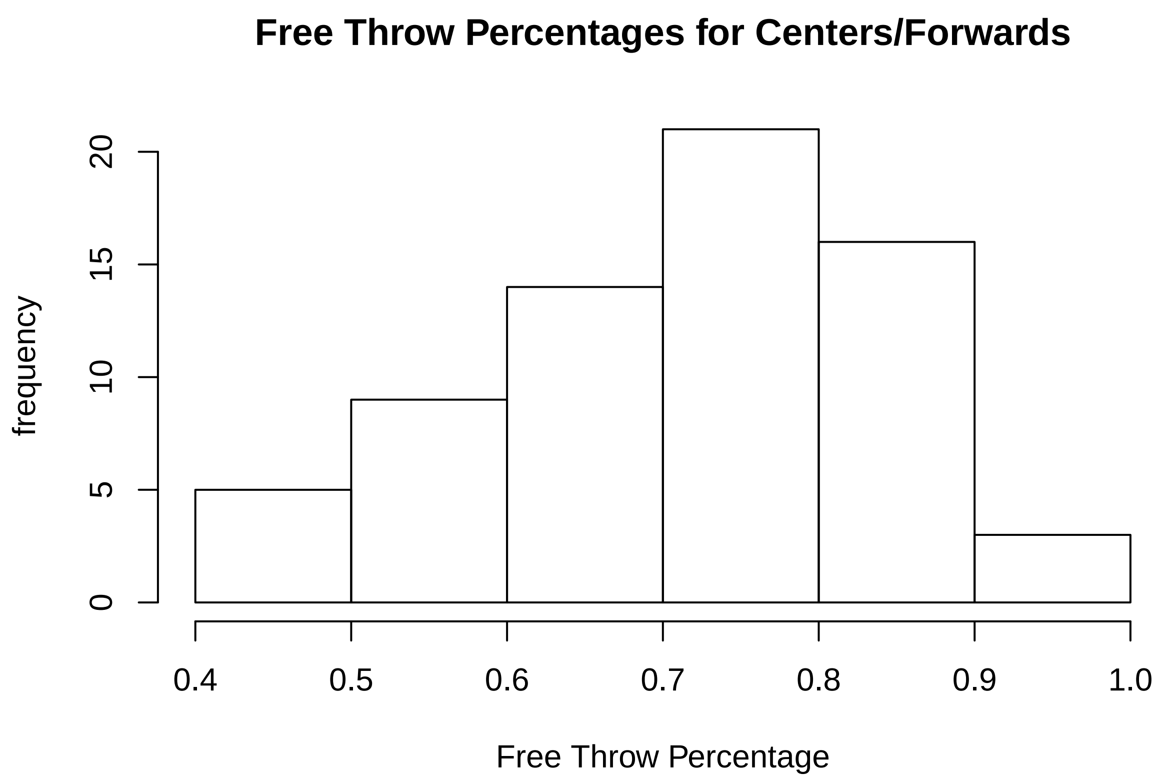
This possibility could be investigated more easily if we had access to information about player free throws broken up by the type of foul. However, we can still investigate this matter using the information we have access to, albeit in a more roundabout way.

We can first confirm our theory about free percentages by position with a two sample t-test to determine whether big men have a lower free throw percentage, on average, than other players. Due to the somewhat smaller size of one of our groups, we use $\alpha = 0.05$. Since we have 383 data points for non-big men, we do not need to worry about the normality assumption. For big men, we apply a Shapiro-Wilkes test and a visual assessment.

```
shapiro.test(Data_Main_Test$FT.[CF])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Data_Main_Test$FT.[CF]
## W = 0.97626, p-value = 0.22
```

```
hist(Data_Main_Test$FT.[CF],
     main = "Free Throw Percentages for Centers/Forwards",
     xlab = "Free Throw Percentage", ylab = "frequency")
```



Free throw percentages for centers/forwards is clearly close to a normal distribution.

We now apply a two-sample t-test.

```r
t.test(Data_Main_Test$FT.[CF], Data_Main_Test$FT.[-CF],
       alternative = "two.sided", var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  Data_Main_Test$FT.[CF] and Data_Main_Test$FT.[-CF]
## t = -1.9846, df = 100.58, p-value = 0.04992
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -6.430216e-02 -1.167135e-05
## sample estimates:
## mean of x mean of y
## 0.7152647 0.7474216
```

There is a singificant difference at the $\alpha = 0.05$ level. The question then becomes whether this apparent difference is sufficient to affect players' offensive ratings.

While we can't gain access to statistics on 3-point foul rates (at least not for free), we do know the league average 3-point foul rate consistently hovers around 1.6%. If we assume that the rate of 3-point fouls is relatively consistent between the two groups, we can estimate the number of points lost.

```r
#Estimated PPG from free throws off three point attempts, centers and center/forwards
PPG_TCF <- Data_Main_Test$`3PA`[CF]*0.016*3*Data_Main_Test$`FT%`[CF]/Data_Main_Test$GP[CF]

#Estimated PPG from free throws off three point attempts, other positions
PPG_TO <- Data_Main_Test$`3PA`[-CF]*0.016*3*Data_Main_Test$`FT%`[-CF]/Data_Main_Test$GP[-CF]

summary(PPG_TCF)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.00000 0.00000 0.01408 0.04376 0.07241 0.28994
```

```r
summary(PPG_TO)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.00000 0.04855 0.09517 0.10787 0.15421 0.55607
```

We can then use this along with regression on our data to estimate the effect on offensive rating.

```r
fit_CF <- lm(Data_Main_Test$ORTG[CF] ~ Data_Main_Test$PPG[CF])

fit_O <- lm(Data_Main_Test$ORTG[-CF] ~ Data_Main_Test$PPG[-CF])

#estimate (very roughly) the change in offensive rating
#from points made off three point fouls
OC_CF <- summary(fit_CF)$coef[2]*PPG_TCF
OC_O <- summary(fit_O)$coef[2]*PPG_TO

#We calculate the
t.test(OC_CF,OC_O,
       alternative = "two.sided", var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  OC_CF and OC_O
```

```
## t = -5.6777, df = 103.25, p-value = 1.263e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.05686927 -0.02742537
## sample estimates:
##  mean of x  mean of y
## 0.03931899 0.08146631
```

While the estimated amount that big men's offensive ratings increased off three point fouls was statistically singificantly less than that of other players, the actual difference, about 0.03, is not large enough to be interesting, and almost certainly not larger than the error induced by our questionable estimation method.

Thus, possibility two cannot be validated at the moment.

Investigation of Possibility 3

We now focus on variation within play styles. While we are most intersted in big men, the nature of the relationship we are investingating justifies the use of all players. We create several nested models measuring the relationship between 3 point attempts, 2 point attempts and effective field goal percentage while controlling for other variables.

Note that both 3 point and 2 point percentage is correlated with effective field goal percentage, so while we can effectively use it as a control, we shouldn't perform inference on it. Since we are using all our data, we use $\alpha = 0.01$.

```r
#Three/Two point attempts standardized by game
TPA_Min_EFF <- 48*Data_Main_Test$`3PA`/(Data_Main_Test$MPG*Data_Main_Test$GP)
TWPA_Min_EFF <- 48*Data_Main_Test$`2PA`/(Data_Main_Test$MPG*Data_Main_Test$GP)

#Effective field goal standardized to 0-100 scale
EFG_Std <- Data_Main_Test$`eFG%` * 100

#nested model with only three point attempts
fit_eff <- lm(EFG_Std ~ TPA_Min_EFF + TWPA_Min_EFF)
#nested model with 3 point percentage
fit_eff_3PP <- lm(EFG_Std ~ Data_Main_Test$`3P%`*TPA_Min_EFF + TPA_Min_EFF)
#nested model with 2 point percentage
fit_eff_2PP <- lm(EFG_Std ~ Data_Main_Test$`2P%`*TWPA_Min_EFF + TPA_Min_EFF)
#parent model 3 point and 2 point percentage
fit_eff_3PP_2PP <- lm(EFG_Std ~ Data_Main_Test$`3P%`*TPA_Min_EFF + Data_Main_Test$`2P%`*TWPA_Min_EFF)


summary(fit_eff)$coef
```

```
##                 Estimate Std. Error    t value      Pr(>|t|)
## (Intercept)  46.04279661  1.6519641 27.8715478 5.640652e-100
## TPA_Min_EFF  -0.06576508  0.1368069 -0.4807146  6.309535e-01
## TWPA_Min_EFF  0.42574921  0.1038419  4.0999768  4.905013e-05
```

```r
summary(fit_eff_3PP)$coef
```

```
##                                   Estimate Std. Error    t value
## (Intercept)                      54.263352  1.0109154  53.677443
## Data_Main_Test$`3P%`             -5.794682  3.5005833  -1.655348
## TPA_Min_EFF                      -4.003450  0.2584959 -15.487481
```

```
## Data_Main_Test$`3P%`:TPA_Min_EFF 10.754541  0.7723572  13.924309
##                                        Pr(>|t|)
## (Intercept)                        3.072182e-197
## Data_Main_Test$`3P%`                9.855432e-02
## TPA_Min_EFF                         1.227481e-43
## Data_Main_Test$`3P%`:TPA_Min_EFF    7.070613e-37
```

```r
summary(fit_eff_2PP)$coef
```

```
##                                      Estimate Std. Error   t value
## (Intercept)                        31.1799108 2.26933764 13.739653
## Data_Main_Test$`2P%`               32.8553864 4.34496038  7.561723
## TWPA_Min_EFF                       -1.6359722 0.26679134 -6.132029
## TPA_Min_EFF                         0.2945059 0.09593397  3.069882
## Data_Main_Test$`2P%`:TWPA_Min_EFF   3.4228312 0.51658120  6.625931
##                                        Pr(>|t|)
## (Intercept)                        4.394448e-36
## Data_Main_Test$`2P%`               2.279659e-13
## TWPA_Min_EFF                       1.909812e-09
## TPA_Min_EFF                        2.271870e-03
## Data_Main_Test$`2P%`:TWPA_Min_EFF 9.946094e-11
```

```r
summary(fit_eff_3PP_2PP)$coef
```

```
##                                     Estimate Std. Error    t value
## (Intercept)                        37.423790  1.0390002  36.019040
## Data_Main_Test$`3P%`               -1.757860  1.3291596  -1.322535
## TPA_Min_EFF                        -3.244811  0.1004228 -32.311490
## Data_Main_Test$`2P%`               24.576608  1.9188923  12.807706
## TWPA_Min_EFF                       -2.132985  0.1179960 -18.076758
## Data_Main_Test$`3P%`:TPA_Min_EFF    9.997763  0.2934162  34.073655
## Data_Main_Test$`2P%`:TWPA_Min_EFF   4.321431  0.2282514  18.932772
##                                         Pr(>|t|)
## (Intercept)                        5.591401e-134
## Data_Main_Test$`3P%`                1.866695e-01
## TPA_Min_EFF                        8.772583e-119
## Data_Main_Test$`2P%`                3.440338e-32
## TWPA_Min_EFF                        3.605566e-55
## Data_Main_Test$`3P%`:TPA_Min_EFF   4.250907e-126
## Data_Main_Test$`2P%`:TWPA_Min_EFF   4.553608e-59
```

The results may seem overwhelming, but keep in mind that we are only interested in TPA_Min_EFF (three point attempts) and TWPA_Min_EFF (two point attempts). We see an interesting result as the nested models progress. Essentially, attempts for a certain type of shot are not significant at the $\alpha = 0.01$ level by themselves. However, when both the shot percetage and the interaction between shot percentage and number of shots are controlled for, the relationship between shot attempts and efficiency becomes significant.

This is strong evidence for possibility three: even when shot percentage and the resulting practices for each player are controlled for, shooting more perimeter shots results in less offensive efficiency than closer shots. This can only be because shooting a greater proportion of two point shots resulted inherently resulted in a higher efficiency. Our last step is to check the extent to which 3-point vs. 2-point shooting is a zero-sum game for centers/forwards. In other words, do big men who take more three point attmepts tend to take fewer two point attempts?

```r
cor(TPA_Min_EFF[CF],TWPA_Min_EFF[CF])
```

```
## [1] -0.462009
```

The negative correlation is substantial. Thus, we have a strong body of evidence supporting possibility 3.

Conclusion

The goal of this paper is explore the effect of 3-point attempts on offensive rating. Our first conclusive discovery is that in the 2018-2019 season, players who took more 3-point attempts tended to have a higher offensive rating. Our second conclusive discovery is that most of this difference can be mostly attributed to big men in ways that don't have to do with three-point shooting percentages.

Further investigation provided compelling evidence that big men who shoot more are less often in s a position to grab offensive rebounds, which may have both a direct and indirect effect on their offensive rating. Investigation also suggested that in the current NBA post-play typically contributes more the offensive rating. Investigation did not suggest that lower free-throw rates by big men resulted in significantly less return on three-point shots.

The next step would be to validate the models used in this paper against data from the 2017-2018 regular season to further test that strength of our results. The models could also be validated against 2019 playoff data to determine whether these relationships hold up under more strenuous competition. This being the case, the investigation provided in this paper could provide valuable insight for coaches looking to improve their players' offensive effectiveness.