title *

Stanley Zheng Eric Lucas †

May 7, 2025

1 Introduction

1.1 Motivation: The Shifting Landscape of the NBA

Over the last couple of decades, anyone watching the NBA has noticed a big change in how the game is played. We see teams shooting way more three-pointers than before. It makes sense to ask why. In the NBA, the basic rule is that shots made inside the arc are worth 2 points, while shots made further out, behind the arc, are worth 3 points. This simple difference in scoring value seems to have encouraged teams to focus more and more on shooting from distance.

This trend is so noticeable that people often talk about the modern NBA as the "Three-Point Era". Famous players like Stephen Curry really showed how valuable long-range shooting can be, making teams rethink their strategies. But is this just a popular trend, or has the actual importance of making three-pointers for winning games really increased compared to the past? For teams trying to win championships, understanding if and how the value of different shots has changed is really important. In today's world with lots of data available, we felt that using statistical analysis could help quantify this change and maybe offer useful insights for teams trying to optimize how they play.

1.2 Research Question and Approach

So, the main question we wanted to investigate in this project is: How has the **statistical association** between key shooting metrics, especially three-point percentage (3P%), and team success, measured by win rate, changed over the last approximately 15-20 years? Our "story" is essentially testing the common idea that the NBA has truly entered a "three-point era" where making threes is more strongly linked to winning than it used to be.

*The whole project is published in this repo

†Equal contribution

We need to be careful, though. We are using data collected from watching games (observational data), not from a controlled experiment like an RCT (Randomized Controlled Trial). This means we can't easily prove that shooting more threes **causes** more wins, because many other things might be happening at the same time (like better players, different defenses, rule changes). So, our approach is to use statistical tools to look for strong patterns or **structural evidence** in the data that show if the association between 3P% and winning has really changed significantly over time, while trying to account for some other factors. Our goal is to describe this change accurately based on the statistics, aiming for an interpretation that acknowledges these limitations but points towards the changing dynamics (a quasi-causal goal).

1.3 Thesis Statement

Using data comparing the 2005-2009 and 2020-2024 NBA regular seasons (source: NBA API https://github.com/swar/nba_api), we find the statistical association between three-point percentage (3P%) and win rate differs significantly between these periods: NBA team-seasons (unit of analysis) in the 2020-2024 period with higher 3P% also have statistically significant higher win rates, while NBA team-seasons in the 2005-2009 period did not have statistically significant higher win rates, even after controlling for two-point efficiency and shot selection proportions.

2 Methods

2.1 Dataset

Our dataset is sourced from the NBA's public API, an online interface that provides comprehensive and standardized statistics for teams and players across multiple seasons. The NBA API offers detailed data covering a wide range of performance categories, including shooting statistics, rebounding, passing, turnovers, fouls, player efficiency metrics, and advanced team analytics.

For our analysis, we primarily focus on shooting-related metrics, including three-point field goal percentage (3P%), two-point field goal percentage (2P%), three-point attempt rate (3PA rate), and two-point attempt rate (2PA rate). Even from the original dataset,

To motivate our investigation into how the impact of three-point shooting on team success has evolved, we begin by comparing data from the 2005-2009 and 2020-2024 NBA regular seasons using two sets of visualizations.

Figure 1 provides a visual summary of the changes in three-point shooting behavior over time. The pie charts (left) show a substantial increase in the proportion of field goal attempts that are three-pointers, rising from 21.1% in 2005-2009 to 39.2% in 2020-2024. Meanwhile, the overlaid

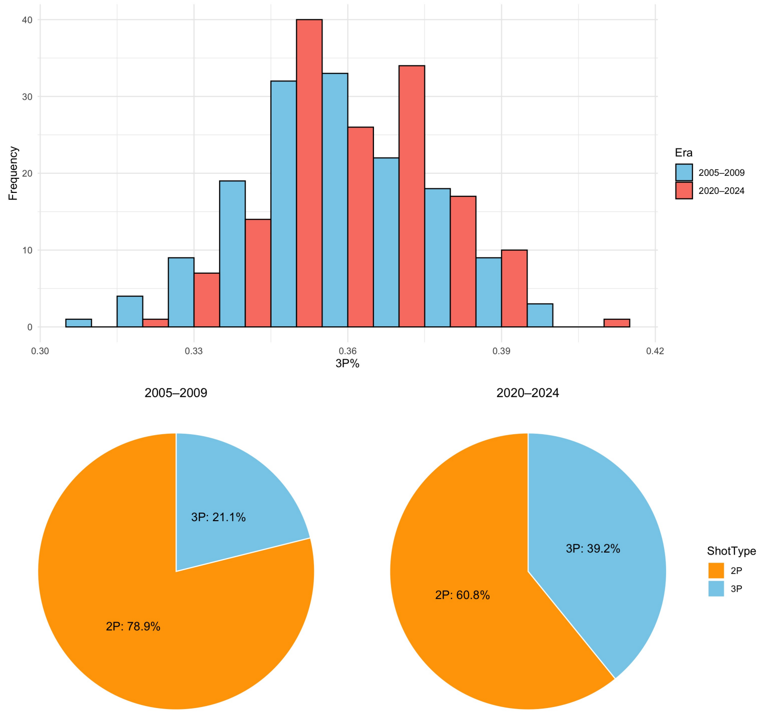


Figure 1: Changes in three-point attempt rate (left) and shooting accuracy (right) between 2005-2009 and 2020-2024.

histograms indicate that teams have also become slightly more accurate, with the distribution of three-point shooting percentages shifting modestly to the right in the modern era. This actually a very strong signal since the accuracy is kind of hard to improve, such a notable shift is already very impressive.

These visual trends motivate our modeling approach: to statistically examine whether the growing reliance on three-point shooting has translated into greater influence on win rate, and whether this shift reflects a broader structural transformation in NBA offensive strategy.

2.2 Data Preprocess

Our data preprocessing involved two main steps. First, we retrieved the raw team-level statistics using the NBA API, where each table corresponds to a single season, and each row records a team's aggregate performance for that year. We combined the annual tables into four period-specific datasets, grouped by five-year intervals (2005-2009, 2010-2014, 2015-2019, and 2020-2024).

Second, we performed variable selection and transformation. A large portion of the original variables were rankings (e.g., team rank in field goal percentage or rebounds), which we excluded, as they are ordinal and not suitable for regression analysis. We also observed that while the raw data included a few basic statistics (e.g., win percentage and FG3_PCT), it lacked several metrics relevant to our analysis. We therefore manually computed key variables, including win rate (WinRate), three-point attempt rate (3P_ratio), three-point shooting percentage (3P%), and two-point shooting percentage (2P%). Since 3P_ratio and 2P_ratio are linearly dependent (they sum to 1), we included only 3P_ratio in our models to avoid multicollinearity.

2.3 Statistical Techniques

We began our analysis by fitting a simple linear regression model using team-level three-point field goal percentage (3P%), two-point field goal percentage (2P%), and three-point attempt rate (3P_ratio) to predict regular season win rate. This baseline model allowed us to examine how different aspects of team shooting efficiency are associated with success across seasons.

To assess whether the impact of three-point shooting has changed over time, we then fit an interactive linear model. In this model, we included interaction terms between the shooting variables and a binary era indicator (early era vs. modern era). This design enables us to statistically test whether the relationship between three-point efficiency and win rate has become significantly stronger in recent years, rather than relying solely on visual or descriptive comparisons.

To investigate the influence of potential confounding variables, we expanded our interactive model by incorporating a range of other team statistics that could plausibly affect win rates. These included measures such as free throw percentage (`FT_PCT`), offensive and defensive rebounds (`OREB`, `DREB`).

To evaluate the validity of our regression models, we also performed standard diagnostic checks. Specifically, we examined residual plots including Residuals vs Fitted, Normal Q-Q, Scale-Location, and Residuals vs Leverage. These plots allow us to assess key assumptions of linear regression such as linearity, homoscedasticity, normality of residuals, and the presence of influential points. The diagnostic results did not reveal major violations, suggesting that the fitted models are reasonably appropriate for the data.

3 Results

3.1 Initial Analysis: Period-Specific Associations

To get a first look at how relationships might have changed, we fitted separate simple linear regression models for different 5-year periods, specified as follows:

$$\text{WinRate} = \beta_0 + \beta_1 \times \text{3P\%} + \beta_2 \times \text{2P\%} + \beta_3 \times \text{3P\_ratio} + \epsilon$$

We focus here on comparing the earliest period with the latest period(2005-2009 vs 2020-2024). Figure 2 displays the summary outputs for these two models.

```
Call:
lm(formula = WinRate ~ X3P. + X2P. + X3P_ratio, data = mydf_2005_2009)

Residuals:
      Min        1Q    Median        3Q       Max
-0.286612 -0.091560  0.009916  0.080742  0.261863

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -2.0312     0.3044  -6.672 4.87e-10 ***
X3P.          1.8842     0.6407   2.941  0.00381 **
X2P.          3.7431     0.6399   5.850 3.11e-08 ***
X3P_ratio     0.2650     0.2554   1.038  0.30118
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1216 on 146 degrees of freedom
Multiple R-squared:  0.3694,	Adjusted R-squared:  0.3564
F-statistic: 28.51 on 3 and 146 DF,  p-value: 1.436e-14
```

(a) Model Summary: 2005-2009 Period.

```
Call:
lm(formula = WinRate ~ X3P. + X2P. + X3P_ratio, data = mydf_2020_2024)

Residuals:
      Min        1Q    Median        3Q       Max
-0.282708 -0.060843 -0.004748  0.068760  0.251019

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -2.3217     0.2431  -9.551  < 2e-16 ***
X3P.          4.7683     0.5728   8.324 5.58e-14 ***
X2P.          2.1837     0.4371   4.996 1.65e-06 ***
X3P_ratio    -0.1782     0.2271  -0.784    0.434
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1022 on 146 degrees of freedom
Multiple R-squared:  0.4985,	Adjusted R-squared:  0.4882
F-statistic: 48.38 on 3 and 146 DF,  p-value: < 2.2e-16
```

(b) Model Summary: 2020-2024 Period.

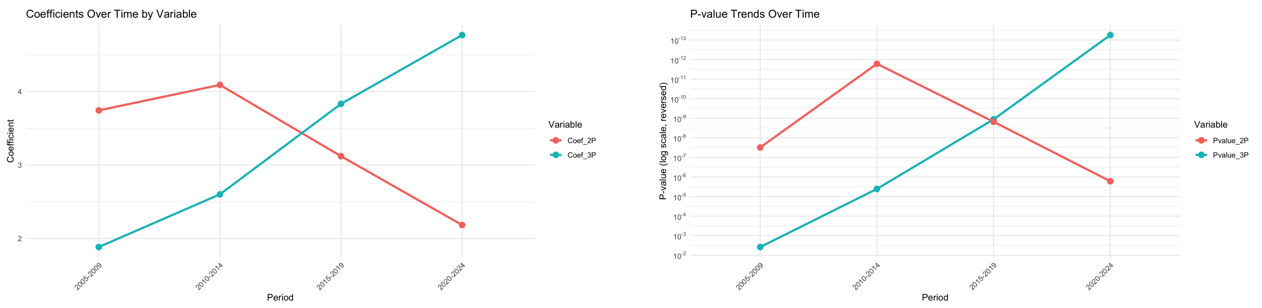
Figure 2: Comparison of Parsimonious Linear Regression Model Summaries for Early and Recent Eras.

Comparing the outputs shown in Figure 2, we observe notable differences between the two periods. In the 2005-2009 model (Figure 2a), both two-point percentage (`X2P.`) and three-point percentage (`X3P.`) show statistically significant positive associations with `WinRate`. However, the estimated coefficient for `X2P.` (approximately 3.74) is considerably larger than that for `X3P.` (approximately 1.88). Furthermore, the p-value associated with `X2P.` is much smaller ($p < 0.001$) compared to the p-value for `X3P.` ($p \approx 0.0038$). In this earlier period, the three-point attempt rate (`X3P_ratio`) was not found to have a statistically significant association with `WinRate`.

In contrast, when looking at the 2020-2024 model (Figure 2b), while both `X2P.` and `X3P.` remain statistically significant and positively associated with `WinRate`, their apparent relative importance seems to have reversed. The coefficient for `X3P.` (approximately 4.77) is now substantially larger than the coefficient for `X2P.` (approximately 2.18). The very small p-value for `X3P.` ($p < 0.001$) indicates a very strong statistical association in this recent period. Similar to the earlier period, `X3P_ratio` remains non-significant. This initial comparison suggests a potential shift: the positive association between three-point efficiency and winning appears stronger in the recent era, while the association for two-point efficiency might have weakened.

To visualize these changes more broadly, we plotted the estimated coefficients and their corresponding p-values (on a reversed log scale) for `X3P.` and `X2P.` across four consecutive 5-year periods. Figure 3 presents these trends.

Figure 3a clearly shows the coefficient for `X3P.` (blue line) steadily increasing over time, while the coefficient for `X2P.` (red line) generally decreases after the 2010-2014 period. Figure 3b complements this by showing the p-value associated with `X3P.` decreasing (indicating strengthening



(a) Trend in Estimated Coefficients. (b) Trend in P-values (Reversed Log Scale).

Figure 3: Trends in Coefficients and P-values for `X3P.` and `X2P.` from Period-Specific Regressions.

statistical significance), while the p-value for `X2P.`, although remaining significant, shows less of a strengthening trend. These descriptive trends strongly indicate a changing relationship and motivate the use of an interaction model to formally test the significance of this observed shift.

3.2 The Interaction Model

To capture the evolving relationship between shooting metrics and team success over time, we fit an interactive linear regression model incorporating both main effects and interaction terms with era. The model is specified as:

$$\begin{aligned} \text{WinRate} = \beta_0 &+ \beta_1 \cdot \text{3P\%} + \beta_2 \cdot \text{2P\%} + \beta_3 \cdot \text{3P}_{\text{ratio}} + \beta_4 \cdot \text{2P}_{\text{ratio}} + \beta_5 \cdot \text{Era} \\ &+ \beta_6 \cdot (\text{3P\%} \times \text{Era}) + \beta_7 \cdot (\text{2P\%} \times \text{Era}) + \beta_8 \cdot (\text{3P}_{\text{ratio}} \times \text{Era}) \\ &+ \beta_9 \cdot (\text{2P}_{\text{ratio}} \times \text{Era}) + \epsilon \end{aligned}$$

Our model summary is shown in Table 1.

Table 1: Summary of Interaction Model Results (Response: WinRate, based on R output)

Predictor	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-2.0312	0.2812	-7.224	4.41e-12 ***
X3P.	1.8842	0.5918	3.184	0.001610 **
X2P.	3.7431	0.5910	6.334	9.02e-10 ***
X3P_ratio	0.2650	0.2359	1.123	0.262187
Era (1 = 2020-24)	-0.2905	0.3879	-0.749	0.454448
X3P.:Era	**2.8842**	**0.8641**	**3.338**	**0.000954** ***
X2P.:Era	-1.5594	0.7616	-2.047	0.041509 *
X3P_ratio:Era	-0.4431	0.3434	-1.290	0.197976

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*
Residual standard error: 0.1123 on 292 degrees of freedom
Multiple R-squared: 0.4301, Adj. R-squared: 0.4165
F-statistic: 31.49 on 7 and 292 DF, p-value: < 2.2e-16

The results in Table 1 provide the crucial statistical evidence for our study. The key term is the interaction between three-point percentage and era (`X3P.:Era`). Its coefficient is approximately 2.8842 and is highly statistically significant ($p \approx 0.000954$, '***'). This positive and significant interaction term statistically confirms our central hypothesis: the positive association between a team's `X3P.` and its `WinRate` is significantly stronger in the 2020-2024 era (Era=1) compared to the baseline 2005-2009 era (Era=0). Specifically, for every one percentage point increase in `X3P.`, the associated increase in `WinRate` is estimated to be about 2.88 points higher in the 2020-2024 era than it was in the 2005-2009 era, holding other predictors constant.

Furthermore, the interaction term `X2P.:Era` has a coefficient of approximately -1.5594 and is statistically significant at the 0.05 level ($p \approx 0.0415$, '*'). This suggests that the positive association between `X2P.` and `WinRate` significantly weakened in the later era compared to the baseline era. These findings align perfectly with the descriptive trends shown in Figure 3a. The main effects for `X3P.` and `X2P.` represent their associations in the baseline (2005-2009) era, both being positive and significant. Other terms, including `X3P_ratio`, its interaction with era, and the main effect of `Era`, did not show statistically significant associations in this model. The overall model provides a reasonable fit (Adjusted R-squared ≈ 0.4165) and is highly significant overall (F-statistic p-value $< 2.2e\text{-}16$). The strong significance of the `X3P.:Era` interaction provides the formal statistical validation for our thesis about the changing relationship.

3.3 Consideration of Confounders

In our primary analysis, we aim to understand the statistical association between three-point shooting — particularly three-point percentage (`3P%`) — and team win rates (`WinRate`). However, various other aspects of team performance might be associated with both a team's reliance on three-point shooting and its overall success, potentially confounding the observed relationship between our primary variables. Understanding and accounting for these potential confounders is important for interpreting the association more clearly.

To address this, we considered including a range of basketball performance metrics as control variables (potential confounders) in our modeling. These included metrics related to other aspects of scoring efficiency, ball control, and defense, such as free throw percentage (`FT_PCT`), offensive and defensive rebounds (`OREB`, `DREB`), turnovers (`TOV`), steals (`STL`), blocks (`BLK`), and fouls committed or drawn (`PF`, `PFD`).

These variables were considered because they represent critical facets of overall team performance beyond just shooting percentages. It is plausible that these factors correlate with both shooting behavior and win rates. For example, teams that secure more defensive rebounds (`DREB`) or generate more steals (STL) might create more fast-break or transition opportunities, potentially leading to different types or qualities of three-point attempts. Similarly, teams committing fewer turnovers (TOV) or drawing more fouls (PFD) might represent more disciplined or skilled teams overall, which could correlate independently with both higher win rates and specific offensive strategies (including three-point efficiency).

By including such variables in the regression model(see Appendix Figure 5 for details), the goal is to statistically adjust for their association with both the predictors and the outcome. This approach helps to isolate the independent statistical association of the primary shooting metrics (3P%, 2P%, etc.) with team success (WinRate), allowing for a more nuanced interpretation of the relationship after accounting for these other performance dimensions. (As noted in the Results section, key findings regarding 3P% remained significant after including controls in exploratory models).

3.4 Model Diagnostics

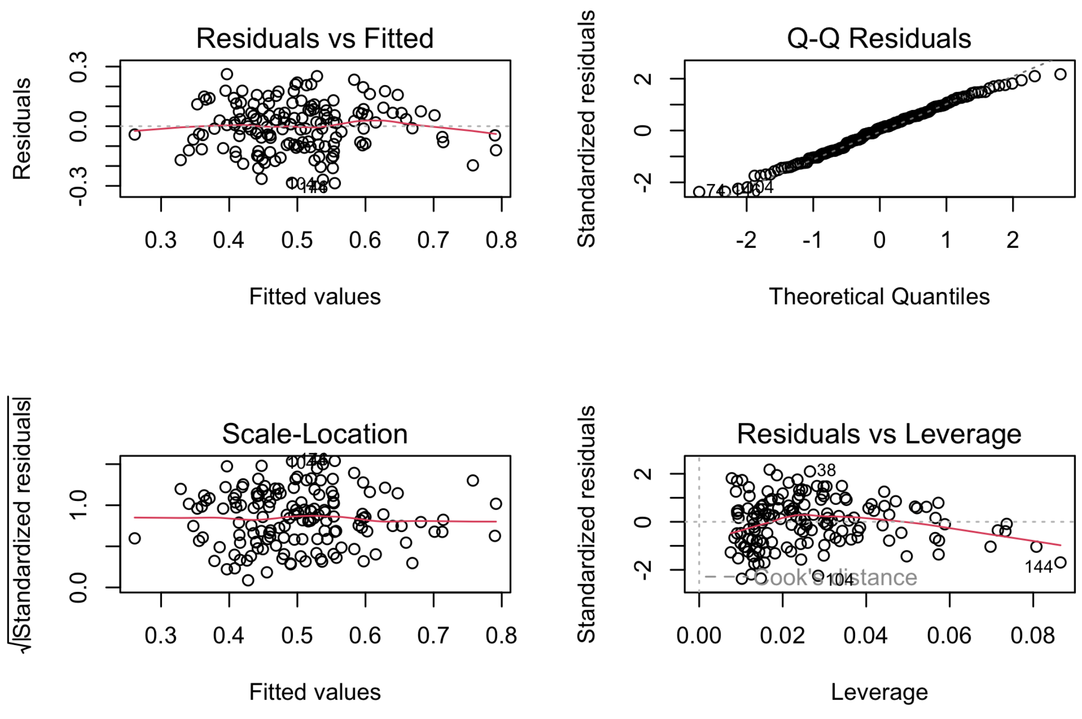


Figure 4: Diagnostic Plots for the Interaction Model.

Finally, we assessed the validity of the assumptions underlying the OLS regression for our final interaction model (results in Table 1). The standard diagnostic plots associated with this model are presented in Figure 4.

Visual inspection of these plots provides confidence in our model. The 'Residuals vs Fitted' plot displays points scattered fairly randomly around the horizontal line at zero, without a clear non-linear pattern or a distinct funnel shape, which supports the assumptions of linearity and constant variance (homoscedasticity). The 'Normal Q-Q' plot shows the standardized residuals falling generally close to the theoretical diagonal line, suggesting that the normality assumption for the residuals is reasonably met, although there might be slight deviations in the tails as is common with real-world data. The 'Scale-Location' plot further examines the homoscedasticity assumption; the points appear relatively evenly spread across the range of fitted values, and the smoothed red line is mostly flat, providing additional support for constant variance. Lastly, the 'Residuals vs Leverage' plot helps identify potentially influential observations. Most points exhibit low leverage. While point 144 shows slightly higher leverage than others, its standardized residual does not appear extreme, and no points seem to significantly exceed typical Cook's distance thresholds (though contours are not explicitly labeled). Therefore, based on this visual assessment, the standard OLS assumptions seem adequately satisfied for our interaction model, lending validity to the statistical inferences we have drawn from it.

4 Conclusion

4.1 Summary of Findings and Implications

This project statistically investigated the perceived emergence of a "Three-Point Era" in the NBA by analyzing how the relationship between shooting efficiency and team success evolved over time. Using data from the 2005–2009 and 2020–2024 NBA regular seasons (source: NBA API), we examined the association between three-point percentage (`3P%`) and team win rate (`WinRate`), controlling for two-point efficiency and shot selection ratios. Our regression results show that while three-point efficiency was not significantly associated with win rate in the earlier period, it became a statistically significant predictor of success in the modern era. The key supporting evidence is the statistically significant interaction term between era and three-point percentage (`Era:X3P.`) in our model (see Table 1), which indicates that the strength of this association increased by approximately 2.88 units in the 2020–2024 period compared to 2005–2009. This provides compelling quantitative support for the claim that the strategic value of three-point shooting has risen over time. Although our observational design cannot establish causality, the magnitude and significance of the shift strongly align with the notion that efficient three-point shooting has become a more decisive factor in team success under modern league dynamics.

4.2 Limitations of the Analysis

Several limitations must be considered when interpreting these results. Primarily, as this study relies on observational data, we cannot make definitive causal claims. The observed associations, and the changes in them, could be influenced by numerous unobserved confounding variables that evolved alongside shooting strategies. Potential confounders include changes in defensive rules and strategies over the years, the overall evolution of player skill sets (especially the rise of prolific shooters), influential coaching philosophies adapting to or driving trends, the impact of specific superstar players on their teams' style and success, uncaptured variations in team defense quality, or even factors like player injuries or strength of schedule within specific seasons. While our model controls for key efficiency and shot selection metrics, it cannot account for all such complex, potentially time-varying confounders. Other limitations include the specific choice of time periods for comparison and the inherent assumptions of the linear regression model framework.

4.3 Future Research

Future research could aim to address these limitations by incorporating data on defensive metrics, coaching changes, or player-level statistics if available. Exploring different time frames or using panel data methods might also provide further insights. Applying more advanced methods aimed at causal inference from observational data could be valuable, though potentially challenging.

Appendix

```
Call:
lm(formula = WinRate ~ ., data = mydf_2020_2024_confounder)

Residuals:
      Min        1Q    Median        3Q       Max
-0.121205 -0.040952  0.000801  0.038556  0.144448

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.681e+00  6.579e-01  -4.075 7.90e-05 ***
MIN         -3.466e-05  6.679e-05  -0.519 0.604676
FGA         -2.265e-04  1.008e-04  -2.247 0.026274 *
FG3A        -6.069e-05  2.445e-04  -0.248 0.804305
FTA         -1.767e-04  5.955e-05  -2.968 0.003561 **
FT_PCT       6.452e-01  1.971e-01   3.274 0.001356 **
OREB         5.827e-04  7.353e-05   7.924 8.46e-13 ***
DREB         5.084e-04  4.398e-05  11.559  < 2e-16 ***
AST         -3.055e-05  4.747e-05  -0.644 0.520944
TOV         -5.046e-04  7.683e-05  -6.568 1.07e-09 ***
STL          7.691e-04  8.876e-05   8.665 1.41e-14 ***
BLK          1.188e-04  8.749e-05   1.358 0.176903
BLKA        -3.797e-04  1.287e-04  -2.950 0.003755 **
PF           5.147e-05  5.677e-05   0.907 0.366277
PFD          3.638e-04  9.694e-05   3.753 0.000261 ***
X3P.         3.088e+00  3.939e-01   7.839 1.34e-12 ***
X2P.         2.708e+00  3.508e-01   7.719 2.57e-12 ***
X3P_ratio    2.348e-01  1.674e+00   0.140 0.888674
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.05362 on 132 degrees of freedom
Multiple R-squared:  0.8751,	Adjusted R-squared:  0.8591
F-statistic: 54.42 on 17 and 132 DF,  p-value: < 2.2e-16
```

Figure 5: Example Regression Model Output Including Confounding Variables (Data: 2020-2024 Period). This model includes variables such as FT_PCT, OREB, DREB, TOV, STL, BLK, PF, PFD etc., in addition to the primary shooting metrics.