WHAT WINS NBA GAMES?

DSA305: Panel Data Analysis

Gangwani, Bharat[1]

Ng, Cliff[1]

Chuang, Jin Lei[1]

Abstract

Forecasting the outcome of an NBA game is a well-studied topic given the financial rewards the betting platforms offer. Building upon existing literature, we identify the key factors which affect the number of wins for a team in the regular season. Using data from the 1996-97 season to the 2022-23 season, this paper estimates several different panel data regressions using exogenous variables selected specifically to alleviate potential endogeneity issues. The regression results suggest that teams with a higher proportion of 3-point attempts, better teamwork, older players, and more experienced coaches tend to perform better in the regular season. The paper also analyses coach effects and concludes that only 30% of coaches have a significant effect on team performance.

Keywords: basketball, NBA, playing style, fixed effects, correlated effects, random effects, coaches

[1] Singapore Management University

1. Introduction

The National Basketball Association (NBA) is the best basketball league in the world, attracting top tier athletes to The United States of America. In addition to physical talent, teams pay top dollar for advertising, health technology, infrastructure as well as operations in hopes to win the title of the World Champion. Predicting or forecasting the winner of an NBA match, whether during the regular season or the playoffs, is a well-studied problem with significant financial rewards for anyone who can outperform the betting markets.

Several studies aim to either create some systematic method to predict the win-rates for a team in a season or the outcome of a particular game. Hu & Zidek (2004) developed and demonstrated a Weighted-Likelihood approach to predict the 1996-97 NBA Final series results incorporating sample-relevant features and the home-field advantage. Other forecasting approaches which have been tried within the literature include dynamic state-space models (Manner, 2016), maximum-entropy based approaches (Cheng et al., 2016), Naive-Bayes, decision-tree, artificial neural networks (Thabtah et al., 2019), and even rules-of-thumb (Mcgivney et al., 2008).

Meanwhile, other studies aim to highlight which factors materially affect the win-rates of a team each season. Teramoto & Cross (2010) argue that the defensive performance of a team becomes more important in post-season playoffs relative to the regular season. They also identify rebounding and fewer turnovers as crucial ingredients for winning teams. Similarly, Chatterjee et al. (1994) analyse 5 seasons of NBA games and attribute 90% of the variation in the win-loss percentage to field goals, free throws, rebounds, and turnovers. Successively, they use the developed model to predict win-loss percentages of teams in the forthcoming season. In contrast to our results, they find that assists, steals, and three-point shots are not significant predictors of the win-loss percentage. However, we believe that their results suffer from endogeneity since their selected covariates are unlikely to be strictly exogenous in a season. We account for the endogeneity issues in our analysis by carefully selecting variables which are unlikely to be correlated to shocks to the dependent variable.

There is also significant literature on the extent to which coaches contribute to the performance of their teams. Ying & Chang (2019) use panel regressions to estimate the relationship between a head coach's managerial experience and the performance of the team and find it to be positive, which is in line with our findings as well. In general, the literature argues that while coaches can have a meaningful impact on their team's outperformance, the number of coaches which do have such an impact are in the minority. A study estimated the coach's contribution to the variation in a team's performance to be between 20% to 30% (Berry & Fowler, 2019). It is difficult to disentangle the effect of a good team managed by a mediocre coach versus a mediocre team being managed by a good coach (Berri et al., 2009). Section 4 of our paper follows in the footsteps of Al-Amine (2020) to utilize coach fixed effects to estimate the coaches' contribution to team performance. Our results are mostly in line with these studies wherein coaches which have a statistically significant impact on their team's performance (either negative or positive) are in the minority.

1.1. Structure of the NBA

The journey to the top is brutal. There are a total of 30 NBA teams, and they are equally divided into the Eastern and the West Conference. Each team then plays a total of eighty-two games during the regular season, only the top eight teams of each conference with the best win-loss record will advance into the playoffs. During the playoffs, the team with the best record plays a series with the worst record (figure 1). The winner of the series is determined by a "best-of-seven" concept, and the winner of the series moves on to the second round and the loser is eliminated. Given the amount of money spent on the pursuit of the title, it could be worthwhile to explore the traits of a team that makes them a contender for the title.

2. Data and Methodology

Statistics for NBA players during games are typically gathered by a team of statisticians who record various metrics such as points, rebounds, assists, steals, and blocks, and shooting percentages. These statistics are recorded using a computer system and then manually verified to ensure accuracy. While most of the data used are sourced from the NBA website itself (NBA Media Ventures, 2024), there remain certain data that was aggregated and sourced third-party websites. The data dates from the 1996-97 season to the 2022-23 season.

The data takes the form of a panel data with TEAM indicating individuals and Season indicating time. It is unbalanced as the NBA had twenty-nine teams in 1996 and added the Charlotte Bobcats as its 30th team in 2004. There were also several changes to the names of the team due to take overs and rebranding processes and this resulted in small variations in the number of teams participating in each season. Overall, each team participated in at least twenty-one seasons across the twenty-seven seasons of data used in this paper.

The dependent variable in this paper will be the number of wins for each team during the regular seasons. The maximum number of wins possible during the regular season is eighty-two, and the best regular season record is seventy-three wins with nine losses, held by the Golden State Warriors in the 2015-16 Season. Covariates used in this paper are listed in the table below, along with their description.

The full data set used for this paper can be found on in this link: https://github.com/bharatgw/DSA305-NBA_Wins_Analysis

Variable	Description
TEAM	The name of the team, used as an index in the panel data. Teams which were historically renamed have been updated to their most contemporaneous name to maintain a sufficiently long time-series for each entity.
Season	The season when the statistics were recorded.
W	Number of Wins recorded during the regular season.
Perc_3PA	Percentage of field goals attempted beyond the 3-point arc. Calculated by taking the number of 3-pointers attempted divided by the total number of shots attempted in the season.
Perc_2PA	Percentage of field goals attempted within the 3-point arc. Calculated by taking the number of 2-pointers attempted divided by the total number of shots attempted in the season. Perc_3PA + Perc_2PA = 1
Perc_AST	Number of Assists divided by the number of field goals made. An assist is recorded by a player after passing the ball to a teammate that results in a field goal made, this variable tells us the percentage of field goals having an assist.
Perc_STL	Number of steals divided by the number of turnovers made by the opponent. This variable measures often the team's defence forces the opponent to turn the ball over.
PFMinusPFD	Total personal fouls committed – total personal fouls committed against them. This variable measures a team's aggression.
OPP_Perc_3PA	Same definition as Perc_3PA, calculated with opposing teams' stats.
OPP_Perc_AST	Same definition as Perc_AST, calculated with opposing teams' stats.
OPP_Perc_STL	Same definition as Perc_STL, calculated with opposing team's stats.
L1_N_Awards_Won	Lagged version of number of awards won by the players on that team. The NBA gives out awards to individual players who excel in a certain aspect. The variable proxies the recent historical performance of the team.
L1_Coach_RS_W_Perc_Overall	Lagged version of win percentage of the coach during the regular season – measures the effectiveness of a good coach on winning.
L1_Coach_P_W_Perc	Lagged version of win percentage of the coach during the playoffs – measures the effectiveness of a good coach on winning.
AVG_PLAYER_AGE	The average age of the players on the team in the current season.
Coach	The name of the team's coach during that season.
Coach_N_Seasons_Overall	The number of seasons the current coach has been in the NBA.
Coach_Perc_Seasons_Team	The number of seasons the current coach has been with the team divided by the number of seasons the coach has been in the NBA – measures how well the coach knows his team.

3. Playstyle of Winning Teams

Teams within the NBA can have a variety of playing styles with some focusing on star players, while others prioritise teamwork. Using the variables identified available, we present three dimensions of playing styles which we consider within our analysis:

1. 2-point shots v. 3-point shots
2. Teamwork v. Star players
3. Aggressive v. Defensive

The first dimension is analysed using the *Perc_3PA* and *Perc_2PA* variables. A comparison of their coefficients will indicate whether teams which shoot more 3-point shots relative to 2-point shots win more games in a season or vice-versa. The second dimension is analysed by observing the coefficient of the variable *Perc_AST* which will suggest the additional number of wins a team can expect if they increase the number of assisted baskets they make each season. We analyse the third dimension by observing coefficients of variables *Perc_STL* and *PFminusPFD* which respectively denote a team's aggression in retaking possession and their propensity to commit fouls relative to their opponent. While we include coach-specific variables within our specifications in this section, the analysis of potential effects of individual coaches on their teams is reserved for Section 4.

In addition to analysing the relationship of the number of wins to a team's play style, we also control for their opposing team's play style in a match by including the respective variables. We also analyse the effect of the average age of players in a team, the coach's record of accomplishment and experience, and his familiarity with the team on the number of wins in a season.

While we control for several observed variables which may affect a team's performance in a season, there may be unobserved team-specific and season-specific variables we do not include within our specifications. To alleviate endogeneity concerns because of such variables, we build and test a series of models within this section assuming different forms for the unobserved effects. We begin with a Pooled OLS model, followed by Random Effects, Correlated Effects, and Fixed Effects models. A summary of the results from each model is presented in Table 1.

TABLE 1: SUMMARY OF REGRESSION OUTPUTS

W	Pooled OLS	Entity RE	Time RE	Two-Way FE
Perc_3PA	30.73* (15.51)	30.48 (15.58)	10.34 (15.61)	20.22 (16.84)
Perc_2PA	-5.49 (13.93)	-5.75 (14.01)	-23.93 (14.12)	-18.48 (15.35)
Perc_AST	37.11*** (10.56)	37.95*** (10.90)	37.35*** (9.57)	42.73*** (9.98)
Perc_STL	64.49*** (13.88)	64.22*** (14.14)	78.0*** (13.33)	80.75*** (13.43)
PFminusPFD	0.0 (0.00)	0.0 (0.00)	-0.0 (0.00)	-0.01*** (0.00)
OPP_Perc_3PA	-11.8 (10.79)	-9.7 (11.14)	-14.46 (12.62)	39.27 (22.47)
OPP_Perc_AST	-90.21*** (14.72)	-93.99*** (15.18)	-85.69*** (14.01)	-110.39*** (17.60)
OPP_Perc_STL	-72.98*** (12.95)	-77.65*** (13.32)	-57.75*** (13.77)	-62.32*** (15.68)
L1_N_Awards_Won	3.34*** (0.62)	3.18*** (0.62)	3.06*** (0.56)	2.55*** (0.58)
L1_Coach_RS_W_Perc_Overall	7.91 (5.88)	6.99 (6.01)	8.47 (5.50)	4.24 (6.04)
L1_Coach_P_W_Perc	-5.58 (3.81)	-5.36 (3.98)	-5.34 (3.53)	-4.02 (3.81)
AVG_PLAYER_AGE	2.5*** (0.32)	2.65*** (0.33)	2.58*** (0.30)	2.6*** (0.33)
Coach_N_Seasons_Overall	0.21*** (0.07)	0.23*** (0.08)	0.19*** (0.07)	0.17* (0.08)
Coach_Perc_Seasons_TEAM	4.05*** (1.30)	4.18*** (1.41)	3.72*** (1.23)	3.24* (1.51)
R^2 Between	0.58	0.57	0.04	0.44
R^2 Within	0.41	0.41	0.47	0.11
R^2 Overall	0.42	0.42	0.42	0.12
N	628	628	628	628

Note: The standard errors presented in the parenthesis are robust to heteroskedasticity and cluster-specific correlations.

3.1. Pooled OLS

Equation (1) models the number of wins of a team *i* in season *t* as a function of the covariates described above. We assume that the error term is normally distributed with heteroskedastic variances and consecutively use robust standard errors. It is estimated using the *PanelOLS* command from the *linearmodels* library in Python with entity and fixed effects set to *False* which uses the regular OLS estimator for the pooled data. The estimation results are presented in Table 2: Pooled OLS Estimator

$$\begin{aligned} W_{i,t} = {} & \beta_1 Perc\_3PA_{i,t} + \beta_2 Perc\_2PA_{i,t} + \beta_3 Perc\_AST_{i,t} + \beta_4 Perc\_STL_{i,t} + \beta_5 PFminusPFD_{i,t} \\ & + \beta_6 OPP\_Perc\_3PA_{i,t} + \beta_7 OPP\_Perc\_AST_{i,t} + \beta_8 OPP\_Perc\_STL_{i,t} + \beta_9 N\_Awards\_Won_{i,t-1} \\ & + \beta_{10} Coach\_RS\_W\_Perc\_Overall_{i,t-1} + \beta_{11} Coach\_P\_W\_Perc_{i,t-1} + \beta_{12} Avg\_PlayerAge_{i,t} \\ & + \beta_{13} Coach\_N\_Seasons\_Overall_{i,t} + \beta_{14} Coach\_Perc\_Seasons\_TEAM_{i,t} + u_{i,t} \end{aligned} \tag{1}$$

Table 2: Pooled OLS Estimator

PanelOLS Estimation Summary

Dep. Variable:	W	R-squared:	0.4241
Estimator:	PanelOLS	R-squared (Between):	0.5588
No. Observations:	628	R-squared (Within):	0.4062
Date:	Mon, Apr 01 2024	R-squared (Overall):	0.4241
Time:	15:57:58	Log-likelihood	-2295.6
Cov. Estimator:	Robust		
		F-statistic:	34.777
Entities:	30	P-value	0.0000
Avg Obs:	20.933	Distribution:	F(13,614)
Min Obs:	15.000		
Max Obs:	26.000	F-statistic (robust):	49.156
		P-value	0.0000
Time periods:	26	Distribution:	F(13,614)
Avg Obs:	24.154		
Min Obs:	21.000		
Max Obs:	27.000		

Parameter Estimates

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
Perc_3PA	30.730	15.515	1.9807	0.0481	0.2623	61.198
Perc_2PA	-5.4855	13.931	-0.3938	0.6939	-32.844	21.873
Perc_AST	37.105	10.563	3.5126	0.0005	16.360	57.850
Perc_STL	64.494	13.883	4.6455	0.0000	37.230	91.758
PFminusPFD	0.0003	0.0007	0.4637	0.6431	-0.0010	0.0017
OPP_Perc_3PA	-11.797	10.793	-1.0931	0.2748	-32.993	9.3980
OPP_Perc_AST	-90.205	14.724	-6.1263	0.0000	-119.12	-61.289
OPP_Perc_STL	-72.976	12.953	-5.6337	0.0000	-98.414	-47.537
L1_N_Awards_Won	3.3434	0.6161	5.4269	0.0000	2.1335	4.5532
L1_Coach_RS_W_Perc_Overall	7.9067	5.8789	1.3449	0.1791	-3.6385	19.452
L1_Coach_P_W_Perc	-5.5770	3.8101	-1.4637	0.1438	-13.059	1.9055
AVG_PLAYER_AGE	2.4955	0.3213	7.7677	0.0000	1.8646	3.1264
Coach_N_Seasons_Overall	0.2146	0.0735	2.9217	0.0036	0.0704	0.3589
Coach_Perc_Seasons_TEAM	4.0503	1.2993	3.1173	0.0019	1.4987	6.6019

The results of the Pooled OLS estimator suggest that 3-point shots are significantly more important to the number of wins a team has in a season relative to 2-point shots. A 10-percentage point increase in the number of 3-point shots attempted by a team lead to 3.5 additional wins in a season on average. The increase is calculated as the coefficient of *Perc_3PA* minus the coefficient of *Perc_2PA* since *Perc_3PA + Perc_2PA = 1*. The coefficient of *Perc_3PA* is statistically significant as well, while the coefficient of *Perc_2PA* is not which would suggest that while an increase in the proportion of 3-point shots by a team does increase the probability of winning, a decrease in the proportion of 2-point shots does not necessarily have the same effect.

The significance and magnitude of coefficients of the variables *Perc_AST* and *OPP_Perc_AST* suggest that teams which work well together have a far greater number of wins each season. Their coefficients suggest that a 10-percentage point increase in a team's Assist ratio (#Assists/#Shots Made) increases their wins in a season by 3.7. Conversely, a 10-percentage point increase in an opposing team's Assist ratio decreases a team's average wins in a season by nine. The threefold increase in the absolute magnitude of the coefficient of the opposing team's Assist ratio would indicate that while better teamwork has a material increase in winning rates, playing against a team which does not play well together is even better.

Meanwhile, coefficients of *Perc_STL* and *OPP_Perc_STL* indicate that teams which are more aggressive in recovering the ball during the opposing team's play tend to have higher win rates with absolute magnitudes of comparable size. However, the benefit of aggression is not very outsized in that the number of excess Personal Fouls committed relative to your opponent does not yield any increase in win rates for a team.

Finally, the coefficients of *L1_N_Awards_Won*, *Coach_N_Seasons_Overall, and Coach_Perc_Seasons_TEAM* are in line with our expectations. Historically, well-performing teams are likely to have higher win-rates in the current season as well. Coach outperformance in the previous seasons does not seem to yield statistically significant benefits but the coach's overall experience and their familiarity with the team (measured by the percentage of their career spent with the team) does lead to statistically significant and positive increases in the win rate. We analyse the effects of individual coaches in greater detail in the subsequent section.

3.2. Random Effects Model

In this section, we assume that entity and time specific errors are uncorrelated with the exogenous variables and estimate the one-way Random Effects models. The conditional Log-Likelihood Ratio tests for these models are presented in Table . They suggest that a two-way Random Effects model is the most appropriate. However, due to the data being unbalanced, we are unable to run a two-way Random Effects model

TABLE 3: LOG-LIKELIHOOD RATIO TESTS

	Conditional Test: Team	*Conditional Test: Season*
Test statistic	24.44	94.87
p-value	0.00	0.00
Distribution	$Chi^2(1)$	$Chi^2(1)$

Entity-specific Effects

Equation (2) models the number of wins of a team *i* in season *t* as a function of the covariates described. We assume that the error term is normally distributed with cluster-specific correlations, use robust standard errors and the unobserved individuals-specific effects are uncorrelated with each regressor. It is estimated using the *RandomEffects* command from the *linearmodels* library in Python. The estimation results are presented in Table Table 4: Entity-Specific RE Estimator.

$$
\begin{aligned}
W_{i,t} = {} & \beta_1 Perc\_3PA_{i,t} + \beta_2 Perc\_2PA_{i,t} + \beta_3 Perc\_AST_{i,t} + \beta_4 Perc\_STL_{i,t} + \beta_5 PFminusPFD_{i,t} \\
& + \beta_6 OPP\_Perc\_3PA_{i,t} + \beta_7 OPP\_Perc\_AST_{i,t} + \beta_8 OPP\_Perc\_STL_{i,t} + \beta_9 N\_Awards\_Won_{i,t-1} \\
& + \beta_{10} Coach\_RS\_W\_Perc\_Overall_{i,t-1} + \beta_{11} Coach\_P\_W\_Perc_{i,t-1} + \beta_{12} Avg\_PlayerAge_{i,t} \\
& + \beta_{13} Coach\_N\_Seasons\_Overall_{i,t} + \beta_{14} Coach\_Perc\_Seasons\_TEAM_{i,t} + \nu_i + u_{i,t}
\end{aligned} \tag{2}
$$

TABLE 4: ENTITY-SPECIFIC RE ESTIMATOR

RandomEffects Estimation Summary

Dep. Variable:	W	R-squared:	0.4109
Estimator:	RandomEffects	R-squared (Between):	0.5391
No. Observations:	628	R-squared (Within):	0.4079
Date:	Mon, Apr 01 2024	R-squared (Overall):	0.4235
Time:	23:40:58	Log-likelihood	-2283.4
Cov. Estimator:	Clustered		
		F-statistic:	32.948
Entities:	30	P-value	0.0000
Avg Obs:	20.933	Distribution:	F(13,614)
Min Obs:	15.000		
Max Obs:	26.000	F-statistic (robust):	47.945
		P-value	0.0000
Time periods:	26	Distribution:	F(13,614)
Avg Obs:	24.154		
Min Obs:	21.000		
Max Obs:	27.000		

Parameter Estimates

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
Perc_3PA	30.478	15.578	1.9565	0.0509	-0.1140	61.071
Perc_2PA	-5.7500	14.010	-0.4104	0.6816	-33.263	21.763
Perc_AST	37.948	10.902	3.4807	0.0005	16.538	59.359
Perc_STL	64.216	14.140	4.5414	0.0000	36.447	91.985
PFminusPFD	0.0004	0.0007	0.6215	0.5345	-0.0009	0.0018
OPP_Perc_3PA	-9.7000	11.136	-0.8710	0.3841	-31.570	12.170
OPP_Perc_AST	-93.995	15.179	-6.1924	0.0000	-123.80	-64.185
OPP_Perc_STL	-77.652	13.322	-5.8288	0.0000	-103.81	-51.489
L1_N_Awards_Won	3.1802	0.6249	5.0891	0.0000	1.9530	4.4074
L1_Coach_RS_W_Perc_Overall	6.9870	6.0066	1.1632	0.2452	-4.8090	18.783
L1_Coach_P_W_Perc	-5.3574	3.9847	-1.3445	0.1793	-13.183	2.4679
AVG_PLAYER_AGE	2.6547	0.3280	8.0934	0.0000	2.0105	3.2988
Coach_N_Seasons_Overall	0.2264	0.0774	2.9255	0.0036	0.0744	0.3784
Coach_Perc_Seasons_TEAM	4.1836	1.4059	2.9758	0.0030	1.4227	6.9445

Similarly to the Pooled OLS estimator, the Entity specific Random Effect model suggests that 3-point shots are significantly more important to the number of wins a team has in a season relative to 2-point shots. A 10-percentage point increase in the number of 3-point shots attempted by a team lead to 3.6 additional wins in a season on average. However, the coefficient of *Perc_3PA* from the Entity Specific Random Effect Model is insignificant.

Overall, the Entity Specific Random Effect Model shares comparable results to the pooled OLS model the magnitude of the coefficient are also similar. This would suggest that the unobserved individual-specific effects have a smaller variance relative to the model residuals.

Time-specific Effects

$$
\begin{aligned}
W_{i,t} = {} & \beta_1 Perc\_3PA_{i,t} + \beta_2 Perc\_2PA_{i,t} + \beta_3 Perc\_AST_{i,t} + \beta_4 Perc\_STL_{i,t} + \beta_5 PFminusPFD_{i,t} \\
& + \beta_6 OPP\_Perc\_3PA_{i,t} + \beta_7 OPP\_Perc\_AST_{i,t} + \beta_8 OPP\_Perc\_STL_{i,t} + \beta_9 N\_Awards\_Won_{i,t-1} \\
& + \beta_{10} Coach\_RS\_W\_Perc\_Overall_{i,t-1} + \beta_{11} Coach\_P\_W\_Perc_{i,t-1} + \beta_{12} Avg\_PlayerAge_{i,t} \\
& + \beta_{13} Coach\_N\_Seasons\_Overall_{i,t} + \beta_{14} Coach\_Perc\_Seasons\_TEAM_{i,t} + \lambda_t + u_{i,t}
\end{aligned} \tag{3}
$$

TABLE 5: TIME-SPECIFIC RE ESTIMATOR

RandomEffects Estimation Summary

Dep. Variable:	W	R-squared:	0.4595
Estimator:	RandomEffects	R-squared (Between):	0.0211
No. Observations:	628	R-squared (Within):	0.4671
Date:	Mon, Apr 01 2024	R-squared (Overall):	0.4183
Time:	23:47:43	Log-likelihood	-2248.2
Cov. Estimator:	Clustered		
		F-statistic:	40.157
Entities:	26	P-value	0.0000
Avg Obs:	24.154	Distribution:	F(13,614)
Min Obs:	21.000		
Max Obs:	27.000	F-statistic (robust):	53.141
		P-value	0.0000
Time periods:	30	Distribution:	F(13,614)
Avg Obs:	20.933		
Min Obs:	15.000		
Max Obs:	26.000		

Parameter Estimates

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
Perc_3PA	10.338	15.606	0.6624	0.5079	-20.309	40.984
Perc_2PA	-23.926	14.123	-1.6941	0.0907	-51.661	3.8089
Perc_AST	37.345	9.5672	3.9035	0.0001	18.557	56.134
Perc_STL	78.004	13.330	5.8519	0.0000	51.827	104.18
PFminusPFD	-0.0008	0.0012	-0.6565	0.5117	-0.0031	0.0015
OPP_Perc_3PA	-14.459	12.620	-1.1458	0.2523	-39.242	10.323
OPP_Perc_AST	-85.689	14.006	-6.1180	0.0000	-113.20	-58.184
OPP_Perc_STL	-57.746	13.772	-4.1930	0.0000	-84.792	-30.700
L1_N_Awards_Won	3.0597	0.5604	5.4604	0.0000	1.9593	4.1602
L1_Coach_RS_W_Perc_Overall	8.4716	5.4968	1.5412	0.1238	-2.3231	19.266
L1_Coach_P_W_Perc	-5.3429	3.5288	-1.5141	0.1305	-12.273	1.5870
AVG_PLAYER_AGE	2.5838	0.3044	8.4893	0.0000	1.9861	3.1815
Coach_N_Seasons_Overall	0.1904	0.0695	2.7397	0.0063	0.0539	0.3269
Coach_Perc_Seasons_TEAM	3.7201	1.2288	3.0275	0.0026	1.3070	6.1332

Results of Time-specific Random Effects estimation are presented in Table and are broadly similar. The most significant differences arise in the coefficients of *Perc_3PA* and *Perc_2PA*. Even these, their signs remain consistent, and the only difference arises in the magnitudes of the two estimates.

3.3. Entity-specific Correlated Random Effects Model

In this section, we assume that the entity-specific effects described in Equation (2) are a linear function of the exogenous variables. Consequently, we can describe the number of wins for team *i* in a season *t* as a function of the exogenous variables and estimate them using OLS for Equation (5). It is crucial to note that since *Perc_3PA* and *Perc_2PA* are linearly dependent; so are their time-averages and hence we cannot estimate an equation with all four covariates simultaneously. Consequently, we drop *Perc_2PA* and its time-mean from this specification to conduct estimation.

$$\begin{aligned} W_{i,t} = {} & \beta_1 Perc\_3PA_{i,t} + \beta_2 Perc\_AST_{i,t} + \beta_3 Perc\_STL_{i,t} + \beta_4 PFminusPFD_{i,t} \\ & + \beta_5 OPP\_Perc\_3PA_{i,t} + \beta_6 OPP\_Perc\_AST_{i,t} + \beta_7 OPP\_Perc\_STL_{i,t} + \beta_8 N\_Awards\_Won_{i,t-1} \\ & + \beta_9 Coach\_RS\_W\_Perc\_Overall_{i,t-1} + \beta_{10} Coach\_P\_W\_Perc_{i,t-1} + \beta_{11} Avg\_PlayerAge_{i,t} \\ & + \beta_{12} Coach\_N\_Seasons\_Overall_{i,t} + \beta_{13} Coach\_Perc\_Seasons\_TEAM_{i,t} \\ & + \beta_{14} Perc\_3PA\_mean_{i,t} + \beta_{15} Perc\_AST\_mean_{i,t} + \beta_{16} Perc\_STL\_mean_{i,t} \\ & + \beta_{17} PFminusPFD\_mean_{i,t} + \beta_{18} OPP\_Perc\_3PA\_mean_{i,t} + \beta_{19} OPP\_Perc\_AST\_mean_{i,t} \\ & + \beta_{20} OPP\_Perc\_STL\_mean_{i,t} + \beta_{21} N\_Awards\_Won\_mean_{i,t-1} + \beta_{22} Coach\_RS\_W\_Perc\_Overall\_mean_{i,t-1} \\ & + \beta_{23} Coach\_P\_W\_Perc\_mean_{i,t-1} + \beta_{24} AVG\_PLAYER\_AGE\_mean_{i,t} + \beta_{25} Coach\_N\_Seasons\_Overall\_mean_{i,t} \\ & + \beta_{26} Coach\_Perc\_Seasons\_TEAM\_mean_{i,t} + u_{i,t} \end{aligned} \tag{5}$$

TABLE 6: ENTITY-SPECIFIC CRE ESTIMATOR

RandomEffects Estimation Summary

Dep. Variable:	W	R-squared:	0.7638
Estimator:	RandomEffects	R-squared (Between):	0.9973
No. Observations:	628	R-squared (Within):	0.4088
Date:	Mon, Apr 01 2024	R-squared (Overall):	0.9545
Time:	22:17:32	Log-likelihood	-2270.8
Cov. Estimator:	Clustered		
		F-statistic:	74.889
Entities:	30	P-value	0.0000
Avg Obs:	20.933	Distribution:	F(26,602)
Min Obs:	15.000		
Max Obs:	26.000	F-statistic (robust):	101.76
		P-value	0.0000
Time periods:	26	Distribution:	F(26,602)
Avg Obs:	24.154		
Min Obs:	21.000		
Max Obs:	27.000		

Parameter Estimates

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
Perc_3PA	36.360	8.9266	4.0732	0.0001	18.829	53.891
Perc_AST	38.500	11.412	3.3737	0.0008	16.088	60.912
Perc_STL	62.225	14.676	4.2400	0.0000	33.403	91.047
PFminusPFD	0.0006	0.0007	0.8427	0.3997	-0.0008	0.0020
OPP_Perc_3PA	-6.8815	11.757	-0.5853	0.5586	-29.971	16.208
OPP_Perc_AST	-99.870	16.211	-6.1608	0.0000	-131.71	-68.034
OPP_Perc_STL	-82.283	14.171	-5.8066	0.0000	-110.11	-54.453
L1_N_Awards_Won	2.9415	0.6546	4.4936	0.0000	1.6559	4.2271
L1_Coach_RS_W_Perc_Overall	5.8167	6.3593	0.9147	0.3607	-6.6724	18.306
L1_Coach_P_W_Perc	-5.3428	4.3152	-1.2381	0.2161	-13.817	3.1318
AVG_PLAYER_AGE	2.8729	0.3407	8.4325	0.0000	2.2038	3.5420
Coach_N_Seasons_Overall	0.2414	0.0833	2.8988	0.0039	0.0778	0.4049
Coach_Perc_Seasons_TEAM	4.3919	1.6312	2.6924	0.0073	1.1883	7.5954
Perc_3PA_mean	-3.8820	61.248	-0.0634	0.9495	-124.17	116.40
Perc_AST_mean	-39.352	85.326	-0.4612	0.6448	-206.93	128.22
Perc_STL_mean	-29.853	119.09	-0.2507	0.8022	-263.74	204.04
PFminusPFD_mean	-0.0016	0.0127	-0.1278	0.8984	-0.0265	0.0233
OPP_Perc_3PA_mean	-1.1706	109.32	-0.0107	0.9915	-215.87	213.52
OPP_Perc_AST_mean	68.428	121.13	0.5649	0.5724	-169.47	306.32
OPP_Perc_STL_mean	111.18	116.81	0.9517	0.3416	-118.23	340.59
L1_N_Awards_Won_mean	1.3424	6.5425	0.2052	0.8375	-11.507	14.191
L1_Coach_RS_W_Perc_Overall_mean	15.241	56.573	0.2694	0.7877	-95.864	126.35
L1_Coach_P_W_Perc_mean	-4.1187	36.653	-0.1124	0.9106	-76.102	67.865
AVG_PLAYER_AGE_mean	-2.7116	2.2441	-1.2083	0.2274	-7.1189	1.6957
Coach_N_Seasons_Overall_mean	0.1385	0.5273	0.2626	0.7930	-0.8972	1.1741
Coach_Perc_Seasons_TEAM_mean	3.4188	11.942	0.2863	0.7748	-20.034	26.872

Unlike the One-Way Random-Effects Model, the Correlated Random Effects Model shows that *Perc_3PA* is significant. Having a coefficient of 36.3 means that for every 10-percent increase in the number of 3-pointers attempted without increasing the total number of shots attempted, a team can win on average three more games.

Otherwise, the Correlated Random Effects Model has similar suggestions to the One-Way Random Effects models.

TABLE 7: TEST OF LINEAR RESTRICTIONS FOR CRE

	F-test	*Wald Test*
Test statistic	1.8579	2.9507
p-value	0.0321	0.9981
Distribution	F(13, 602)	$Chi^2(13)$

The two tests for Correlated Random Effects have very differing results and hence are inconclusive.

3.4. Fixed Effects Model

In the fixed effects section of the paper, we construct two one-way fixed effects model, one for entity and one for time fixed effects, as well as a two-way fixed effects model. We assume that the v_{it} error terms are normally distributed with heteroskedastic variances. The models are estimated with the *PanelOLS* command from the *linearmodels* library in Python with entity and fixed effects set to *True* in each respective model. The joint, conditional, and marginal F-tests are presented in Table . They suggest that a two-way fixed effects model is the most appropriate. Estimation results are presented in Table-Table .

TABLE 8: F-TESTS FOR FIXED EFFECTS

	Joint F-test	*Conditional F-test: Season*	*Conditional F-test: Team*	*Marginal F-test: Season*	*Marginal F-test: Team*
Test statistic	3.97	5.77	1.82	6.02	2.14
p-value	0.00	0.00	0.00	0.0	0.00
Distribution	F(54, 560)	F(25, 589)	F(29, 585)	F(25, 560)	F(29, 560)

Entity-specific Effects

The results of the One-Way Entity-specific Fixed Effect Model reflects that *Perc_AST* and *OPP_Perc_AST* are still statistically significant. This is in line with the results of the previous model that teams who have great chemistry together in passing the ball showed greater win percentages. In fact, a 10-percentage point increase in a team's Assist ratio (#Assists/#Shots Made) increases their wins in a season by 3.8. Conversely, a 10-percentage point increase in an opposing team's Assist ratio decreases a team's average wins in a season by ten.

Meanwhile, coefficients of *Perc_STL* and *OPP_Perc_STL* indicate that teams which are more aggressive in recovering the ball during the opposing team's play tend to have higher win rates with absolute magnitudes of comparable size. An increase in 10-percentage point of the *Perc_STL* ratio increases the average number of wins by 6.2 games, a sizeable amount.

L1_N_Awards_Won indicates the number of outstanding players in each team. The results show that with every addition of an accoladed player, the number of wins increases by three games, and it is not a surprising result.

The coefficient of *AVG_PLAYER_AGE* suggests that teams who have an average age of one additional year will win an average of three more games per season, proving that experience makes a greater impact compared to a younger athletic ability.

Lastly, the coach's experience in the NBA and with the team are both statistically significant, suggesting that management teams do play a part in the success of an NBA team.

TABLE 9: ENTITY-SPECIFIC FE ESTIMATOR

PanelOLS Estimation Summary

Dep. Variable:	W	R-squared:	0.4088
Estimator:	PanelOLS	R-squared (Between):	0.5084
No. Observations:	628	R-squared (Within):	0.4088
Date:	Mon, Apr 01 2024	R-squared (Overall):	0.4210
Time:	23:47:42	Log-likelihood	-2268.6
Cov. Estimator:	Clustered		
		F-statistic:	31.120
Entities:	30	P-value	0.0000
Avg Obs:	20.933	Distribution:	F(13,585)
Min Obs:	15.000		
Max Obs:	26.000	F-statistic (robust):	42.397
		P-value	0.0000
Time periods:	26	Distribution:	F(13,585)
Avg Obs:	24.154		
Min Obs:	21.000		
Max Obs:	27.000		

Parameter Estimates

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
Perc_3PA	30.993	15.821	1.9590	0.0506	-0.0789	62.066
Perc_2PA	-5.3666	14.417	-0.3722	0.7099	-33.682	22.949
Perc_AST	38.500	11.602	3.3183	0.0010	15.713	61.287
Perc_STL	62.225	14.749	4.2191	0.0000	33.259	91.192
PFminusPFD	0.0006	0.0007	0.8331	0.4051	-0.0008	0.0020
OPP_Perc_3PA	-6.8815	11.841	-0.5811	0.5614	-30.138	16.376
OPP_Perc_AST	-99.870	16.428	-6.0794	0.0000	-132.13	-67.606
OPP_Perc_STL	-82.283	14.365	-5.7278	0.0000	-110.50	-54.069
L1_N_Awards_Won	2.9415	0.6614	4.4477	0.0000	1.6426	4.2404
L1_Coach_RS_W_Perc_Overall	5.8167	6.4842	0.8971	0.3701	-6.9184	18.552
L1_Coach_P_W_Perc	-5.3428	4.3555	-1.2267	0.2204	-13.897	3.2116
AVG_PLAYER_AGE	2.8729	0.3420	8.4006	0.0000	2.2012	3.5446
Coach_N_Seasons_Overall	0.2414	0.0843	2.8614	0.0044	0.0757	0.4070
Coach_Perc_Seasons_TEAM	4.3919	1.6389	2.6798	0.0076	1.1731	7.6106

F-test for Poolability: 1.8155
P-value: 0.0061
Distribution: F(29,585)

Included effects: Entity

Time-specific Effects

The results of One-Way Time-specific Fixed Effects are like the One-Way Entity-specific Fixed Effects. One noteworthy change is the significance of *PFminusPFD,* but the absolute value of the coefficient is extremely small, which suggests insignificance.

TABLE 10: TIME-SPECIFIC FE ESTIMATOR

PanelOLS Estimation Summary

Dep. Variable:	W	R-squared:	0.4799
Estimator:	PanelOLS	R-squared (Between):	0.4783
No. Observations:	628	R-squared (Within):	-0.5280
Date:	Mon, Apr 01 2024	R-squared (Overall):	-0.4182
Time:	23:47:43	Log-likelihood	-2226.9
Cov. Estimator:	Clustered		
		F-statistic:	41.810
Entities:	30	P-value	0.0000
Avg Obs:	20.933	Distribution:	F(13,589)
Min Obs:	15.000		
Max Obs:	26.000	F-statistic (robust):	51.901
		P-value	0.0000
Time periods:	26	Distribution:	F(13,589)
Avg Obs:	24.154		
Min Obs:	21.000		
Max Obs:	27.000		

Parameter Estimates

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
Perc_3PA	18.046	16.409	1.0998	0.2719	-14.181	50.274
Perc_2PA	-18.709	15.116	-1.2377	0.2163	-48.397	10.979
Perc_AST	41.416	9.3877	4.4117	0.0000	22.978	59.853
Perc_STL	83.110	12.974	6.4061	0.0000	57.630	108.59
PFminusPFD	-0.0115	0.0030	-3.8531	0.0001	-0.0174	-0.0056
OPP_Perc_3PA	19.629	19.812	0.9907	0.3222	-19.283	58.540
OPP_Perc_AST	-99.104	15.120	-6.5547	0.0000	-128.80	-69.409
OPP_Perc_STL	-54.488	14.471	-3.7653	0.0002	-82.909	-26.067
L1_N_Awards_Won	2.8494	0.5521	5.1611	0.0000	1.7651	3.9337
L1_Coach_RS_W_Perc_Overall	8.1720	5.4996	1.4859	0.1378	-2.6291	18.973
L1_Coach_P_W_Perc	-4.7874	3.3989	-1.4085	0.1595	-11.463	1.8881
AVG_PLAYER_AGE	2.3045	0.3078	7.4861	0.0000	1.6999	2.9091
Coach_N_Seasons_Overall	0.1510	0.0695	2.1733	0.0302	0.0145	0.2874
Coach_Perc_Seasons_TEAM	3.4948	1.2384	2.8221	0.0049	1.0626	5.9269

F-test for Poolability: 5.7664
P-value: 0.0000
Distribution: F(25,589)

Included effects: Time

Two-way Effects

Results from the Two-Way Fixed Effects Estimator are like the other Fixed Effects Estimators.

TABLE 11: TWO-WAY FE ESTIMATOR

PanelOLS Estimation Summary

Dep. Variable:	W	R-squared:	0.4689
Estimator:	PanelOLS	R-squared (Between):	0.3846
No. Observations:	628	R-squared (Within):	-0.8638
Date:	Tue, Apr 02 2024	R-squared (Overall):	-0.7278
Time:	00:24:50	Log-likelihood	-2193.9
Cov. Estimator:	Clustered		
		F-statistic:	38.028
Entities:	30	P-value	0.0000
Avg Obs:	20.933	Distribution:	F(13,560)
Min Obs:	15.000		
Max Obs:	26.000	F-statistic (robust):	45.688
		P-value	0.0000
Time periods:	26	Distribution:	F(13,560)
Avg Obs:	24.154		
Min Obs:	21.000		
Max Obs:	27.000		

Parameter Estimates

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
Perc_3PA	20.225	16.839	1.2010	0.2302	-12.851	53.301
Perc_2PA	-18.481	15.349	-1.2040	0.2291	-48.629	11.668
Perc_AST	42.730	9.9833	4.2802	0.0000	23.121	62.340
Perc_STL	80.755	13.433	6.0116	0.0000	54.369	107.14
PFminusPFD	-0.0125	0.0033	-3.8249	0.0001	-0.0188	-0.0061
OPP_Perc_3PA	39.273	22.470	1.7478	0.0811	-4.8637	83.410
OPP_Perc_AST	-110.39	17.601	-6.2714	0.0000	-144.96	-75.812
OPP_Perc_STL	-62.324	15.677	-3.9755	0.0001	-93.117	-31.531
L1_N_Awards_Won	2.5518	0.5794	4.4039	0.0000	1.4137	3.6900
L1_Coach_RS_W_Perc_Overall	4.2433	6.0422	0.7023	0.4828	-7.6247	16.111
L1_Coach_P_W_Perc	-4.0249	3.8111	-1.0561	0.2914	-11.511	3.4608
AVG_PLAYER_AGE	2.5984	0.3321	7.8235	0.0000	1.9461	3.2508
Coach_N_Seasons_Overall	0.1682	0.0791	2.1273	0.0338	0.0129	0.3236
Coach_Perc_Seasons_TEAM	3.2444	1.5114	2.1466	0.0323	0.2756	6.2131

F-test for Poolability: 3.9688
P-value: 0.0000
Distribution: F(54,560)

Included effects: Entity, Time

3.5. Hausman Test

TABLE 12: HAUSMAN TESTS

	Entity	*Season*
Test statistic	1.26	14.62
p-value	0.99	0.40
Distribution	$Chi^2(14)$	$Chi^2(14)$

Entity Effects

The test statistic for the Hausman Test is 1.26, following a chi-square distribution with 14 degrees of freedom. The corresponding p-value is 0.99, which suggests we cannot reject the null hypothesis. This suggest that both fixed-effects and random-effects estimate are consistent, but the random-effects estimator is more efficient.

Time Effects

Conducting a second Hausman Test on the time effects, the test statistic is 14.62, following a chi-square distribution with 14 degrees of freedom. The corresponding p-value is 0.40, which is not statistically significant. Both fixed-effects and random-effects estimate are consistent, but the random-effects estimator is more efficient compared to the fixed-effects estimator.

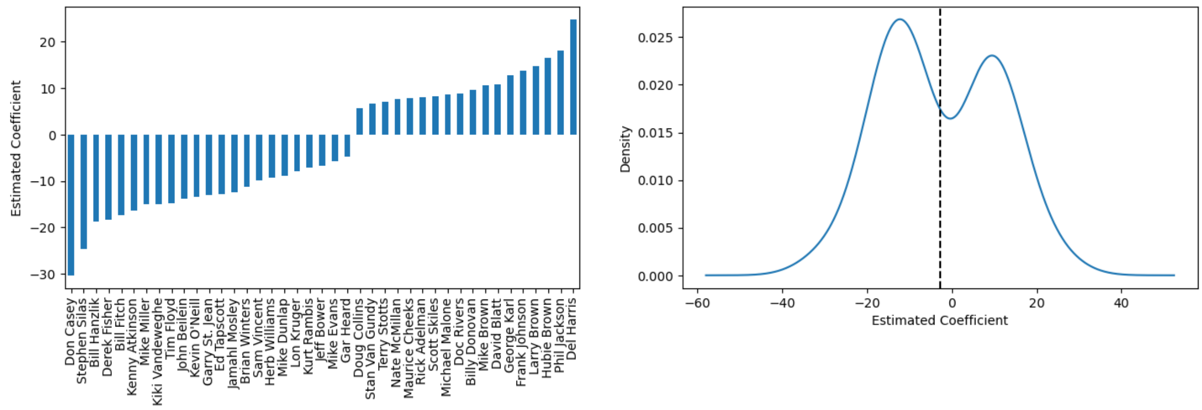
4. Coach Effects

In this section, we estimate the impact of individual coaches on their team's performance using *Coach* dummy variables within the Pooled OLS specification. Specifically, our Pooled OLS regression is specified in Equation (6).

$$\begin{aligned} W_{i,t} &= \beta_1 Perc\_3PA_{i,t} + \beta_2 Perc\_2PA_{i,t} + \beta_3 Perc\_AST_{i,t} + \beta_4 Perc\_STL_{i,t} + \beta_5 PFminusPFD_{i,t} \\ &+ \beta_6 OPP\_Perc\_3PA_{i,t} + \beta_7 OPP\_Perc\_AST_{i,t} + \beta_8 OPP\_Perc\_STL_{i,t} + \beta_9 N\_Awards\_Won_{i,t-1} \\ &+ \beta_{10} Coach\_RS\_W\_Perc\_Overall_{i,t-1} + \beta_{11} Coach\_P\_W\_Perc_{i,t-1} + \beta_{12} Avg\_PlayerAge_{i,t} \\ &+ \beta_{13} Coach\_N\_Seasons\_Overall_{i,t} + \beta_{14} Coach\_Perc\_Seasons\_TEAM_{i,t} + \sum_{j=1}^{136} \delta_j d_j + u_{i,t} \end{aligned} \tag{6}$$

Specifically, our coefficients of interest are $\delta_j \ \forall j \in \{1, \dots 136\}$ where j is an index variable for the 136 unique coaches within our dataset. Post-estimation, we filter the coefficients to keep coach effects which are statistically significant at the 5% level. Notably, only 30% of all coach effects are statistically significant. The distribution of these effects is presented in Figure 1: Significant Coach EffectsFigure 1.

FIGURE 1: SIGNIFICANT COACH EFFECTS



The distribution of significant coach effects appears quite symmetrical with the two peaks of the distribution at +- 20 wins on either side. Our estimate of 20/82 ≈25% is in line with previous studies on the topic of coach contribution to variations in team performance. Among the significant estimated coach effects, Del Harris appears to have had the most positive impact on any team while Don Casey appears to have had the most negative impact.

5. Conclusion

In this paper, we develop several panel data models to explain the effect of different playing styles on the number of wins for an NBA team within a regular season. Our results are qualitatively comparable across the different models. However, the statistical significance of the covariates does vary across the specifications. Using the Hausman test we identify the Random Effects Model as the most appropriate and efficient. In both Entity and Time Random Effects models, the coefficient of Perc_3PA is positive and suggests that a higher proportion of 3-point shots increases the number of wins for a team in a season. Additionally, teams which collaborate with each other to score baskets, steal the ball to initiate turnovers, have older players and coaches with more experience and familiarity with the team tend to outperform teams which do not. Orthogonal to coaching experience, individual coaches which have a significant impact on a team's number of wins in a regular season seem to be in the minority with only 30% of the coach effects in our sample being statistically significant. Conditional on a coach being effective, evidence suggests that they contribute 25% to the overall (out/under) performance of the team.

While we have been thorough in ensuring that the variables included in our analysis are strictly exogenous, one could still make the argument that shocks to any of the contemporary exogenous variables in our specification, such as the percentage of assists, are correlated with shocks to a team's wins in a regular season. If they are negatively correlated, such as in the instance when a team scores many field goals and hence the percentage of assists mathematically shrinks, then its coefficient estimate would be biased downwards. One limitation of our analysis is that we do use any instruments to address such concerns.

Despite these limitations, our analysis yields insights into the different important dimensions of play styles for teams within the NBA and how these contribute to their success. We also evaluate the effectiveness of coaches using coach effects within our dataset and provide metrics to evaluate the extent of their contribution.

References

Al-Amine, R. (2020, October 7). *Quantifying the Contribution of NBA Coaches using Fixed Effects*. Medium. https://towardsdatascience.com/quantifying-the-contribution-of-nba-coaches-using-fixed-effects-56f77f22153a

Berri, D., Leeds, M., Leeds, E., & Mondello, M. (2009). The Role of Managers in Team Performance. *International Journal of Sport Finance*, *4*, 75–93.

Berry, C. R., & Fowler, A. (2019). *How Much Do Coaches Matter?*

Chatterjee, S., Campbell, M. R., & Wiseman, F. (1994). Take that jam! An analysis of winning percentage for NBA teams. *Managerial and Decision Economics*, *15*(5), 521–535. https://doi.org/10.1002/mde.4090150514

Cheng, G., Zhang, Z., Kyebambe, M. N., & Kimbugwe, N. (2016). Predicting the Outcome of NBA Playoffs Based on the Maximum Entropy Principle. *Entropy*, *18*(12), Article 12. https://doi.org/10.3390/e18120450

Hu, F., & Zidek, J. V. (2004). Forecasting NBA Basketball Playoff Outcomes Using the Weighted Likelihood. *Lecture Notes-Monograph Series*, *45*, 385–395.

Manner, H. (2016). Modeling and forecasting the outcomes of NBA basketball games. *Journal of Quantitative Analysis in Sports*, *12*(1), 31–41. https://doi.org/10.1515/jqas-2015-0088

Mcgivney, K., Mcgivney, R., & Zegarelli, R. (2008). Light it up: Predicting the winner of an NBA game before the end. *Chance*, *21*(4), 45–50. https://doi.org/10.1007/s00144-008-0039-x

NBA Media Ventures. (2024). *NBA STATS*. NBA. https://www.nba.com/stats/teams

Teramoto, M., & Cross, C. L. (2010). Relative Importance of Performance Factors in Winning NBA Games in Regular Season versus Playoffs. *Journal of Quantitative Analysis in Sports*, *6*(3). https://doi.org/10.2202/1559-0410.1260

Thabtah, F., Zhang, L., & Abdelhamid, N. (2019). NBA Game Result Prediction Using Feature Analysis and Machine Learning. *Annals of Data Science*, *6*(1), 103–116. https://doi.org/10.1007/s40745-018-00189-x

Ying, Y.-H., & Chang, H.-C. C. (2019). *Who Helps the Teams to Win the Games? - The Case of the NBA*. The Asian Conference on the Social Sciences. https://papers.iafor.org/wp-content/uploads/papers/acss2019/ACSS2019_51780.pdf