A GAME-THEORETIC ANALYSIS OF OPTIMAL OFFENSIVE STRATEGY IN BASKETBALL

Ruifeng Li[*] Mingyang Yan[†] Zihan Zhang[‡]

January 4, 2019

Abstract

Game Theory could be applied in Basketball playing in terms of selecting the offense and defense strategy. We develop "The Offense and Defense Strategies Models" to analyze the optimal strategy under different circumstances. We simplify the strategies to shoot 2 and shoot 3. Then, we use linear regression models to estimate the payoff which is the winning probability. After taking in the practical data for each team, we solve the Mixed Strategy Nash Equilibrium. Theoretically, it is the optimal distribution of shooting opportunities. Finally, we compare it with the team's actual performance in NBA 2016-2017 regular season and try to provide some potential interpretations.

Keywords:

Game Theory, Optimal strategy, Mixed Strategy Nash Equilibrium, Empirical performance

[*]Student ID: 15220162202189
[†]Student ID: 15220162202432
[‡]Student ID: 15220162202518

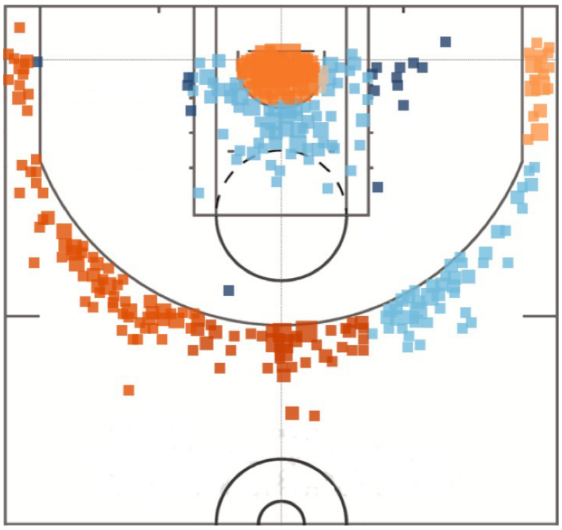
1 Introduction

1.1 Motivation

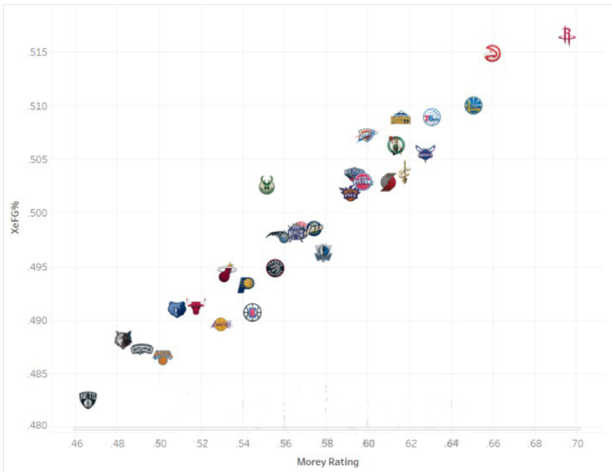
In the field of professional sports, game theory could provide a theoretically optimal solution. And more professional the league is, the players or teams' performance tends to be closer to the hypothetical Nash Equilibrium. And there were some researches about tennis and soccer. Therefore, we want to make attempt on basketball. NBA, being the most professional basketball association in the world, also has an available and very detailed statistical data records. All the factors above make it a good option.

1.2 Data visualization

We collect the data from the website www.espn.com of NBA 2016-2017 Regular Season. We try to use this pool data to roughly overview the shoot efficiency. Then the question here is "How to evaluate the quality of a shoot?" Intuitively, it should relate to the distance and the defense you face. And we draw a heat map, which verifies this intuition. We can find that the efficiency in two areas is much higher, meaning that there could be an optimal distribution of your shoot opportunities.



Then we have a closer look at each specific team. Even though Different ideas on basketball and the different players allocation could give different results of such distribution, we still observe a relatively strong relationship between the Morey Rating and the expected percentage of EFG. (i.e. $MoreyRating = \frac{(3points+under\ basket)\ attempts}{total\ shoot\ attempts}$ and EFG is effective field goal)



2 Literature Review

IGNACIO PALACIOS-HUERTA(2003) (1), reveals that professional soccer players would always endeavor to maximize their own payoffs on the assumption that their rivals would choose to hurt them as much as possible. In other words, professional players would only focus on their goals in every round, and do not care about other complicated strategies in the reality. For simplicity, we also suppose that the two basketball teams in our model would always choose an action that maximize their payoffs in every single round, which coincides with most of the basic game theoretical assumptions.

Besides, IGNACIO PALACIOS-HUERTA (2003) also claims that professional sports players are ideal participants in theoretical studies, since their performance would be closer to rational people. According to this claim, we use NBA data from 2016 to 2017 as our sample data.

Mark Walker and John Wooders (2001) (2)suggest that being unpredictable is vital for professional players, which means they prefer to keep their opponents guessing so that winning is more realizable when facing to a rival who is at a comparable or even higher level than themselves. Based on their findings, we use mixed strategy Nash equilibrium to construct our initial model, which would be described in the following sections.

Justin Kubatko et al. (2007) (3) argue that the offensive and defensive rates could be calculated by the ratio of hits to attempted shootings. Inspired by this method, we regard offensive ratios of 2 and 3 points shootings as our independent variables and the winning rate as the dependent variable in our regression model.

Oliver A. Entine and Dylan S. Small(2008) (4) adopt linear regression models to estimate all the factors that could do something with the visiting team's performance. Identical to their method, we try to use regression model to estimate theoretical payoffs. Due to the properties of our dependent and independent variables, we choose non-linear probit model. Similarly, we include almost all the factors recorded by official website and could affect the dependent variable in our model.

3 Theoretical Framework

3.1 Normal form representation of the game

We use a simplified model here, making assumption that we are the offensive side. The strategy profile of the offensive team is Shoot 2,Shoot 3. For the defense counterpart, it is {Defense 2,Defense 3}. They move simultaneously.

We denote the payoff for (Shoot 2, Defense 2) as $S_{2,2}$, and so on so forth.

Table 1: Payoff matrix

		Defence	
		Defend 2	Defend 3
Offense	Shoot 2	$S_{2,2}$	$S_{2,3}$
	Shoot 3	$S_{3,2}$	$S_{3,3}$

$S_{a,b}$ where $a \in \{2, 3\}, b \in \{2, 3\}$ is the winning rate estimated by a Probit model. Mathematically, it a real number in the interval (0,1). It is easy to understand the following inequalities hold:

$$S_{2,2} < S_{2,3}, \quad S_{3,2} > S_{3,3}, \quad S_{2,3} > S_{3,3}, \quad S_{2,2} < S_{3,2} \tag{1}$$

If your opponent exactly aims at your strategy, your winning probability could be lower than if there is a mismatch between the offense and defense strategies. Such an assumption implies that no Pure Strategy Nash Equilibrium exists. If the assumption is violated, there will be dominant strategy. And we do not take that into our consideration.

3.2 Solving the Mixed Strategy Nash Equilibrium

With the model framework defined in the previous section, we can solve for the Mixed Strategy Nash Equilibrium with following equations.

$$pS_{2,2} + (1 - p)S_{3,2} = pS_{2,3} + (1 - p)S_{3,3} \tag{2}$$

$$p^* = \frac{S_{3,3} - S_{3,2}}{S_{2,2} + S_{3,3} - S_{3,2} - S_{2,3}} \tag{3}$$

3.3 Probit regression model

Probit model was adopted to estimate the payoffs in the previous payoff matrix since the payoff in each cell is the estimated winning probability under each kind of strategy. It contains two major variables, including the two points and three points goals, and some other variables to make the estimation more precise. Moreover, for each team, they have different coefficients in the model due to their different characteristics. The model we used is as follows.

$$
\begin{aligned}
Pr(Y_i = 1) = & \phi(\beta_{0i} + \beta_{1i}P_{2i} + \beta_{2i}P_{3i} + \gamma_{1i}DRB_i + \gamma_{2i}TOV_i + \gamma_{3i}STL_i + \gamma_{4i}BLK_i \\
& + \gamma_{5i}FOL_i + \gamma_{6i}Home_i + \epsilon)
\end{aligned} \tag{4}
$$

Here in the model, P_{2i} represents the 2 points goals for team i, while P_{3i} representing the 3 points goals. The other variables include defense rebounds (DRB), turnovers (TOV), steals (STL), blocks (BLK) and fouls (FOL) for each team. In addition, a dummy variable Home is also included in the model, representing whether the game is played in home or as guests.

We collected the data for all the teams of 2016-2017 regular season and run the regression for each team separately to settle all the coefficients in the model and we would like to estimate for the payoffs with those known coefficients. For those other variables we just take the average performance in the whole season of the team back to the equation. And we adjust the value of the major variables, trying to describe the four payoffs in the matrix.

$$
S_{a,b} = \phi(\hat{\beta_0} + \hat{\beta_1}\hat{P_2}i + \hat{\beta_2}\hat{P_{3i}} + \hat{\gamma_1}_i D\bar{R}B_i + \hat{\gamma_3}_i T\bar{O}V_i + \hat{\gamma_3}_i S\bar{T}L_i + \hat{\gamma_4}_i B\bar{L}K_i + \hat{\gamma_5}_i F\bar{O}L_i + \hat{\gamma_6}_i Home_i) \tag{5}
$$

Now is the question that how do we adjust the value of$\hat{P_2}iand\hat{P_3}i$to satisfy our strategies. Intuitively, the number of goals is the product of goal percentages and goal attempts, so we have:

$$
\hat{P} = PA \times P\% \tag{6}
$$

Therefore, we would like to adjust the value of goal percentages and goal attempts to adjust the value of goals. Noticed that the strategies we defined are assumed to be extreme, (i.e. if offense team choose playing on 2 points, we will consider that they'll shoot on 2 points as much as possible, while shooting on 3 points rarely. Similar for defense team, if they choose to play on two points, we will consider that they'll defend on 2 points majorly, while defending on 3 points rarely) which will have an impact on the goal percentages of offense team directly. As a result, if both sides play on 2 points, we will take the highest 2 points attempts and 3 points percentages as well as the lowest 3 points attempts and 2 points percentages back to the model.

$$
\hat{P_{2i}} = (highest2PA_i) \times (lowest2P\%_i) \tag{7}
$$

$$
\hat{P_{3i}} = (lowest3PA_i) \times (highest3P\%_i) \tag{8}
$$

Furthermore, we can calculate the payoff of $S_{2,2}$ as well as other three payoffs in this way.

4 Empirical Evidence

4.1 Example of payoff matrix

Based on the model, we have calculated the payoffs for each team. Here is an example of the payoff matrix, which is the result of New York Knicks.

		Defence	
		Defend 2	Defend 3
Offense	Shoot 2	$0.202, 0.798$	$0.753, 0.247$
	Shoot 3	$0.757, 0.243$	$0.198, 0.802$

We can see if Knicks plays on 2 points as offense team and the opponent plays on 2 points as defense side, Knicks have a probability of 20.2% to win the game. Intuitively, the winning probability is 79.8% correspondingly. Moreover, the payoffs in the matrix satisfies our initial assumptions shown in equation 1, which means there's no any PSNE as well as dominated strategies in the game and there must be a mixed-strategy Nash equilibrium in the game, representing the optimal distribution of 2 points and 3 points shots.

Generally, we find that over 80% of teams in that season satisfy our initial assumption, verifying the settings on strategies are reasonable.

4.2 Real-world evidence

Then we would like to investigate whether the predicted optimal allocation between 2 points and 3 points attempts matches the real one. We will illustrate with some examples first.

Here is the result of New Orleans Pelicans, the 10th place in western. We can see from the last row of the table, the accuracy of prediction increased after adding those other variables, reaching to 80%. The prediction accuracy measures how much do predictions on victory or defeat by the model harmonize with the real ones, which is a good indicator showing the quality of the model. Here, the prediction accuracy is 86.6%, representing the model is reliable to some extent. The result is shown in Figure 1.

Then, we calculate the mixed-strategy Nash Equilibrium in this game. The left column represents the predicted allocation of goal attempts from our model while the right column represents the actual one. It is obvious to see that prediction are close to the reality in Figure 2.

Some results also show a great match for eastern teams. This is the result of Atlantic Hawks, which took the 5th place in eastern. Similarly, the prediction is accurate and the predicted allocation is quite close to the real one. The results are shown in figure 3 and 4.

Figure 1: coefficients of New Orleans Pelicans

VARIABLES	(1)	(2)	(3)
Point_2	0.0810**	0.126***	0.131***
	(0.0335)	(0.0413)	(0.0405)
Point_3	0.214***	0.262***	0.272***
	(0.0576)	(0.0633)	(0.0666)
DRB		0.205***	0.220***
		(0.0651)	(0.0692)
TOV		-0.112*	-0.116*
		(0.0625)	(0.0619)
STL		0.235***	0.237***
		(0.0589)	(0.0583)
BLK		-0.218***	-0.215***
		(0.0750)	(0.0743)
FOL		0.00516	0.00876
		(0.0470)	(0.0476)
Home			-0.317
			(0.380)
Constant	-4.670***	-13.06***	-13.73***
	(1.371)	(2.931)	(3.068)
Observations	82	82	82
Pseudo R-squared	0.157	0.481	0.486
Prediction Accuracy	0.695	0.866	0.573

Figure 2: predicted and real allocation for New Orleans Pelicans

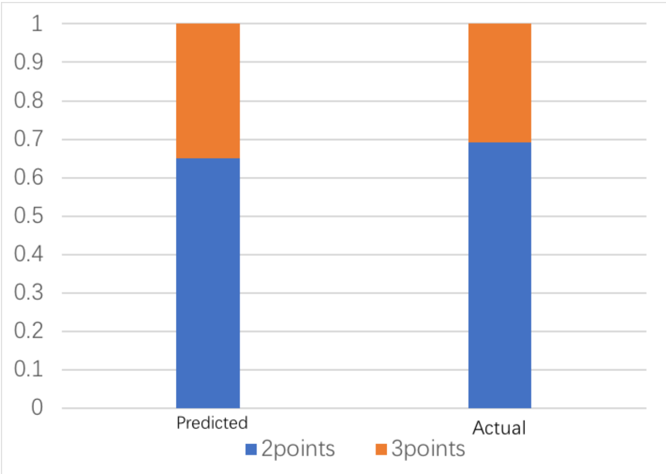
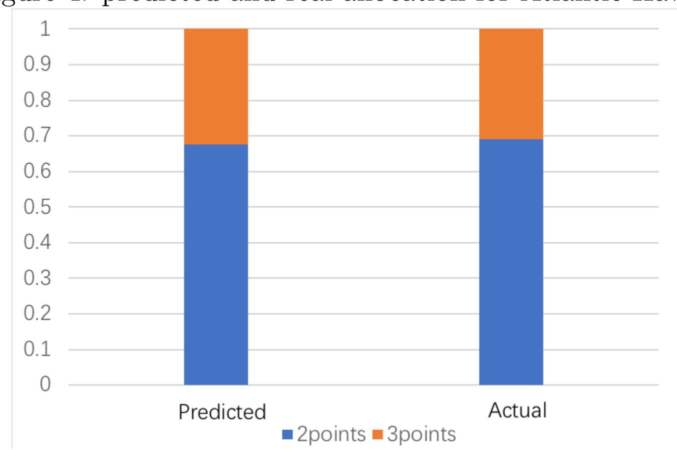


Figure 3: coefficients of Atlantic Hawks

VARIABLES	(1)	(2)	(3)
Point_2	0.137***	0.122***	0.119***
	(0.0374)	(0.0411)	(0.0419)
Point_3	0.158***	0.287***	0.286***
	(0.0551)	(0.0711)	(0.0708)
DRB		0.134***	0.137***
		(0.0429)	(0.0437)
TOV		-0.0235	-0.0226
		(0.0483)	(0.0482)
STL		0.205***	0.206***
		(0.0665)	(0.0681)
BLK		0.0956	0.0936
		(0.0780)	(0.0772)
FOL		0.0470	0.0447
		(0.0546)	(0.0555)
Home			0.159
			(0.362)
Constant	-5.325***	-13.23***	-13.30***
	(1.318)	(2.675)	(2.683)
Observations	82	82	82
Pseudo R-square	0.205	0.383	0.385
Prediction Accuracy	0.732	0.817	0.805

Figure 4: predicted and real allocation for Atlantic Hawks



However, some prediction doesn't have access to the reality. Here is the result of Houston Rockets, which took the third place in the regular season in western. The result shows there's a huge mismatch between the real and predicted allocation even if the prediction accuracy was over 90%. The results are shown in figure 5 and 6.

Figure 5: coefficients of Houston Rockets

VARIABLES	(1)	(2)	(3)
Point_2	0.152***	0.184***	0.204***
	(0.0437)	(0.0667)	(0.0753)
Point_3	0.274***	0.349***	0.358***
	(0.0520)	(0.0801)	(0.0911)
DRB		0.328***	0.330***
		(0.0656)	(0.0693)
TOV		-0.0972	-0.0721
		(0.0667)	(0.0598)
STL		0.295***	0.283***
		(0.105)	(0.106)
BLK		0.212*	0.187*
		(0.111)	(0.108)
FOL		0.179***	0.190***
		(0.0688)	(0.0717)
home			0.504
			(0.537)
Constant	-7.214***	-24.97***	-26.22***
	(1.564)	(4.864)	(6.057)
Observations	82	82	82
Pseudo R-square	0.317	0.656	0.663
Prediction Accuracy	0.756	0.878	0.903

In fact, we have calculated the predicted optimal allocation for all the teams in the season of 2016-2017, trying to investigate which kind of teams have a better match between the prediction and the reality. The mismatch between our prediction and reality for all the teams that satisfies our initial assumption is shown in figure 7.

The horizontal axis represents the rank of real winning rate in the season and "23" means the team had the highest winning rate in the season. The height of the bar represents the mismatch for teams.

We can see from the figure that mismatch for teams with higher winning rate is relatively greater than other groups. Besides, we discover that those teams with better match were middle ranked. Therefore, we speculate that probably intermediate teams match better. Thus, we take eight intermediate teams out as a subsample, calculating the mixed-strategy Nash Equilibrium within this group and comparing them with the previous results.

In figure 8, we can see the mismatch of majority of intermediate teams declines, indicating the intermediate teams match better.

Figure 6: predicted and real allocation for Houston Rockets

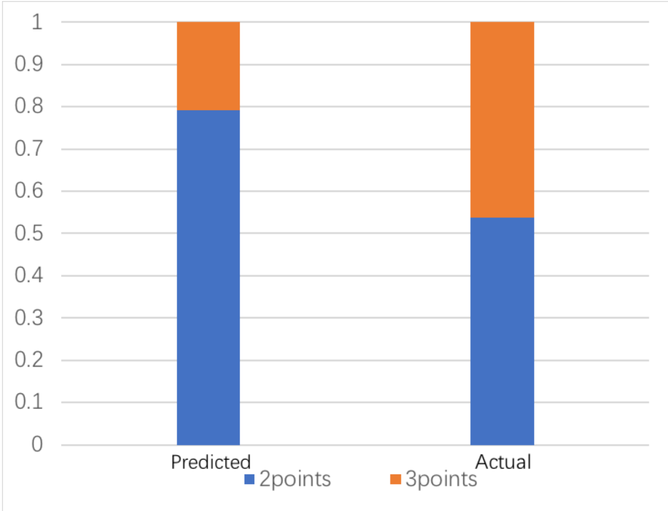


Figure 7: mismatch for teams in the season of 2016-2017

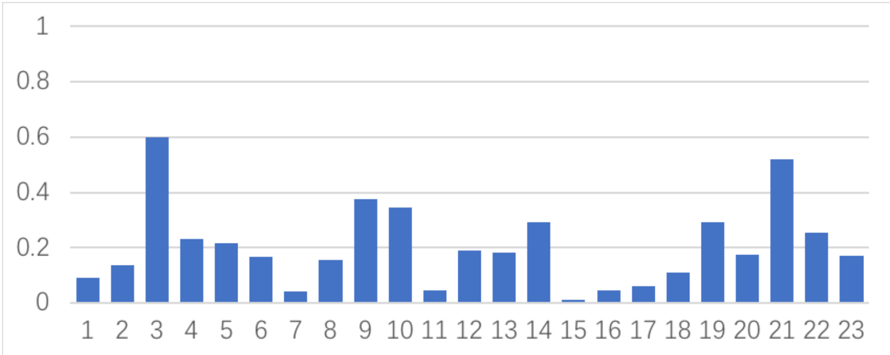
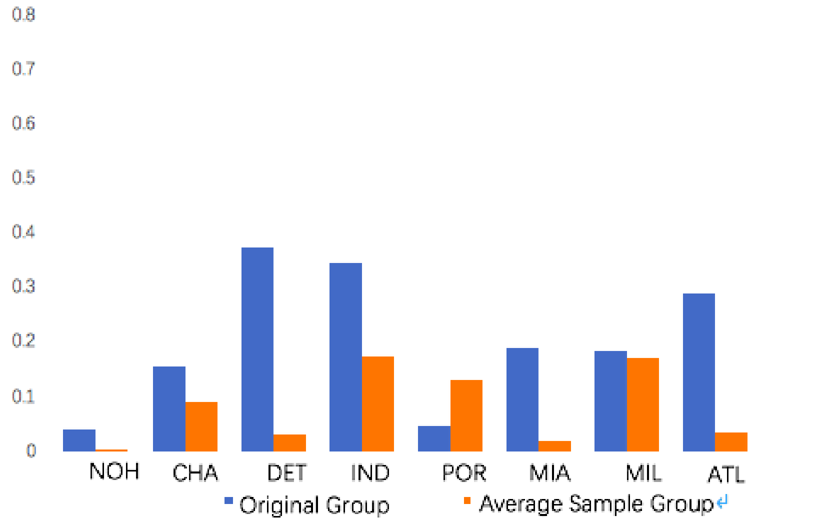


Figure 8: mismatch for eight intermediate teams within the average sample group



4.3 Potential interpretation

Based on previous results, we found that our model is more suitable for intermediate-level teams. Since we mainly focus on offensive side, trying to deal with the distribution of 2 and 3 points shots, we assume that the defensive side is at an average level, which ensures the robustness of our model. Besides, out of symmetry of the two parts, the outcomes would be better when the two teams possess more comparable capabilities. For simplicity, we assume both sides have no access to their rivals' strategies, since the effect of the information they get is hard to be quantified in such games. However, they know each other very well in the real world.

Making advantage of regression above, the strategy used in basketball competitions could be quantified.

According to practical data we get, most of the teams' 2 points shot constitute 70% of the total number of shootings. As for some teams with certain features such as superstars and conventional strategies, our model may be not accurate enough. As mentioned before, Houston rockets prefers to throw lots of 3 points, leading to certain mismatch between our prediction and real situation.

5 Conclusion

Our aim is to find the equilibrium of 2 and 3 points shots allocation in professional basketball games, at which the winning rate could be maximized. According to previous researches and empirical experience, we take mixed strategy theorem as the base of our proposal. Despite the complexity of completed basketball games, we mainly focus on the offensive side, dividing the whole competition into one round games. Inspired by the offensive rate defined by Justin Kubatko et al. (2007), the ratio, the number of achieved shootings to the number of attempted shootings is used as our interested independent variable in the regression model to predict theoretical winning rate, that is, the dependent variable. Finally, the theoretical mixed strategy Nash equilibrium could be figured out with probit regression model and from data of NBA regular season 2016-2017. Here are conclusions we draw from our work.

The shooting strategy used in basketball could be quantified. Our model successfully uses the defined independent variable to estimate the expected winning rate with the help of empirical data and probit regression model. The result of sensitiveness test shows an accuracy of 80% or so, which differs across different teams, implying that shooting strategy could be quantified and our method is feasible.

The best shooting strategy for intermediate-level teams is to have 2 points vs 3 points shooting roughly equals to 7:3. In general, intermediate-level teams do not have their own special strategies which could almost ensure their wins, and adopting this method could help them maximize their winning rate when facing to an intermediate-level team.

Teams with outstanding skills should use them in an appropriate way. Our model also suggests that not all the teams should have a 7:3 distribution of 2 and 3 points shots. For teams with unique feats, sufficiently using their advantages is also maximizing their winning rate. For instance, Houston rockets does well in throwing 3 points, which helps them reach lots of goals, and this feat is hard to be duplicated in a short time, which ensures their absolute advantages in games.

Besides, our research could be regarded as an innovative one, which offers a new angle to study basketball strategies. To the best of our knowledge, there is few models using estimated winning rate to study the theoretically reasonable distribution of 2 and 3 points shootings. Most of the papers merely use the shooting efficiency, which is defined as the number of achieved shots to the number of attempted shootings. The methodology of this paper may be a threshold for future researches.

6 Prospective work

To make this model more accurate, several details needs to be revised.

As far as we are concerned, more factors should be taken into consideration when constructing regression models, such as players' physical condition, confidence, or even weather and the enthusiasm of audience, which are likely to act essential roles in every basketball game. However, they are hard to be quantified for now, which asks for more advanced econometric skills.

Moreover, this model could be adjusted to do some dynamic analysis. Since the pace of basketball games varies a lot, in time adjustment is necessary, which needs more inner study in the future. From our present work, the changing score difference could be taken into consideration to accomplish this purpose. The research could be more fruitful, with our proposal as a starting point.

References

[1] I. Palacios-Huerta, “Professionals play minimax,” *The Review of Economic Studies*, vol. 70, no. 2, pp. 395–415, 2003. [Online]. Available: http://dx.doi.org/10.1111/1467-937X.00249

[2] M. Walker and J. Wooders, “Minimax play at wimbledon,” *American Economic Review*, vol. 91, no. 5, pp. 1521–1538, December 2001. [Online]. Available: http://www.aeaweb.org/articles?id=10.1257/aer.91.5.1521

[3] J. Kubatko, D. Oliver, K. Pelton, and T. R. Dan, “A starting point for analyzing basketball statistics,” *Journal of Quantitative Analysis in Sports*, vol. 3, no. 3, pp. 1–1, 2007.

[4] O. A. Entine and D. S. Small, “The role of rest in the nba home-court advantage,” *Journal of Quantitative Analysis in Sports*, vol. 4, no. 2, pp. 6–6, 2008.