The Effectiveness of the Elo Rating System in the NBA

Neil Weingarten – Graduating May 2023

KENNESAW STATE UNIVERSITY
COLLEGE OF COMPUTING AND SOFTWARE ENGINEERING
School of Data Science and Analytics

Advisors: Professor Michael Frankel, Dr. Marla Bell

INTRODUCTION

The Elo Rating system is a method of calculating the relative skill levels of players that began being used to rate players in chess. The system was adopted in favor over the Harkness system due to its ability to compare players from evaluating underlying variables as opposed to the players' individual achievements. Over time, Elo has been adopted as a rating system in several different sports, including basketball. These Elo ratings were then used to predict the outcome of NBA games. Over time through the NBA's history, two other rating systems would be created to rival Elo in becoming better predictors, those being the Carmelo rating and the Raptor rating. As sports continue to evolve, finding an effective rating system is important in determining the outcome of games.

METHODS

Binary Logistic Regression was used to evaluate which predictor between Elo, Carmelo, and Raptor was more accurate in determining whether the home team wins a game.

Simple Linear Regression was used to view the relationship between the total score of a match and the quality rating of a match.

Two-Way ANOVA was used to see if there was an interaction effect on the Elo change of the home team by if the home team won the game and if it was a playoff game.

Stratified Boxplots were used to view the descriptive statistics and the variability of Elo, Carmelo, and Raptor in the 5 teams with the most games played.

Chi-Square was used to determine if Elo changes were different in playoff games compared to regular season games.

Multiple Linear Regression was used to determine the strongest quantitative predictors of the total score. The dataset was broken into subsections to allow for accurate predictions using Carmelo and Raptor. Furthermore, multicollinearity was explored to see which predictors in the dataset were related to each other.

Line Charts were used to explore the Elo history of NBA teams and visualize these changes in Elo.

An **Overlayed Regression Plot** was used to visualize the Elo changes between the Golden State Warriors and the Cleveland Cavaliers during the 2016 season.

CODE

GitHub

Figure 1: ROC Curves for Three Logistic Regression Models Predicting a Home Team Win

Left: Elo; Middle: Carmelo; Right: Raptor

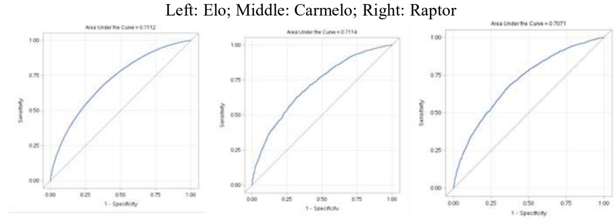


Table 1: Model for Score Total from Quality Estimate

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	208.32911	0.18019	1156.17	<.0001
quality	1	-0.03284	0.00336	-9.77	<.0001

Table 2: Two-Way ANOVA

The GLM Procedure

Dependent Variable: elo_change

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	376591.1737	34235.5612	1454.46	<.0001
Error	4618	108700.2179	23.5384		
Corrected Total	4629	485291.3916			

R-Square	Coeff Var	Root MSE	elo_change Mean
0.776010	1829.237	4.851636	0.265227

Source	DF	Type III SS	Mean Square	F Value	Pr > F
playoff	5	153.87320	30.77464	1.31	0.2576
homewin	1	23417.08645	23417.08645	994.85	<.0001
playoff*homewin	5	890.57927	178.11585	7.57	<.0001

Figure 2: Tukey's Post Hoc Test

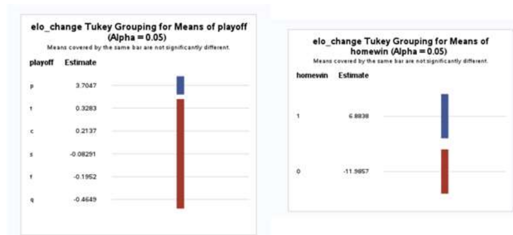


Table 3: Chi-Square Test Results

The FREQ Procedure

Frequency / Expected / Deviation

Table of elo_changecat by playoffstatus			
	playoffstatus		
elo_changecat	n	y	Total
Large gain	7494 7532.6 -38.65	555 517.35 38.649	[illegible]
Large loss	12064 12121 -57.48	890 832.52 57.483	12954
Small gain	33891 34005 -114.5	2450 2335.5 114.47	36341
Small loss	13964 13753 210.6	724 944.6 -210.6	14688
Total	67413	4630	72043

Statistics for Table of elo_changecat by playoffstatus

Statistic	DF	Value	Prob
Chi-Square	3	[illegible]	<.0001
Likelihood Ratio Chi-Square	3	[illegible]	<.0001
Mantel-Haenszel Chi-Square	1	[illegible]	<.0001
Phi Coefficient		0.0297	
Contingency Coefficient		0.0297	
Cramer's V		0.0297	

Sample Size = 72043

Figure 3: Stratified Boxplots of Rating Predictors by 5 Teams with the Most Games Played

Left: Elo; Middle: Carmelo; Right: Raptor

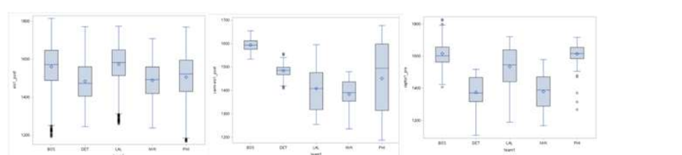


Table 4: Multiple Linear Regression Output for Predicting Total Score

From Top to Bottom: Entire Dataset, Seasons 2015-2019 (where Carmelo was tracked), Seasons 2019+ (where Raptor was tracked)

Number in Model	C(p)	R-Square	Adjusted R-Square	Parameter Estimates: Intercept	elo_diff	playoffstatus	season	quality
4	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Number in Model	C(p)	R-Square	Adjusted R-Square	Parameter Estimates: Intercept	elo_diff	playoffstatus	season	quality	carmelo_diff
3	2.5000	[illegible]	[illegible]	-10322		-4.56823	5.22138	0.02163	

Number in Model	C(p)	R-Square	Adjusted R-Square	Parameter Estimates: Intercept	elo1_pre	elo_diff	playoffstatus	season	quality	carmelo_diff	carm-elo1_pre	raptor_diff	raptor1_pre
5	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]					[illegible]	[illegible]

Figure 4: Line Chart of Golden State Warriors' Elo Changes over Time

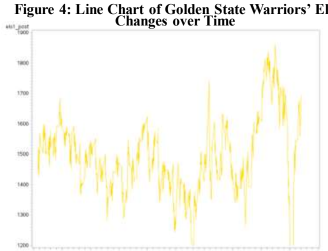
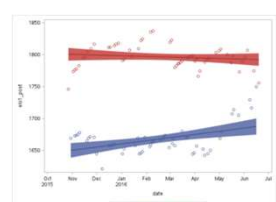


Figure 5: Line Chart of Chicago Bulls' Elo Changes over Time

Figure 6: Overlayed Regression Plot of Golden State Warriors' Elo Changes and Cleveland Cavaliers' Elo Changes during 2016 NBA Season



RESULTS

Binary Logistic Regression: Figure 1 shows the ROC curves for the three predictors of a home team win individually. Carmelo proved to be the strongest predictor of the three, with Elo coming in a close second and Raptor a more distant third.

Two-way ANOVA: Table 2 shows that the home team win itself and the interaction between a home team winning and the kind of playoff game were significant in determining the change of Elo in the home team. The Tukey's Post-Hoc Test revealed that Elo losses are typically larger for the home team than Elo gains. Also, Play-in games cause larger Elo changes than any other kind of NBA playoff game.

Simple Linear Regression: Table 1 shows that the predicted quality rating of a matchup tended to generate smaller scoring totals.

Stratified Boxplots: Figure 3 shows that the variability between the 5 teams with the most games played with regards to Elo are less pronounced than the variability with regards to Carmelo and Raptor. Of the 5 teams selected for analysis here, it appears that the Boston Celtics rank higher than the other teams in long-term and short-term ratings.

Chi-Square: Table 3 shows that changes in Elo were indeed different depending on whether the game was a playoff game or not. Large changes in Elo were more prevalent in playoff games as well as small Elo gains. This shows that the home team advantage during playoff games were more pronounced, and when the home team loses, the losses in Elo they suffer are greater during the playoffs.

Multiple Linear Regression: As seen in Table 4, with multicollinearity considered, when looking at the entire dataset (with Elo present in all observations), the playoff status, season, and quality rating all caused the total score to lower whereas total score increased when the Elo difference increased (favoring the home team more). However, the strongest model came from looking at the dataset from 2015-2019 seasons (with Carmelo present in all observations). In this subsection, total score decreased during the playoffs, whereas total score increased in later seasons and when the quality rating was larger.

Line Charts: Figures 4 and 5 show the Elo changes of the two teams with the highest peaks in the NBA: the Golden State Warriors and the Chicago Bulls. Shortly after they reached their peak, they soon after reached their lowest point.

Overlayed Regression Plot: Figure 6 shows an overlayed regression plot of the Elo changes between the Golden State Warriors and the Cleveland Cavaliers during the 2016 season, the same season that Cleveland pulled a notable comeback during the NBA finals against the Warriors. For almost the entire season, the Warriors' Elo was much higher than the Cavaliers' Elo up until the Finals where their Elo's were almost even.

CONCLUSIONS

The use of the Elo, Carmelo, and Raptor rating systems can be effective predictors, but the consideration of other factors such as injuries and schedule should be considered in order to make more accurate predictions of the outcomes of NBA games.

- Predicting both the winner of the game and the total score of the game is different between **playoff games** and **regular season games**. The home team is generally more favored to win playoff games and these games generally produced a lower score total.
- The higher the **quality rating** of a game, the lower **total score** will be. Quality ratings can be useful indicators of whether it will be a fair match or closer to a complete blowout.
- **Raptor Ratings** are a less effective predictor of outcome than either **Elo** or **Carmelo**, however, seeing as how the Raptor ratings are a newer implementation, it is likely that this rating system will be adjusted to be more effective.
- While **Carmelo** can be useful in **predicting the winner** of a match, it is not as useful in **predicting the total score** like **Elo** and **Raptor** are.