From Pole to Podium: Adjusting Elo Method to Separate Car and Driver in Formula One Racing

Zijian Xun*

Project Advisor: Timothy P. Chartier+

Author Note

*Zijian Xun, first year student at Carleton College, harryxun2022@163.com

+Timothy P. Chartier, Joseph R. Morton Professor of Mathematics and Computer Science at Davidson College, tichartier@davidson.edu

Abstract

This article presents a novel approach to separating and quantifying the effect of a car's performance and a driver's skill on Formula 1(F1) race outcomes. By analyzing data from the past decade, we propose a formula to measure F1 drivers' ability. This approach could be used to predict race outcomes for a given driver in cars with different performance levels, thereby aiding teams in optimizing resource allocation for car development.

Keywords: **Elo ranking system, Formula one, Linear Regression**

From pole to podium: Adjusting Elo method to separate car and driver in Formula One racing

Ⅰ. Introduction

Formula 1 racing (F1) is a sport that heavily relies on interactions with drivers' skill and cars' technological prowess. Despite the rigorous selection process and the intensive training of F1 drivers, the cars' technology often plays a critical role in the outcome of a race. The significant variations in car performance, budget, and developer's competency among teams emphasize the need to optimize resource allocation for car development.

This article presents a modified Elo method to isolate the effects of the car and driver in F1 racing and rate a driver's true ability. The Elo method is a rating system for calculating the relative skill levels of players in zero-sum games. It's widely used in chess[1], e-sports[2], etc. Performance in the Elo system is inferred from wins, losses, and draws against other players. Players' ratings depend on the ratings of their opponents and the results of the game. After every game, the winning player takes points from the losing one, and the number of points is determined by the difference in the two players ratings. If the higher-rated player wins, a few points are taken from the lower-rated player. If the lower-rated player wins, a lot of points are taken from the higher-rated player. If it's a draw, the lower-rated player gains a few points from the higher-rated player.[3]

Ⅱ. Literature Review

Reiner Eichenberger and David Stadelmann in 2009[4] modeled a driver's classification based on six factors: drivers' effect, car-year effect, numbers of drivers finishing the Grand Prix, technical failure, weather condition and the length of the track. Our adjusted Elo method incorporates two of these variables - driver performance and car-year effect - though the Elo method necessitates the estimation of cars and drivers after each race, as opposed to an annual basis. Given the performance variations of the same car across different races, an annual estimation of cars is not suitable for the Elo method. Also, as the number of drivers is a constant after 2004, weather condition is considered as luck in Elo method, and track length is factored in as part of car's performance, these three factors are not independently considered in our method.

The previous multilevel modeling study[5] conducted by Andrew Bell, James Smith, Clive E. Sabel and Kelvyn Jones analyzed what affects a driver's position. However, the article uses car effect and driver effect on an annual basis, making it unsuitable for the Elo method. The article points out that drivers who are competing against better drivers will tend to perform worse than those competing against worse drivers. This variable will be eliminated by Elo. By definition of the Elo method above, a driver competing against superior drivers would gain more Elo points upon winning, thereby nullifying the effects of competition quality.

Ⅲ.Methodology

The primary focus of this paper is a novel rating adjustment methodology that combines a car's rating with a driver's Elo rating, thereby providing a more comprehensive evaluation of a driver's winning probability.

To achieve this, we first establish a system to rate cars based on the average of the fastest lap times during qualifying rounds. The system is validated using data from the 2011-2013 seasons. We then construct an Elo rating system for drivers, using race results as a series of 1-on-1 tournaments. The accuracy of this system is tested using root mean square error (RMSE). To combine the two ratings, we conduct a linear regression analysis on the difference in car and driver ratings across consecutive seasons. This analysis yields a model that effectively ties together the performance of the car and the driver.

By using the results of the linear regression analysis, our revised Elo rating system now incorporates the car's performance into a driver's rating.This final rating adjustment methodology offers a more holistic estimation of a driver's winning probability. This paper will demonstrate that this methodology significantly enhances the predictive accuracy and fairness of the ranking system.

Ⅳ. Car Rating

The data for this article is collected from pitwall.app, a professional online Formula 1 database containing data and statistics from seasons 1950 to 2023.

The aim of our study is to track the performance of race cars. Car performance is a dynamic metric contingent upon various variables that differ greatly on each track. As such, we assess the performance of cars in each Grand Prix, rather than a yearly analysis. To cater to these variables, we propose a new rating method, represented by the formula:

$$Rc_i = \frac{T_{driver} - T_{fastest}}{T_{fastest}}$$

Rc_i represents the car rating in a track. This value varies for the same car on different tracks. Notably, a higher rating signifies a slower car.

T_{driver} represents the average of the fastest lap times of each of the two drivers in the same team in the qualifying round.

$T_{fastest}$ represents the fastest lap time recorded among any driver in the corresponding qualifying round.

The proposed formula has multiple benefits. First, it adjusts for track length by using lap times rather than average speeds, which can significantly vary across different tracks. Hence, it allows for more reliable comparisons of car performances across different tracks. Secondly, by taking the fastest lap time during qualifying as the baseline, the car rating provides a readily interpretable measure of a car's performance relative to the best achievable performance under similar conditions.

To further assess the effectiveness of the formula, for each season, we calculated the average Rc_i for the same team on each track, written as $\overline{Rc}$. Then we compared each team's earned points with their $\overline{Rc}$. We gathered data from season 2011 to season 2013. The results are as follows.

Team	$\overline{Rc}$ in 2011	Points in 2011	$\overline{Rc}$ in 2012	Points in 2012	$\overline{Rc}$ in 2013	Points in 2013
Caterham			7.7449325	0	12.387532	0
Ferrari	2.141768737	375	1.93963375	400	1.9281430	354
Force India	6.754664737	44	2.55701145	109	4.8519726	77
HRT	17.32771053	0	12.8682115	0		
Lotus	12.20876421	0	2.05247755	303	2.4354868	315
Manor			10.5801255	0	12.677149	0
McLaren	1.944798263	497	1.00864617	378	3.7300731	122
Mercedes	3.529433158	165	2.1757438	142	1.0549817	360
Red Bull	0.646128163	650	1.10434288	460	0.8650631	596
Renault	5.00788	73				
Sauber	6.624633158	69	2.72293741	126	6.9545678	57
Toro Rosso	7.306819474	41	4.20052	26	6.2960221	33
Virgin	15.86462632	0				
Williams	7.720824211	5	2.7844538	76	8.2951342	5

Table 1-Average car ratings and team standings in season 2011 to 2013

Through simple linear analysis of the data in the chart, we used the average car rating ($\overline{Rc}$) as the independent variable and the team's annual points as the dependent variable. Our analysis yielded the equation $y = -66.335x + 459.29$. This indicates that for every one-point increase in a team's $\overline{Rc}$ in a season, their annual points decrease by 66.335. This finding aligns with our initial hypothesis that the faster a team's car, the lower its $\overline{Rc}$ would be, providing robust evidence of the effectiveness of our car rating system.

Ⅴ. Original Elo Rating

In this section, our objective is to develop and implement an Elo rating system designed to rank Formula 1 drivers. This process will be evaluated based on the root mean error (RMSE) and the driver rankings on the scoreboard.

Firstly, we clarify how we view and represent each race. We model each race as if it were a round-robin 1-on-1 tournament. For instance, a driver who finishes second out of 20 drivers is viewed as having an 18-1 record in this tournament --- having lost to the first-place finisher but defeated all the rest.

Secondly, we explain how the rating is initiated. To avoid negative Elo ratings, we assign every driver an initial Elo rating of 2000. It's important to note that the starting Elo ratings do not impact the final rankings or the accuracy of the Elo model, as this system is fundamentally focused on the differences in drivers' ratings.

Next, we detail our measure of accuracy for the Elo system: the root mean sqaured error (RMSE):

$$RMSE = \sqrt{\frac{1}{n}\sum (prediction - actual)^2}$$

Prediction signifies the predicted outcome of a race, represented as a value from 0 to 1, indicating the likelihood of a driver winning.

Actual signifies the factual outcome, which can be 1 (win) , 0 (lose), or 0.5 (draw).

In this system, a lower RMSE is indicative of a more accurate Elo system.

We then provide a practical illustration of the Elo ratings using the 2021 season's data, and the RMSE we calculated is 0.46. In other words, about 68% of the drivers' Elo ratings are within 0.46 points from what the model predicts. However, we noticed some errors caused by the Elo ratings.

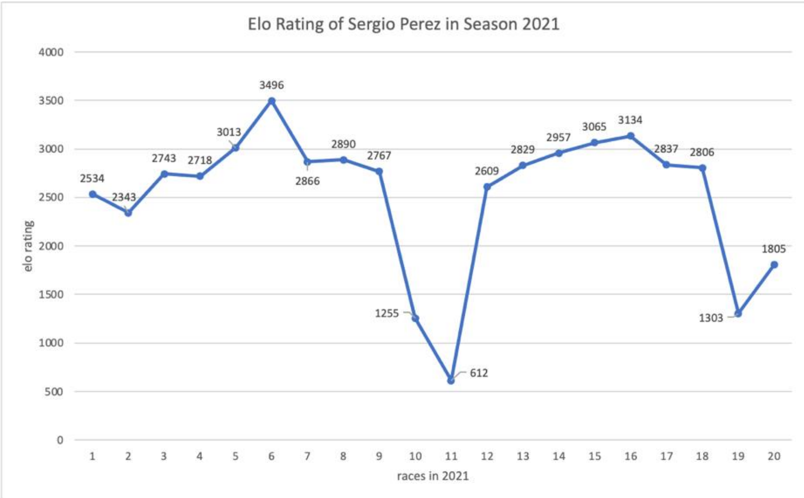

Figure 1-Elo rating of Sergio Perez in Season 2021 after each race

Figure 1 illustrates the Elo rating progression of Sergio Perez throughout the 2021 season. The x-axis represents the consecutive races, while the y-axis represents his Elo rating after each race based on the performance of Sergio Perez relative to other drivers, following the Elo rating rules. If he dropped out, he was accounted for by being ranked the last driver, and the data was plotted chronologically to generate the graph.

Despite his 4th rank on the 2021 Formula 1 scoreboard, dropouts in races 11 and 19 negatively impacted Perez's final rating.

Recognizing the unfair penalization caused by dropouts, we modified the Elo system to ignore races where the driver failed to finish. This adjustment resulted in a more sensible ranking and did not compromise the accuracy of the model, as evidenced by the unchanged RMSE of 0.44.

However, we observed another limitation in our Elo rating, with figure 2, showcasing George Russel's Elo ratings.

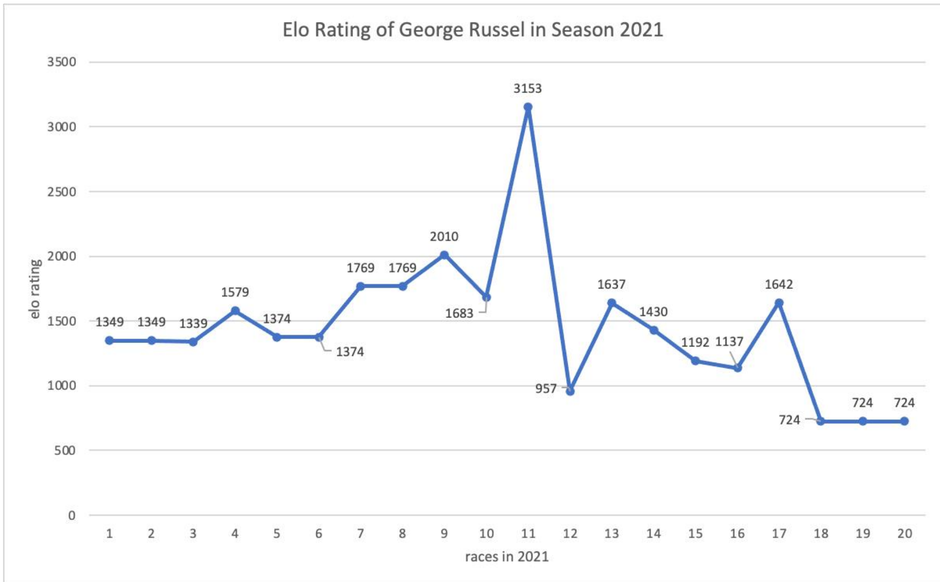

Figure 2-Elo rating of George Russel in season 2021 after each race

Figure 2 follows a similar structure to figure 1, but focused on George Russel. Due to his success in the middle of the season, his Elo rating is 3153 after the 11th race, but because of his failure in the last couple of races, he is only rated 724, which is the lowest rating among all 20 drivers. It is unfair that a poor performance towards season's end exerts a stronger negative impact on the final rating than an equally poor performance at the season's start.

To address this issue, we propose the use of the average Elo rating across all races. Each driver's Elo rating of the seasons would be the average of his Elo ratings after each race. This approach ensures a more balanced reflection of a driver's performance throughout the season, thereby providing a more accurate assessment.

VI. Data Collection and Linear Regression

In this section, we discuss the application of linear regression analysis to investigate the relationship between F1 driver ratings and car ratings. The dataset includes data from F1 seasons across 2011 to 2021, excluding 2020 due to the global pandemic.

To achieve this, two distinct formulas were devised:

1. Difference in Cars: For each driver and across every pair of consecutive seasons, the average car rating ($\overline{Rc}$) calculated by the method mentioned in section IV from the latter season was subtracted from the former. This calculated the 'Difference in Car Rating'. Mathematically, this was represented as:

$$\overline{\Delta \mathrm{Rc}} = \overline{\mathrm{Rc}}(\text{year n}) - \overline{\mathrm{Rc}}(\text{year n} - 1)$$

2. Difference in Driver: Similarly, for the 'Difference in Driver Rating', the Elo rating of the driver calculated by the method mentioned in section V from the preceding season was subtracted from the Elo rating of the same driver for the subsequent season. This was calculated as:

$$\Delta \text{Elo Rating} = \text{Elo Rating}(\text{year n}) - \text{Elo Rating}(\text{year n} - 1)$$

These computations yield a total of 143 data points, each encapsulating the differences in driver and car ratings for a specific driver over two consecutive seasons. This data formed the core input for the subsequent linear regression analysis. A detailed representation of this data is provided in Table 2.

1		difference in car	difference in driver
2	Max Verstappen	-0.91173	628.92
3	Lewis Hamilton	0.14897	111.42
4	Valtteri Bottas	0.14897	-521.41
5	Sergio Pérez	-4.0586	681.91
6	Carlos Sainz	-2.01194	58.29
7	Lando Norris	-1.96781	373.57
8	Charles Leclerc	1.25447	-601.58
9	Daniel Ricciardo	-1.81347	78.56
10	Pierre Gasly	-1.98374	353.59
11	Sebastian Vettel	3.00974	-880.35
12	Lance Stroll	-1.39664	236.89
13	George Russell	-4.39577	464.156
14	Kimi Räikkönen	0.96042	-245.66
15	Antonio Giovinazzi	0.96042	193.25
16	Lewis Hamilton	-0.04419	-405.96
17	Valtteri Bottas	-0.04419	238.32
18	Max Verstappen	0.19701	-113.28
19	Charles Leclerc	-4.86095	1270.32
20	Sebastian Vettel	0.28727	-338.86
21	Carlos Sainz	-0.29189	624.8
22	Pierre Gasly	-1.57756	633.55
23	Daniel Ricciardo	2.04957	-826.84
24	Sergio Pérez	0.69961	41.64

Table 2-Examples of the data

‘Difference in Car’ means the difference between the average car ratings of two consecutive seasons, which is the $\overline{\Delta \mathrm{Rc}}$ in the formula of car rating. ‘Difference in driver’ means the difference between the Elo ratings of two consecutive seasons, which is the $\Delta\text{Elo Rating}$. In some cases, $\overline{\Delta \mathrm{Rc}}$ is the same for two drivers, such as Lewis Hamilton and Valtteri Bottas in the table. This is because these two drivers were in the same team for two consecutive seasons, and thus their differences in car ratings are the same.

To facilitate a graphical interpretation of the relationship $\overline{\Delta \mathrm{Rc}}$ and $\Delta\text{Elo Rating}$, this data is plotted in a rectangular coordinates system. Each driver in a consecutive season corresponds to a point, the x-coordinate of which symbolizes the difference in car rating , $\overline{\Delta Rc}$ and the y-coordinate of which symbolizes the difference in driver rating, $\Delta \text{Elo Rating}$.

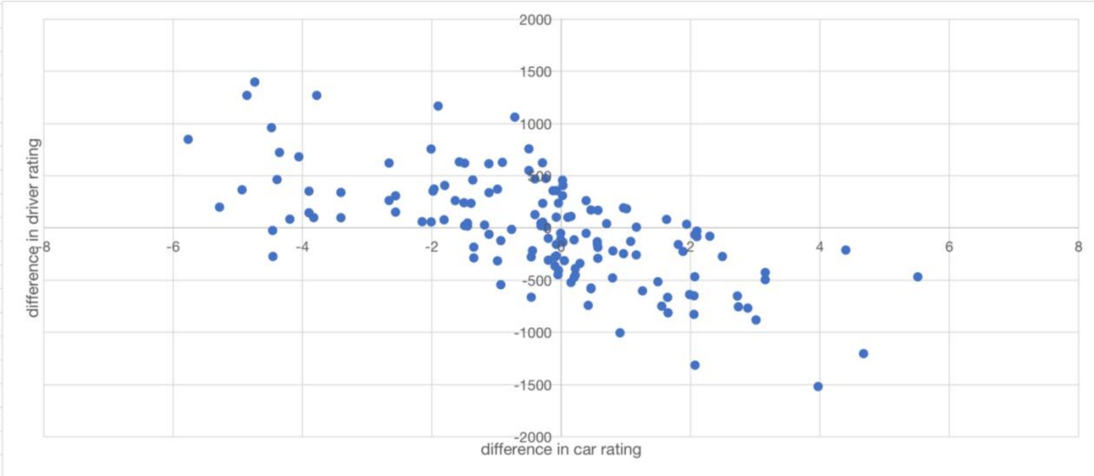

Figure 3-Rectangular coordinate system

The scatter plot reveals a clear negative linear correlation between car and driver rating differences, paving the way for the application of linear regression analysis.

The linear regression analysis is conducted to predict the difference in driver rating based on the difference in car rating. The sample size is 143 (n=143).

Overall the model fits the data, with an R Square of 0.44, suggesting that the model explains 44% of the variance in the difference in driver rating. However, this also means that a substantial portion, approximately 56%, of the variability in drivers' ratings remains unexplained by the model. This seems to be surprising, but considering the nature of the dataset, it would be more understandable. Our data spans across the past ten seasons, which is a substantial timeframe with many potential changes and variability. Moreover, the Elo method, which is used to measure driver ratings, evaluates the relative differences between drivers. Therefore, a driver's rating does not solely depend on their individual performance, but it is also influenced by the performance of other drivers. Thus, the criteria used to determine driver ratings can vary significantly from one season to another. This inconsistency in rating criteria might contribute to the unexplained variance in car ratings.

The regression equation is found to be:

$$\Delta \text{Elo Rating} = -69.4679 - 156.974 \times \overline{\Delta \text{Rc}}$$

The intercept (b0) is -69.4679 and the slope (b1) is -156.974. Both the intercept and slope were statistically significant, indicating a significant negative linear relationship.

The residuals of the model are examined to check the assumptions of linear regression.

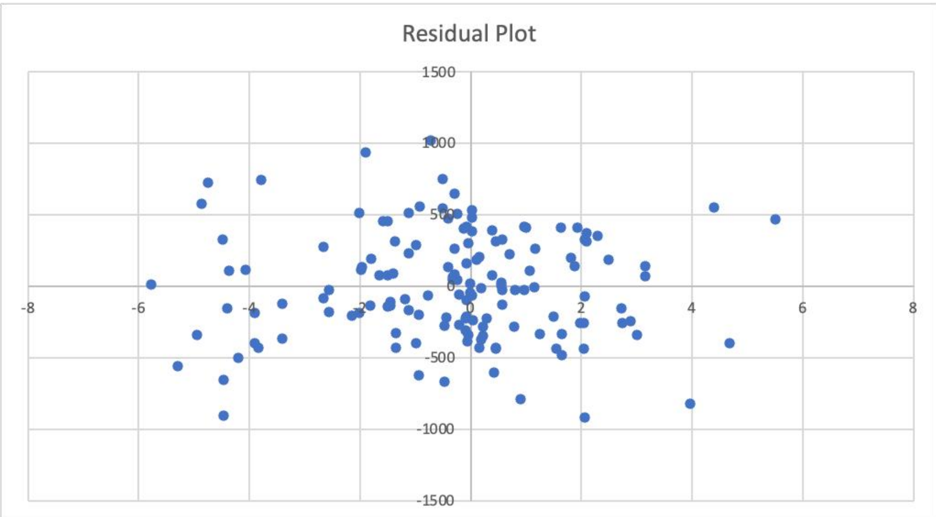

Figure 4-Residual plot

In figure 4, the x-coordinate is the differences in average car rating, $\overline{\Delta \text{Rc}}$, while the y-coordinate is the residual. These residuals are determined by subtracting the predicted differences in driver rating, as calculated by the regression equation, from the actual differences observed in the driver rating.

The residual plot shows a random pattern, indicating that the assumption of linearity is met.

Ⅶ. Adjusting Rating Method

The culmination of this meticulous statistical analysis is a proposition to modify the existing Elo system. The basic principles remain unaltered; driver ratings still commence at 2000, and dropout handling remains consistent. The key modification lies in the computation of winning probabilities, which now considers the influence of car ratings. This leads to an updated Elo rating formula (R), which merges the current Elo rating of the driver (R_{DRIVER}) and the specific race's car's qualifying rating (Rc_i):

$$R = R_{DRIVER} - 156.976 \times Rc_i$$

By integrating both driver ability and car performance, the revised Elo rating offers a more holistic estimation of a driver's winning probability. This innovative approach, underpinned by rigorous linear regression analysis, displays the interplay between driver and car in the realm of Formula 1.

Ⅷ. Analyzing Result

Table 3 showcases the adjusted ratings for various drivers across three distinct seasons, highlighting the effectiveness of the proposed method.

2021		2019		2018	
#33 Max Verstappen	3513	#44Lewis Hamilton	2817.97	#44Lewis Hamilton	3007.93
#44 Lewis Hamilton	3050	#77Valtteri Bottas	2859.26	#5Sebastian Vettel	2653.55
#77 Valtteri Bottas	2513	#33Max Verstappen	2918.7	#7Kimi Räikkönen	2630.33
#11 Sergio Pérez	2521	#16Charles Leclerc	2802.95	#33Max Verstappen	2869.51
#55 Carlos Sainz	2301	#5Sebastian Vettel	2397.61	#77Valtteri Bottas	2420.66
#4 Lando Norris	2364	#55Carlos Sainz	2464.49	#3Daniel Ricciardo	2591.26
#16 Charles Leclerc	2343	#10Pierre Gasly	2286.44	#27Nico Hülkenberg	2095.87
#3 Daniel Ricciardo	2140	#23Alexander Albon	1872.74	#11Sergio Pérez	1875.27
#10 Pierre Gasly	2537	#3Daniel Ricciardo	1980.82	#20Kevin Magnussen	1614.26
#14 Fernando Alonso	2014	#11Sergio Pérez	2061.93	#55Carlos Sainz	1826.56
#31 Esteban Ocon	1987	#4Lando Norris	2113.11	#14Fernando Alonso	1955.06
#5 Sebastian Vettel	2098	#7Kimi Räikkönen	1580.12	#31Esteban Ocon	2049.74
#18 Lance Stroll	2008	#26Daniil Kvyat	1812.96	#16Charles Leclerc	1939.08
#22 Yuki Tsunoda	1800	#27Nico Hülkenberg	1752.22	#8Romain Grosjean	1780.46
#63 George Russell	1435	#18Lance Stroll	1649.5	#10Pierre Gasly	1709.42
#7 Kimi Räikkönen	1365	#20Kevin Magnussen	1186.12	#2Stoffel Vandoorne	1547.5
#6 Nicholas Latifi	1086	#99Antonio Giovinazzi	1349.54	#9Marcus Ericsson	1667.43
#99 Antonio Giovinazzi	1569	#8Romain Grosjean	1263.27	#18Lance Stroll	1240.64
#47 Mick Schumacher	765	#88Robert Kubica	1214.68	#28Brendon Hartley	1308.95
#9 Nikita Mazepin	591	#63George Russell	1615.59	#35Sergey Sirotkin	1216.51

Table 3-Examples of Adjusted Drivers Ratings

In the 2019 season, the rating of Verstappen exceeded Hamilton's by 101 points, indicating comparable capabilities between these two drivers. However, when examining their earned scores, Hamilton surpasses Verstappen by 48%, suggesting Hamilton's dominance. Interestingly, during the 2021 season, Verstappen triumphed in the final Grand Prix when both their cars exhibited similar ratings. While traditional scoring could not anticipate this result, the enhanced Elo method provided accurate predictions.

Moreover, the 2019 season ratings for Russell in the improved Elo method stood at 1615, compared to a significantly lower 997 in the traditional Elo rating. By the 2022 season, Russell had transitioned to Mercedes and outscored Hamilton, a feat underestimated by the traditional Elo system but appropriately recognized by our improved method.

However, the adjusted rating system appears biased towards aggressive drivers, as it does not factor in the implications of driver dropouts. Aggressive drivers often secure higher positions, despite an increased likelihood of dropping out, leading to inflated final ratings. To counteract this, one solution is to employ a smaller K-factor, an element that either amplifies or dampens the impact of a win or loss on a player's Elo rating, for adjusting Elo ratings when drivers drop out due to their own mistakes.

This methodology can be employed to predict race outcomes, aiding teams in the development and fine-tuning of their cars. Given the driver ratings and a formula to calculate race-winning probabilities, teams can determine the necessary car speed for a driver to claim the championship. For example, if Alpine desire Ocon to stand a 50% chance against Vettel, Ocon's car should be rated 0.64 points lower than Vettel's. More explicitly, to secure a win, the car should be 0.52 seconds faster per lap on the Monza racetrack. This grants teams a definitive benchmark for resource allocation.

The recent introduction of the Budget Cap in F1, which mandates that all teams maintain expenses below 130 million dollars annually, has significantly constrained resources for top-tier teams such as Ferrari and Red Bull. Consequently, optimal resource allocation is critical for a team's success. Balancing the immediate need for car adjustments with long-term car development becomes paramount, and the results of this study could help guide teams towards making optimal decisions.

IX. Conclusion

This article presents a novel approach to separating and quantifying the effect of a car’s performance and a driver’s skill on Formula 1(F1) race outcomes. By analyzing data from the past decade, we propose a formula that measures F1 drivers’ ability. This approach could be used to predict race outcomes for a given driver in cars with different performance levels, aiding teams in optimizing resource allocation for car development.

References

[1] Glickman, M. E., & Jones, A. C. (1999). *Rating the chess rating system*. *CHANCE-BERLIN THEN NEW YORK-*, 12, 21-28.

[2] Conroy, E., Kowal, M., Toth, A. J., & Campbell, M. J. (2021). *Boosting: Rank and skill deception in esports*. Entertainment Computing, 36, 100393.

[3] Raghav Mittal. What is an ELO Rating?. Sep 11, 2020. https://medium.com/purple-theory/what-is-elo-rating-c4eb7a9061e0

[4] Eichenberger, R., & Stadelmann, D. (2009, December). *Who Is The Best Formula 1 Driver? An Economic Approach to Evaluating Talent*. Economic Analysis and Policy.

[5]Bell, A., Smith, J., Sabel, C. & Jones, K. (2016). *Formula for success: Multilevel modelling of Formula One Driver and Constructor performance*, 1950–2014. Journal of Quantitative Analysis in Sports, 12(2), 99-112.