Received: 23 August 2022 | Revised: 2 October 2022 | Accepted: 17 December 2022

DOI: 10.1002/asmb.2745

RESEARCH ARTICLE

WILEY

On the probability of Magnus Carlsen reaching 2900

Sohan Bendre[1] | **Shiva Maharaj**[1] | **Nick Polson**[2] | **Vadim Sokolov**[3]

[1]Chess Ed, Chicago, Illinois, USA

[2]Booth School of Business, University of Chicago, Chicago, Illinois, USA

[3]Department of Systems Engineering and Operations Research, George Mason University, Fairfax, Virginia, USA

Correspondence
Nick Polson, Booth School of Business, University of Chicago, Chicago, IL, USA.
Email: ngp@chicagobooth.edu
Vadim Sokolov, Department of Systems Engineering and Operations Research, George Mason University, Fairfax, VA, USA.
Email: vsokolov@gmu.edu

Abstract

How likely is it that Magnus Carlsen will achieve his goal of a 2900 Elo rating? At what level of play does Magnus have a reasonable chance of reaching the 2900 goal? These two questions are of great current interest to Magnus and the chess community. The probabilistic properties of Elo's rating system are well known, and together with a Brownian motion model of rating evolution, we use simulation-based methods to address these questions. Our model assesses that Magnus has a 4.5% chance of reaching 2900 if he continues his 2020–2022 level of play. However, this increases dramatically to 80% chance if he can repeat his hot streak performance of 2019 which is not an easy undertaking. The probabilities are intimately related to Elo's choice K-factor used for grandmaster chess play. Finally, we conclude with a discussion of the policy issues involved with the choice of K-factor.

KEYWORDS

Bayes, chess, Elo, grandmaster, K-factor, Magnus Carlsen

1 | INTRODUCTION

Chess grandmaster GM Magnus Carlsen recently announced his newest goal to reach a 2900 FIDE rating. Carlsen (2022, personal communication) described his challenge in the following manner: what level of play or *"hot streak"* would he have to play at to have a reasonable chance of achieving 2900? Moreover, what is the probability of achieving 2900 if he continues to play at his current level? To address the first question, we analyze Magnus' hot streak at the beginning of 2019 where his rating jumped from 2835 to 2881 in a matter of months to analyze how well he would have to play to achieve his goal of 2900 with a reasonable success. He intuitively realized the need to repeat a hot streak and he was very curious as to how probabilistic methods could help him understand how his Elo[1] rating would change. Our article uses a probabilistic model and simulation methods to address these questions. Specifically, we find that Magnus has a 4.5% chance of reaching 2900 if he continues his 2020–2022 level of play. However, this increases to 80% chance if he can repeat his hot streak performance of 2019 matching his intuition.

Many argue that the current implementation of Elo's system is the reason that Carlsen has such a difficult time increasing his rating. Many believe that the problem lies in the value of the K-factor. For grandmaster (GM) play, $K = 10$, the lowest possible K-value for a FIDE rated chess player. The K-factor governs the variability in rating changes. Carlsen (2022, personal communication) suggested that a $K = 15$ would give him a reasonable chance of achieving his goal. Moreover, he suggested that he thought he was currently fairly rated by the system and that a $K = 20$ would be too large for GM play.

As of August 2022, Carlsen's official rating is 2861, far above second ranked Ding Liren with a rating of 2806. At first sight, it might seem that Carlsen's goal is very achievable to the average chess player, whose rating might change by 50 points or more in a single tournament. But for Carlsen it has proven to be quite difficult, as his rating has not fluctuated by more than 20 points in the past 2 years. In reality, it seems virtually impossible that Carlsen can ever reach 2900, due to the constraints of the current chess rating system. Specifically, the choice of K-factor used in Elo's rating system for chess GM play is only $K = 10$, which governs how many points he can gain with a win. Intuitively, as he is already the best player in the world (by a wide margin) if he draws (which is common in GM play) he will lose points and, of course, any loss will lead to a large setback. For example, after an incredible performance at the Tata Steel tournament (9.5/13) he only gained 3 points only then to lose this in a game against a much lower-rated player.

1.1 | Elo rating system in chess

The basic Elo probability model[1] for assessing the probability of A's play against B is

$$P(A) = W(R_A - R_B),$$

where $P(A)$ is the probability of A wining and R_A is the strength of A.

The function W maps the difference in scores into $(0, 1)$ interval and is symmetric, monotonically increasing. Further, it approaches one as difference in ratings grows. The Elo rating system is used by the World Chess Federation (FIDE) and numerous other chess organizations to determine how many points need to be assigned to each player after a game. The Elo rating system is a statistical model for estimating the expected score of a player after a game. Elo's system was adopted by FIDE in 1970 and has been in use ever since. A few other scoring systems have been recently proposed, including the TrueSkill, created by Microsoft Research Dangauthier et al.,[2] the Glicko[3] rating system created by the statistician Mark Glickman and the Chessmetrics rating system created by Jeff Sonas. Elo's system continues to be the prevalent chess rating system given its large success.

In Elo's system the difference of G points is equivalent the odds ratio of $K : 1$. The K-factor used by FIDE is equal to 10, and the logistic growth rate G is 400. A difference of 400 means higher scored player is likely to win 10 out 11 games. Hence, the FIDE Elo formula for calculating the odds of a game between A and B with corresponding ratings of R_A and R_B is as follows

$$\text{Odds}(A \text{ beats } B) = \frac{P(A)}{P(B)} = \frac{P(A)}{1 - P(A)} = K^{\frac{R_A - R_B}{400}},$$

where $P(A)$ is the probability that player re-expressing this gives

$$P(A) = \frac{1}{1 + 10^{-\frac{R_A - R_B}{400}}}.$$

After a game, the Elo rating of each player is updated to adjust for the observed outcome of win (1), loss (0), or draw (0.5).

The Elo rating of the player A is updated as follows

$$R_A^+ = R_A + K(S_A - P(A)) \quad \text{where} \quad S_A \in \{0, 1, 1/2\},$$

is the result of the game. At the end of a tournament, adjustments for all the games will be applied to update the post tournament rating. Our approach builds on the Elo ratings analysis of Aldous.[4]

Elo's system is considered to be efficient and become the standard in many games of play. For example, it was shown that when it was applied to outcomes of tennis matches, it performed poorly due to the inefficiency how points get assigned to ATP (men) and WTA (women) players Williams et al.[5] However, it is still a topic of debate whether this system can be improved. Our view is that a discussion of the K-factor and the implication for player's ratings is the more pressing question.

Figure 1 shows the distribution of Elo ratings of players registered with FIDE. It shows how hard it is to become a grand master (Elo greater than 2500) or super GM (Elo greater than 2700).

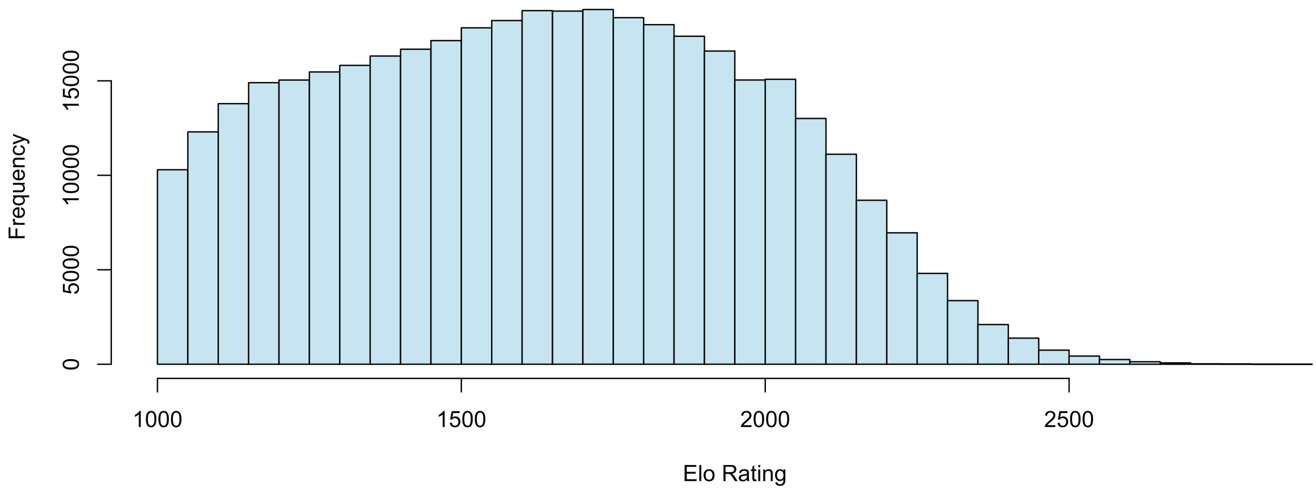


FIGURE 1 Histogram of Elo ratings of player registered with FIDE as of August 2022

Other key features of the Elo rating system are

1. Dynamic ratings that are updated after each game or tournament.
2. Implicitly weight recent games higher and apply exponential smoothing.
3. The updating model is straightforward to understand and implement.

However, there is still a question of how accurate the rating is and does it truly reflect a current ability of a player. Aldous[4] provides several asymptotic theoretical results. He shows that, the typical error of predicted win probabilities will not be substantially less than 10%, regardless of number of matches played. To a large extent, as Elo's model makes clear, the larger values of K will allow for greater fluctuation in rating. Due to this, it is quite difficult for Carlsen's rating to increase by a large amount, but it also does ensure that his rating won't immensely decrease either. If the K factor was increased to allow greater rating changes, theoretically if Carlsen won all his games he may be able to reach the 2900 threshold. However, drawing or losing a game would have a negative impact on his rating, because he is so much more highly rated than all other players. For example, with the current K factor, if Carlsen drew against the second-ranked player in the world, Ding Liren (rating 2806), his rating would decrease by about 0.83 points. If instead he lost to Ding Liren, his rating would drop by a dramatic 5.83 points. However, increasing K to 20 (too large in Carlsen's view), a draw would cause his rating to drop about 1.65 points and a loss would mean a drop of about 11.65 points. Considering that draws at the GM level are occurring 70%–75% of the time nowadays, the situation seems tough for Carlsen any way you look at it due to the dramatic increase in competition. Reaching such a goal would be one of the greatest sporting achievements.

The rest of the article is outlined as follows. Section 2 provides a simple Brownian motion model for assessing the two fundamental questions of how likely is it that Magnus Carlsen will reach 2900, and at what level of play would he have to achieve to have a reasonable chance of achieving 2900? To do this we use the Brownian motion model for sports scores as originally developed by Stern[6] and extended by a number of authors, Feng et al.[7] for EPL, Polson and Stern[8] for implied volatility of a sports game. Then we provide an empirical analysis of Magnus Carlsen's games and changes in Elo ratings. In particular, we analyze Magnus' hot streak at the beginning of 2019. We use this to analyze how well he would have to play to achieve his goal of 2900 with reasonable success. Finally, Section 3 concludes with a discussion of the choice of K-factors in chess.

2 | THE PROBABILITY THAT MAGNUS CARLSEN WILL MAKE 2900

In order to address the question of how likely is it for Magnus to make 2900 we build a probability model and then use simulation. Stern[9] provides a model for addressing the probabilities of winning a sports contest. Feng et al.[7] extend this model to discrete outcomes using a Skellam process for EPL football matches. This is more appropriate given that chess has a similar outcome distribution profile with (*win*, *draw*, *loss*). While this leads to more accurate assessments, it has the caveat of interpretability. Stern's Brownian motion model is easily interpretable and can be viewed as a limit of the underlying process which due to small changes in the Elo ratings will be an excellent approximation.

2.1 | Brownian motion model

Let X_t denote Magnus Carlsen's current rating at time t. Let X_0 denote the Elo rating at the beginning of the time period. We assume that X_0 is a fair assessment at the beginning of the period and the Elo system is in equilibrium. Let μ denote the instantaneous skill level over the period $(0, t)$. This might result in a change in performance due to extra effort. Stern[6] shows that changes in outcome approximately follow a Brownian motion model of the form

$$X_t = X_0 + \mu t + \sigma B_t,$$

where σ denotes the volatility of the Elo rating and B_t is a standard Brownian motion. In particular, $B_0 = 0$ and $B_t \sim N(0, t)$.

Let $X_{\text{Magnus Carlsen}}$ denote the random variable that measures the changes in Magnus Carlsen's ratings. Then we can estimate mean, $\hat{\mu}_X$ and standard deviation, $\hat{\sigma}_X$, via the formulas

$$\hat{\mu}_X = \sum_{i=1}^{3} \hat{e}_i \hat{p}_i \quad \text{and} \quad \hat{\sigma}_X^2 = \sum_{i=1}^{3} \hat{e}_i^2 \hat{p}_i - \hat{\mu}_X^2,$$

summed over the three possibilities of .(*win*, *draw*, *loss*). Put simply, we estimate the mean and standard deviation from empirical averages.

Following Stern, then we can calculate the desired probability—how likely is it for Magnus to reach his goal of 2900, starting at $X_0 = 2861$ with his current skill and volatility of his Elo rating from the last 2 years, namely

$$\mathbb{P}(X_T > 2900 | X_0 = 2861) = \mathbb{P}\left(Z > \frac{39 - \hat{\mu}_X T}{\hat{\sigma}_X \sqrt{T}} \right) = 1 - \Phi\left(\frac{39 - \hat{\mu}_X T}{\hat{\sigma}_X \sqrt{T}} \right),$$

where Φ denotes the normal cumulative distribution function.

What level of play is required to achieve 2900? Polson and Stern[10] define the implied volatility of a sports game and this allows us to use Elo's formula to infer an implied skill level (a.k.a. performance) to achieve a desired probability of success of achieving their goal. Let p denote the probability of achieving the goal. Let μ denote the required level of skill to achieve this goal.

From Stern's calculation, with $X_T - X_0 = 40$ where X_T is the desired goal

$$1 - p = \Phi\left(\frac{X_T - X_0 - \mu_X T}{\hat{\sigma}_X \sqrt{T}} \right). \tag{1}$$

This can be inverted to address the level of play, denoted by μ_{imp} for implied level of play, required to guarantee the given p.

$$\mu_{imp} = \frac{1}{T}(X_T - X_0) - \frac{1}{\sqrt{T}} \Phi^{-1}(1 - p)\, \hat{\sigma}_X.$$

This is an instantaneous level of play whose advantage compounds with T. Now we apply our methodology.

2.2 | Empirical analysis of Magnus Carlsen's games

An important feature of the Elo system is the exponential decay of increase in probabilities. In other words, the probability of win (and thus adjustment to the Elo rating) increases exponentially slow as the difference in ratings goes up. Implying that the gap between Carlsen and the number two is large, it is very hard for Carlsen to improve the rating. To show this fact empirically, we plot the histogram of Carlsen's rating changes (per game) over the last 2 years 2020–2022 with a total of 110 classical games.

Figure 2 plots his rating changes for the whole period and clearly shows the left skew. The recent sample of 110 classical games is used to estimate our model parameters.

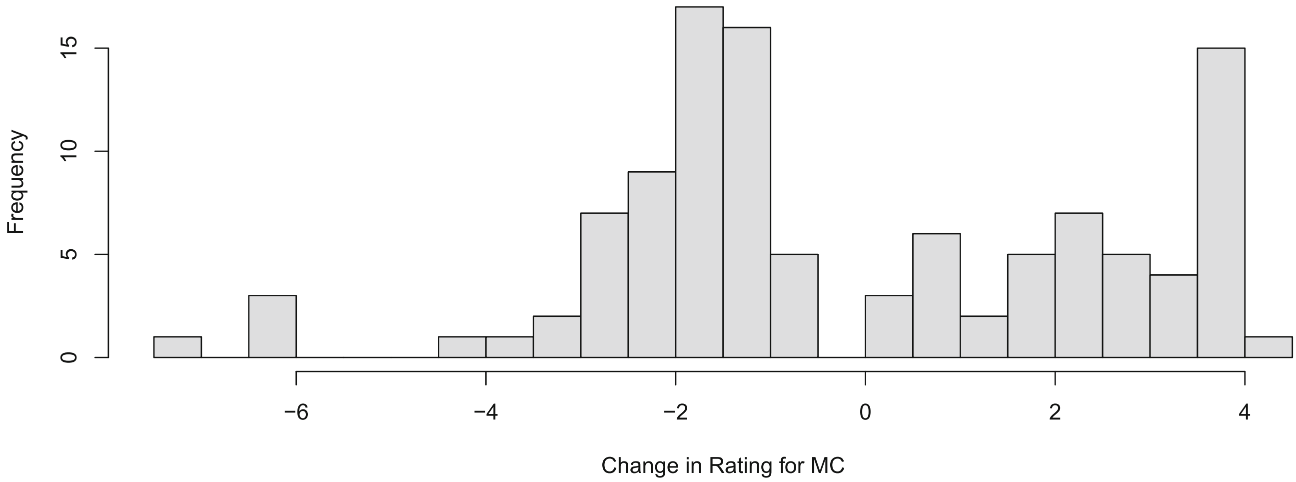


FIGURE 2 Histogram of Magnus Carlsen's change in rating in 2020–2022 period

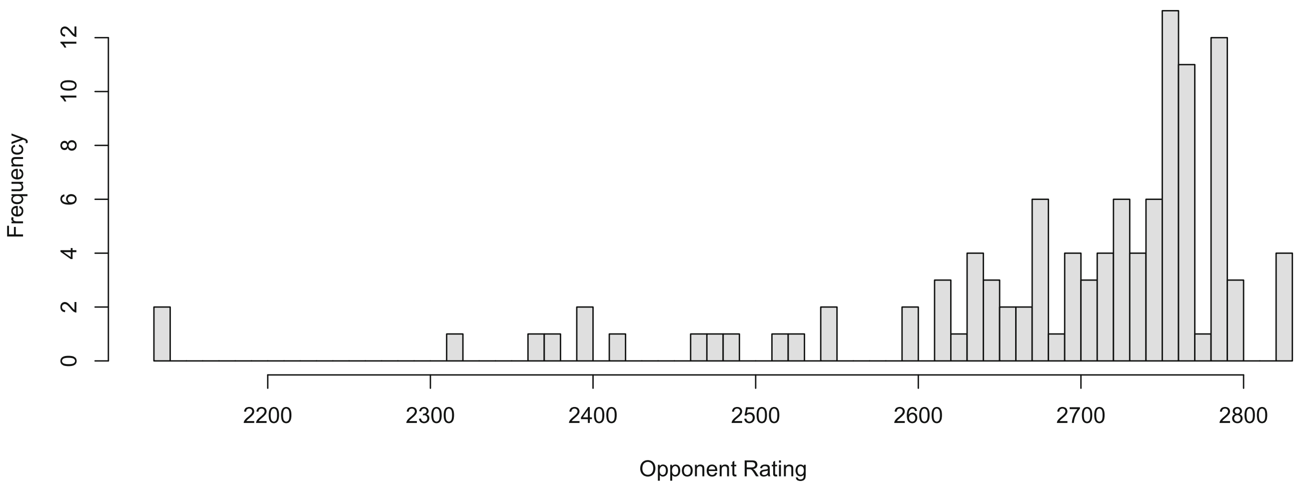


FIGURE 3 Opponent ratings for 2020–2022 period.

Notice that as Carlsen is the best player in the world, he loses Elo rating points when he draws even though this is the most likely outcome. This skewness in the distribution is what makes his goal of 2900 extremely hard.

2.3 Simulated paths

We simulate two paths for Magnus Carlsen. First simulation is based on his most recent performance during the 2020–2022 period. Our data set contains 110 classical games from this period, starting from Tata Steel Masters 2020 and ending with World Chess Olympiad 2022.

Figure 3 shows the histogram of the ratings of Magnus Carlsen's opponents during this period.

Figure 4 compares the predicted outcome by the Elo system (blue dots) and the actual outcome (black dots).

Table 1 summarizes the data set the observed changes in rating after each game. We have a total of 110 games in this period with average change in rating being −0.11 with a standard deviation of 2.67.

We simulate 2000 times the 200 future games by Magnus Carlsen. First, we assume that the performance (likely outcome) and the mix of opponents is the same as in 2020–2022 period.

In this simulation. Magnus reached the rating in 90 simulated trajectories out of 2000. If Magnus Carlsen continues showing the same performance as he did during the 2020–2022 period, he has 4.5% chance of reaching 2900 (Figures 5 and 6).

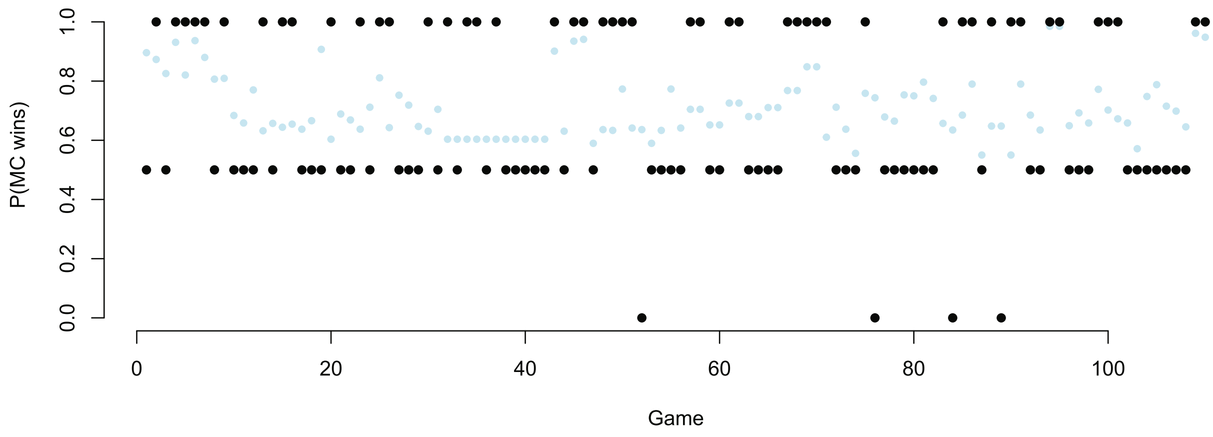


FIGURE 4 Game outcomes for 2020–2022 period.

TABLE 1 Summary statistic for change in rating per game during the 2020–2022 period.

N	μ	σ	Min	Max
110	−0.11	2.67	−7.44	4.5

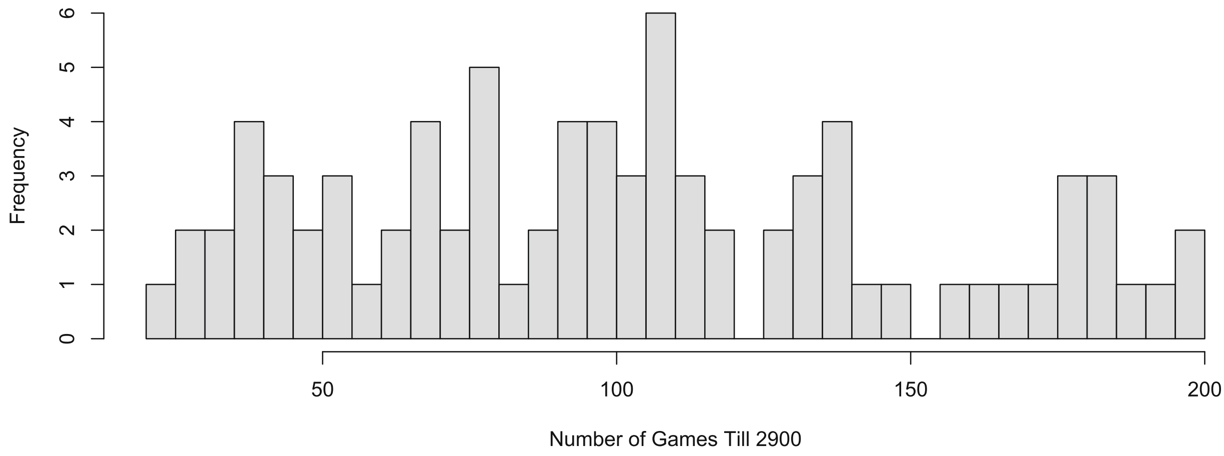


FIGURE 5 Histogram showing when the rating of 2900 is first reached over the simulated 1000 games.

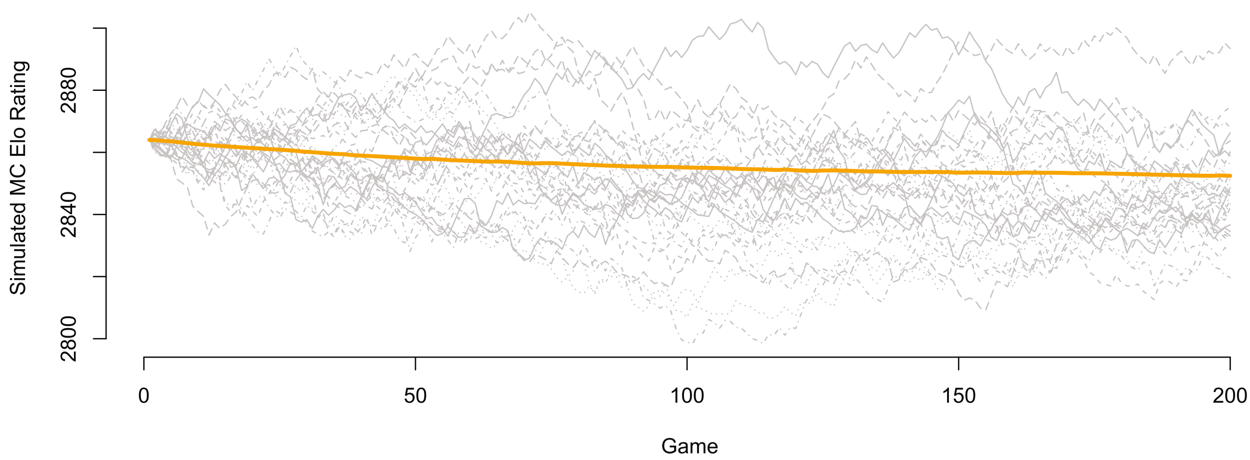


FIGURE 6 Several simulated trajectories over the next 200 games. The solid line shows the average across 2000 simulations for this specific game

2.4 | Hot streak of 2019

Now we perform same simulations using the data from 2019's Magnus' hot streak period. During this period the rating went from 2835 to 2872 and peaked in August at 2882.

Figure 7 plots his rating changes for each of the months in 2019, the hot streak period. Table 2 provides the same set of statistics, but for his hot streak in 2019. Notice the difference in the estimates. His extremely high win percentage (hot streak) allowed him to gain over 50 points in a short period of time.

Table 2 summarizes the observed changes in rating after each game. We have a total of 78 games in this period with average change in rating being 0.48 with standard deviation of 2.42.

Figure 8 shows the histogram of the ratings of Magnus Carlsen's opponents during this period.

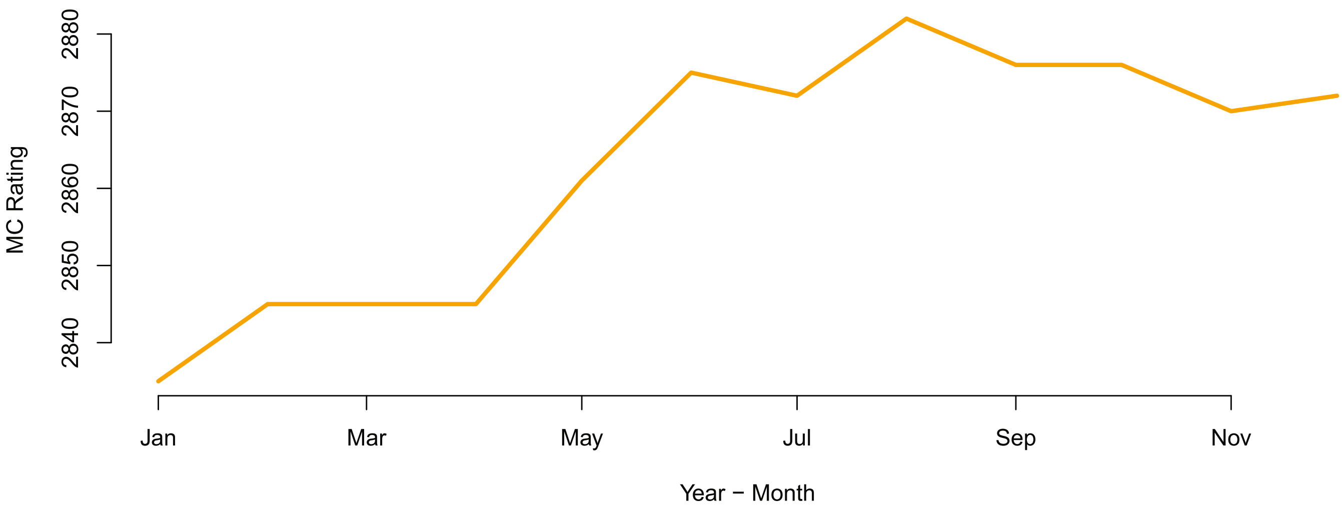


FIGURE 7 Elo rating for every month of 2019.

TABLE 2 Summary statistic for change in rating per game during the 2019 period.

N	μ	σ	Min	Max
78	0.48	2.42	−2.8	4.7

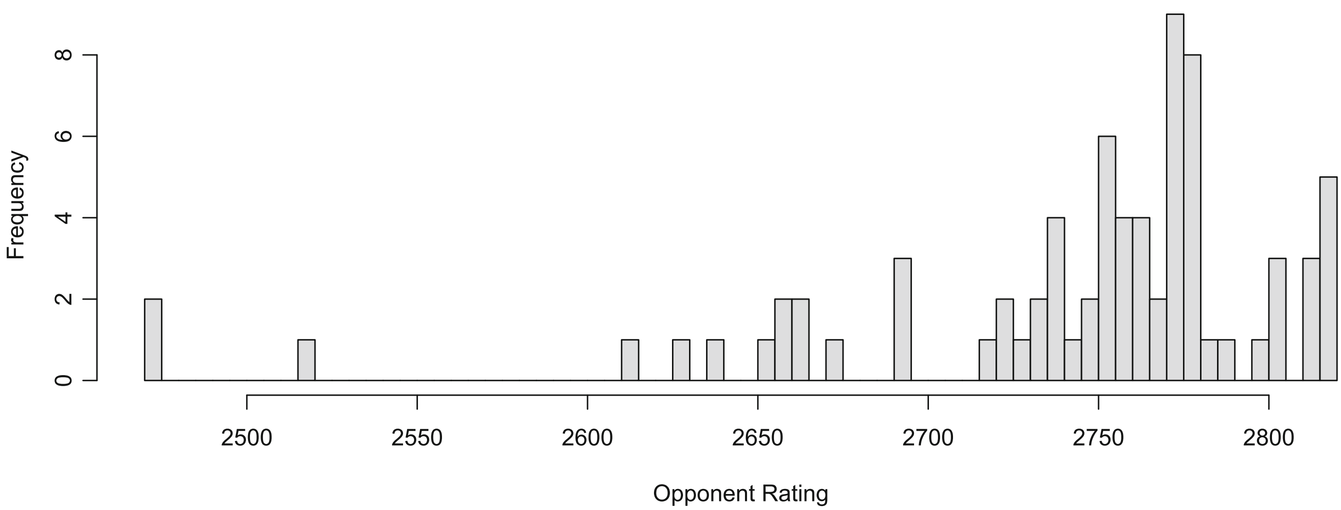


FIGURE 8 Opponent ratings

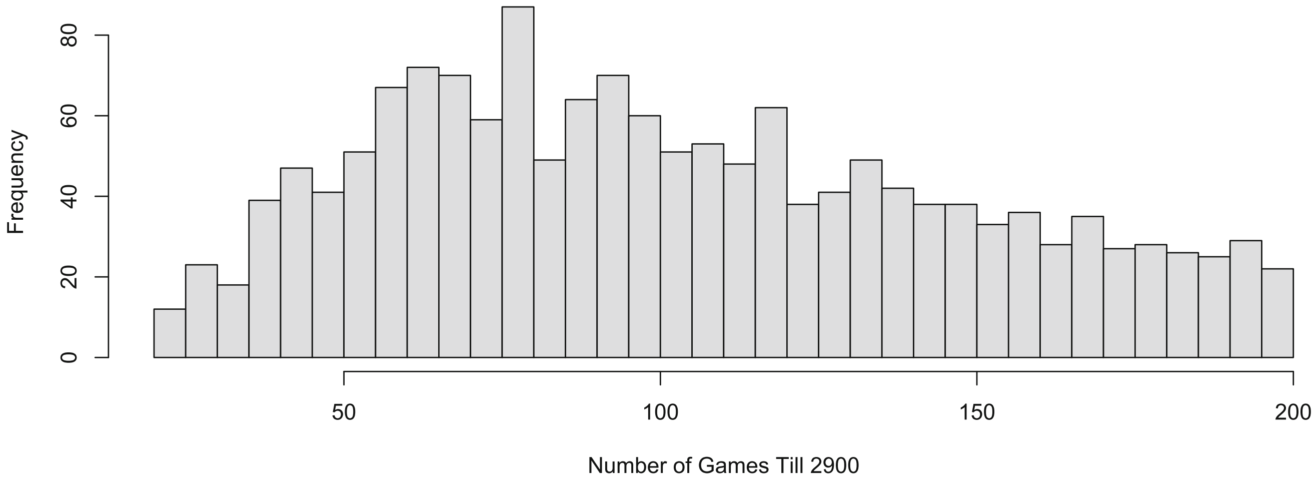


FIGURE 9 Histogram of how many games till the rating of 2900 is first reached over the simulated 1000 games.

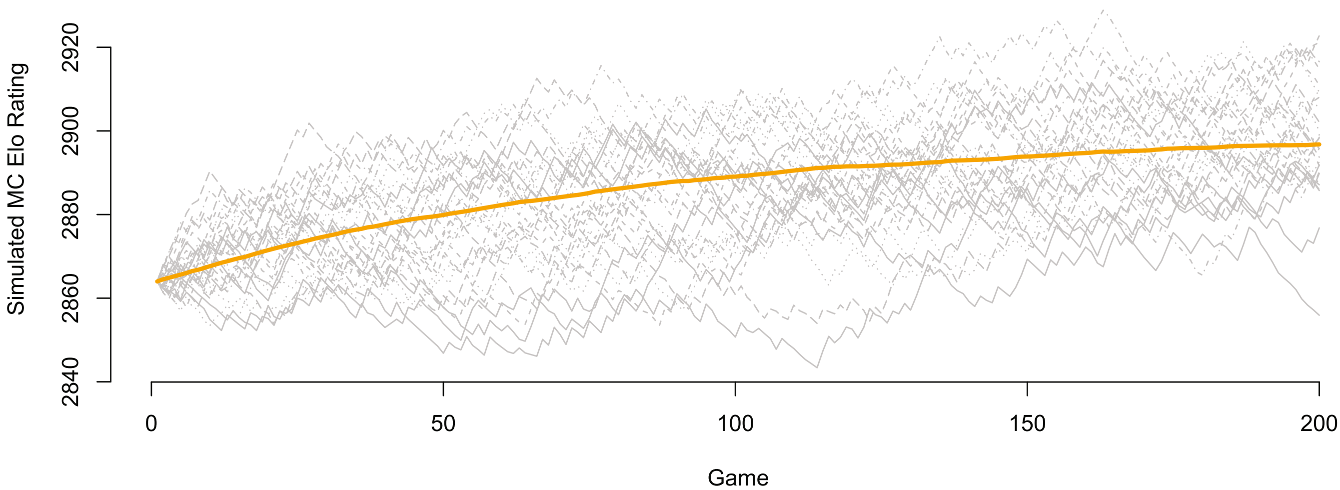


FIGURE 10 Several simulated trajectories over the next 200 games. The solid line shows the average across 2000 simulations for this specific game

Now we simulate 2000 times the 200 future games by Magnus Carlsen using the 2019 data. Again, we assume that the performance (likely outcome) and the mix of opponents is the same as in 2019 period.

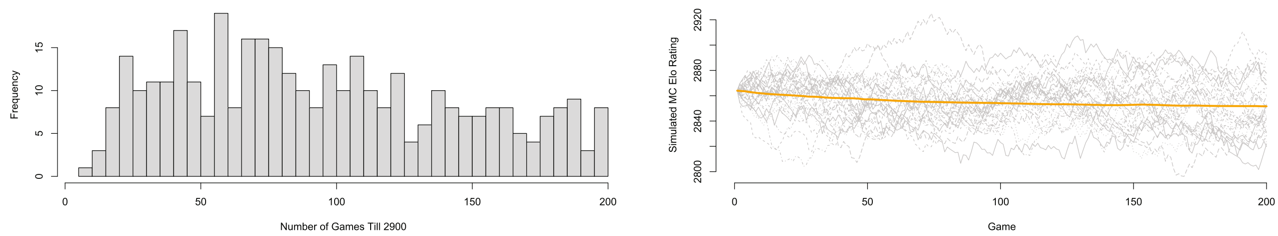
In this simulation, Magnus reached the rating in 1600 simulated trajectories out of 2000. If Magnus Carlsen continues showing the same performance as he did during the 2019 period, he has 80% chance of reaching 2900 (Figures 9 and 10).

2.5 | What happens if we increase the *K*-factor?

Increasing the *K*-factor, say to $K = 15$, essentially changes the volatility in our model. From Equation (1), we see that this increases the probability of a tail probability, namely the probability of getting to 2900. The other way of increasing your probability is to increase your skill level, μ, as described above. Here, we assume that the skill level does not change and the average number of wins is the same, but consider a scenario, when K-factor is 15, rather than 10, and using 2020–2022 data (Figure 11):

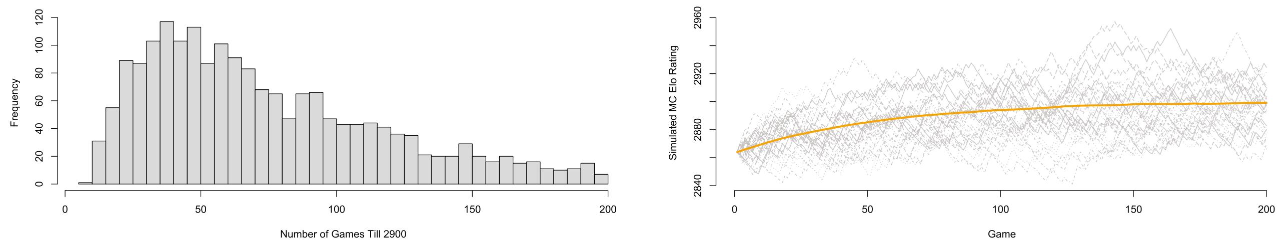
If Magnus Carlsen continues showing the same performance as he did during the 2019 period, and K-factor is 15, he has 18% chance of reaching 2900.

When the K-factor is 15, and we are using the 2019 data (Figure 12):



(A) (B)

FIGURE 11 Left panel: Histogram of when the rating of 2900 is first reached over the simulated 1000 games. Right Panel: Several simulated trajectories over the next 200 games. The solid line shows the average across 2000 simulations for this specific game. (A) How many games till 2900; (B) simulated trajectories



(A) (B)

FIGURE 12 Left panel: Histogram of when the rating of 2900 is first reached over the simulated 1000 games. Right Panel: Several simulated trajectories over the next 200 games. The solid line shows the average across 2000 simulations for this specific game. (A) How many games till 2900; (B) simulated trajectories

TABLE 3 Chances of reaching 2900 in the next 200 games under different future performance assumptions (2019 vs 2020–2022) and different K factors (10 vs 15)

	$K = 10$	$K = 15$
2020–2022 data	4.5%	18%
2019 data	80%	95%

If Magnus Carlsen continues showing the same performance as he did during the 2019 period, and K-factor is 15, he has 95% chance of reaching 2900.

Table 3 summarizes our findings of the simulation studies and shows the probability of reaching the rating of 2900 under two different assumptions about the skill level (2019 performance vs 2020–2022 performance) and the K-factor (10 vs 15).

3 | DISCUSSION

There is a large literature on the choice of K-factor in chess and in other sports[6–8] as the Elo rating system is used extensively to great effect. The problem of lack of volatility in a players' rating changes when $K = 10$ has been the center of much debate. As noted by a referee, the first question to address is the overall goal of rating chess players which clearly is not solely based on optimizing the chances of Carlsen making 2900. Aldous[4] talks about scaling the Elo rating and its relationship with the waiting time for the tournament design to reach some form of equilibrium. Elo in his seminal book describes the limiting distribution of the best player and the dynamics involved.

Glickman and Jones[3] also showed empirically that the quality of prediction differs for player of different rank. He used maximum likelihood estimator to identify the adjusting factor α

$$P(A) = \frac{1}{1 + 10^{-\alpha \frac{R_A - R_B}{400}}}.$$

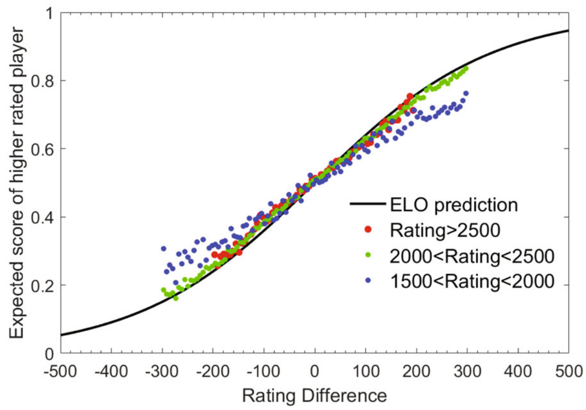


FIGURE 13 Comparing sample distribution of results for different rating groups. *Source*: Viswanath[11]

They found that for all groups the best estimate of $\alpha < 1$, meaning that the growth rate has to be more than 400. This suggests, that rating need to be updated "faster." This also can be achieved by increasing the K-value. Glickman and Jones[3] estimated that $\alpha = 0.59$ for player with the rating between 1400 and 1600, and it is 0.95 for players with rating between 2200 and 2700. Thus, the Elo performs most poorly on the middle-range players, that happen to be the largest group.

Furthermore, as shown in the figure below, the calibration of K needs to be done differently for different skill levels (Figure 13).

There are a number of areas for future research. The policy decision of changing the K-factor is also hotly debated. Nunn[12] asked for a proof of why it would be a good idea citing that the Elo system has worked well for 50 years. Our point is that it is not the Elo system per se which has worked very well however, a subjective choice of $K = 10$ has implications for how high the current best player can achieve his goals. Our research outlines a path to answer Nunn's question and to influence policy. Put simply, we have demonstrated the effect of changing K-factor on the probability of Carlsen reaching rating of 2900 in the next 200 games. Specifically, we have shown that it goes from 4.5% under the current system to 18% when is K-factor is 15. This provides a probabilistic benchmark for wider discussion.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in ChessGames at https://www.chessgames.com.

ORCID

Vadim Sokolov https://orcid.org/0000-0002-6618-2965

REFERENCES

1. Elo AE. *The Rating of Chessplayers, Past and Present*. Arco Pub; 1978.
2. Dangauthier P, Herbrich R, Minka T, Graepel T. TrueSkill through time: revisiting the history of chess. *Advances in Neural Information Processing Systems*. Vol 20. Curran Associates, Inc; 2007.
3. Glickman ME, Jones AC. Rating the chess rating system. *Chance*. 1999;12:21-28.
4. Aldous D. Elo ratings and the sports model: a neglected topic in applied probability? *Stat Sci*. 2017;32(4):616-629.
5. Williams LV, Liu C, Dixon L, Gerrard H. How well do elo-based ratings predict professional tennis matches? *J Quant Anal Sports*. 2021;17(2):91-105.
6. Stern H. On the probability of winning a football game. *Am Stat*. 1991;45(3):179-183.
7. Feng G, Polson NG, Xu J. The market for English premier league (EPL) odds. *J Quant Anal Sports*. 2016;12(4):167-178.
8. Polson N, Stern H. The implied volatility of a sports game. *J Quant Anal Sports*. 2015;11:145-153.
9. Stern HS. A brownian motion model for the progress of sports scores. *J Am Stat Assoc*. 1994;427:1128-1134.
10. Polson NG, Stern HS. The implied volatility of a sports game. *J Quant Anal Sports*. 2015;11:145-153.
11. Viswanath G. Elo rating system: how underrated are the kids? 2016.
12. Nunn J. On the K-factor: show me the proof! April 2009. https://en.chessbase.com/post/nunn-on-the-k-factor-show-me-the-proof.

How to cite this article: Bendre S, Maharaj S, Polson N, Sokolov V. On the probability of Magnus Carlsen reaching 2900. *Appl Stochastic Models Bus Ind*. 2023;39(3):372-381. doi: 10.1002/asmb.2745