Spring 5-10-2014

Ranking World Class Chess Players Using Only Results From Head-To-Head Games

Sterling Swygert

University of South Carolina - Columbia

Director of Thesis: Edsel Pena

Abstract

The purpose of this paper is to determine whether the world's current top ten chess grandmasters can be ranked appropriately based only on their head-to-head matches. The solution was based on the likelihood technique in mathematical statistics and was derived by assuming that the outcomes of chess games follow a logistic function with meaningful parameters. After defining the likelihood function specific to the problem, the likelihood function is maximized using numerical methods contained in the statistical package R. The ranking approach was then applied to simulated results following the logistic model to examine its applicability. Finally, the ranking approach was applied to the data of head-to-head matches among the world's current top ten chess grandmasters.

Contents

1 Thesis Summary

This honors thesis explores a method of ranking the world's top ten chess grandmasters using only the outcomes of games containing only players in that very set. This method allows for players in a single era to be quickly ranked via algorithmic and numerical means, including very specific information, from a statistical standpoint. Furthermore, unlike the rating systems that are commonly used, the Elo and the Glicko systems, this method is Classicist in its statistical approach, rather than Bayesian. Finally, this ranking method also differs from others as it limits the information to games between the individuals being ranked.

Some of the main topics utilized in this thesis are mathematical statistics through the use of the likelihood approach, statitical inference in the use of the logit distibution, and algorithmic design through the formulation of the solution to our given problem. The statistical package "R" was used in order to code all of the programs used during this research endeavor. This area of research used all of these tools in order to find an adequate mechanism for ranking a given set of players in a game with a win, draw, and loss as possible outcomes.

Many of the techniques used in this research are extremely valuable in terms of further research. For instance, the likelihood approach that is used is a common statistical method that is often utilized in order to estimate parameters of a given model. Furthermore, algorithmic design and mathematical modelling are commonplace in statistical research and are used to statistically infer estimates and to verify the validity of models.

Initially assuming a logistical model as the distribution governing the outcomes of players games, an algorithm was constructed in order to rank a set of players based on outcomes of games. This function was then tested in single-simulation tests that followed the logistical model. Finally, after the algorithm was tested, it was applied to the outcomes of games between the world's top ten chess grandmasters in order to rank-order them.

2 Introduction

The purpose of this paper is to examine an alternative to the current paradigm of chess player ranking systems. Today, chess players are often rated using the Elo rating system. Because of some of its own defficiencies, Professor Mark Glickman developed his own rating system, the Glicko system [**1**]. Each of these rating systems uses complex formulae whose origin can be difficult to understand. They also rely on the accuracy and precision of past ratings produced by the same formulae. It is the goal of this paper to explore a different method that uses probability theory, statistical inference, and is computational in nature.

To provide perspective to this research, consider the set of the world's current top ten chess players, ranked based on their Elo ratings (see www.365Chess.com). These ten grandmasters were labeled 1 through 10, and their names, player numbers, current world ranks based on their Elo ratings, and Elo ratings are displayed in Table 1.

Table 1: Name, country, player number, world rank, and Elo rating of the current top 10 internationally ranked chess players. Source: Data obtained from www.365Chess.com

Name	Country	Player No.	World Rank	Elo Rating
GM Magnus Carlsen	Norway	1	1	2881
GM Viswanathan Anand	India	2	3	2785
GM Levon Aronian	Armenia	3	2	2812
GM Hikaru Nakamura	USA	4	9	2772
GM Vladimir Kramnik	Russia	5	4	2783
GM Fabiano Caruana	Italy	6	5	2783
GM Alexander Grischuk	Russia	7	6	2777
GM Sergey Karjakin	Ukraine	8	7	2772
GM Veselin Topalov	Bulgaria	9	8	2772
GM Shakhriyar Mamedyarov	Azerbaijan	10	10	2760

These ten grandmasters are representatives of their respective countries, and

Figure 1: From left to right, top: GM Carlsen, GM Anand, GM Aronian, GM Nakamura, GM Kramnik, bottom: GM Caruana, GM Grischuk, GM Karjakin, GM Topalov, GM Mamdeyarov

some are considered to be superstars in their countries with many having a multitude of followers. This is because of the aesthetic traits of the game of chess, and because its mastery demonstrates a higher level of intelligence, which some countries utilize as a way of demonstrating the superiority of their culture and political system. Figure 1 contains snapshots of these grandmasters. Below are some characteristics and tidbits regarding these grandmasters.

Grandmaster Magnus Carlsen of Norway (age of 23), the current World Champion, is the highest rated chess player in the world with an Elo rating of 2881. This is the highest rating ever achieved by any chess player in the history of the Elo rating scheme. Known as a "grinder," a player who is willing to play more evenly without necessarily aiming to gain an advantage in the early parts of games in favor of drawing games out. Carlsen is often finds success in fatiguing his opponents, leading to them making mistakes in the later parts of games [**2**].

Grandmaster Levon Aronian (31), from Armenia, is the second rated chess player. His success and greatness are sometimes overlooked, even though he was only the sixth person to break the 2800 barrier with an Elo of 2812. His personality and demeanor tend to be calm and easy going, but his playstyle is aggressive and dynamic, leading him to be one of the most dominant players during the middle portions of games [**3**].

Rated third internationally, Grandmaster Viswanathan Anand (44) of India, a former World Champion, has an Elo rating of 2785. Being older than most of the other top-rated players, he has been considered to be on the decline in his chess-playing career [**4**]. Like Aronian, he also tends to fly under the radar because of his more quiet personality. Though he began as a fast and tactical player, he has matured and settled into a chess style that demonstrates a broad array of tactics [**4**].

Grandmaster Vladimir Kramnik (38) of Russia with an Elo rating of 2783, is the fourth ranked chess player in the world. Being heavily influenced by both former World Champions Grandmaster Anatoly Karpov and Grandmaster Garry Kasparov, two of Russia's finest champions who had extremely different playing styles, Kramnik followed the philosophy that he must study a broad range of playing styles if he were to become a well-rounded player, then find the style that fits him best, and ultimately, become a chess master. This has led him to great success as he is also a former World Champion [**5**].

Grandmaster Fabiano Caruana (21) of Italy, is rated fifth with an Elo Rating of 2783, fractions of a point behind Kramnik. He plays a polar opposite style when compared to Carlsen. Classically trained, Caruana is one of the most precise players in regards to his openings. Ambitious and single-minded in his approach, Caruana attempts to use his openings to lead him to tightly defined endgames. Though he has been noted as having a high level of endurance, as he plays more tournaments than any of the other top 10 chess grandmasters, he has been considered to make errors when it comes to the later part of particularly important games [**6**].

Grandmaster Alexander Grischuk (30), another Russian player, is rated sixth globally with an Elo rating of 2777. Being an aggressive player with an attacking style, he is often prone to being exposed in games against top competitors. However, his style tends to lead to success when his opponents have little time to prepare for his aggression. Despite his exciting play, he is sometimes viewed as a “neglected genius” in the world of chess [**7**].

Grandmaster Sergey Karjakin (24), of Ukraine and Russia, has an Elo rating of 2772, which rates him as the seventh best chess player in the world. He is a phenomenally rapid chess and blitz chess player, so naturally his play tends toward aggression and an attacking style. He has often been considered to be one of Carlsen’s chief rivals as they are each strong players and of comparable age and experience [**8**].

Grandmaster Veselin Topalov (39) of Bulgaria, is one of the most controversial chess players of all time. He is the eighth rated chess player with an Elo rating of 2772. His resurgence during his thirties caused many to question his means of winning, sparking cheating allegations that have haunted him since [**9**].

Grandmaster Hikaru Nakamura (24) of the United States is rated ninth overall with an Elo rating of 2772. Another contemporary of Carlsen, he plays with a highly aggressive style that he employs in order to intimidate opponents into making early blunders. Being a player who hates drawing almost as much as he hates losing, he attempts to gain any small advantage he can in order to eventually exploit it into a win [**10**].

Grandmaster Shakhriyar Mamedyarov (29) of Azerbaijan is the tenth rated player with an Elo Rating of 2760. He is known for his aggressive playing style and his unusual openings [**11**].

Our goal is to rank these 10 players based **only** on the outcomes of the games that these players played against each other over their respective careers. Using the regularly updated data base www.365Chess.com, we gathered information regarding their head-to-head matches throughout their careers. This information regarding their head-to-head matches are contained in Tables 2, 3, 4, and 5. Table 6 provides a summary of the percentages of wins, draws, and losses of each of the 10 players in the games that were played against each of the other nine players.

Table 2: Number of games played against each other for the 10 grandmasters.

Player	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	0	61	47	40	37	18	23	30	22	17
2	61	0	43	18	154	15	20	19	87	16
3	47	43	0	31	47	13	43	38	20	26
4	40	18	31	0	27	20	19	23	9	13
5	37	154	47	27	0	14	31	24	69	23
6	18	15	13	20	14	0	7	16	6	7
7	23	20	43	19	31	7	0	43	14	26
8	30	19	38	23	24	16	43	0	25	33
9	22	87	20	9	69	6	14	25	0	11
10	17	16	26	13	23	7	26	33	11	0

Table 3: Number of wins of column player versus row player.

Player	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	0	14	8	14	10	9	9	13	11	6
2	10	0	8	0	24	2	5	5	27	8
3	8	10	0	11	10	6	8	10	4	7
4	5	6	8	0	9	9	6	5	3	7
5	8	17	4	6	0	4	6	4	23	7
6	4	5	3	2	4	0	1	3	2	1
7	5	3	8	6	8	1	0	9	2	11
8	5	0	10	4	9	3	13	0	7	10
9	1	11	1	2	10	1	4	9	0	2
10	2	1	7	5	7	2	5	8	3	0

Thus, the basic questions that we would like to address are: Based on these head-to-head matches, how should these 10 players be ranked among themselves? Should they be ranked based simply on their winning percentages? Or, their losing percentages? See, for instance, Table 6. The game of chess is somewhat unique among games (though in the game of soccer [international football], draws occur quite often as well, though now tie-breakers have been included in order to attempt to determine

Table 4: Number of draws of column player versus row Ppayer.

Player	1	2	3	4	5	6	7	8	9	10
1	0	37	31	21	19	5	9	12	10	9
2	37	0	25	12	113	8	12	14	49	7
3	31	25	0	12	23	4	27	18	15	12
4	21	12	12	0	12	9	7	14	4	1
5	19	113	23	12	0	6	17	11	36	9
6	5	8	4	9	6	0	5	10	3	4
7	9	12	27	7	17	5	0	21	8	10
8	12	14	18	14	11	10	21	0	9	15
9	10	49	15	4	36	3	8	9	0	6
10	9	7	12	1	9	4	10	15	6	0

Table 5: Number of losses of column player versus row player.

Player	1	2	3	4	5	6	7	8	9	10
1	0	10	8	5	8	4	5	5	1	2
2	14	0	10	6	17	5	3	0	11	1
3	8	8	0	8	14	3	8	10	1	7
4	14	0	11	0	6	2	6	4	2	5
5	10	24	10	9	0	4	8	9	10	7
6	9	2	6	9	4	0	1	3	1	2
7	9	5	8	6	6	1	0	13	4	5
8	13	5	10	5	4	3	9	0	9	8
9	11	27	4	3	23	2	2	7	0	3
10	6	8	7	7	7	1	11	10	2	0

a winner and a loser) in the sense that the outcome of a game could end in a draw, aside from either a win or a loss. At the highest level of the game, a drawn outcome actually occurs more often than wins or losses because of the almost equal parity of the abilities of the very top players. Furthermore, Chess is a highly sophisticated strategy game that their are many facets to its mastery. Thus, Player A could be dominant against Player B, Player B could be dominant against Player C, but Player C could be dominant against Player A. That is, a non-transitive type of relationship could occur among players, thus complicating the ranking problem when there are

Table 6: Percentages of the players wins, draws, and losses among themselves. NGames is the number of games played, NWins, NDraws, and NLosses are, respectively, the number of wins, draws, and losses. The last three columns converts these wins, draws, and losses into percentages.

	Player	NGames	NWins	NDraws	NLosses	PercWins	PercDraws	PercLosses
1	1	295	94	153	48	31.86441	51.86441	16.27119
2	2	433	89	277	67	20.55427	63.97229	15.47344
3	3	308	74	167	67	24.02597	54.22078	21.75325
4	4	200	58	92	50	29.00000	46.00000	25.00000
5	5	426	89	246	91	20.89202	57.74648	21.36150
6	6	116	25	54	37	21.55172	46.55172	31.89655
7	7	226	53	116	57	23.45133	51.32743	25.22124
8	8	251	61	124	66	24.30279	49.40239	26.29482
9	9	263	41	140	82	15.58935	53.23194	31.17871
10	10	172	40	73	59	23.25581	42.44186	34.30233

several players that need to be ranked. Thus, the relevance of the question that we are posing in this research.

3 Methods

3.1 Postulated Probabilistic Model

In the ranking approach that we consider in this research, it is postulated that each chess player has an intrinsic chess playing ability denoted by θ, which is an unobservable. Given two players, i and j, whose abilities are θ_i and θ_j, respectively, when they play a game, the probabilities of each of the three possible outcomes of the game are modeled according to a logistic function which takes as main input the difference between the abilities. Both the Elo and Glicko rating systems also use the same basic postulate. The outcome probabilities for the two players are given by:

$$P_{i,j} = \frac{e^{\alpha+\beta(\theta_i-\theta_j)}}{1+e^{\alpha+\beta(\theta_i-\theta_j)}+e^{\alpha+\beta(\theta_j-\theta_i)}}$$

$$Q_{i,j} = \frac{e^{\alpha+\beta(\theta_j-\theta_i)}}{1+e^{\alpha+\beta(\theta_i-\theta_j)}+e^{\alpha+\beta(\theta_j-\theta_i)}}$$

$$R_{i,j} = \frac{1}{1+e^{\alpha+\beta(\theta_i-\theta_j)}+e^{\alpha+\beta(\theta_j-\theta_i)}},$$

where $P_{i,j}$ is the probability of player i winning against player j, $Q_{i,j}$ is the probability of player i losing against player j, and $R_{i,j}$ is the probability that the game will be drawn. In these formulae, aside from the chess playing abilities θ_i and θ_j, α and β are two parameters encoding the way in which draws occur and the multiplier (regression coefficient) for the impact of the difference in playing abilities of the players regarding the outcome of the game. However, observe that there is a certain non-identifiability problem with this model because β could be multiplied by a constant c ($c > 0$), and the θ_i's could be divided by the same constant c, and the probabilities would remain the same. Thus, we shall simply recode the parameters by letting

$$\gamma_i = \beta\theta_i \quad \text{and} \quad \gamma_j = \beta\theta_j.$$

In terms of the re-parametrization, the outcome probabilities become

$$P_{i,j} = \frac{e^{\alpha+(\gamma_i-\gamma_j)}}{1+e^{\alpha+(\gamma_i-\gamma_j)}+e^{\alpha+(\gamma_j-\gamma_i)}}$$

$$Q_{i,j} = \frac{e^{\alpha+(\gamma_j-\gamma_i)}}{1+e^{\alpha+(\gamma_i-\gamma_j)}+e^{\alpha+(\gamma_j-\gamma_i)}}$$

$$R_{i,j} = \frac{1}{1+e^{\alpha+(\gamma_i-\gamma_j)}+e^{\alpha+(\gamma_j-\gamma_i)}}.$$

Even with the re-parametrization, the model remains non-identifiable with respect to the γ_i's because the probabilistic functions depend only on the differences of the the θ_i's. To address this problem, additional constraints are imposed on the γ_i's, such as, for example, the constraint that the γ_i's for all the players should average out to be equal to some fixed value, e.g., 1500.

So suppose now that we have a group consisting of M players, such as the current top ten grandmasters in the world. If we could determine their chess-playing abilities $\gamma_i, i = 1, 2, \ldots, M$, then we could rank these M players, with the player having the highest chess-playing ability ranked number 1 among the group. The problem, however, is that their chess-abilities are not observed. Rather, we are only able to observe the outcomes of their games against each other. Statistics therefore enters the picture because on the basis of the outcomes of their head-to-head matches, we will try to discover their (relative) chess-playing abilities, and then consequently rank them among each other.

3.2 Statistical Inference

Suppose then that the M players play a total of N games and that in game k, player i plays against player j. Define the outcome of game k to be X_k where $X_k = 1$ corresponds to player i winning, $X_k = 0$ corresponds to player i and player j drawing, and $X_k = -1$ corresponds to player j winning. The likelihood function of the parameters α, γ_i, and γ_j from the outcome of game k is

$$L_k(\alpha, \gamma_i, \gamma_j | X_k) = \frac{e^{I[X_k=1](\alpha+(\gamma_i-\gamma_j))} e^{I[X_k=-1](\alpha+(\gamma_j-\gamma_i))}}{1 + e^{\alpha+(\gamma_i-\gamma_j)} + e^{\alpha+(\gamma_j-\gamma_i)}},$$

where $I[\cdot]$ is the indicator function that is equal to 1 when its input is true and 0 when its input is false.

Observe that if we consider the first player in game k to be player A and the second to be player B, then we can relabel $i = A_k$ and $j = B_k$ for each k. This provides the likelihood function of game k to be:

$$L_k(\alpha, \gamma_{A_k}, \gamma_{B_k} | X_k) = \frac{e^{I[X_k=1](\alpha+(\gamma_{A_k}-\gamma_{B_k}))} e^{I[X_k=-1](\alpha+(\gamma_{B_k}-\gamma_{A_k}))}}{1 + e^{\alpha+(\gamma_{A_k}-\gamma_{B_k})} + e^{\alpha+(\gamma_{B_k}-\gamma_{A_k})}}$$

Then if we take the natural logarithm of both sides, we see that

$$\ell_k(\alpha, \gamma_{A_k}, \gamma_{B_k}|X_k) = log(L_k(\alpha, \gamma_{A_k}, \gamma_{B_k}|X_k))$$

so

$$\begin{aligned}\ell_k(\alpha, \gamma_{A_k}, \gamma_{B_k}|X_k) &= I[X_k = 1](\alpha + (\gamma_{A_k} - \gamma_{B_k})) + I[X_k = -1](\alpha + (\gamma_{B_k} - \gamma_{A_k})) - \\ &\quad log(1 + e^{\alpha+(\gamma_{A_k} - \gamma_{B_k})} + e^{\alpha+(\gamma_{B_k} - \gamma_{A_k})}).\end{aligned}$$

Next, we assume that the N games are independent of each other, though we admit that this may not be an accurate assumption due to the possible impacts of player momentum and presence of streaks. Nevertheless, the independence assumption is an approximate one, enabling us to proceed with the estimation of the players' abilities. With this independence assumption, we are able to obtain the likelihood function of the parameters α and $\gamma_i, i = 1, 2, \ldots, M$, by multiplying the likelihoods from each of the N games.

Let the overall likelihood function be

$$L(\alpha, \underline{\gamma}|\underline{X}) = \prod_{k=1}^{N} L_k(\alpha, \gamma_{A_k}, \gamma_{B_k}|X_k),$$

which is the product of the likelihood functions of all N games, where $\underline{X}$ is the vector of outcomes of all N games and $\underline{\gamma}$ is the vector of all M players' abilities. Then, notice that

$$\sum_{k=1}^{N} \ell_k(\alpha, \gamma_{A_k}, \gamma_{B_k}|X_k) = \ell(\alpha, \underline{\gamma}|\underline{X}),$$

the log-likelihood function based on the entire set of games and the entire set of players. The goal is to maximize this likelihood or log-likelihood function. The maximizers will be the maximum likelihood estimates of α and the γ_i's.

3.3 ML, Newton-Raphson, & Direct Maximization

Continuing the likelihood approach, let $U(\alpha, \underline{\gamma}) = [\frac{\partial \ell}{\partial \alpha} \frac{\partial \ell}{\partial \underline{\gamma}}]^T$, where $\frac{\partial \ell}{\partial \underline{\gamma}}$ is the vector of partial derivatives of the log-likelihood function with respect to each γ value, i.e. the gradient of the log-likelihood function. Note that this is the score funcion of the likelihood equation. When the score function is equal to the zero vector, a local extreme point, presumably the global maximum, has been obtained. Similarly, let

$$I(\alpha, \underline{\gamma}) = -\begin{bmatrix} \frac{\partial^2 \ell}{\partial \alpha^2} & \frac{\partial^2 \ell}{\partial \alpha \partial \gamma_1} & \cdots & \frac{\partial^2 \ell}{\partial \alpha \partial \gamma_N} \\ \frac{\partial^2 \ell}{\partial \gamma_1 \partial \alpha} & \frac{\partial^2 \ell}{\partial^2 \gamma_1} & \cdots & \frac{\partial^2 \ell}{\partial \gamma_1 \partial \gamma_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 \ell}{\partial \gamma_N \partial \alpha} & \frac{\partial^2 \ell}{\partial \gamma_N \partial \gamma_1} & \cdots & \frac{\partial^2 \ell}{\partial^2 \gamma_N} \end{bmatrix}$$

This matrix is the matrix of negative second partial derivatives of the log-likelihood function, which is in this case a symmetric matrix. Note that this is the information matrix for the log-likelihood function.

Using both the score function and the information matrix, one potential approach could be to use the Newton-Raphson method in order to numerically approximate the parameters that maximize the likelihood function. First, an initial guess, call it, $g_{old} = [\alpha_{seed}, \underline{\gamma}_{seed}]^T$, must be made in order to start the algorithm. Then, the vector $U_{seed} = U(\alpha_{seed}, \underline{\gamma}_{seed})$ is multiplied by the inverse of information matrix, $I_{seed} = I(\alpha_{seed}, \underline{\gamma}_{seed})$. The result of this matrix-vector multiplication is of course a vector which is then added to g_{old}. Call this final result $g_{new} = [\alpha_{new}, \underline{\gamma}_{new}]^T$. This can be written simply as $g_{new} = ((I_{seed})^{-1} * U_{seed}) + g_{old}$. This process is then repeated by replacing U_{seed} with U_{new}, I_{seed} by $I(\alpha_{new}, \underline{\gamma}_{new})$, and g_{old} by g_{new}. The algorithm ends when the Euclidean distance between g_{seed} and g_{new} is less than a given tolerance for an iteration of the algorithm.

The success of this algorithm, in the context of our problem, depends on two assumptions: 1) that the log-likelihood function has one maximum that the algorithm can easily determine, and 2) that the parameters are identifiable. Consider the second assumption. The term *identifiable* essentially pertains to the ability to separate the parameters and determine them using the approach above. Notice that in the likelihood function, for players of game k, A_k and B_k, their playing abilities are always paired together as $(\gamma_i - \gamma_j)$, up to a change in sign. This makes it extremely difficult to separate the two parameters, meaning that the playing ability parameters are considered to be unidentifiable in this situation. One solution to this predicament could be to put a constraint on the ability values in the Newton-Raphson method, using tools such as Lagrange multipliers. However, this approach was discontinued in what appears to be a more viable and more convergent approach.

Because the computation for this project was carried out in the statistics package R, there was a plethora of functions available from which to select, along with ample capability to code new programs to fit the specifics of this problem. The function that was essential to approximating the maximum of the log-likelihood function was the *optim* function. The *optim* function attempts to find a minimum or a maximum of a given function, given parameters and a maximum number of iterations. It also determines whether an algorithm has in fact converged. The *optim* function was used in accordance with $\ell(\alpha, \underline{\gamma})$ in order to find the parameters that maximize the log-likelihood function. Furthermore, these maximizers were used as input to the score function to determine whether an approximate maximum was actually found by the algorithm.

3.4 Testing the Methods

The next thing to consider is whether the presumed solution will actually work when it is supposed to. The term “supposed to” here applies to the fact that the model is specifically meant to be applied when a simple logistical model is assumed to fit the game outcomes.

First, a set of 5 players with distinct labels, 1 though 5, and with randomly generated relative playing abilities $\gamma_j \sim N(\mu = 0, \sigma^2 = .5)$, were created. The simulation then randomly paired these players together in a series of 500 total games and outcomes were determined according to the logistic model discussed earlier with an α value of -1. This resulted in the sets of data regarding mumber of games, wins, and draws among the different pairs of players, displayed in Table 7, Table 8, and Table 9.

Table 7: Number of games of column player versus row player

Player	**1**	**2**	**3**	**4**	**5**
1	0	32	52	42	63
2	32	0	51	52	56
3	52	51	0	50	55
4	42	52	50	0	47
5	63	56	55	47	0

Table 8: Number of wins of column player versus row player. **Note:** The table corresponding to the number of losses of the column player versus the row player is just the transpose of this table.

Player	**1**	**2**	**3**	**4**	**5**
1	0	10	24	11	40
2	6	0	11	7	16
3	3	11	0	7	19
4	3	10	15	0	15
5	1	7	4	4	0

After this data was properly recorded in R, the function described earlier was applied in order to estimate their relative playing abilities (γ_j), as well as to rank the players based on the ordering of relative abilities. This resulted in the following output displayed in Table 10.

Based on the estimation of the intrinsic abilities, γ_j, and the estimation of the parameter α, the values were used as input in order to estimate $P_{i,j}$ and $Q_{i,j}$. Each

Table 9: Number of draws of column player versus row player

Player	**1**	**2**	**3**	**4**	**5**
1	0	16	25	28	22
2	16	0	29	35	33
3	25	29	0	28	32
4	28	35	28	0	28
5	22	33	32	28	0

Table 10: Estimation of Playing Abilities and Ranks of Players

Player	**1**	**2**	**3**	**4**	**5**
γ_j	0.627	-0.211	-0.181	0.096	-0.705
$\hat{\gamma}_j$	0.752	-0.0756	-0.174	0.113	-0.726
Real Rank	1	4	3	2	5
Estimated Rank	1	3	4	2	5

of these probabilities were then multiplied by the corresponding number of games played between players i and j in order to estimate the number of wins, draws, and losses that players i and j would have if they played the same number of games again. This resulted in the following output displayed in Table 11 and Table 12:

Table 11: Estimated number of wins of column player versus row player. **Note:** The table corresponding to the estimated number of losses of the column player versus the row player is just the transpose of this table.

Player	**1**	**2**	**3**	**4**	**5**
1	0	12.4	21.5	14.2	35.4
2	2.4	0	10.8	8.2	19.1
3	3.4	8.9	0	7.1	17.4
4	4	12	12.6	0	18.3
5	1.8	5.2	5.8	3.4	0

First, the estimation of the γ_j-values should be considered. As discussed earlier, there is an element of non-idenifiability regarding the values, so what is truly important here is that the *relative* abilities, i.e. the differences between the γ_j values is close

Table 12: Estimated number of draws of column player versus row player

Player	**1**	**2**	**3**	**4**	**5**
1	0	17.2	27.1	23.9	25.7
2	17.2	0	31.3	31.7	31.7
3	27.1	31.3	0	30.2	31.9
4	23.9	31.7	30.2	0	25.2
5	25.7	31.7	31.9	25.2	0

to the actual differences. This is in fact the case. Next, the estimation of the rank ordering should be examined. Three out of the five players have an estimated ranking that corresponds exactly to their actual rankings. Each of the other two players have only a difference of 1 in terms of the discrepancy between the actual ranking and the estimated ranking. If we take a closer look at these players, player 2 and player 3, we see that their abilities are extremely close, which is likely the cause of the small discrepancy.

Next, the same set of players was randomly paired together again in a series of 500 total games with outcomes determined according to the logistic model with an α value of 0. This resulted in the following sets of data, regarding wins, draws, and losses between the different pairs of players, displayed in Table 13, Table 14, and Table 15.

Table 13: Number of games of column player versus row player

Player	**1**	**2**	**3**	**4**	**5**
1	0	51	51	52	57
2	51	0	53	37	49
3	51	53	0	51	55
4	52	37	51	0	44
5	57	49	55	44	0

Based on the estimations of the intrinsic abilities, γ_j, and the estimation of the

Table 14: Number of wins of column Player versus row player. **Note:** The table corresponding to the number of losses of the column player versus the row player is just the transpose of this table.

Player	**1**	**2**	**3**	**4**	**5**
1	0	34	37	25	48
2	5	0	15	8	27
3	6	16	0	15	33
4	12	22	21	0	21
5	4	7	10	5	0

Table 15: Number of draws of column player versus row player

Player	**1**	**2**	**3**	**4**	**5**
1	0	12	8	15	5
2	12	0	22	7	15
3	8	22	0	15	12
4	15	7	15	0	18
5	5	15	12	18	0

Table 16: Estimation of playing abilities and ranks of players

Player	**1**	**2**	**3**	**4**	**5**
γ_j	0.627	-0.211	-0.181	0.096	-0.705
$\hat{\gamma}_j$	0.667	-0.258	-0.182	0.102	-0.761
Real Rank	1	4	3	2	5
Estimated Rank	1	4	3	2	5

parameter α, the values were used as input in order to estimate $P_{i,j}$ and $Q_{i,j}$. Each of these probabilities were then multiplied by the corresponding number of games played between players i and j in order to estimate the number of wins, draws, and losses that players i and j would have if they played the same number of games again. This resulted in the following output displayed in Table 17 and Table 18.

Once again, we observe that the differences between any two estimated γ_j values are very similar to the actual differences between those two γ_j values. This suggests

Table 17: Estimated number of wins of column player versus row player. **Note:** The table corresponding to the estimated number of losses of the column player versus the row player is just the transpose of this table.

Player	**1**	**2**	**3**	**4**	**5**
1	0	33.9	32.7	28.6	45
2	5.3	0	17	8.6	25.8
3	6	19.8	0	13	30.5
4	9.2	17.6	22.9	0	28.4
5	2.6	9.5	9.6	5.1	0

Table 18: Estimated number of draws of column player versus row player

Player	**1**	**2**	**3**	**4**	**5**
1	0	11.8	12.3	14.2	9.5
2	11.8	0	16.1	10.8	13.7
3	12.3	16.1	0	15.1	15
4	14.2	10.8	15.1	0	10.5
5	9.5	13.7	15	10.5	0

that though the γ_j values may not be estimable, the differences are estimable. This implies that the algorithm seems to be working in this case as well, considering the unidenifiability of the γ_j values. As can be seen in Table 16, the estimated rankings are exactly aligned with the actual rankings based on the intrinsic abilities.

Finally, the same set of players was randomly paired together again in a series of 500 total games with outcomes determined according to the logistic model with an α value of 1. This resulted in the following sets of data, regarding wins, draws, and losses between the different pairs of players, shown in Table 19, Table 20, and Table 21.

Based on the estimations of the intrinsic abilities, γ_j, and the estimation of the parameter α, the estimates were used as input in order to estimate $P_{i,j}$ and $Q_{i,j}$. Each of these probabilities were then multiplied by the corresponding number of games played between players i and j in order to estimate the number of wins, draws,

Table 19: Number of games of column player versus row player

Player	**1**	**2**	**3**	**4**	**5**
1	0	49	52	55	52
2	49	0	54	57	35
3	52	54	0	51	46
4	55	57	51	0	49
5	52	35	46	49	0

Table 20: Number of wins of column player versus row player. **Note:** The table corresponding to the number of losses of the column player versus the row player is just the transpose of this table.

Player	**1**	**2**	**3**	**4**	**5**
1	0	36	42	30	40
2	8	0	22	11	19
3	8	22	0	20	33
4	12	32	27	0	36
5	7	5	6	9	0

Table 21: Number of draws of column player versus row player

Player	**1**	**2**	**3**	**4**	**5**
1	0	12	8	15	5
2	12	0	22	7	15
3	8	22	0	15	12
4	15	7	15	0	18
5	5	15	12	18	0

Table 22: Estimation of playing abilities and ranks of players

Player	**1**	**2**	**3**	**4**	**5**
γ_j	0.627	-0.211	-0.181	0.096	-0.705
$\hat{\gamma}_j$	0.581	-0.204	-0.115	0.149	-0.673
Real Rank	1	4	3	2	5
Estimated Rank	1	4	3	2	5

and losses that players i and j would have if they played the same number of games again. This resulted in the following output displayed in Table 23 and Table 24.

Table 23: Estimated number of wins of column player versus row player **Note:** The table corresponding to the estimated number of losses of the column player versus the row player is just the transpose of this table.

Player	**1**	**2**	**3**	**4**	**5**
1	0	35	35.6	32.4	43.3
2	7.3	0	20.3	15.7	21.1
3	8.9	24.3	0	15.7	29.3
4	13.7	31.9	26.7	0	35.5
5	3.5	8.3	9.6	6.9	0

Table 24: Estimated number of draws of column player versus row player

Player	**1**	**2**	**3**	**4**	**5**
1	0	6.7	7.5	8.9	5.2
2	6.7	0	9.4	9.4	5.6
3	7.5	9.4	0	8.6	7.1
4	8.9	9.4	8.6	0	6.6
5	5.2	5.6	7.1	6.6	0

Finally, we observe that differences between any two estimated γ_j values is about the same as the actual difference between the two γ_j values. Furthermore, all the estimated rankings are exactly the same as the actual rankings, based on the playing abilities. This suggests that the solution continues to be adequate when $\alpha = 1$.

4 Ranking of the Top 10 Players

When the same algorithm is applied to the chesscdata set shown in Tables 2, 3, 4, and 5. We are able to estimate the relative rankings, which are show in Table 25.

Also, the estimated value of α was $\hat{\alpha} = -0.867775$. By looking at the logistic model for this problem, we can consider α to be a weighting of players' abilities. As

Table 25: Ranking of the the world's top ten players based on the ranking approach in this research.

Name	Player	World Rank (Elo-Based)	Estimated Rank
GM Carlsen	1	1	1
GM Anand	2	3	3
GM Aronian	3	2	4
GM Nakamura	4	9	2
GM Kramnik	5	4	5
GM Caruana	6	5	8
GM Grischuk	7	6	6
GM Karjakin	8	7	7
GM Topalov	9	8	10
GM Mamedyarov	10	10	9

can be seen in the simulations when α is positive, it corresponds to a smaller number of draws. When α is negative, it corresponds to a greater number of draws. Finally, when $\alpha = 0$, it corresponds to a uniform distribution of total wins, losses, and draws in the simulation. Further analysis of the model could more quantitatively explain this phenomenon, but this was not pursued any deeper in this paper.

Quickly comparing the actual rankings to the estimated rankings, it can be seen that most players' estimated rankings are within either 1 or 2 of their actual rankings. However, there is one player who has a significant difference in ranking versus estimated ranking based on the model. Nakamura has the ninth highest Elo rating, yet the model estimated that he is the second best player based on his games against other top 10 players. Analyzing this qualitatively, this could be because of Nakamura's disdain for drawing matches and his aggressive playing style. He has even gone as far as to say, "There is no point in taking draws" [**United States Chess Federation**]. This means that the model may favor players who are more willing to aim for a high winning percentage instead of a high drawing percentage.

Another possible reason for the discrepancies between the actual and estimated rankings is the restriction of the data used. The only data considered in this paper were the games played between the top 10 players with each other, while the actual rankings and ratings used internationally take into account every competition-level game that a player has played. However, it is possible that using all games could skew the ratings, and thus the rankings. For instance, a competitor could play a greater number of matches, but with weaker opponents on average. As discussed earlier, Caruana plays more competitive chess than any of the other top 10 players. He has the fifth highest Elo rating from this sample, yet the modelling technique determined he was the eighth best player of the group. This is potentially because he has a slightly inflated Elo rating that may be a result of playing weaker opponents.

5 References

[**1**] M. Glickman, E.: The Glicko system.

[**2**] L. Barden: Magnus Carlsen: The greatest of all time or too much of a grinder?, The Guardian (February 28, 2014).

[**3**] D. McClain, L.: Overlooked, Except at the Board, The New York Times, (January 28, 2012).

[**4**] S. Ninan,: Anand must play his best chess this time: Gelfand, The Times of India, (November 6, 2013).

[**5**] M. Franett, J.: Vladmir Kramnic, Britannica, (September 11, 2013).

[**6**] L. Barden: Fabiano Caruana reaches world No4 ranking and closes in on Vladmir Kramnik, The Guardian, (October 18, 2013).

[**7**] D. Monokroussos : Alexander Grischuk, the neglected genius, Chess News, (June 21, 2007).

[**8**] L. Barden : Collapse by Magnus Carlsen benefits Serget Karjakin at World Rapid, The Guardian, (July 13, 2012).

[**9**] L. Barden : Experts divided over cheating allegation against Topalov, The Guardian, (January 28, 2007).

[**10**] NPR Staff : Chess Champ Hikaru Nakamura: Next Bobby Fischer?, NPR, (May 11, 2012).

[**11**] FIDE, FIDE Grand Prix Series: Mamedyarov, Shakhriyar, FIDE, (December 5, 2010).