Predicting Grandmaster Performances: Skill Analysis in Offline Chess

Tathagata Banerjee[1] & Diganta Mukherjee[2]
[1]Indian Institute of Technology, Kanpur [2]Indian Statistical Institute, Kolkata,

1 Introduction

Chess is an immensely popular strategic board game, said to have originated in India around the 6th century AD. Initially known as "Chaturanga," it eventually spread to Persia, where it evolved into "Shatranj." In the 15th century, the rules for the present game were formed. Despite being played for several centuries among all classes of people, chess is still a timeless game, Chess symbolizes intellect, strategy, and foresight, attracting enthusiasts worldwide. From all-time greats like Magnus Carlsen and Gary Kasparov, streaming sensations like Hikaru Nakamura and Levy Rozmann, and advanced engines like Stockfish and Komado, chess has many dimensions. Though variants like Fischer Random, Antichess and Horde are gaining popularity, standard chess is still by far the most popular.
Chess is played in many time formats, which can be broadly classified into four categories in descending order of time, **Classical, Rapid, Blitz and Bullet**. There are multiple time controls within the same category as well. Today, Chess is popular in both **online** and **offline (OTB)** formats.
In this study, we will be considering Classical and Rapid OTB chess games.
For classical chess, the time format in these games are 90/100 minutes for the first 40 moves, followed by 30/50 minutes for the remaining game with an increment of 30 second per move.
For rapid games, the time format is 15 minutes per player, with an increment of 10 seconds per move.

2 Data Description

2.1 Classical Chess

For this study, data has been collected from multiple editions of two tournaments, Tata Steel Chess Tournament and the Fide World Cup.
The Tata Steel Chess Tournament is played in two brackets, called the Tata Steel Masters and Tata Steel Challengers. 14 top players play in each bracket, and the winner is decided by a Round Robin system within each bracket.
The Fide World Cup is a knockout tournament across two legs, consisting of a maximum of 128 players.

The dataset contains game level data for Tata Steel Masters '24,'23 and '22 tournaments, Tata Steel Challengers '24 and '23, and Fide World Cup '23 and '21 editions. The following information for each game is presented in the dataset-

- Tournament Name

- Name of White Player
- Age of White Player: For ease of analysis, looking at the distribution of age for the players in the sample, we have scaled Age dividing by 6.5.
- Rating of White Player: For similar reason as for age, Rating is transformed through a shift of origin (set at 2000) and scaling by 100.
- Title of White Player
- Name of Black Player
- Age of Black Player
- Rating of Black Player
- Title of Black Player
- Result of the game (with respect to white) - 1, 0.5 and 0 denote a win, draw and loss for white respectively

Descriptives	Age (Scaled)	Rating (Scaled)
Minimum	1.846	-0.02
5th percentile	2.615385	3.893
1st Quartile	3.385	5.82
Median	4.308	6.4
Mean	4.351	6.294
3rd Quartile	5.077	7.14
95th percentile	6.46	7.78
Maximum	9.846	8.65
S.D.	1.26	1.2

Table 1: Descriptive Statistics for Age (scaled) = Age/6.5 and Rating (scaled) = (Rating - 2000)/100 of Players in Classical Chess

The descriptive statistics for age and rating (see table 1) and the plots (histograms in figures 1 for age and 2 for rating, Scatter plot in figure 3) have been constructed using White Age and White Rating columns of the database, since all players play with white pieces at least once.

It is observed that the correlation coefficient between age and rating is 0.013. The scatter plot 3 has an initially positively sloped part where learning and hence rating increases with age. But then there is a downward bend at around the age of 30 - 35 when fatigue sets is and rating starts to deteriorate. This explains the low correlation. This is further corroborated by the rather thin presence of relatively older players in this sample, as observed in histogram (figure 1).

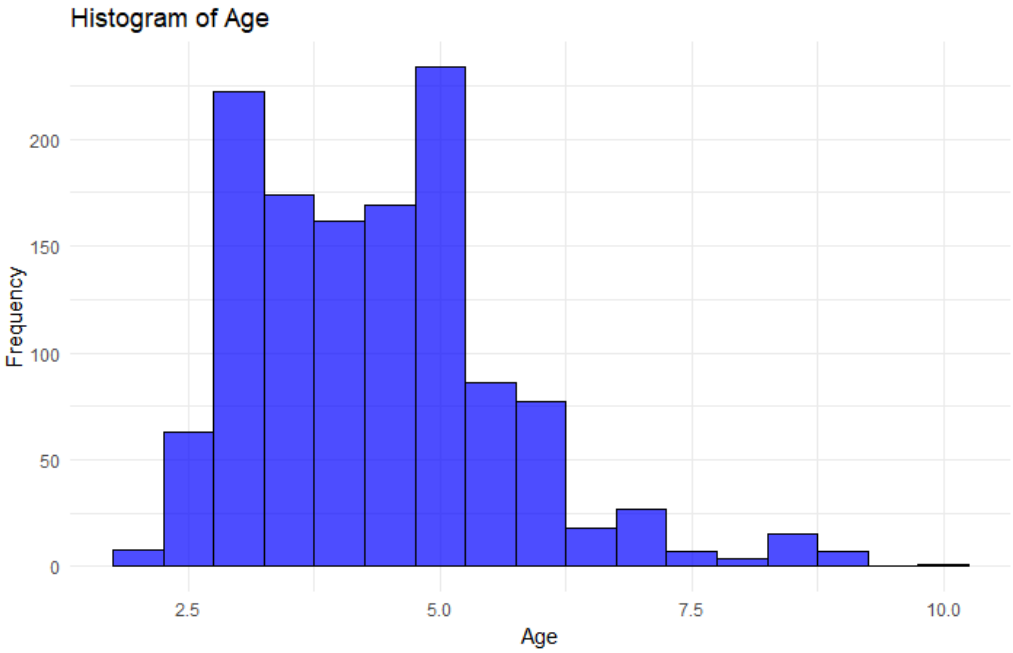

Figure 1: Histogram of Age for Classical Chess

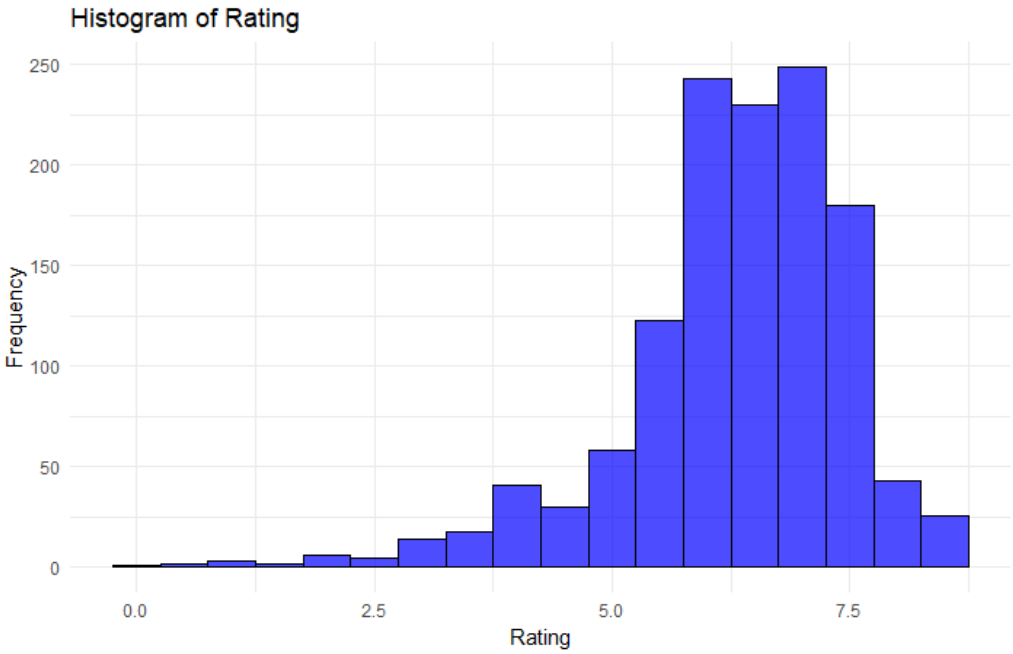

Figure 2: Histogram of Ratings for Classical Chess

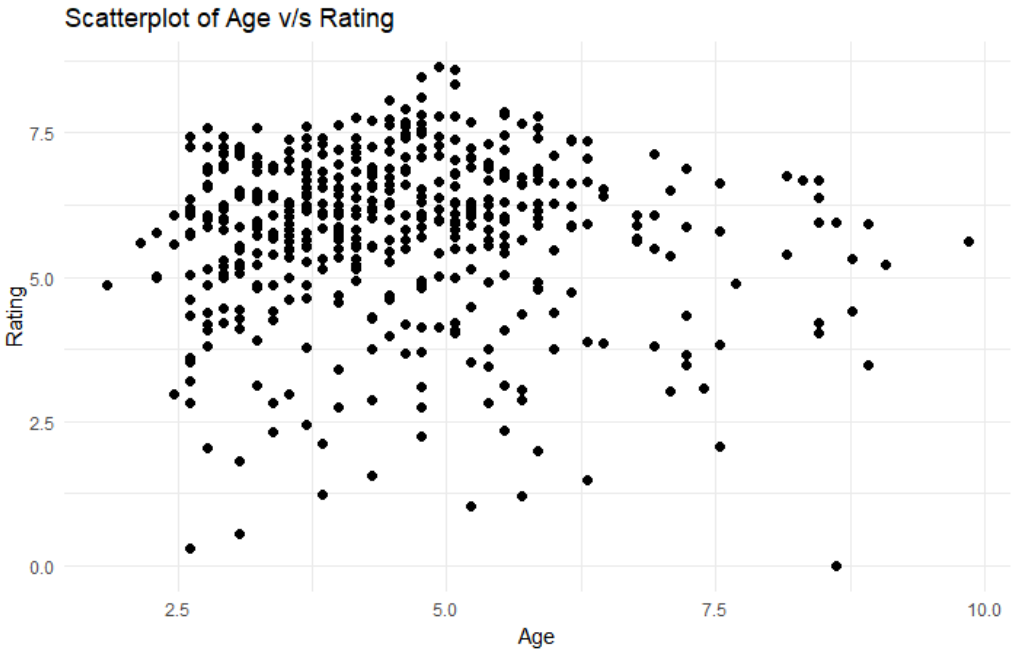

Figure 3: Scatterplot of Age v/s Rating for Classical Chess

2.2 Rapid Chess

For this part of the study, data has been collected from Fide World Rapid Championships 2023 - Open Rapid section. The tournament is played by 202 players across 13 rounds in the Swiss system. Game level data of 1298 games from this tournament has been presented in the dataset in the same format as the Classical one. Again, looking at the distribution, the scaling for Age has been done dividing by 6 and for Rating origin is shifted to 1700 and scale is 120. It is seen that Rapid Chess is played by relatively younger players (with possibly somewhat lower ratings than the classical version).

For this Rapid version data, similar descriptive statistics (table 2) and the plots (histograms in figures 4 for age and 5 for rating, Scatterplot in figure 6) have been constructed using White Age and White Rating columns of the database, as for the classical case above.

Descriptives	Age (Scaled)	Rating (Scaled)
Minimum	1.333	0.375
5th percentile	2.167	3.417
1st Quartile	3.167	5.642
Median	4.167	6.742
Mean	4.38	6.443
3rd Quartile	5.5	7.492
95th percentile	7	8.508
Maximum	9.33	9.317
S.D.	1.54	1.59

Table 2: Descriptive Statistics for Age (scaled) = Age/6 and Rating (scaled) = (Rating - 1700)/120 of Players in Rapid Chess

Now the correlation coefficient between age and rating turns out to be 0.467, much higher than for Classical. This is explained by looking at the scatter plot 6 where we no longer see the downward bending segment (fatigue). This is possibly due to the preponderance of younger players in this format. This is also supported by the histogram 4 which rarely populated by players older than 35. So the fatigue segment almost does not appear.

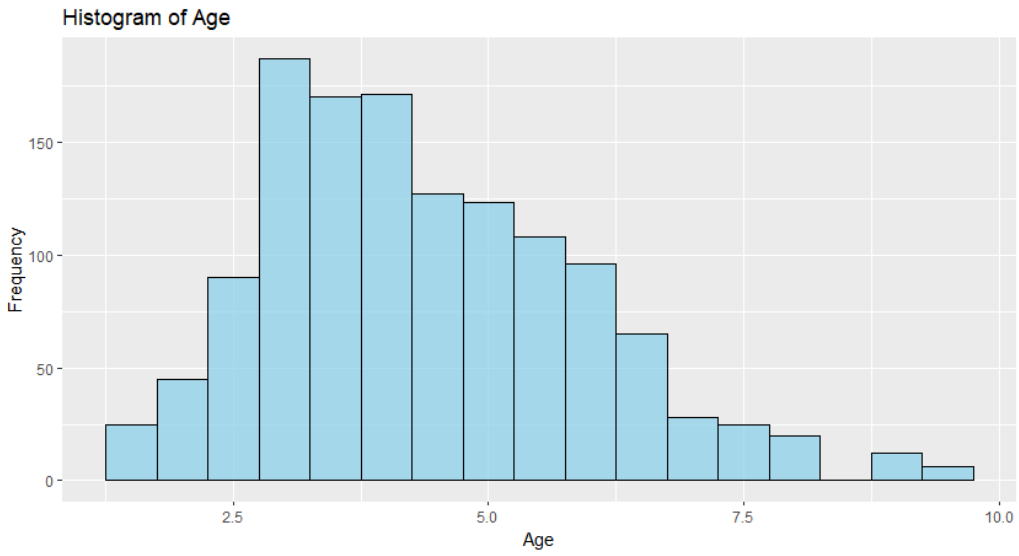

Figure 4: Histogram of Age for Rapid Chess

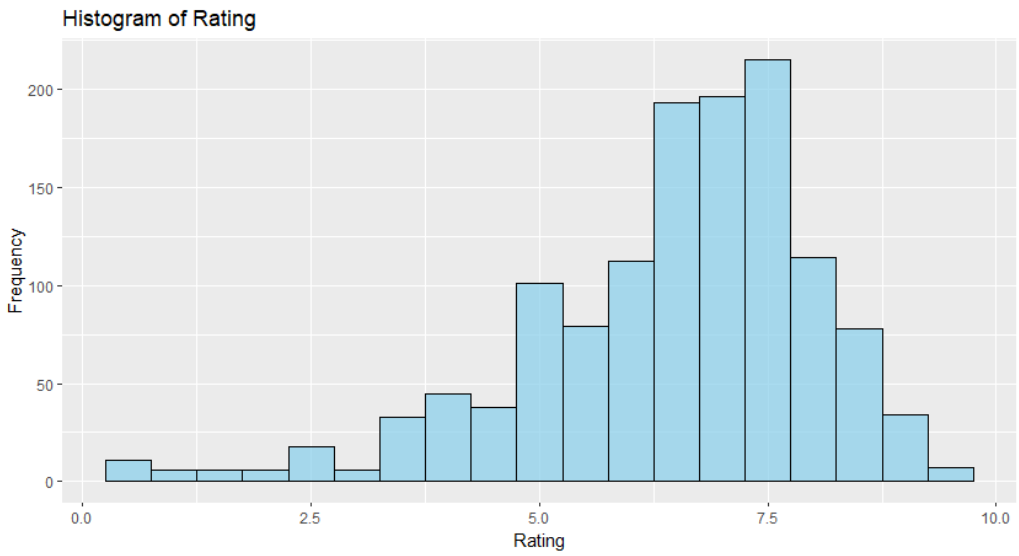

Figure 5: Histogram of Ratings for Rapid Chess

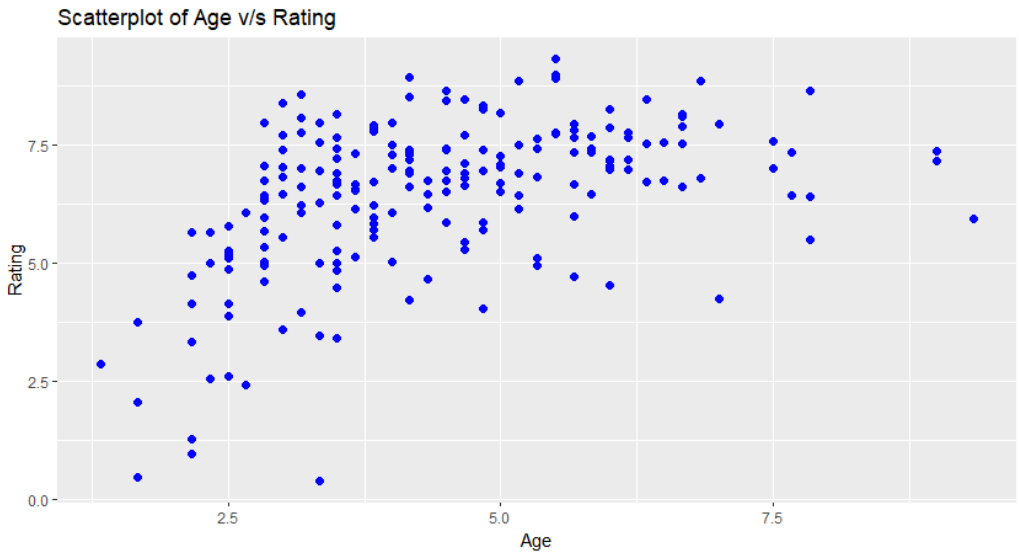

Figure 6: Scatterplot of Age v/s Rating for Rapid Chess

3 Analysis for Classical Chess: Model and Results

The multinomial logistic regression model was fitted to analyze the relationship between predictors and response across multiple categories. The model was estimated using maximum likelihood estimation.

Model Information:

Response Variable: Result of White in the game. 1 denotes a win, 0.5 denotes a draw and 0 denotes a loss.

Predictor Variables: White Age, White Rating, Black Age, Black Rating and the corresponding squared and cross product terms, after scaling. Recall that Age is scaled by a factor of 1/6.5 and rating by a factor of 1/100 and shifted origin by -2000.

The Log-Likelihood: -2262.023

The estimates are presented in table 3.

Predictor	Estimate	Std. Error	Test Statistic	p-value
(Intercept):0.5	5.086677	2.960317	1.718	0.086
. (Intercept):1	0.703244	3.477806	0.202	0.84
WhiteAge:0.5	-0.639427	0.525353	-1.217	0.24
WhiteAge:1	0.464168	0.458002	-1.913	0.048
WhiteRating:0.5	0.401650	0.514189	0.781	0.57
WhiteRating:1	0.293208	0.428493	0.684	0.506
WhiteAgeSquared:0.5	0.077655	0.030084	2.581	0.009
WhiteAgeSquared:1	0.022810	0.031281	0.729	0.47
WhiteRatingSquared:0.5	0.079319	0.043011	1.844	0.06 .
WhiteRatingSquared:1	0.011578	0.036558	0.317	0.75
WhiteAgeRatingCP:0.5	0.039170	0.062488	2.1788	0.029
WhiteAgeRatingCP:1	0.115133	0.053012	0.06	0.98
BlackAge:0.5	-0.247743	0.434162	0.571	0.57
BlackAge:1	-0.284396	0.705262	0.403	0.68
BlackRating:0.5	-0.849957	0.491061	-1.731	0.08 .
BlackRating:1	-0.314878	0.706404	-0.446	0.65
BlackAgeSquared:0.5	0.002649	0.027131	0.098	0.92
BlackAgeSquared:1	-0.021793	0.043973	-0.496	0.62
BlackRatingSquared:0.5	0.016559	0.040382	0.410	0.68
BlackRatingSquared:1	0.067894	0.058541	1.160	0.24
BlackAgeRatingCP:0.5	0.026230	0.052194	-0.503	0.61
BlackAgeRatingCP:1	0.051807	0.084612	-0.612	0.54

Table 3: Full Model Summary for Classical Chess

This model is not used for further analysis since only three of the 20 predictors are significant, which is due to severe multicollinearity in the predictors.

A more parsimonious model with only linear terms is estimated instead, the result is presented in table 4.

Log Likelihood for this model = -1138.676

This has all the desirable properties and is used for our analysis.

Predictor	Estimate	Std.Error	Test Statistic	p-value
(Intercept):0.5	-1.5358239	0.733	-2.09	0.036
(Intercept):1	0.0866663	0.841	0.193	0.84
WhiteAge:0.5	-0.03022938	0.059	-0.513	0.62
WhiteAge:1	-0.14232809	0.072	-1.981	0.047
WhiteRating:0.5	0.6413002	0.068	9.477	0
WhiteRating:1	0.9973328	0.093	10.716	0
BlackAge:0.5	0.2408411	0.071	3.347	0.0008
BlackAge:1	0.3518401	0.079	4.43	0
BlackRating:0.5	-0.3526729	0.096	-3.669	0.002
BlackRating:1	-1.0827602	0.105	-10.27	0

Table 4: Parsimonious Model Summary

Note: The model is symmetric with respect to the predictor and the responses, i.e, with Result of Black as the response and Black Age, Black Rating, White Age and White Rating as the predictors, the fitted values are same as the previous model.

3.1 Interpretation of the Parsimonious Regression Model

We illustrate the effect of rating and age combination, for both players, on the probability of win based on the Parsimonious model (table 4). The results for a few combinations of age and rating (which are well represented in the sample) are presented in table 5.

Age		Rating		Probability			Age		Rating		Probability	
White	Black	White	Black	Win (W)	Draw		White	Black	White	Black	Win (W)	Draw
20	20	2500	2500	0.331	0.424		35	20	2500	2500	0.272	0.450
20	20	2500	2600	0.171	0.455		35	20	2500	2600	0.134	0.460
20	20	2500	2700	0.077	0.425		35	20	2500	2700	0.059	0.418
20	20	2500	2800	0.032	0.363		35	20	2500	2800	0.024	0.351
20	20	2600	2500	0.461	0.413		35	20	2600	2500	0.394	0.457
20	20	2600	2600	0.273	0.507		35	20	2600	2600	0.221	0.532
20	20	2600	2700	0.138	0.533		35	20	2600	2700	0.108	0.537
20	20	2600	2800	0.062	0.500		35	20	2600	2800	0.047	0.491
20	20	2700	2500	0.579	0.363		35	20	2700	2500	0.512	0.416
20	20	2700	2600	0.385	0.501		35	20	2700	2600	0.323	0.544
20	20	2700	2700	0.218	0.590		35	20	2700	2700	0.175	0.612
20	20	2700	2800	0.109	0.610		35	20	2700	2800	0.084	0.612
20	20	2800	2500	0.677	0.298		35	20	2800	2500	0.617	0.351
20	20	2800	2600	0.495	0.451		35	20	2800	2600	0.429	0.506
20	20	2800	2700	0.311	0.588		35	20	2800	2700	0.256	0.628
20	20	2800	2800	0.170	0.668		35	20	2800	2800	0.135	0.686
20	35	2500	2500	0.431	0.427		35	35	2500	2500	0.365	0.468
20	35	2500	2600	0.249	0.511		35	35	2500	2600	0.200	0.532
20	35	2500	2700	0.123	0.525		35	35	2500	2700	0.095	0.526
20	35	2500	2800	0.055	0.484		35	35	2500	2800	0.041	0.474
20	35	2600	2500	0.551	0.382		35	35	2600	2500	0.484	0.435
20	35	2600	2600	0.358	0.515		35	35	2600	2600	0.298	0.555
20	35	2600	2700	0.198	0.592		35	35	2600	2700	0.158	0.611
20	35	2600	2800	0.097	0.601		35	35	2600	2800	0.075	0.601
20	35	2700	2500	0.653	0.317		35	35	2700	2500	0.591	0.372
20	35	2700	2600	0.467	0.471		35	35	2700	2600	0.402	0.525
20	35	2700	2700	0.287	0.601		35	35	2700	2700	0.235	0.637
20	35	2700	2800	0.154	0.669		35	35	2700	2800	0.122	0.684
20	35	2800	2500	0.737	0.251		35	35	2800	2500	0.683	0.301
20	35	2800	2600	0.570	0.402		35	35	2800	2600	0.505	0.461
20	35	2800	2700	0.383	0.562		35	35	2800	2700	0.323	0.613
20	35	2800	2800	0.224	0.681		35	35	2800	2800	0.181	0.712

Table 5: A profile of Win & Draw probability for White against Black with specified age (20 or 35) and rating (2500, 2600, 2700 or 2800) for Classical Chess

The following salient points emerge from a study of table 4 and 5.

- From the model parameters in table 4, it is clear that rating (current level of skill) improves

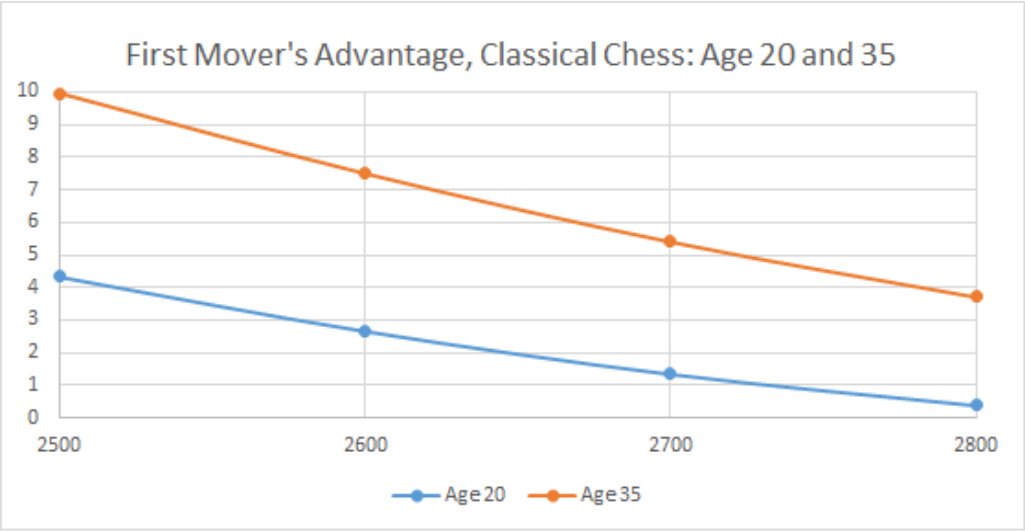

Figure 7: First Mover's Advantage, Classical Chess

chances of winning, but not in a perfect way. Also, age (as a proxy of fatigue) is a detrimental factor.

- The above result is further illustrated by the pairwise comparisons in table 5 where a clear edge for rating and disadvantage for age is documented. E.g. look at results in each block of age pair, and within that if we consider progressive rating gap, the win probability strictly increase. Also, for the same rating pair, younger players enjoy a higher winning probability.

- From table 5 it is clear that first mover's advantage exists and is increasing with age and decreasing with rating. At age 20, average advantage is 2.2% which becomes 6.6% at age 35 when considering games between equally rated players. This is illustrated in figure 7.

- Outcomes are certainly not foregone as even for a rating difference of 300 (2800 playing against 2500), the win probability ranges between 79% and 86% (averaging about 83%) when the higher ranked player plays white. The uncertainty is more when older players face off against younger ones.

4 Analysis for Rapid Chess: Model and Results

Similar analysis is performed for the Rapid chess data. The response and predictor variables are defined similarly, with appropriate change in origin and scale for age and rating, as described in table 2. A multinomial logistic model is fitted to analyze the relationship between the response and the predictors.

The log likelihood = -1290.94

The log likelihood for the alternate (parsimonious) model = -1308.67

Predictor	Estimate	Std. Error	Test Statistic	p-value
Intercept:0.5	-0.9896	1.3077	-0.7567	0.45
Intercept:0.1	-2.1513	1.1835	-1.8177	0.07
WhiteAge:0.5	-0.5332	0.3429	-1.5549	0.12
WhiteAge:1	-0.1730	0.3160	-0.5475	0.58
WhiteRating:0.5	-0.0831	0.2264	-0.3670	0.71
WhiteRating:1	0.1892	0.2178	0.8687	0.39
WhiteAgeSquared:0.5	0.0051	0.0282	0.1822	0.86
WhiteAgeSquared:1	-0.0083	0.0269	-0.3088	0.76
WhiteRatingSquared:0.5	0.0272	0.0270	1.0073	0.31
WhiteRatingSquared:1	0.0291	0.0249	1.1658	0.24
WhiteAgeRatingCP:0.5	0.0615	0.0494	1.2450	0.21
WhiteAgeRatingCP:1	0.0130	0.0449	0.2895	0.77
BlackAge:0.5	0.4928	0.3695	1.3338	0.18
BlackAge:1	0.9767	0.3197	3.0548	0.00
BlackRating:0.5	-0.0925	0.2349	-0.3939	0.69
BlackRating:1	0.0570	0.1985	0.2872	0.77
BlackAgeSquared:0.5	-0.0161	0.0278	-0.5804	0.56
BlackAgeSquared:1	-0.0620	0.0266	-2.3277	0.02
BlackRatingSquared: 0.5	0.0156	0.0256	0.6102	0.54
BlackRatingSquared:1	-0.0531	0.0238	-2.2309	0.03
BlackAgeRatingCP:0.5	-0.0390	0.0472	-0.8265	0.41
BlackAgeRatingCP:1	-0.0169	0.0448	-0.3770	0.71

Table 6: Full Model Summary for Rapid Chess

Since only three of the 20 predictors are significant (refer table 6), possibly due to severe multicollinearity in the predictors, we fit a more parsimonious model with only the linear terms. Results are presented in table 7.

Predictor	Estimate	Std.Error	Test Statistic	p-value
Intercept:0.5	-2.4704	0.4784	-5.1634	0
Intercept:1	-0.1383	0.4165	-0.3320	0.74
WhiteAge:0.5	-0.0686	0.0550	-1.2470	0.21
WhiteAge:1	-0.1788	0.0534	-3.3500	0
WhiteRating:0.5	0.3979	0.0589	6.7550	0
WhiteRating:1	0.5286	0.0561	9.4269	0
BlackAge:0.5	0.0584	0.0558	1.0476	0.29
BlackAge:0.5	0.2137	0.0524	4.0763	0
BlackRating:0.5	-0.0104	0.0617	-0.1688	0.87
BlackRating:1	-0.4909	0.0557	-8.8094	0

Table 7: Parsimonious Model Summary for Rapid Chess

4.1 Interpretation of the Parsimonious Regression Model (Rapid Chess)

We again illustrate the effect of rating and age combination, for both players, on the probability of win based on the Parsimonious model (table 7). The results for a few combinations of age and rating (which are well represented in the sample) are presented in table 8. Note that the selected age and rating levels are lower than that for Classical Chess, as expected from our earlier discussion.

Age		Rating		Probability		Age		Rating		Probability	
White	Black	White	Black	Win (W)	Draw	White	Black	White	Black	Win (W)	Draw
18	18	2300	2300	0.427	0.208	30	18	2300	2300	0.353	0.215
18	18	2300	2500	0.248	0.270	30	18	2300	2500	0.195	0.264
18	18	2300	2600	0.181	0.292	30	18	2300	2600	0.139	0.281
18	18	2300	2700	0.128	0.309	30	18	2300	2700	0.097	0.293
18	18	2500	2300	0.573	0.224	30	18	2500	2300	0.501	0.245
18	18	2500	2500	0.374	0.326	30	18	2500	2500	0.309	0.336
18	18	2500	2600	0.285	0.371	30	18	2500	2600	0.230	0.373
18	18	2500	2700	0.210	0.408	30	18	2500	2700	0.166	0.402
18	18	2600	2300	0.633	0.222	30	18	2600	2300	0.567	0.248
18	18	2600	2500	0.435	0.340	30	18	2600	2500	0.368	0.359
18	18	2600	2600	0.339	0.396	30	18	2600	2600	0.280	0.408
18	18	2600	2700	0.255	0.445	30	18	2600	2700	0.206	0.448
18	18	2700	2300	0.684	0.216	30	18	2700	2300	0.624	0.245
18	18	2700	2500	0.491	0.345	30	18	2700	2500	0.425	0.372
18	18	2700	2600	0.392	0.411	30	18	2700	2600	0.331	0.432
18	18	2700	2700	0.301	0.471	30	18	2700	2700	0.248	0.484
18	30	2300	2300	0.522	0.186	30	30	2300	2300	0.446	0.198
18	30	2300	2500	0.327	0.260	30	30	2300	2500	0.263	0.261
18	30	2300	2600	0.244	0.290	30	30	2300	2600	0.192	0.284
18	30	2300	2700	0.177	0.314	30	30	2300	2700	0.137	0.302
18	30	2500	2300	0.658	0.189	30	30	2500	2300	0.592	0.212
18	30	2500	2500	0.462	0.296	30	30	2500	2500	0.393	0.313
18	30	2500	2600	0.364	0.348	30	30	2500	2600	0.301	0.359
18	30	2500	2700	0.277	0.394	30	30	2500	2700	0.223	0.397
18	30	2600	2300	0.711	0.183	30	30	2600	2300	0.652	0.209
18	30	2600	2500	0.523	0.300	30	30	2600	2500	0.455	0.325
18	30	2600	2600	0.423	0.362	30	30	2600	2600	0.358	0.382
18	30	2600	2700	0.329	0.420	30	30	2600	2700	0.271	0.432
18	30	2700	2300	0.754	0.174	30	30	2700	2300	0.702	0.202
18	30	2700	2500	0.577	0.297	30	30	2700	2500	0.512	0.329
18	30	2700	2600	0.477	0.367	30	30	2700	2600	0.412	0.395
18	30	2700	2700	0.379	0.434	30	30	2700	2700	0.319	0.456

Table 8: A profile of Win & Draw probability for White against Black with specified age (18 or 30) and rating (2300, 2500, 2600 or 2700) for Rapid Chess

The following salient points emerge from a study of table 7 and 8.

- From the model parameters in table 7, it is clear that rating (current level of skill) improves

chances of winning, but not in a perfect way. Also, age (as a proxy of fatigue) is a detrimental factor. In fact, now the significance of these terms are slightly less (comparing z-statistics with the corresponding ones in table 4. Thus, in this Rapid version, the role of chance is higher.

- The above result is further illustrated by the pairwise comparisons in table 5 where a clear edge for rating and disadvantage for age is documented. E.g. look at results in each block of age pair, and within that if we consider progressive rating gap, the win probability strictly increase. Also, for the same rating pair, younger players enjoy a higher winning probability.

- From table 8 it is clear that first mover's advantage exists and is increasing with age. But, surprisingly in the Rapid version this is not related to rating. At age 18, average advantage is 3.5% which becomes 4.8% at age 30 when considering games between equally rated players. So this advantage is more equalised in Rapid chess than in the Classical case where the corresponding range was 2.2% to 6.6%. But the trend against rating is non-monotonic.

- Outcomes are certainly not foregone as even for a rating difference of 400 (2700 playing against 2300), the win probability ranges between 75% and 84% (averaging about 80%) when the higher ranked player plays white. The uncertainty is again more when older players face off against younger ones. In fact, the variability is higher in Rapid compared to Classical.

5 Concluding Remarks

Chess is considered to be one of the highest skill based table game. In this report we analyse Chess outcomes among very competent players (Rating 2300 to 2800) for both Classical and Rapid versions of the game. The win probability is strongly dependent on rating of player (positively) and age (fatigue factor making it negative). We also see progression of win probability with age and rating and distinct presence of a first mover's advantage (as documented elsewhere also).

While all these relations point towards a skill premium in Chess, the outcomes are not close to deterministic. In fact there are 20% or higher chances of adverse outcomes, which provides a starting point for future analysis regarding comparison of skill component across games.

References

1. Diganta Mukherjee & Subhamoy Maitra (2023), Unveiling the potential and scope of the Online Skill Gaming Industry: Study with technology students and professionals, report submitted to E-Gaming Federation.

2. Tomašev, N., Paquet, U., Hassabis, D., & Kramnik, V. (2020). Assessing Game Balance with AlphaZero: Exploring Alternative Rule Sets in Chess. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2009.04374

3. Prajit K. Dutta (1999), Strategies and Games: Theory and Practice, The MIT Press.

4. Diganta Mukherjee, Subhamoy Maitra, and Swagatam Das (2023), Role of Skill in the Game of Online Rummy: A Statistical Analysis, March 2023, DOI:10.13140/RG.2.2.12826.11203

5. Jeff Sonas (2002), The Sonas Rating Formula — Better Than Elo?. ChessBase.com. 22 October 2002.

6. Barthelemy, M. (2023). Statistical analysis of chess games: space control and tipping points. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2304.11425

7. Duersch, P., Lambrecht, M., & Oechssler, J. (2020). Measuring skill and chance in games. European Economic Review, 127, 103472. https://doi.org/10.1016/j.euroecorev.2020.103472

8. Rudd, D., & Mills, R. (2019). CAN YOU WIN: GAMES OF SKILL. ResearchGate.