Columbia University

Where Are the Draws:

Sorting out the result discrepancies between official chess competitions and amateur play

Abstract:

Draws in chess are an interest to many in the sport. Some consider draws inevitable as players get better, due to their belief that the best outcome any player who plays perfectly can hope to achieve if her opponent plays perfectly as well is a draw. Although many anecdotal comments and opinions about draws in chess have been talked about by fans, there has not been much academic research into draws. This paper examines the difference in draw rate between amateur and competitive chess play, looking into reasons why the results in competitive chess end up in draws much more often than amateur play. The findings show that amateur players do not draw as often as competitive chess players due to amateurs making more mistakes than competitive chess players. Further research must be done to discern the impact drawing less has on player ratings in amateur chess.

Jonathan McClement

Sports Economics Seminar

Professor Gulati

I. Introduction

In *On an Application of Set Theory to the Theory of the Game of Chess* (1913), Ernest Zermelo provided one of the first formal theorems and proofs in the academic study of games which has now become known as game theory. Zermelo tried to answer two questions: first, what does it mean to have a winning position[1], and can this position be determined mathematically; second, if a player is determined to be in a winning position, can it also be mathematically shown how many moves are needed to force a win? (Schwalbe et al. 1997). According to Schwalbe et al., the second question was of primary interest for Zermelo, with the first making up only a small portion of his paper. However, it is from the first result we obtain an interesting implication for the theory of games:

> In every 2-player combinatorial game[2] of perfect information, either player 1 has a winning strategy, player 2 has a winning strategy, or both players have a strategy which guarantee them a draw (Devos and Kent 2016).

As an example, applying this theorem to the game of Tic-Tac-Toe from the initial position, both players have a strategy to force a draw. Thus, when two players play a game of tic-tac-toe, and neither makes a strategical mistake, it will be impossible for any one player to win.

Chess is also defined as a 2-player combinatorial game and was the game of Zermelo's interest. As chess falls under Zermelo's theorem, it is known that from the initial position, a game of chess must fall under either the White player having a winning strategy, the Black player having a winning strategy, or either player being able to force a draw. The difference between chess and tic-tac-toe, and what makes chess an interesting game for those who pick it up, is that the answer to what category chess falls under,

---

[1] A player who has a winning position or strategy is said to have a sequence of moves that win a game despite what strategy his opponent takes.

[2] Devos and Kent defines a combinatorial game as follows:

> A *combinatorial game* is a 2-player game satisfying the following criteria:
> (1) A set of possible *positions* (states of the game)
> (2) A *move rule* which defines how a player can move given a position from the set defined above
> (3) A *win rule* which is a set of *terminal positions* where the game ends. Further, each terminal position must define an outcome such that Player 1 wins and Player 2 loses, Player 2 wins and Player 1 loses, or the result is a draw (no one wins or loses).

and thus a possible solution to the game, is unknown, due to the sheer number of how many possible games exist – an estimated 10^{120} possible games. To put this number into perspective, it is estimated that there are 10^{80} atoms in the observable universe; if a different game of chess was assigned to every atom, there would be so many games left over that it is still rather impossible to comprehend (Numberphile 2015).

Looking at the results of chess between top players over the past 50 years, it has been said that the probable solution to chess is the same as the solution to a solved game such as tic-tac-toe – if both players play perfectly, the best anyone of them can do is draw. According to *lichess.org*, an open-source online chess server, 43% of all tournament games (played in officially recognized competition) in its database, from what it calls master-level players, have resulted in a draw. Similarly, a study commissioned by *chessbase.com* (a company which collects and sells chess data) showed that 53.1% of official games from the last 40 years has resulted in a draw (Zhou 2018). It is from this statistical overview that the claim a probable solution to chess has been made. The number of draws has also been the ire of chess fans; known as the "draw problem," games which result in draws are considered uninteresting, and fans are looking for more wins (chessbase.com 2008). Some view this problem as inevitable. Jeff Sonas, a chess statistician, has shown that as the quality of players increases, the rate of draws does as well (2011). To complicate matters further, many top chess players utilize what is known as a "short draw:" before the game is in an official terminal position, players will agree with each other to end a game in a draw. Players have made various claims as to why they agree to short draws and why they should stay[3] (chessbase.com 2008).

Attempts have been made to make to make chess more interesting for spectators. Some tournament organizers have suggested making rule changes, such as banning the ability to agree to a draw

[3] An interesting one I've read comes from the early career of chess grandmaster Bobby Fischer, who stated after agreeing to a draw with Russian Yuri Averbakh, "I was afraid of losing to a Russian grandmaster and he was afraid of losing to a kid" (Evans 1970).

or as Clint Ballard has suggested, adjusting the points rewarded to players for certain outcomes, such as giving the White player no points for drawing, incentivizing players to play for a win (Edwards 2006).

However, the draw problem seems to be nonexistent outside of the official chess competition circuit – of over 360 million games played on the chess server *lichess.org*, only 5% have ended in a draw. Further, of games between the top players on the same server, still only 12% of games have ended in a draw. This comparison in draw rate between amateur players online and professional players in competition has resulted in the following question: if chess tends towards draws as players get better, where are all the draws in top-level amateur play?

In this paper, I will be examining chess data to see if the difference in draws between amateur and competitive play is the result of a skill differential (i.e., amateur players commit more errors) or if another factor is at play. I will use empirical data test my main hypothesis: the difference in the number of draws between top-level amateur and competitive chess is not the result of a difference in skill between the two groups. The findings show that amateur players draw less in chess due to committing more errors. In the next section, I will briefly give a background what constitutes a draw in chess, followed by a literature review, and then I will present my data, estimation, analysis, and conclusion.

II. Drawing in Chess

Fédération Internationale des Échecs (FIDE) is the internationally recognized governing body of chess, and their standards regarding chess play are recognized as the official rules of chess. These are the rules which all official competitions follow and so do many popular chess servers.

a. Winning a Chess Game

FIDE Rules of Chess Article 5.1 says that a chess game can be won only under 2 conditions:

(1) A player has checkmated her opponents' king; this means that the opponents king is being attacked and there is no legal recourse preventing or avoiding the attack.

(2) The opposing player has resigned, giving the player the win.

Besides these conditions, there are no other avenues for a player to claim victory in a standard chess match.

b. The Drawn Game

FIDE Rules of Chess Article 9 describes the scenarios under which the game is determined a draw:

(1). Stalemate: one or both players are not in check but have no legal moves on their turn. This state ends the game as a draw.

(2) Threefold Repetition: A particular position in a game as occurred three times. This repetition signifies that there has been no meaningful progress in the game by either player, ending the game in a draw.

(3) Fifty-Move Rule: If players have made 50 consecutive moves and no piece has been taken by either player, it is deemed that there has been no meaningful progress by either player, ending the game in a draw.

(4) Impossibility of Checkmate: If players do not have enough material left on the board to put each other in checkmate, the game is considered a draw.

(5) Draw by agreement: at any point in the game, a player can offer the other player a draw. If the other player accepts, the game is over and is considered a draw.

Etiquette further suggests that if the game appears to be a *theoretical draw*, meaning that it appears the game is going to be drawn as described in 1-4, that the stronger player offer a draw to the weaker player. Other practical situations include agreeing to a draw to conserve energy in a tournament, or to advance another player to the next round if she happens to be on the cusp of placing (Schiller 2003).

III. Literature Review

Despite the interest in the “draw problem” among the chess community, there are not many academic articles discussing draws in chess. Of the existing papers dealing with draws, one that particularly stands out analyses Soviet championship play between the years 1940-1978, trying to find evidence of supposed Soviet collusion.

Moul and Nye found that the success the Soviets had in chess during the years after the Second World War would happen with a 60% probability if the Soviet players acted as a cartel, compared with only a 25% probability otherwise. The supposed Soviet cartel was using the rules of chess to advance further than their opponents in tournaments – using the ability to end the game in a draw, Soviet players almost always drew against another soviet player when faced against each other in international competition. This would allow Soviet players to conserve their energy when faced with another Soviet player, while also guaranteeing that both would earn points towards advancing in the tournament. When Soviets played against non-Soviets, this tendency to draw disappeared. The Soviets were accused by other tournament contestants during this time, however there was no firm evidence that it was taking place (Moul and Nye 2007).

To perform the analysis, Moul and Nye reasoned that games that ended in a positional draw due to realistic competition between both players would last much longer than games drawn on purpose. Further, they reasoned that if one player had a much stronger position than the other and the game still ended in a draw that such a game would be an indication of the stronger player purposely throwing the game away. They also compared Soviet player performance in national competitions where only Soviets participated (URS Championships) against Soviet performance in the international FIDE tournaments. What they found was that soviets were drawing much more frequently in FIDE tournaments, at a rate of 60%, versus URS Championships, where the draw percentage was a more standard 46%.

From their dataset, they then performed an ordered probit regression to try to predict the likelihood of a draw based on whether a tournament was FIDE sponsored, whether two Soviets or two non-Soviets played each other, or if a Soviet and non-Soviet played each other (the Soviet as white). From this model, they found that the predicted probability that a Soviet drew against another Soviet in a FIDE tournament was around 14% higher than the predicted probability of a draw against a non-Soviet.

Moul and Nye took their research one step further to see if the extra draws Soviets took versus other Soviets had an appreciable impact on the outcome of a tournament. Using Monte Carlo simulations[4], they found that there was an appreciable difference in the chance a Soviet won a tournament when a drawing cartel was involved versus when one was not. For example, the Monte Carlo simulation predicted the chance Bobby Fischer would have won the Curacao tournament in 1962 to have been 25% under non-Collusion by the Soviets, versus 6% with collusion. From these simulations, Moul and Nye concluded that a Soviet cartel would have had an appreciable effect on tournament outcome if it were to have existed.

Moul and Nye recognize that there has been no firm evidence that the Soviets purposely colluded, only accusations (most notably by Bobby Fischer), and since many chess players of the era have passed away, there may never be an official answer to the collusion question. However, there is evidence that Soviet players drew more often against other Soviet players and that the increased draw rate would have had an appreciable positive effect on tournament outcome for Soviet players.

In the Chessbase research article *Has the number of draws in chess increased?* researcher Qiyu Zhou performed an analysis to determine whether chess players drew more games in recent years than they have in the past, and whether the seeming increase in the number of draws was due to the increasing amount of information available to researchers (2018). Through her research, she did not find definitive evidence based on statistical metrics that the number of draws seen in recent years was due to the amount of data which is now available. However, she did find that the percentage of games drawn has increased with the number of games played, leveling out around 50%, which may indicate a trend that, on average, given top-level players and enough matches, the percentage of games drawn will be 50%.

Another paper of note was published by Robin Hankin (2020). In his paper, Hankin adjusted the Bradley-Terry model used to predict the predict the outcome of a paired comparison for chess. Due to

[4] A Monte Carlo simulation uses repeated random sampling to determine the probability of different outcomes. The average outcome can be used to make predictions on the probability of a certain event occurring. It is named after the famous Monte Carlo casino due to a suggestion by Nicholas Metropolis, whose uncle frequently visited the locality to gamble.

chess having three possible outcomes, the standard Bradley-Terry model could not apply. However, as it is useful to find evidence of certain outcomes such as collusion, an adjustment could prove useful to future research. The essence of his adjustment involved using a three-way competition between the two standard players and an entity which won if the game was a draw. Using his adjusted model, Hankin, like Moul and Nye, found similar evidence of collusion by the Soviets in the 1962 Curacao tournament.

IV. Dataset

To perform the analysis, raw chess data was gathered and there was a need to convert it into a quantifiable form. Raw chess data is contained in a PGN file, which stands for portable game notation, a universal standard for recording chess games created by Steven Edwards in 1993. A PGN file records every move a player made during a game, along with metadata including when the game was played, where the game was played, who played in the game, their ratings, and the result. However, in this form, the data is not very useful for humans to perform analysis on. Therefore, based on an evaluation model used on *lichess.org*, a method was devised to take the raw chess data and evaluate it based on errors made, which will be discussed later.

Raw data was collected from two main sources: *lichess.org*, where amateur chess data was collected, and *pgnmentor.com*, which had a database of official chess games from competitive play dating back to 1970. Due to the overwhelming ease technology has granted chess enthusiasts in recording games played, there was a lot of data to sift through – over 10 million datapoints were available. Limitations based on computing power available made this an impractical amount of data to collect and analyze. Several criteria were used to limit the possible data set even further. To be part of the dataset, games had to exhibit several of the following qualities:

(1) each player’s rating had to be above 2200

(2) the game mode must be *rapid chess*

(3) a complete PGN file must be available

These criteria narrowed the scope of possible datapoints, and from this narrow list, a random 1200 games were chosen from amateur chess data and another 1200 games were chosen from competitive data, making the complete dataset contain 2400 games in total.

As to why these criteria were chosen, it has to do with the quality of play. A chess player's rating indicates how well players perform *relative* to their competition pool. A higher score indicates a better player. Ratings are determined on a per-game basis; every game lost reflects poorly on a player's score, and every win reflects positively. Draws generally have no effect. How much a player's score changes is based on how they compare to the other player in rating. For instance, a player with a high rating will lose more of her rating when they lose against a lower ranked player than if they were to win against the same player. Ratings are calculated by standard methods, the most popular one having been the Elo rating system designed by Arpad Elo, but in recent years new methods have been devised, such as Glicko-2 by Mark Glickman (*lichess.org*). These metrics are similar, the main difference coming from Glicko-2's use of standard deviation and confidence intervals which adjust how many points a player should gain or lose given an outcome. However, this difference is ignored as it has little appreciable effect on the rating itself.

Rapid chess is a game mode whereby chess players of a time limit between 10 and 60 minutes per player to reach the end of the game (FIDE Handbook). Each player's clock only counts down on their turn. Although a preferable dataset would contain only *classical chess* games, where players have 60 minutes or more to complete the game, there is not enough data available from high-level chess players in amateur play to make a sizeable dataset. Given the smaller time constraint, data collected is expected to have more mistakes than would otherwise be found in classical games where players would have more time to think through their moves. This should not discount the analysis, as all things being equal, the players of the amateur and competitive levels face the same constraint.

Some PGN files had to be excluded due to missing data, such as one player not having an Elo rating, or the game result was not recorded. This ended up being a higher bar than the other criteria, resulting in several fewer observations than 1200 for amateur and competitive play. The number of removed games was small and new games were not found to replace them.

Evaluation of each game was done based on a model from *lichess.org*. Using the Stockfish[5] chess engine, the difference in centipawn loss between the best move available and the move a player made was calculated for every move for each game in the dataset. Centipawn loss is a chess metric which estimates how much material (i.e., chess pieces) a player loses in making a certain move. A centipawn is regarded to be worth 1/100 of a pawn. In general, the more material that is lost, the worse of a position a player is in. Centipawn loss can also be a positive value if a player makes a move deemed to have taken material from her opponent. As it is an estimated metric, recorded centipawn loss varies between different chess evaluation engines. The *best move* is the move considered by the engine to lose the least amount of material of all moves available. By calculating the difference between the best move and the actual move, an estimation of the quality of a move is given. The greater the difference, the worse a move is considered.

After evaluating this estimation of quality for each move, a second pass over the moves made marked whether a move was an inaccuracy, mistake, or blunder. These are well known chess metrics. An inaccuracy is a move which is considered a minor error, resulting in a small but appreciable difference in centipawn loss between the best move and the move made, but overall does not affect the evaluation of a position. A mistake is considered an error with a noticeable effect on the evaluation of a position, putting a player who has made a mistake in a weaker position. A blunder is considered a major error resulting in a much weaker position, the outcome of which can lose a player the game, if it is utilized by her opponent. For this evaluation, if the difference in centipawn loss between the best move and the move made is greater than 70, it is considered an inaccuracy, if it is greater than 170, it is considered a mistake, and if it is greater than 270, it is considered a blunder. The average centipawn loss (ACL) per move was also calculated through this second pass, which is an indication of move quality on average.

[5] Stockfish is an open-source chess engine which can be found at https://stockfishchess.org. It uses several heuristics to evaluate the strength of a position and to come up with the best move for a given player in any position. For this project, Stockfish 14.1 was utilized.

After marking each move in this way, each game was quantified based on the following explanatory variables: each player's rating; each player's ACL; the number of moves made; the number of inaccuracies, mistakes, and blunders by both players; and whether the game result was a draw. Summary statistics are included in Tables 4.1, 4.2, and 4.3.

Table 4.1: Summary of Amateur Explanatory Variables

Variable	*Observations*	*Mean*	*Standard Deviation*	*Min*	*Max*
Black Rating	1,199	2398	95	2201	2824
Black ACL	1,199	12	12	0	91
White Rating	1,199	2394	95	2200	2752
White ACL	1,199	12	11	0	111
Inaccuracies	1,199	7	4	0	25
Mistakes	1,199	2	2	0	13
Blunders	1,199	3	3	0	22
Moves	1,199	80	37	3	233

The final dataset contains 2317 observations. Table 4.3 confirmed that our dataset had a mix of games which showed amateur matches drawing many times less than competitive matches. Amateur matches most likely drew more often the overall *lichess.org* statistic of 5% due to the selected pool of games only containing higher level players. For competitive games, the number of draws was slightly less than the *lichess.org* statistic of 43%, but they are of similar multitudes. From looking at the summary statistics provided in Tables 4.1 and 4.2, amateurs tend to make more errors on average than competitive players. Of particular interest, amateurs have an ACL of around 12, which is about double the ACL of competitive players. This could support the rejection of the hypothesis, as it could indicate that amateurs commit more errors which lead to less draws, as they do not play perfectly.

Table 4.2: Summary of Competitive Explanatory Variables

Variable	*Observations*	*Mean*	*Standard Deviation*	*Min*	*Max*
Black Rating	1,118	2467	128	2200	2838
Black ACL	1,118	7	9	0	119
White Rating	1,118	2469	126	2200	2745
White ACL	1,118	6	7	0	79
Inaccuracies	1,118	5	4	0	19
Mistakes	1,118	1	2	0	9
Blunders	1,118	1	2	0	22
Moves	1,118	84	34	1	249

Table 4.3: Percentage of Draws

Competition Type	*% Draws*
Amateur	13.76 %
Competitive	41.14 %

However, the difference between the two datasets appears to be small, so this cannot be taken as conclusive evidence rejecting the hypothesis.

Another point of interest in both datasets is that inaccuracies and blunders tend to be more common than mistakes. Looking at the *Max* column of Tables 4.1 and 4.2, there are games with many inaccuracies or many blunders, but games with many mistakes don't seem to exist. This may be due to the nature of the evaluation; if the mistake cutoff of 170 centipawn loss is set too high or the blunder cutoff of 270 centipawn loss set too low, errors that would be deemed mistakes by another evaluation method are instead recorded as inaccuracies and blunders. Although this may indicate that there is finetuning to be done in the evaluation method used, as this trend appears in both the datasets it should not affect the overall analysis.

One datapoint which would make the overall analysis stronger is whether a drawn game ended due to a draw by agreement or a draw due to board positioning. This may be important due to the existence of short draws – if competitive players are drawing at a higher rate due to agreeing to a draw more often than amateur players, it may indicate that more competitive matches would end with a winner if matches were to continue. Data would also be needed as to whether players believed that their game was headed toward a theoretical draw, as that may influence how may draws by agreement exist. However, given the PGN format, there is no reliable way to discern the difference between a draw by agreement or a positional draw. Thus, this datapoint was excluded from the dataset.

V. Estimation and Results

a. Estimation

A probit model is used to estimate the dependent variable *isDraw*. The model is given in Equation 5.1.

Equation 5.1: The Probit Model

$$\Pr(isDraw) = \beta_0 + \beta_1 \text{diffRating} + \beta_2 \text{acpl} + \beta_3 \text{inaccuracies} + \beta_4 \text{mistakes} + \beta_5 \text{blunders} + \beta_6 \text{moves}$$

The explanatory variables used are as follows. *diffRating* is calculated to be the absolute difference in each player’s rating. The difference in rating was chosen above including both ratings independently due to the idea that if the difference in rating is close to zero, the game should be expected to draw more often versus if the difference was great, such as 100 or more. Thus, the difference in rating between two players is a more reasonable choice as an explanatory variable versus including both ratings independently. This reasoning could be modified if the sample of games included datapoints outside of “elite” level chess players, as better chess players should draw more often, but as only players from the same skill pool are included the absolute ratings for each player are ignored. *acpl* represents the average ACL for both players during a game. The reasoning for averaging the two players instead of including them separately is like the reasoning for *diffRating*. As a higher centipawn loss indicates poorer performance, averaging the two players ACLs will give a decent indicator for overall game performance throughout any one specific match. As discussed before, *inaccuracies*, *mistakes*, and *blunders* represent the total number of errors in a match at varying levels of intensity. For similar reasons as discussed above, these are the combined totals for each error throughout a match. Finally, *move* represents how many moves were made in a game. Based on the reasoning Moul and Nye give, the more moves that are made in a game, the greater the likelihood of a draw as players lose more pieces and checkmate becomes harder to come across. A large amount of moves in a game also indicates that players are evenly matched, as one cannot get the upper hand on the other to win the game (a reason for the 50-move rule discussed in Section II.b.3).

No control variables or interaction variables were used in this model. Control variables were omitted as there are no foreseeable values which would need to be kept constant throughout the analysis. While interaction variables would appear to be necessary, as a player's rating directly relates to how many errors a player commits, due to the study limiting the dataset to only elite players it can be assumed that a player's rating would have little appreciable effect on the number of errors they commit, as errors will most likely be due to misestimation of a particular position not due to a lack of skill.

Three regressions will be run with this model. The first, the results of which will be given in Table 5.1, will be a regression run on the complete dataset (i.e., amateur data and competitive data combined. The second and third regressions, the results given in Tables 5.2 and 5.3 respectively, will run regressions either on the amateur dataset or the competitive dataset only. These regressions are of interest, as if either can predict the probability of its opposing dataset, then there is an argument that there are either too many draws, in the case of competitive chess, or not enough, in the case of amateur chess. If the predictions from these regressions are differentiated enough from the actual draw percentages found in the dataset, the hypothesis can be rejected.

b. Results

Results from the regressions can be found in Tables 5.1, 5.2, and 5.3 below.

Table 5.1: Regression 1 on Complete Dataset

Pr(isDraw)	*Coefficient*	*Robust Std. Err.*	*P-value*
diffRating	-.0026	.0004	0.000
acpl	-.0299	.0163	0.067
inaccuracies	-.0816	.0119	0.000
mistakes	-.1191	.0271	0.000
blunders	-.0865	.0375	0.021
moves	.0118	.0012	0.000
constant	-.2180	.1165	0.061

Pseudo R-Squared: .1868 Obs.: 2317

Table 5.2: Regression 2 on Amateur Dataset

Pr(isDraw)	*Coefficient*	*Robust Std. Err.*	*P-value*
diffRating	-.0027	.0006	0.000
acpl	-.0536	.0227	0.018
inaccuracies	-.0404	.0160	0.011
mistakes	-.0723	.0342	0.034
blunders	-.0611	.0358	0.088
moves	.0184	.0022	0.000
constant	-1.192	.2394	0.000

Pseudo R-Squared: .2337 Obs.: 1199

Table 5.3: Regression 3 on Competitive Dataset

Pr(isDraw)	*Coefficient*	*Robust Std. Err.*	*P-value*
diffRating	-.0012	.0005	0.026
acpl	-.0079	.0131	0.547
inaccuracies	-.1010	.0165	0.000
mistakes	-.1536	.0406	0.000
blunders	-.1360	.0656	0.038
moves	.0049	.0016	0.003
constant	.4385	.1217	0.000

Pseudo R-Squared: .1598 Obs.: 1118

From the results it can be said that most regressors are statistically significant. Moreover, the values of Pseudo R-Squared for every regression falls between .15 and .24. While these numbers would indicate that the data was a poor predictor for the outcome variable in a normal R-Squared statistic, due to Pseudo R-Squared being a logistic statistic, these numbers show that the data overall fits well within the regression. However, it is noticeable that for Regressions 1 and 3, *acpl* was not statistically significant. This is interesting because ACL should be indicator of game quality and thus should influence the outcome variable. This result may be due to many games, especially among the competitive players, having a low ACL due to high-level play, unfortunately causing problems in the regression. However, this variable shall be kept for analysis as does provide insight into the results in Regression 2. As expected, difference in rating, ACL, inaccuracies, mistakes, and blunders all have negative coefficients. This makes sense due to the intuitive cost of an error – the more errors a player makes, the more likely that a player will lose. Similarly, as reasoned before, the more moves that occur in a game, the more likely the game will end in a draw as more moves in a game can indicate an even matchup of players. Across all three regressions, the difference in rating had the most consistent coefficient, coming in at around -.002. This could indicate the quality of player rating formulas used, as the consistency across regressions suggests that evenly rated players will draw more often, showing that the players are equal in skill, the general goal for any rating system in use. It appears that inaccuracies, mistakes, and blunders of an outsized effect on the probability of a draw, with most of their coefficients across the regressions having an absolute difference of greater than .1. However, this may be due to the absolute number of these errors being less than that of other errors, such as the difference in rating, so the impact of one extra blunder is higher on average.

These regressions were then used to predict outcomes for the amateur and competitive datasets. The results of the predictions are provided in Table 5.4 below.

Table 5.4: Predicted Chances of Drawing

	Averaged % Chance of Drawing
	Regression 1
$Pr(isDraw \mid amateur)$	19%
$Pr(isDraw \mid \overline{amateur})$	36%
	Regression 2
$Pr(isDraw \mid amateur)$	14%
$Pr(isDraw \mid \overline{amateur})$	28%
	Regression 3
$Pr(isDraw \mid amateur)$	25%
$Pr(isDraw \mid \overline{amateur})$	42%

After looking at the predicted values and comparing them with the actual values in Table 4.3, each regression does not accurately predict the actual values in some way. Regression 1 overestimates the likelihood that an amateur draws, whereas it underpredicts when a competitive player draws. Regressions 1 and 2 also show similar results; Regression 2 accurately predicts the data which it is based on (amateur's chance of drawing) whereas it underpredicts the chance that a competitive player draws, and Regression 3 can predict if a competitive game will draw, but overpredicts if an amateur game will draw. From these results, the hypothesis that the difference in draw rate between amateur and competitive play is not due to a difference in the quality of play can be rejected. The regressions have shown that the few more mistakes amateur players make (as can be seen in Table 4.1 and 4.2) have an appreciable difference in a match's outcome. The regression may also indicate that amateur player's do not draw often enough, as evidenced by the large overpredictions by Regressions 1 and 3. However, there is not enough supporting evidence to make that claim confidently.

VI. Conclusion

The research conducted in this paper has rejected the hypothesis, the difference in the number of draws between top-level amateur and competitive chess is not the result of a difference in skill between the two groups, as the difference in draw rate between amateur and competitive play does seem to be impacted by the skill differential. The errors that top-level amateur players make, although slight when compared to the errors made by competitive players, has a significant impact on how many games are drawn by amateurs.

These results may have an unfortunate bearing on the "draw problem" discussed in the introduction: top-level chess players draw due to their outsized skill leading to a small likelihood of making a mistake. Beyond putting greater time pressure or other methods of drawing out mistakes onto players, there seems to be little that rule changes can do to bring about less draws in competitive play. Although the results support the continuation of the "draw problem," they also support the theory that chess falls under the "either player can force a draw" category from Zermelo's theorem. This is a hopeful result as this is further evidence that a strategy exists from the opening position which would allow any player to force a draw. Given the limitations of humans and technology, such a result most likely will not be found soon.

Three major limitations were found during the conduction of this study. The first was that there are no strict standards as to what constitutes an error in chess, only vague ideas of severity exist. Due to this, another evaluation model than the one used may prove to be more accurate in this regard. The second is that the type of draw which ended the game is not included in PGN files. This information could be important as it could give an indication as to the real effect of short draws, and whether short draws influence the ultimate game outcome. The last major limitation was that there was not enough computing power available given the length of this study to process more data, even though more data was available. Although more data shouldn't influence the outcome of this study, as the sample size was significantly large, it would have led to more precise results during the estimation phase of this study.

An interesting area of future research would be to look at whether amateur players should be taking draws more often. As most online chess servers allow players to end the game by a draw by agreement, it would be interesting to see if this ability is underutilized by amateur players and why. From a practical standpoint, as little or no change occurs to a player's rating when a game is drawn, rating maximizing players may utilize a strategy to draw to avoid loses when possible. Another interesting question is whether "white player advantage" exists. If chess falls under the same category of combinatorial games as tic-tac-toe, then "white player advantage" should not exist, as the black player has a "winning" strategy.

Although chess may theoretically be heading to the gloomy state of tic-tac-toe, the complexity of the game has at least guaranteed its playable into the near future. As this study has shown, even the best players still make mistakes – as long as that is true, chess will continue to be a game enjoyed by many people worldwide.

VII. Works Cited

DeVos, Matthew Jared, and Deborah A. Kent. *Game Theory: A Playful Introduction*. American Mathematical Society, 2016.

Edwards, Haley. "It Worked for Poker: Can Chess Make It on TV?" *The Seattle Times*, The Seattle Times Company, 10 Oct. 2006, https://www.seattletimes.com/entertainment/it-worked-for-poker-can-chess-make-it-on-tv/.

"The Great Draw Debates Summarized: Definitions, Causes, Effects." *Chess News*, 20 May 2008, https://en.chessbase.com/post/the-great-draw-debates-summarized-definitions-causes-effects.

Hooper, David, and Kenneth Whyld. *The Oxford Companion to Chess*. Oxford University Press, 1988.

"Lichess Elite Database." *Lichess.org*, https://lichess.org/team/lichess-elite-database.

Moul, Charles C., and John V.C. Nye. "Did the Soviets Collude? A Statistical Analysis of Championship Chess 1940–1978." *Journal of Economic Behavior & Organization*, vol. 70, no. 1-2, 10 Feb. 2009, pp. 10–21., https://doi.org/10.1016/j.jebo.2009.01.009.

Numberphile, director. *How Many Chess Games Are Possible? YouTube*, YouTube, 24 July 2015, https://www.youtube.com/watch?v=Km024eldY1A. Accessed 1 Dec. 2021.

"PGN Files." *PGN Mentor*, https://www.pgnmentor.com/files.html.

Schiller, Eric. *The Official Rules of Chess: Professional, Scholastic & Internet Chess Rules*. Cardoza Pub., 2003.

Schwalbe, Ulrich, and Paul Walker. "Zermelo and the Early History of Game Theory." *Games and Economic Behavior*, vol. 34, no. 1, Aug. 1997, pp. 123–137., https://doi.org/10.1006/game.2000.0794.

"Sonas: What Exactly Is the Problem?" *Chess News*, 6 Sept. 2011, https://en.chessbase.com/post/sonas-what-exactly-is-the-problem-.

Zhou, Qiyu. "Has the Number of Draws in Chess Increased?" *ChessBase*, 21 Mar. 2018, https://en.chessbase.com/post/has-the-number-of-draws-in-chess-increased.