A Statistical Analysis of Material vs. Mobility in Expert-Level Chess

Kurt S. Godden, Ph.D.

Simi Valley, CA

Abstract

Using random samples from a database of 1,696,727 expert level chess games (ELO >= 2000), I performed a statistical analysis to compare the significance of material vs. mobility in determining the outcome of a game. Five thousand randomly selected games provided baseline point estimates of these two parameters to describe the population, regardless of outcome. Then, two separate random samples of one thousand games each for wins by black and wins by white provided additional data for hypothesis tests of significance against this baseline. Finally, I calculated and compared the relative significance of material vs. mobility between the two players for wins by black and separately for wins by white. Perhaps not surprisingly, the winning side tends to have higher mean material and mobility compared to the baseline; however, the hypothesis tests indicate that this increase is not always significant. The more interesting result is found when doing a direct comparison of the winning side to the losing side. While the winning side has an advantage in both material and mobility, the significance of these advantages differs substantially when considering black wins vs. white wins.

*Keywords:* chess, material, mobility, statistics

Author Note

Please direct correspondence regarding this research to the author at kurtgodden@gmail.com.

A Statistical Analysis of Material vs. Mobility in Expert-Level Chess

The analysis presented here is an attempt to answer the question: Which is a more important factor in determining the winner of a chess game - material or mobility?

I have long suspected that mobility, i.e. the average number of moves of either player in a game, is more important in determining a game’s result than material, or the point value of the pieces excepting the King, whose Majesty exceeds finite value.

This paper describes my initial foray into analyzing both material and mobility for expert-level players (ELO >= 2000) using game data available for download at [http://www.kingbase-chess.net](http://www.kingbase-chess.net/). I downloaded 1,696,727 games from that site on November 16, 2015 and preprocessed it using J. Kunst’s *01\_pgn\_parser.R* code, using only the PGN and result data in the following analysis. https://github.com/jbkunst/chess-db

Then hit the tab key to indent and then continue typing the paper. In APA any source that you use in your paper must have an in-text citation. In APA these citations include the author’s last name and the year of the publication in parentheses (Barrett, 2002). If the source you are using does not identify an author, use a shortened version of the source title rather than the author name (“E-Portfolios for Education,” 2006).

This is your introduction. It doesn’t have a heading that says “Introduction,” but this is it. You may go on for several pages, but when you get ready for the next section, “Method,” you don’t start a new page.

By the way, APA 6.0 supports five levels of headings. They are different and more sane than previous editions of the style guide. (Enough said.) I have set the heading styles in this document to correspond with those heading levels 1 through 5. The next heading is a Heading 1. It will be followed by a Heading 2.

# Earlier Research

## Google Query: “research on chess material and mobility”

Hits:

<http://spider.sci.brooklyn.cuny.edu/~kopec/Publications/P_index.html>

<http://spider.sci.brooklyn.cuny.edu/~kopec/Publications/R_index.htm>

<https://chessprogramming.wikispaces.com/Mobility>

<https://chessprogramming.wikispaces.com/Material>

<https://books.google.com/books?id=Ao6qCAAAQBAJ&pg=PA31&lpg=PA31&dq=research+on+chess+material+and+mobility&source=bl&ots=47q3j_jZKG&sig=PDF3OdflmQNB1wxJ73xyuO5hRtM&hl=en&sa=X&ved=0ahUKEwiz3LKrjI_KAhVI62MKHQ2aDf4Q6AEIOjAF#v=onepage&q=research%20on%20chess%20material%20and%20mobility&f=false>

Both material and mobility are discussed in a famous paper by Claude Shannon (Shannon, 1950), wherein he proposes a basic evaluation function, f(P), that computes a value for a given board position. Positive values of this function indicate an advantage for white while negative values are to black’s advantage. Shannon used material values for the pieces that are still generally taught to beginners, i.e. pawns are worth one point, bishops and knights are worth three, rooks five, and queens nine.

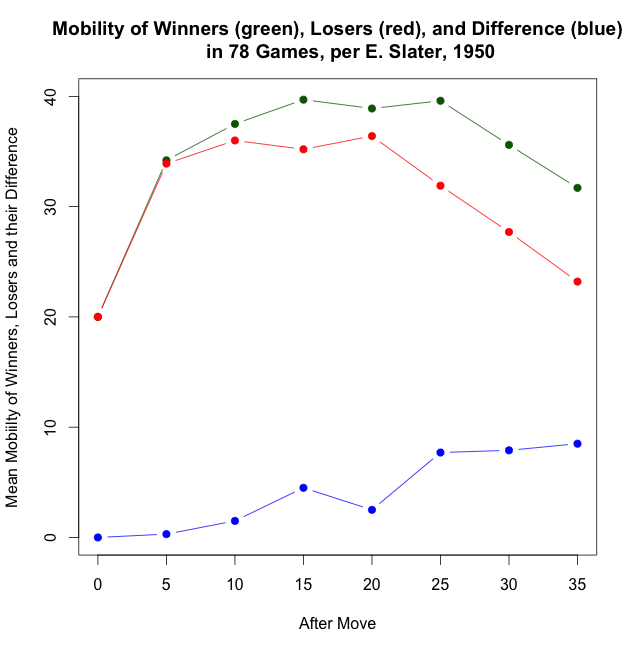
Shannon’s evaluation function summed three factors: the difference of the material values, subtracting black material from white, another factor that related to pawn structure, and a third factor for the difference between white and black mobility. The most interesting part of Shannon’s function for our present purpose is in the fact that his equation values a mobility advantage of 10 moves as equivalent to a material advantage of one pawn by virtue of the coefficient ‘1’ as the multiplier for the difference between white and black pawns, while a coefficient of 0.1 is used as the multiplier for the difference in mobility. However, Shannon does qualify these coefficient choices by saying they are “merely the writer’s rough estimate”.

The first statistical investigation of mobility appears to be by Eliot Slater in the same year as Shannon’s paper (Slater, 1950). Slater supplied two data tables, shown here as Tables 1 and 2.

|  |  |  |
| --- | --- | --- |
| **After move** | **Winners** | **Losers** |
| 0 | 20.0 | 20.0 |
| 5 | 34.2 | 33.9 |
| 10 | 37.5 | 36.0 |
| 15 | 39.7 | 35.2 |
| 20 | 38.9 | 36.4 |
| 25 | 39.6 | 31.9 |
| 30 | 35.6 | 27.7 |
| 35 | 31.7 | 23.2 |

*Table 1: Slater’s Data on Mobility*

Table 1 describes “Means taken from 78 arbitrarily selected games which ended with a decision on or before the 40th move,” and Slater goes on to say how the data indicate that mobility for both players rises as pieces are developed, falls with piece exchanges, and reveal an increasing advantage to the winner. I have created Figure 1 from Table 1 to show these trends graphically.



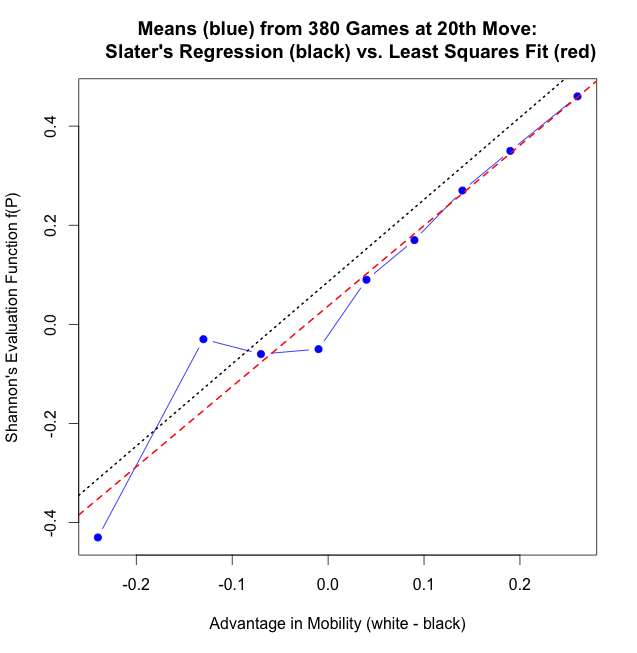
*Figure 1: Plot of Slater’s data from Table 1*

The second dataset from Slater is used to predict the values of Shannon’s evaluation function from a player’s advantage in mobility, using mean values computed from 380 games.

|  |  |
| --- | --- |
| **Mobility Advantage** | **Value of Shannon’s Function** |
| + 0.26 | + 0.46 |
| + 0.19 | + 0.35 |
| + 0.14 | + 0.27 |
| + 0.09 | + 0.17 |
| + 0.04 | + 0.09 |
| - 0.01 | - 0.05 |
| - 0.07 | - 0.06 |
| - 0.13 | - 0.03 |
| - 0.24 | - 0.43 |

*Table 2: Slater’s Data for Regression Analysis*

Slater wrote that “If graphed these points approximate to a straight line, and to the equation f(P) = +.086 + 1.658M”, where ‘M’ represents the statistic of mobility advantage. In Figure 2 I provide the graph not given by Slater, along with his regression line from this equation as well as the least squares regression line I computed in R using his data.



*Figure 2: Plot of Slater’s Regression Data*

Slater did not provide discussion into his calculations to determine the linear equation coefficients, but the equation computed by the ‘lm()’ function in R is: f(P) = +.0369 + 1.62182M. The reader will note that the least squares equation has slightly smaller coefficients for both the intercept and the slope.

# Statistical Analysis of Material and Mobility

## Analytical Methods

My first step was to establish baseline statistics for both material and mobility from a random sample of expert-level games, followed by further analysis to determine the significance, if any, of any departures from those baselines in samples taken from games ending in a win by white or a win by black.

For game data I downloaded 1,696,727 chess games in portable game notation (pgn) format from KingBase ([http://www.kingbase-chess.net](http://www.kingbase-chess.net/)). All games were played by players with ELO ratings of at least 2000, considered expert level and above by the United States Chess Federation. Computer memory limitations prevented me from computing true population means for material and mobility, so I took a random sample of 5,000 games from this dataset to compute point estimates of the population statistics. All code was written in the R statistical programming language, augmented with the *rchess* package written by Joshua Kunst, which provides R tools to load and process pgn chess data. This package is available at *The Comprehensive R Network* (<https://cran.r-project.org>). Before performing the analysis I preprocessed the entire game database using J. Kunst’s *01\_pgn\_parser.R* script available at <https://github.com/jbkunst/chess-db/tree/master/scripts>. I then wrote my own R code to analyze both material and mobility from random samples selected from this preprocessed data. The point estimates thus computed appear in Table 3.

To compute material I used the ‘standard’ values of pieces (1, 3, 3, 5, 9) as cited above for Shannon’s early research. Since at most given moves the material for either side can change via promotions and captures, I computed the total material for both white and black for every board position in a game, and then after the final position, I divided each by the number of board positions (i.e. half moves) to arrive at a *game mean material* for each player. Then each of those 5000 game means for each player were used to arrive at the population mean and standard deviation values shown here. A similar computation was done for mobility, except at each board position only the legal moves for the side to move were counted so that the *game mean mobility* for each side was the average number of legal moves for that player in the game.

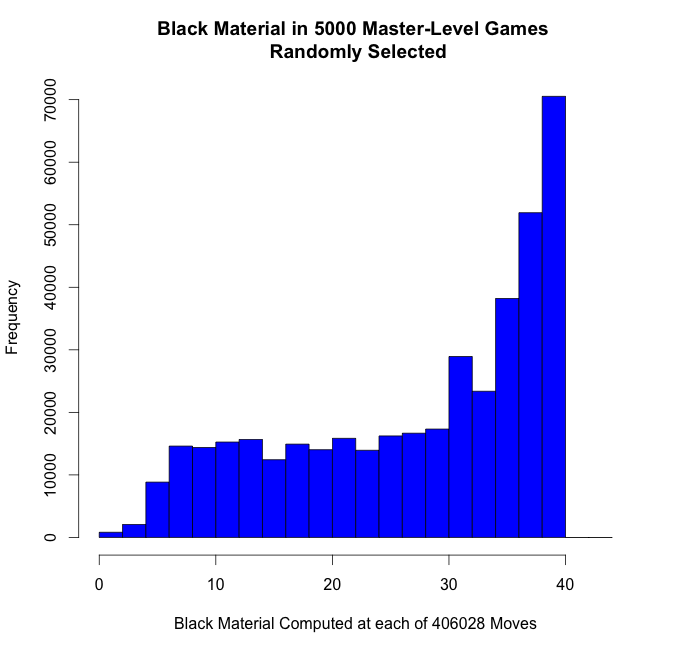
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Statistic** | **White Material** | **Black Material** | **White Mobility** | **Black Mobility** |
| **Mean** | 29.574640 | 29.553987 | 33.13604 | 30.839317 |
| **St. Dev.** | 5.650526 | 5.680532 | 4.86725 | 4.354868 |

*Table 3: Population Statistics from Point Estimates of 5000 Games*

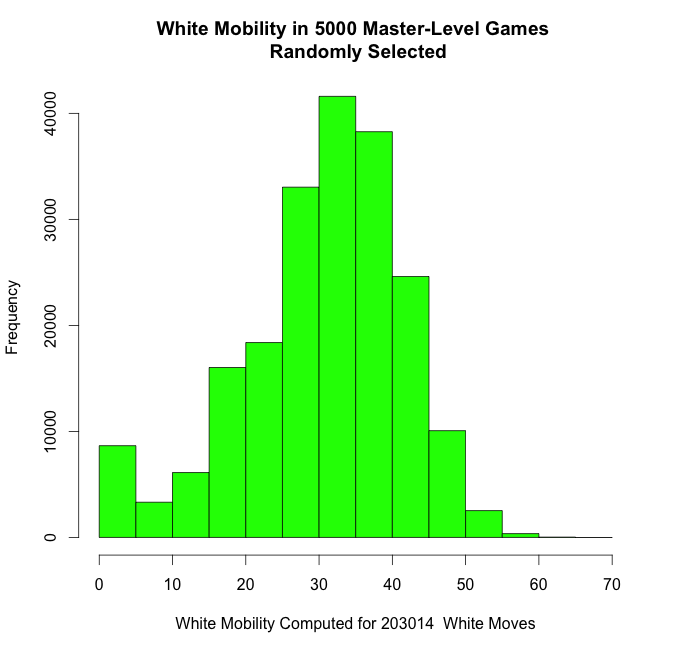
In these population estimates we see that white enjoys only a tiny advantage over black in material, but an advantage in mean mobility of almost 2.3 moves. Although the standard deviations are larger than these advantages, in the random sample there were 1821 white wins, 1388 black wins, 1790 draws and one game in the sample of 5000 had a result of ‘\*’, indicating an indeterminate outcome at the time the game data was collected.

The histograms in Figures 3 and 4 graphically illustrate the distribution of black material and white mobility using all 406,028 half-moves moves played in the 5000 games. The shapes of the histograms for white material and black mobility are not substantially different.

*Figure 3: Black Material in Population*



*Figure 4: White Mobility in Population*



## Hypothesis Tests: Departure from Population Mean when Game is Won

One may conjecture that when a game ends with one side winning, then that winning side may have game mean material/mobility values that are, on average, greater than those values in the population which were estimated in Table 3 not only from wins, but also losses and draws. We can state this belief in the form of four formal hypothesis tests, beginning with black.

Let us use the notation *B.matb*to refer to black’s mean material in games won by black and *B.matp* to represent black’s mean material in the population regardless of the outcome, i.e. the value shown in Table 3 above. Similarly, we use *B.mobb* and *B.mobp* to refer to the same concepts regarding black mobility in games won by black and black’s mobility in the population (Table 3). Thus, for our first test the null hypothesis, H0, which is assumed true unless proven false beyond a reasonable doubt, is that black’s mean material in games won by black is the same as black’s mean material in the population. The alternative hypothesis, Ha, represents our conjecture above, but cannot be accepted unless there is overwhelming evidence of its truth. That conjecture is that black’s mean material is greater in games won by black than black’s mean material in the population as a whole. Expressing these formally results in:

|  |
| --- |
| H0: B.matb = B.matp |
| Ha: B.matb > B.matp |

*Hypothesis Test 1: Black Material in Black Wins vs. the Population*

Hypothesis test 2, also for black, takes the same form except that the statistic of interest is for mobility:

|  |
| --- |
| H0: B.mobb = B.mobp |
| Ha: B.mobb > B.mobp |

*Hypothesis Test 2: Black Mobility in Black Wins vs. the Population*

The third and fourth hypothesis tests are exactly like these two, except stated for white material and mobility in the presence of white wins vs. the population:

|  |
| --- |
| H0: W.matw = W.matp |
| Ha: W.matw > W.matp |

*Hypothesis Test 3: White Material in White Wins vs. the Population*

|  |
| --- |
| H0: W.mobw = W.mobp |
| Ha: W.mobw > W.mobp |

*Hypothesis Test 4: White Mobility in White Wins vs. the Population*

To determine the results of these four tests, we need to compute point estimates of B.matb, B.mobb, W.matw, and W.mobw from a random sample of black wins and a random sample of white wins. These point estimates are among those shown in Tables 4 and 5, using sample sizes of 1,000 games.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Statistic** | **White Material** | **Black Material** | **While Mobility** | **Black Mobility** |
| **Mean** | 28.663276 | 29.038162 | 31.747328 | 32.336095 |
| **St. Dev.** | 5.214587 | 5.073105 | 4.688989 | 4.410876 |

*Table 4: Mean Material and Mobility in 1000 Black Wins*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Statistic** | **White Material** | **Black Material** | **While Mobility** | **Black Mobility** |
| **Mean** | 29.45825 | 29.142364 | 34.362652 | 29.786741 |
| **St. Dev.** | 5.34346 | 5.487446 | 4.900061 | 4.323617 |

*Table 5: Mean Material and Mobility in 1000 White Wins*

This is where I discuss the tests.

# Future Research

I want to do a regression of mobility by material for each side, and mobility of one side by the other, same for material. Could also regress one side mobility by the other side’s material. Another thing would be this: for large set of games with known outcome, mark every board position (fen string) by outcome. Some positions will get marked multiple times, and some only once. If large enough, this could be used to predict a game outcome by looking up a given position, and doing KNN classifier based on fen string distance. Would be interesting to try out. Would be best to do using games all from the same ECO, so as to maximize multiple tags for board positions.

**Heading three is the beginning of a paragraph ending with a period.** It may be best just to bold the normal style rather than using the Heading 3 style. Nulla vitae lectus sit amet justo mollis ultricies at ac enim. Maecenas ullamcorper bibendum consequat. Pellentesque ultrices, eros eu tincidunt pretium, magna leo volutpat libero, non bibendum diam nunc eget urna. Vivamus eu tortor et dui aliquam vestibulum at vel augue. Vivamus elit dui, porttitor eget egestas at, rhoncus in justo. Curabitur tristique, elit ac venenatis volutpat, eros mauris iaculis diam, vitae rhoncus erat metus vitae eros. Maecenas nunc urna, molestie nec faucibus et, imperdiet vitae velit. In tellus est, scelerisque ut aliquet et, sagittis ac justo. Sed a mi id orci congue iaculis fermentum dapibus nisi. Nullam porttitor vehicula dignissim. In ac lectus purus, id mattis metus.

**Second Heading 3 level in the section.** If you use one Heading 3, you should use a second, just like when you create an outline, if you have an “A.” you should have a “B.” Proin tempus est et felis ultrices vitae vestibulum massa tempus. Nulla at lacinia nibh. Integer vel sapien ut enim dignissim semper. Donec eleifend dui nec neque convallis vitae pretium quam ornare. Etiam hendrerit urna velit, ut consectetur risus.

***First Heading 4 level in the section.*** It may be best just to bold and italicize the normal style rather than using the Heading 4 style. Nulla congue egestas ante, id ultricies orci dignissim commodo. Fusce placerat, libero eu pharetra pulvinar, lorem dui pulvinar nisi, et semper orci orci vitae magna. Nullam sodales, felis id feugiat scelerisque, tortor nulla interdum mauris, ac porttitor odio dolor eget eros.

***Second Heading 4 level in the section.*** If you use one Heading 4, you should use a second, just like when you create an outline, if you have an “A.” you should have a “B.”Duis sit amet ipsum pretium erat accumsan iaculis vitae eget risus. Donec ut dui in lorem volutpat fermentum bibendum pulvinar libero. Nunc imperdiet eros et mi posuere pellentesque. Donec tincidunt ipsum eget nisl ullamcorper eu placerat libero ullamcorper. Maecenas id luctus ligula. Cras condimentum eleifend nibh sit amet iaculis. Suspendisse placerat sollicitudin mi, vel ornare augue hendrerit ac. Nulla sed suscipit sapien. Cras pellentesque orci lectus, eu consequat enim.

*First Heading 5 level in the section.* It may be best just to italicize the normal style rather than using the Heading 4 style. Nulla congue egestas ante, id ultricies orci dignissim commodo. Fusce placerat, libero eu pharetra pulvinar, lorem dui pulvinar nisi, et semper orci orci vitae magna. Nullam sodales, felis id feugiat scelerisque, tortor nulla interdum mauris, ac porttitor odio dolor eget eros.

*Second Heading 5 level in the section.* If you use one Heading 5, you should use a second, just like when you create an outline, if you have an “A.” you should have a “B.”Duis sit amet ipsum pretium erat accumsan iaculis vitae eget risus. Donec ut dui in lorem volutpat fermentum bibendum pulvinar libero. Nunc imperdiet eros et mi posuere pellentesque. Donec tincidunt ipsum eget nisl ullamcorper eu placerat libero ullamcorper. Maecenas id luctus ligula. Cras condimentum eleifend nibh sit amet iaculis. Suspendisse placerat sollicitudin mi, vel ornare augue hendrerit ac. Nulla sed suscipit sapien. Cras pellentesque orci lectus, eu consequat enim.

# Discussion

# Results

Maecenas id luctus ligula. Cras condimentum eleifend nibh sit amet iaculis. Suspendisse placerat sollicitudin mi, vel ornare augue hendrerit ac. Nulla sed suscipit sapien. Cras pellentesque orci lectus, eu consequat enim.

## Outcome 1

Maecenas id luctus ligula. Cras condimentum eleifend nibh sit amet iaculis. Suspendisse placerat sollicitudin mi, vel ornare augue hendrerit ac. Nulla sed suscipit sapien. Cras pellentesque orci lectus, eu consequat enim.

## Outcome 2

Maecenas id luctus ligula. Cras condimentum eleifend nibh sit amet iaculis. Suspendisse placerat sollicitudin mi, vel ornare augue hendrerit ac. Nulla sed suscipit sapien. Cras pellentesque orci lectus, eu consequat enim.

# Conclusion

Maecenas id luctus ligula. Cras condimentum eleifend nibh sit amet iaculis. Suspendisse placerat sollicitudin mi, vel ornare augue hendrerit ac. Nulla sed suscipit sapien. Cras pellentesque orci lectus, eu consequat enim.

References

Shannon, C. E. (1950). Programming a computer for playing chess. *Philosophical Magazine.* Ser. 7, Vol. 41 (314). Available online at <http://vision.unipv.it/IA1/ProgrammingaComputerforPlayingChess.pdf>

Slater, E. (1950). Statistics for the chess computer and the factor of mobility. *Proceedings of the Symposium on Information Theory.* London. Published in 1953 under the same title in *Transactions of the Institute of Radio Engineers Professional Group on Information Theory,* 150-152, 198-200. London: Ministry of Supply. Available online at <http://eliotslater.org/index.php/chess/147-statistics-for-the-chess-computer-and-the-factor-of-mobility>

Lastname, C. (2008). Title of the source without caps except Proper Nouns or: First word after colon. *The Journal or Publication Italicized and Capped*, Vol#(Issue#), Page numbers.

Lastname, O. (2010). Online journal using DOI or digital object identifier. *Main Online Journal Name*, Vol#(Issue#), 159-192. doi: 10.1000/182

Lastname, W. (2009). If there is no DOI use the URL of the main website referenced. *Article Without DOI Reference*, Vol#(Issue#), 166-212. Retrieved from <http://www.mainwebsite.org>

For more information and example related to references, visit the presentation   
*Writing In Style* at:

<https://www.sworps.utk.edu/training/APA_6_0/player.html>