

Risk Aversion and Gender in Chess

Charlie Mayville

November 5, 2022

Question

There is a lot of economic literature about the effect of gender on risk aversion, most studies pointing to an increased risk aversion among women.¹ An interesting question is whether the selection effect that occurs in male dominated spaces lends itself to any preference towards personality that may be measurable via risk aversion. Chess is a good candidate for this study since it is easily quantifiable and extensively recorded.

I intend to follow a 2010 study by Gerdes and Gränsmark² to determine the causal effect of gender on playing risky chess. In the study they use a simple OLS model, which could be replaced with a more robust statistical method learned in this class.

Dataset and Methods

ChessTempo³ has a free database of games with pgn, rating, and FIDE id. Cross-checking this against FIDE's⁴ list of rated players gives us gender along with all the covariates used in Gerdes and Gränsmark with a handful of others. That study uses only the top 100 female and male chess players. Selecting with some cutoff only top players seems like a reasonable thing to do.

Measuring risky play is a bit trickier. Gerdes and Gränsmark ask a handful of players above master level to measure whether a opening position is “solid” or “aggressive”. To me this seems utterly absurd. First of all depending on players not incredibly far from my own skill level to help evaluate some of the best players in the world feels silly. But more importantly, most of the game is played after the opening. While some openings may really lead to solid games no matter what (like the Grünfeld), it is very unclear

Instead of doing this, I intend to attempt to measure risky play as how much a move increased the probability of a draw, all else held equal. Modern chess engines that run on neural networks evaluate positions based on WDL likelihood, which makes this easier to measure.

$$RiskScore = \sum 1 - \min(\Delta win\ chance, \Delta draw\ chance)$$

1. Gary Charness and Uri Gneezy, “Strong Evidence for Gender Differences in Risk Taking,” *Gender Differences in Risk Aversion and Competition*, *Journal of Economic Behavior and Organization* 83, no. 1 (2012): 50–58, ISSN: 0167-2681, <https://doi.org/https://doi.org/10.1016/j.jebo.2011.06.007>, <https://www.sciencedirect.com/science/article/pii/S0167268111001521>.

2. Christer Gerdes and Patrik Gränsmark, “Strategic behavior across gender: A comparison of female and male expert chess players,” *European Association of Labour Economists 21st annual conference*, Tallinn, Estonia, 10-12 September 2009, *Labour Economics* 17, no. 5 (2010): 766–775, ISSN: 0927-5371, <https://doi.org/https://doi.org/10.1016/j.labeco.2010.04.013>, <https://www.sciencedirect.com/science/article/pii/S0927537110000588>.

3. “Chess Games Database,” 2022, <https://old.chesstempo.com/game-database.html>.

4. “FIDE Ratings Database,” 2022, https://ratings.fide.com/download_lists.phtml.

I have written a Perl script, given a list of pgn's, to calculate this score using Lc0⁵, an engine far above human strength on virtually any hardware. Running it on a few sample games seems to give a positive gut-check. If it does not work out I can always fall back on Gerdes and Gränsmark's method.

Possible Problems

The first problem I could foresee with this project is my estimator of risk not capturing what I expect it to measure. The tricky part about risky play is it has more to do with intent than with the technical evaluation of the resulting position. Using only high rated games could control this problem, as it is conceivable that the better you are the more likely your intent to be realized in the position, but this could only be a loose corollary, especially in positions complicated for humans to evaluate.

The second issue I am not quite sure how to address is how to adjust for how poor of a measurement Elo can often be. Younger players often gain 200 to 300 rating points in one year as their play matures; older players with titles may hit rating floors; and players often come off of hot streaks or tough patches. Players may choose different strategies against players perceived to be above or below their rating, or players less confident. There does not seem to be a good way to adjust for this—using age as a confounder helps—but I would really love to adjust for Elo K-factor, but I cannot find a nice dataset that contains it.

5. “Leela Chess Zero,” 2022, <https://lczero.org/>.

References

- Charness, Gary, and Uri Gneezy. “Strong Evidence for Gender Differences in Risk Taking.” *Gender Differences in Risk Aversion and Competition, Journal of Economic Behavior and Organization* 83, no. 1 (2012): 50–58. ISSN: 0167-2681. <https://doi.org/https://doi.org/10.1016/j.jebo.2011.06.007>. <https://www.sciencedirect.com/science/article/pii/S0167268111001521>.
- “Chess Games Database,” 2022. <https://old.chesstempo.com/game-database.html>.
- Dilmaghani, Maryam. “A matter of time: Gender, time constraint, and risk taking among the chess elite.” *Economics Letters* 208 (2021): 110085. ISSN: 0165-1765. <https://doi.org/https://doi.org/10.1016/j.econlet.2021.110085>. <https://www.sciencedirect.com/science/article/pii/S0165176521003621>.
- “FIDE Ratings Database,” 2022. https://ratings.fide.com/download_lists.phtml.
- Gerdes, Christer, and Patrik Gränsmark. “Strategic behavior across gender: A comparison of female and male expert chess players.” European Association of Labour Economists 21st annual conference, Tallinn, Estonia, 10-12 September 2009, *Labour Economics* 17, no. 5 (2010): 766–775. ISSN: 0927-5371. <https://doi.org/https://doi.org/10.1016/j.labeco.2010.04.013>. <https://www.sciencedirect.com/science/article/pii/S0927537110000588>.
- “Leela Chess Zero,” 2022. <https://lczero.org/>.