

Prediction Modeling of International Figure Skating Scores

Statistics 98

Diana Zhu

Harvard University

2022/04/19

Intro to the Research Topic

My final project seeks to analyze figure skating competition data from the past Olympic cycle in order to predict the 2022 World Championship results and assess potential judging bias.

Research Question

- What are the most significant predictors that determine a skater's total score and rank, and how can we use this to predict the 2022 World Championships?
- Do skaters receive a home-town advantage in scores when competing in their home country?

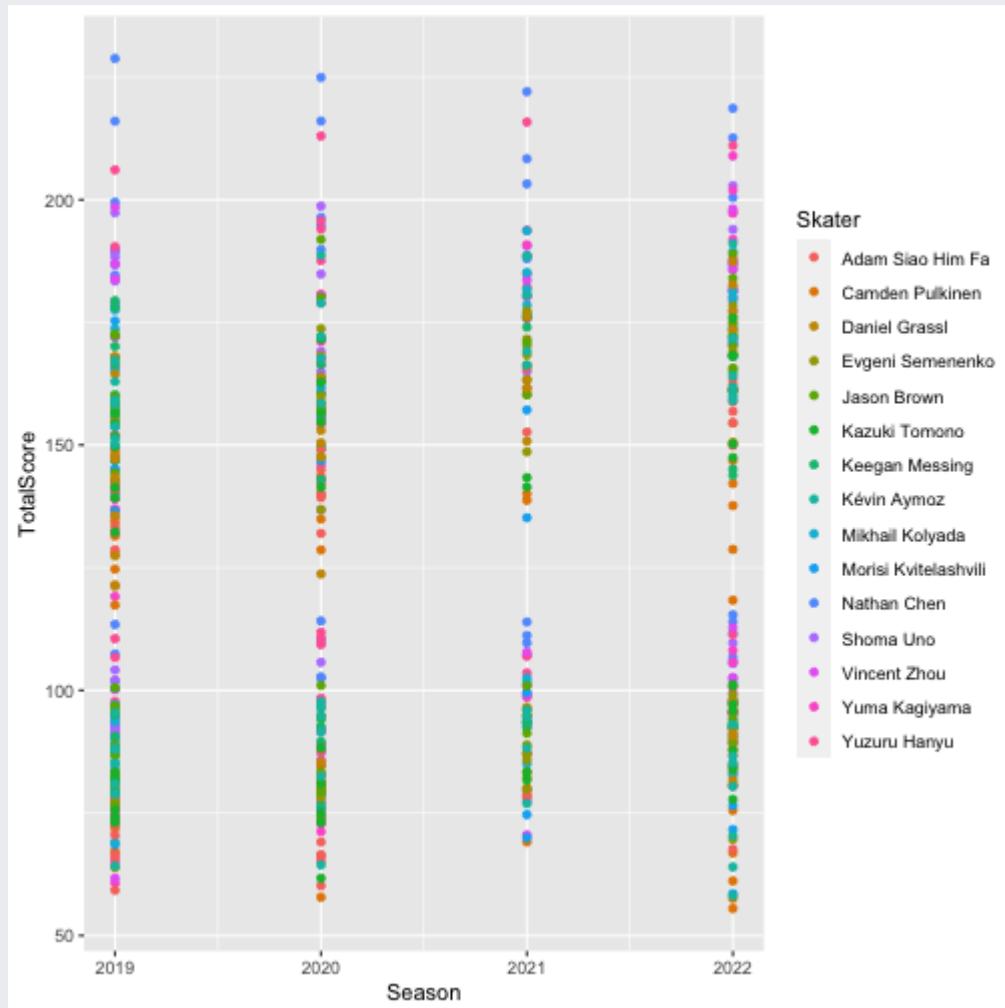
Data Collection Process

I scraped data using the `rvest` package from skatingscores.com, which uploads scoring protocols from all international and major domestic competitions under the IJS (International Judging System).

- I chose skaters who earned the top 15 free skate scores at the past two senior world championships (2022 and 2021).
- Selected skaters who competed at the world championships because oftentimes large countries like Russia will have several top scoring skaters, but each country can only send a maximum of three skaters to the world championships.
- Due to the volatile nature of the sport and short-lived seasons of skaters, I chose skaters who have been performing well in the past two years.

Due to the Russia-Ukraine conflict in recent months, the International Skating Union banned Russian athletes from participating in any international figure skating competitions on or after March 1, 2022, including the 2022 World Championships.

Initial EDA



Modeling Plan

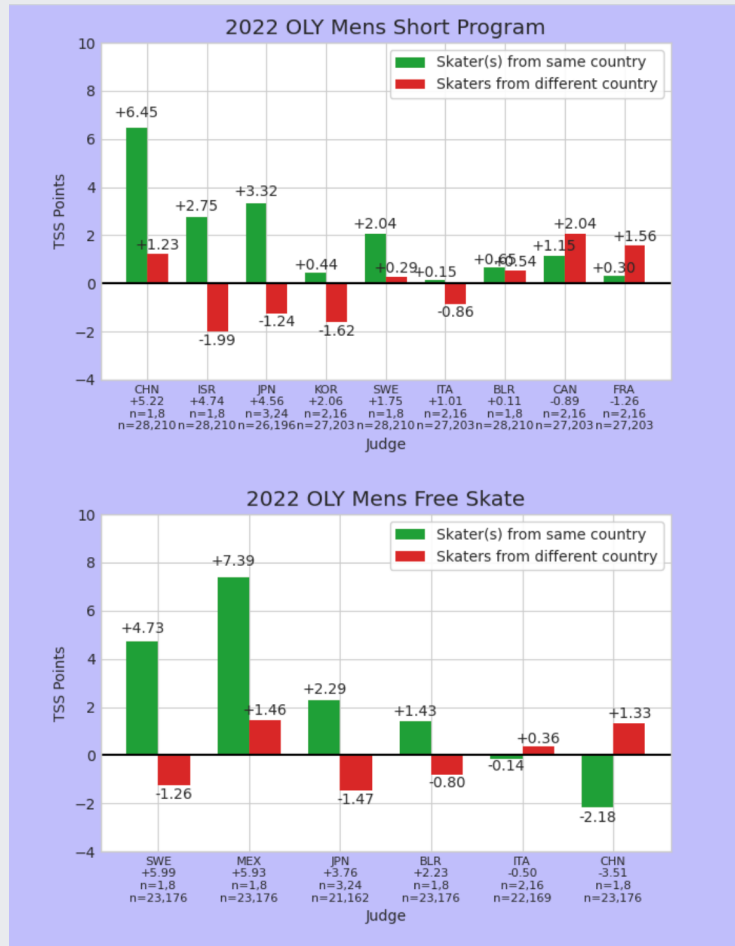
Prediction

- OLS: outcome is the final score, predictors: technical segment score, program component score
- mixed effects model

Judging Bias

- Use Spearman's Rank Correlation to assess how judges' rankings line up with the final skater rankings
- Hypothesis testing for a difference in scores given to skaters of their own country vs foreign skaters

Next Steps and Questions



- Hypothesis: There is more variance between scores in singles skating than in ice dance and pairs. Run t-tests for significance on TSS and mean GOE.
- Plan to scrape the actual dates for the events so the data can follow a more continuous time scale