Utilizing Correlational Analysis to Identify Traits of Successful Forecasters

Hannah Liu[1]; Cory Clark[2]; Igor Grossmann[3]; Philip Tetlock[4]

[1]University of Pennsylvania CAS '25, [2]University of Pennsylvania Dept. of Psychology, [3]University of Waterloo Canada, [4]University of Pennsylvania Dept. of Psychology CAS & Wharton

Funded by Grants for Faculty Mentoring Undergraduate Research

Abstract

The Human Progress Forecasting Tournament is a two-year longitudinal study that was designed to help identify traits of individuals who were capable of successfully predicting future outcomes regarding a set of four domains: non-state conflicts, US poverty rate, global infant mortality, and global CO2 concentrations. During Year 1 of the tournament, participants first made predictions by themselves for each domain. After making predictions for all four domains, they were then able to see other participants' predictions and rationales, and participants had the opportunity to change their answers if they chose to. Utilizing correlational analyses between variables derived from the tournament dataset, we were able to statistically infer two conclusions. First, we found no significant relationship between the amount of change in participants' Year 1 predictions between individual and team stages and prediction error. These findings were also supported by an auxiliary correlational analysis between average change across all predictions within a domain and prediction error, which yielded no significant correlation. Second, we found a small significant positive correlation between the size of the range of estimates and prediction error, indicating that an increase in estimate ranges is related to an increase in prediction inaccuracy. Because more change in predictions between the two stages might indicate open-mindedness to external opinions, and a larger range of estimates might indicate flexibility to multiple possible outcomes, we expected there to be a significant relationship between prediction change and prediction error and for a negative significant relationship between range and prediction error. However, we found opposite patterns, and future research is needed to understand these results.

Introduction

There are individuals in this world-whether they are ordinary people, superforecasters, or experts in a specific field-who excel at making extremely accurate predictions about future events. However, what makes these individuals the way they are? If we are able to identify indicative traits of highly accurate forecasters, it will aid in society's capability to make predictions about the future and decisions regarding those predictions. This research question became a topic of interest for our team, and we became interested in investigating what sort of characteristic traits and behaviors were related to these excellent predictors. Prior research has suggested that forecasters with intellectual humility and data-driven predictions were the most accurate. To investigate our research question, we created a longitudinal forecasting tournament taking place over 2 years that had participants make predictions about the future of human progress, awarding monetary incentives for answers that were accurate and precise.

Methods and Materials

Participants were asked to make predictions about four different domains: non-state conflicts, US poverty rate, global infant mortality, and global CO2 concentrations. For each individual, we collected their lowest estimate, best estimate, and highest estimate for three different time points for each domain. For each year, there were two stages. In Stage 1, participants made independent predictions. In Stage 2, participants had a chance to change their estimates after seeing other people's predictions and rationales. For this study, we focus on Year 1 estimates. Using the tournament dataset, we wanted to identify what prediction behaviors the successful individuals in this tournament share.

We computed the amount that Year 1 predictions changed between Stage 1 and Stage 2 and the range of each individual's Stage 2 estimates. We also computed the absolute error between an individual's Stage 2 prediction and the actualized outcome one year later, with higher error values indicating more inaccurate predictions. After obtaining these numbers, we ran a correlational analysis utilizing Python between the change in prediction variable and the absolute error variable, and between the range variable and the absolute error variable.

Results

We generally found no significant correlation between the amount of prediction change and prediction error, rs = 0.00 to 0.01 (see Figure 1). However, we found small to medium correlations between the range of estimates and prediction error, rs = 0.15 to 0.30, thus indicating that larger ranges are positively associated with more inaccurate predictions (see Figure 2).

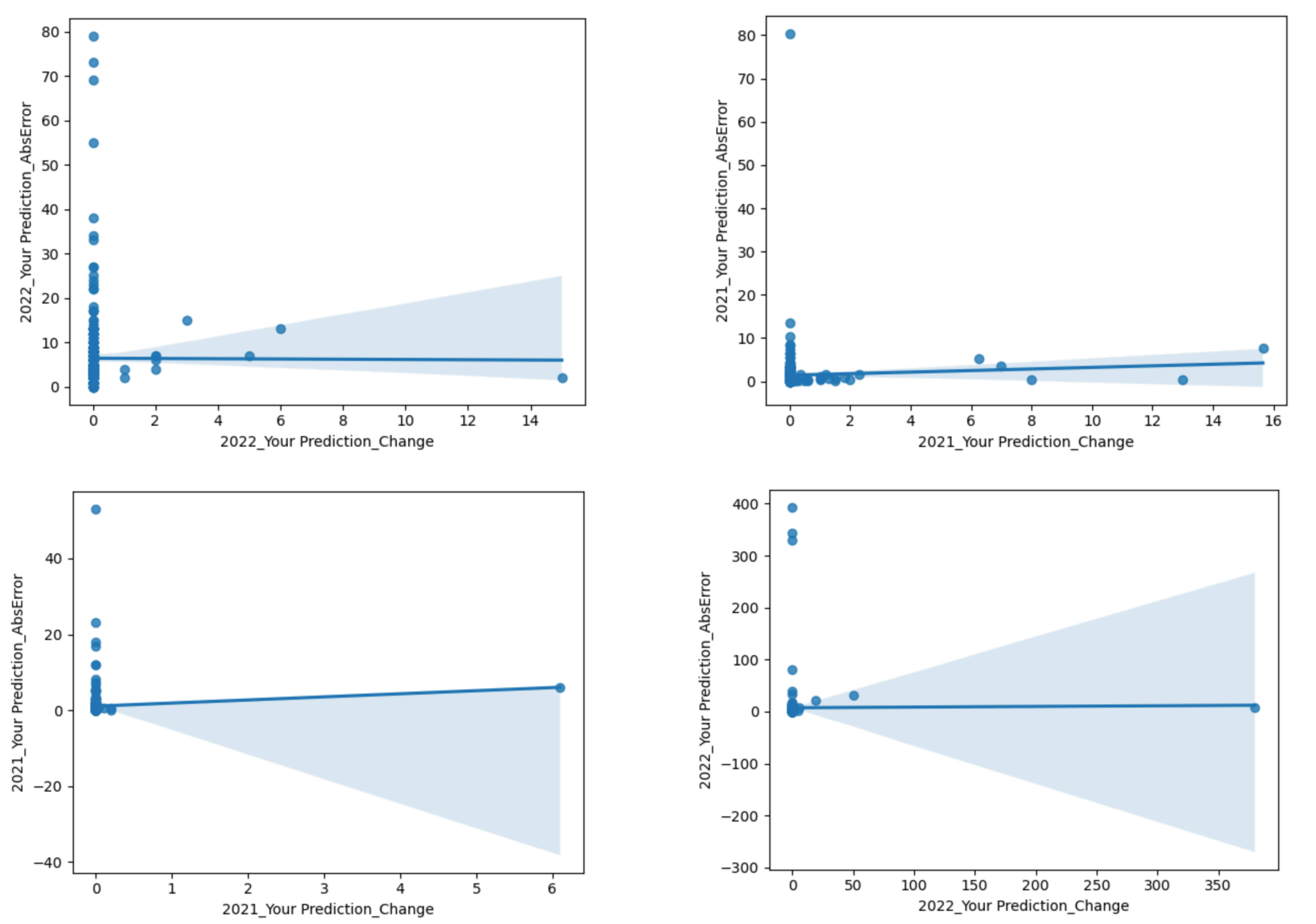

Figure 1. Scatterplots for Change in Prediction v.s Absolute Error Correlation. Top left is Non-State Conflicts, top right is US Poverty Rate, bottom left is Global Infant Mortality, bottom right is Global CO2 Concentrations

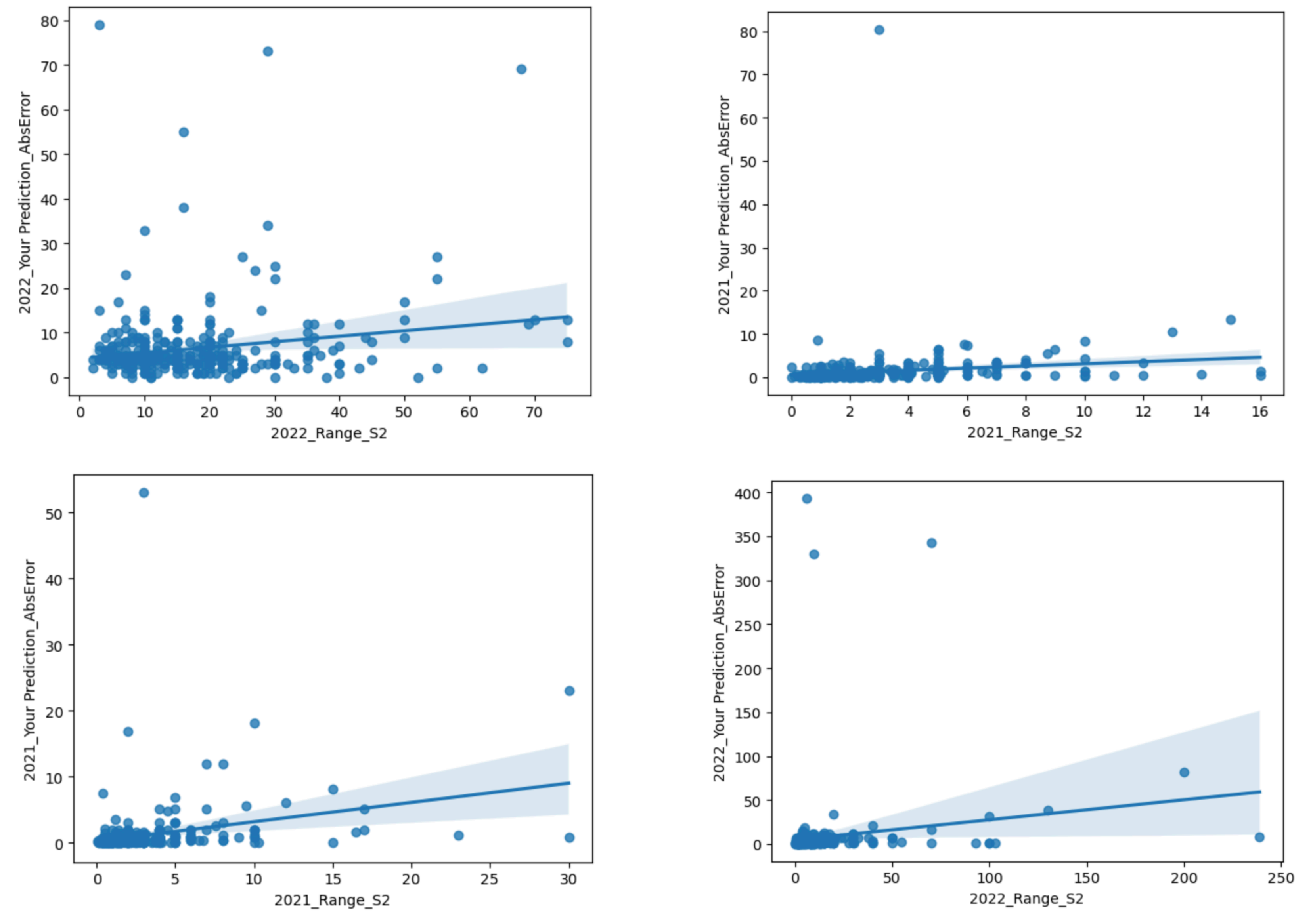

Figure 2. Scatterplots for Range in Estimates v.s Absolute Error Correlation. Top left is Non-State Conflicts, top right is US Poverty Rate, bottom left is Global Infant Mortality, bottom right is Global CO2 Concentrations

Discussion

As seen by the correlation coefficients above, we can observe two results. First, we can see no significant correlation between individuals changing their predictions between Stage 1 and Stage 2 and absolute error of individuals' Stage 2 predictions. We can interpret this as a change in predictions after the team mode stage having little to no relationship with accuracy. As secondary analysis to confirm these (non)relationships, we also tested the relationships between the average change across all time points within a domain and the absolute error of individuals' Stage 2 predictions and similarly found no significant associations, rs = -.02 to .02. This suggests that changing one's answer after being exposed to additional external opinions does not affect a prediction's accuracy, which conflicts with our initial hypothesis of how a willingness to change tends to lead to more accuracy.

Second, we can see a slightly significant positive correlation between an individual's range of estimates during Stage 2 and their absolute error with their Stage 2 prediction. We can interpret this as the size of an individual's range of estimates having a positive relationship with error. Again, this conflicts with our initial hypothesis, since we assumed that a larger range will encompass more values, thus increasing the chances that the actualized outcome will be included in the range.

Conclusions

We created a 2-year long forecasting tournament that tested participants' ability to predict future outcomes regarding domains related to human progress to help us understand what kinds of characteristic traits successful predictors have. From our correlational analyses, we observed two main patterns.

First, we found that belief updating was not a reliable predictor of forecasting accuracy. Initially, we assumed that belief updating could be an indicator of intellectual humility, and therefore thought that belief updating might predict accuracy. However, we observed no relationship between these variables. One possible explanation for this observation is that few participants updated their beliefs, and thus we had little variation to test for relationships here. Participants may have been hyperconfident in their initial answers or were not sufficiently engaged with the research materials. In the future, we can test this explanation by making it harder for participants to keep the same answer (ie. they have to manually type in their answer) or forcing participants to change their answers in team mode. This ensures that individuals are not simply clicking through the survey mindlessly or overconfidently, which can help increase the number of occurrences that predictions change in our data.

Second, we found that range of estimates was associated with more inaccurate predictions. Initially, we assumed that larger answer ranges could indicate one's flexibility to multiple outcomes, and therefore would help with accuracy. However, our results showed an opposite scenario. One possible explanation for this observation is that individuals with larger ranges were generally more "clueless" or did not care enough to aim for precision. In the future, we can use observations from qualitative data to see if individuals with larger answer ranges reflect these traits in their comments to help prove this explanation.

Along with executing additional correlational analyses on variables computed from the tournament dataset to confirm our previous findings and discover new correlations, our future research will also dive deeper into the implications of our initial findings. Because our findings in this study did not align with what our team initially assumed as common sense, more discussions will be had to understand how these findings could potentially change how we think about forecasting.

References

1. Ville A. Satopää, Marat Salikhov, Philip E. Tetlock, Barbara Mellers (2021) Bias, Information, Noise: The BIN Model of Forecasting. Management Science 67(12): 7599-7618. https://doi.org/10.1287/mnsc.2020.3882
2. The Forecasting Collaborative. Insights into the accuracy of social scientists' forecasts of societal change. *Nat Hum Behav* 7, 484–501 (2023). https://doi.org/10.1038/s41562-022-01517-1