How does training reduce miscalibration? nsights from the Good Judgment Project

Vahid Karimi Motahhar, Thomas Gruca

Tippie College of Business, University of Iowa

Tippie College of Business, The University of Iowa

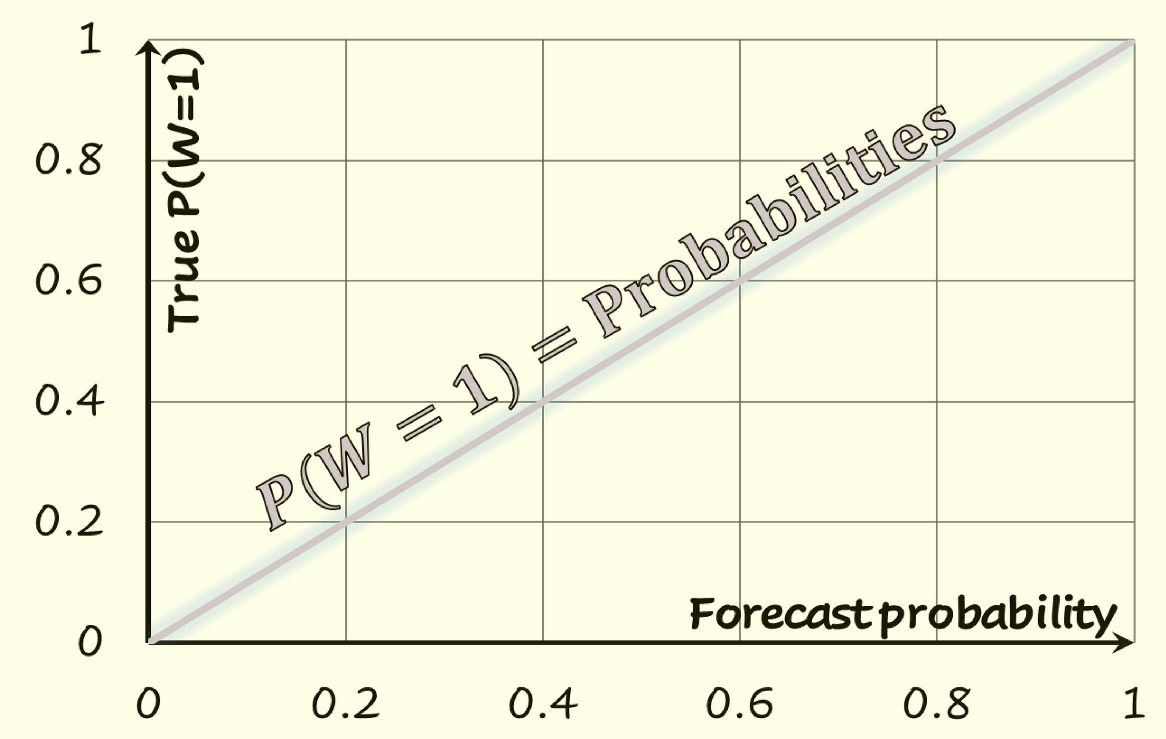
Society for Judgment and Decision Making

Introduction

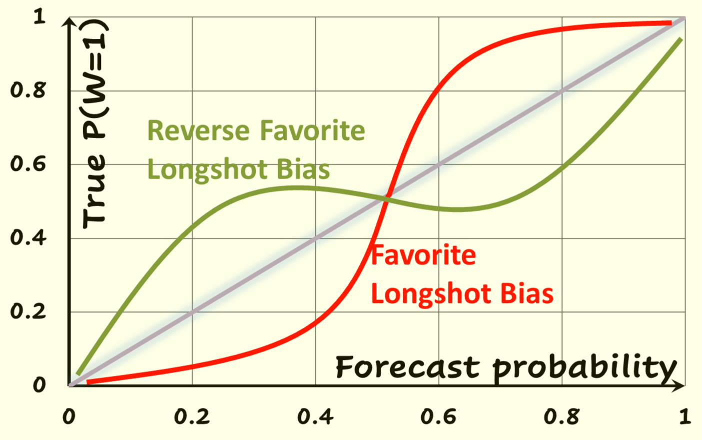
Zoom Meeting: https://uiowa.zoom.us/j/98646196961

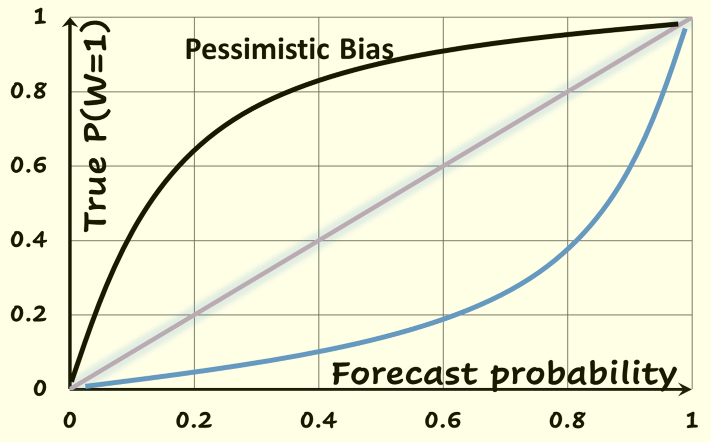
- ***Probability estimation training*** **has been shown to help individuals to produce better forecasts:**
  - **It reduces forecast error, i.e., the Brier Score** (Mellers etl al. 2014, Moore et al. 2017).
  - **It reduces the miscalibration component of the Brier Score = forecaster's average absolute divergence from the "well-calibrated" 45-degree line** (Chang et al. 2016).

- **Set of well-calibrated forecasts:**



- **Research Question: Consider these two types of miscalibration biases (Left: Compensatory, Right: Non-compensatory). How impactful is training in reducing either of these biases?**





Overview of Data and Methods

- **Publicly available data set called the "Good Judgment Project", winner of a multi-year geopolitical forecasting tournament.**
- **An example of forecasted question:**
  - **Will Greece remain a member of the EU through 1 June 2012?**
- **We explore biases among 307 individual forecasters (with at least 22 forecasts) who randomly were assigned to either go through a probability training (n=137) or not (n=170).**
- **Initial forecasts in the first year of the tournament are studied.**
- **Logit model** (Berg and Rietz 2019)
  - **Dependent Variable = Actual outcome (0 or 1),**
  - **Independent Variables: Log ratio transformation of the forecast, training (1: gone through training or 0:otherwise)**

Logit Method: $\frac{P(W=1)}{1-P(W=1)} = \exp\left(b_0 + b_1 * \log\frac{F}{1-F} + b_2 * training + b_3 * training * \log\frac{F}{1-F}\right)$

Notes:
- Log ratio of forecasts as an IV
- Dummy training IV
- Individual differences are controlled

$$\frac{P(W=1)}{1-P(W=1)} = \exp\left((b_0 + b_2 * training) + (b_1 + (b_3 * training))\log\frac{F}{1-F}\right)$$

Case of No Bias: $True\ P(W=1) = F\ (Forecast)$

Estimation greater than 0: Pessimistic Bias
Estimation less than 0: Optimistic Bias

Estimation greater than 1: Favorite Longshot Bias
Estimation less than 1: Reverse Favorite Longshot Bias

Results

Parameter	**Coefficient (Std. Error)**	**Bias Detection**
Intercept (Null: b_0=0)	**-1.18 (0.27) *****	*Non-compensatory type*: Optimistic bias
Training (versus non-trained)	0.55 (0.43)	
Forecasts' Log ratio (Null: $b_1 = 1$)	**0.54 (0.22) ****	*Compensatory type*: Reverse Favorite-Longshot bias (marginally helped by probability training)
Forecasts' Log ratio * training	**0.65 (0.39) ***	
McFadden pseudo-r2	0.166	

Conclusions

- **We use a novel model to detect both *non-compensatory* and *compensatory* biases in a set of repeated probabilistic forecasts.**
- **Probability estimation training helps by (marginally) reducing the extent of the reverse favorite longshot bias.**

Important References

- Berg, J. E., & Rietz, T. A. (2019). Longshots, overconfidence and efficiency on the Iowa Electronic Market. *International Journal of Forecasting*, *35*(1), 271-287.
- Chang, W., Chen, E., Mellers, B., & Tetlock, P. (2016). Developing expert political judgment: The impact of training and practice on judgmental accuracy in geopolitical forecasting tournaments. Judgment & Decision Making, 11(5).
- Mellers, B., Ungar, L., Baron, J., Ramos, J., Gurcay, B., Fincher, K., ... & Tetlock, P. E. (2014). Psychological strategies for winning a geopolitical forecasting tournament. *Psychological science*, *25*(5), 1106-1115.
- Moore, D. A., Swift, S. A., Minster, A., Mellers, B., Ungar, L., Tetlock, P., ... & Tenney, E. R. (2017). Confidence calibration in a multiyear geopolitical forecasting competition. *Management Science*, *63*(11), 3552-3565.