

Report of the CAS Working Party on Superforecasting

David L. Ruhm, FCAS and Liza Wong, ACAS

*CAS Annual Meeting
Orlando, Florida
November 2016*

Overview

- Superforecasting by Tetlock & Gardner (2015)
 - Chronicles research into “the science of prediction”
 - Must-read for all actuaries: buy the book
- CAS Working Party (2016)
 - Actuaries doing prediction science activities & learning
- Today’s report
 - Book highlights, to introduce topic (a small part of what the book covers)
 - WP activities and future plans, invitation to join

Book discussion

- Superforecasting by Tetlock & Gardner (2015)
 - Introduces new phrases & terminology, some mentioned in this presentation
- Philip Tetlock
 - Wharton professor, psychology and political science
 - Government began research into forecasting
 - Tetlock's methods were enormously successful
- Dan Gardner
 - Journalist, author of books on related topics

Everyone forecasts

- Forecasting is part of daily life
- Some forecasts are easy
 - Driving / meetings / restaurant hours / astronomical events
- Some forecasts are very difficult
 - Rainfall in NYC 4 months from today
- Most important questions are in-between

Political pundits

- Widely followed, widely respected
- Lots of opinions and post-hoc explanations
- Predictions not rigorously measured
- “Forecasts” usually very flexible & can’t be measured by outcomes

Scientific test of pundits and academics

- Experiment: Pundits and academics made thousands of forecasts
- Rigorous measurement was applied
 - Specific future event
 - Specific timeframe
 - Subjects asked to estimate probability of event (not what will happen)
- Result: Predictions were no better than a dart-throwing chimp

How did we do?

- The Experiment
 - Before the session started, we asked everyone to estimate the weight of an object. The rule was not to discuss or share your estimate with others.
- What was mean square error?

Superforecasters

- Top 2% of volunteer forecasters, using rigorous measurement
- Ordinary people from diverse professions
 - Very numerate / comfortable with numbers / sense of probabilities
 - News junkies
- Usually have no knowledge about a question when asked to forecast

Researcher view: The optimistic skeptic

- Important questions usually have some mix of predictability and unpredictability
- Hypothesis: Set “accuracy” as the goal, measure well, and we can learn how to get better at forecasting
- Nat’l Security: 20,000 US intelligence analysts forecasting every day
 - Getting better at forecasting, and learning how to improve, are vital

Forecasting Science – Milestones

- 2002: Intelligence community concludes Iraq stockpiling WMDs
- 2006: IARPA begins (Intelligence Advanced Research Projects Activity)
 - Analogous to successful DARPA predecessor
- 2008: NRC committee synthesizes “good judgment” research
- 2010: IARPA sponsors 4-year tournament
 - Who can invent the best forecasting methods for intelligence questions

NRC Committee – research synthesis

- Analytical methods require evidence that they are actually effective
 - Intuitive appeal is not enough by itself to support using a method
 - As in medicine for new medicines and procedures
- Rigorously test current & proposed analysis methods
 - Realistic conditions
 - Measure accuracy of forecasts with metrics (Brier score)
- Promote continuous learning through the testing process – necessary

NRC Committee – research synthesis

- “Meaningful accountability”
 - Track forecasting accuracy scientifically
 - Hold intelligence community accountable for overall forecasting accuracy
 - Don’t blame when something bad happens (some forecasts will be off)
- Checklists for organizing analysis procedures – pros and cons
 - May be useful for ensuring methods implemented consistently
 - Don’t measure results
 - Don’t measure whether correctly-implemented methods actually work

IARPA tournament

- Who can invent the best forecasting methods, for questions like:
 - Will Tunisia's president flee into exile within the next month?
 - Will the euro drop below US\$1.20 within the next 12 months?
 - Will H5N1 influenza kill > 10 people in China within the next 6 months?
- Well-defined questions are essential / no ambiguity
 - Not easy but not impossible / stated briefly
- Researchers formed teams, intelligence community also participated
 - Tetlock's team "The Good Judgment Project"

The Good Judgment Project Team

- 3,200 applicants answered ads & passed screening process
- Team required to make forecasts on many IARPA questions
- Top 40 forecasters identified by results after first group of questions
- Team's average consensus forecasts were enhanced:
 - Forecasts by the top 40 were given more weight
 - "Extremization" applied (for example, 40% → 30%, 65% → 75%)

Lessons from competition

- Tetlock's team won IARPA's forecasting tournament
 - Sept 2011 – June 2015
 - Beat U Michigan, MIT and intelligence analyst teams, by a lot
 - Analysts had access to classified information
- Important key findings from the competition:
 - Some people are consistently better forecasters ("superforecasters")
 - Forecasters can improve their forecasting ability

The most accurate forecasters: skill or luck?

- Great coin-flippers are lucky
 - Different people each year – reversion to the mean
- Superforecasters are mostly the same people each year
 - 70% retention rate from one year to the next / improvement over time
 - More accurate 300 days out than other forecasters 100 days out
 - Best one beat the extremized consensus and a prediction market (by a lot)
- Extremization: Those with more information enter a more certain (extreme) forecast. Simulates whole team being as well-informed.

Profile of a top superforecaster

- Retired IBM programmer
- Reads NYTimes regularly
- Knows world geography
- Answered over 100 questions in first year (~1,000 total forecasts)
- Beat the team's extremized consensus and a prediction market

What makes a superforecaster (top 2%)?

- Superior results appear to come from differences in:
 - What they do (processes)
 - How they think (reasoning)
- Some key personal characteristics:
 - Committed to self-improvement
 - Open-minded
 - Curious
 - Careful
 - Smart / Often not geniuses
 - More about how they use their intelligence

“Fermi-ize”: Take it apart, analyze the pieces

- How many piano tuners in Chicago?
- Most people just guess
- Better results by estimating (model) pieces, then combining:
 - How many pianos?
 - How often?
 - How long for one piano?
 - How many hours / tuner in a year?

Start with base rates, then use the details

- The Robinson family lives in a small house in a US suburb:
 - Bookkeeper & homemaker, 5-year-old son, widowed grandmother
- How likely is it that the Robinsons have a pet?
- Usual approach: Think about picture, guess what feels right
- More effective approach: What % of US households have a pet?
 - Then refine with details later – if they’re actually relevant
 - “Outside” followed by “Inside”: first the world, then the unique features

Improving your view's accuracy

- Synthesize the outside and inside Views
 - Essentially Bayesian, without using math
 - Judgment-based weighting – an art of forecasting skill
- Add others' logical arguments, but ignore detritus
 - Value facts
 - Value good reasoning / synthesis of facts
 - Discount fluff
- Ask other people to critique your logic

Improving your view's accuracy

- Assume you're wrong, and ask "How does that happen?"
 - Make another estimate based on that scenario
 - Combine with first estimate
- Re-word the question, and you'll get a new angle, a new view:
 - "Will South Africa grant the Dalai Lama a visa within 6 months?"
 - Think of: Reasons they might want to grant a visa
 - "Will South Africa deny the Dalai Lama a visa for 6 months?"
 - Think of: Why would they want to deny a visa?
 - New ideas / different thinking kicks in

How did we do?

- The Experiment
 - The second experiment: we asked everyone to estimate the weight of another object but encouraged everyone to discuss, and share your reasons with others.
- What was mean square error?

Superforecasters synthesize: “dragonfly eye”

- Dragonfly’s compound eye: many views, wide angle, whole picture
 - Brain synthesizes together into a perspective
- Synthesize many valid possible views
 - As many as you can get, if valid and add something new
 - Mentally demanding
- Superforecasters enjoy this:
 - Love puzzles / curious about the subject / very interesting & fun

Actively open-minded

- Beliefs are hypotheses to be tested
 - We tend to treasure our beliefs and guard them – the opposite
- Question an emotionally-charged belief, rationally
- Look for people and evidence that challenge your beliefs
- Changing your mind isn't a weakness (flexibility / search for truth)
 - Goes against usual behavior & institutional political dynamics

Qualitatively numerate

- All superforecasters are comfortable with numbers
 - Nuanced sense of probabilities
- Forecasting doesn't use complex math
- Numeracy and sense of probabilities informs subjective judgment
- Good at assigning relevance to information subjectively
 - Bayesian, without the math (even those who know the math)

Dial-thinking vs. shades of "maybe"

- "Dial thinking" – 3 settings:
 - "It's going to happen"
 - "It might happen"
 - "Everything's 50-50: it will happen or it won't"
 - "50-50" is a common default conclusion, generating inaccuracy
 - "It isn't going to happen"
 - Tell someone "80% confidence level" and they'll expect it
- Better approach: Accept that uncertainty exists
 - Everything is "maybe" with a probability

Granular probability estimates

- The probability is the whole story
 - Boils down all the information into a reality-based result
- Granularity (1%) predicts accuracy
 - More granular forecasters are better forecasters
 - Rounding to nearest 5% reduces supers' accuracies
- Make your probability estimates as granular as you can justify

Update your estimate frequently

- Very important to performance
- Superforecasters update much more often than others
 - Small adjustments (subjective-Bayesian)
- Monitor the news carefully
- Identify relevant, subtle information accurately
 - Don't react to irrelevant noise
- Don't overreact or underreact – evaluating correctly as % is a major art

Watch out for common biases

- Superforecasters are aware of main cognitive biases
 - Sensitive to their internal objectivity level – introspective
- Asking a substitute question
 - Usually an easier question (audience: search Sir Joshua Reynolds expedient)
 - Emotionally-charged question, e.g., polonium question
- Being “married” to a belief (“belief perseverance”)
 - “It can't be true” / “I need it to be true” / “Everything I believe in...”
 - Already told everyone it's true

Growth mindset vs. fixed mindset

- Fixed mindset is common / leads to discouragement, failure
 - Tutors often hear: "I'm not good at math" – fixed mindset
- Growth mindset is effective, and essential for superforecasters
 - "I can grow if I work hard enough and learn along the way."
- Use the growth mindset
 - Actuaries and exams – probably both pre-existing and reinforced

Teams improve forecast accuracy

- Teams were much more accurate than individuals
 - Diversity of perspectives is a major driver
 - Superforecasters did even better when on a team
- Robust & respectful debates
 - Precision questioning, logical walk-through
- Split up research work
 - Each team member gets more data than they could gather on their own
- Culture of sharing / common purpose / givers / variety of perspectives

Additional key concepts

- Confidence and humility are about context
 - "I'm good, but reality's very challenging"
- Plan for surprises – adaptability, Taleb's anti-fragility engineering
 - Relevant for actuarial work
- If it's Pareto instead of Normal, "1-in-10,000" might become 1-in-100
 - Wealth distribution / other phenomena?
- Desired results and power (politics) affect forecasts & accuracy
 - Relevant for actuaries / However, bias can increase utility-scored "accuracy"

Computers and forecasting

- Deep Blue (Jeopardy), chess-playing computers
- Statistical algorithms usually beat experts' judgment at prediction
- Future: Blended computational-subjective forecasting
 - Computer-generated forecasts / human override
 - Other variations possible

CAS Working Party on Superforecasting

- Launched in Spring 2016 by Jessica Leong
 - Approximately 25 members
 - Includes a Good Judgment Project member
- Teams and forecasting competitions
 - E.g., total US medals at Rio Summer games
- Current questions have Jan 2017 completion
- The WP currently is accepting new members
 - Contact Dave Core at the CAS (e-mail is on website)

Summary

- We hope you found this interesting and learned something new
- Prediction Science is new, exciting, and here to stay
 - Complements Data Science as a separate discipline
- Get the book ([Superforecasting](#), Tetlock / Gardner) and enjoy it
 - Probably relevant for most actuaries, across practice areas and societies
- Join us to explore forecasting science
 - Be ready to participate on a team
- What are some topics you would like to see us forecasting?

And now, it's time for...

Questions

Thanks to our audience today

Thank you for attending!