The Inference-Forecast Gap in Belief Updating

Authors: Tony Q. Fan, Stanford; Yucheng Liang, CMU; Cameron Peng, LSE

Reviewer: Pedro Vogt, PUC-Rio

October 17, 2023

A Brief Experiment

- There is a pool of 20 firms.
- The pool has 10 Good Firms and 10 Bad Firms.
- Good Firms have an average monthly profit of USD 100.
- Bad Firms have an average monthly profit of USD 0.

A Brief Experiment

- The distribution of monthly profits for Good and Bad Firms is given by:

Good Firms							
Monthly Profit (USD)	25	50	75	100	125	150	175
Probability (%)	4%	8%	16%	44%	16%	8%	4%

Bad Firms							
Monthly Profit (USD)	-75	-50	-25	0	25	50	75
Probability (%)	4%	8%	16%	44%	16%	8%	4%

A Brief Experiment

- We have randomly chosen Firm X from the pool
  - What is the (unconditional) probability that Firm X is Good?
  - What is the expected value of its monthly profit?

- We now observed a monthly profit of USD 75 for Firm X
  - Conditional on observing this profit, what is the probability that Firm X is Good?
  - Conditional on observing this profit, what is the expected value of its next month profit?

Agents Seem to Make Mistakes when Forming Expectations

- Ample research on agents failing to apply Bayes' rule correctly when processing new information.

- In some cases, individuals tend to **overreact** to recent news when asked to make **forecasts**
  - Overreaction has been linked to excess volatility (Barberis et al, 2015), boom-bust cycles (Maxted, 2020) and other puzzles

- Another strand of literature shows that agents **underreact** to signals when **inffering** underlying states
  - Underreaction may drive momentum, post-earnings announcement drift and slow response to macro conditions

What Drives Under and Overreaction?

- What makes people underreach in some contexts and overreact in others?
- Is there a tension between those two literatures? Can they be reconciled?
- Hard to say in practice as we do not know the true data-generating processes.

Paper Conducts a Simple Experiment in Belief Updating

- Paper runs a series of online experiments
- Participantes are given a more general version of the above problem (always with a known DGP)
- Given an observed signal, they are asked to:
  - Update their belief about the underlying state - the likelihood of firm being Good (”inference problem”)
  - Update their belief about future outcomes that depend on the state - the expected profit next month (”forecast-revision problem”)

Inference and Forecast-Revision Problems are Closely Tied

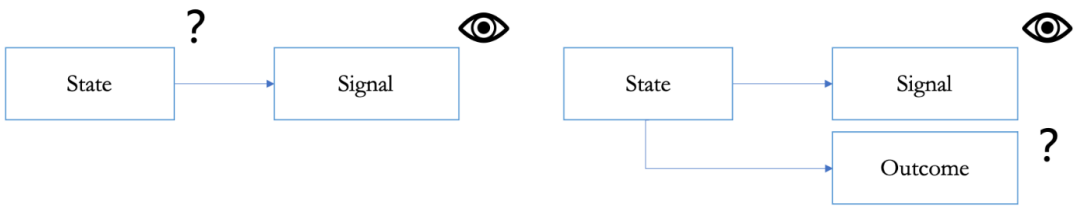

Figure 1: Inference problem (left) and forecast-revision problem (right)

Notes: In an inference problem, people observe a signal and then update their beliefs about the underlying states. In a forecast-revision problem, people revise their forecasts about outcomes in response to a realized signal.

However, Paper Finds a Big Difference in How Participants Approach Each Problem

- Participants underreact to signals when making inferences - 61% of underreactions against 24% of overreactions
- And overreact when revising forecasts - 40% of underreactions against 54% of overreactions
- They define this discrepancy in belief updating as an “inference-forecast gap”
- The gap may be a result of different decision procedures or heuristics selected for each problem

Roadmap of Talk

Experiment

Results

Decision Procedures

Mechanisms

External Validity

Conclusion

The Baseline Problem

- A firm is randomly drawn from a pool of 20
- The firm's (unknown) state θ is either G (Good) or B (Bad)
- The composition of the pool is known - which pins down $Pr(\theta = G)$
- The signal variable s_t is framed as either the stock price or profit growth in month t
- Signals from the Good Firms are drawn from $N(100, \sigma^2)$
- Signals from the Bad Firms are drawn from $N(0, \sigma^2)$
- Participants are given s_0 and asked to update beliefs over the state - $Pr(\theta = G|s_0)$ - and next month's signal forecast - $E(s_1|s_0)$

In This Setting the Inference and Forecast-Revision Problems are Tightly Linked

- Optimal Update to Inference Problem:

$$Pr(\theta = G|s_0) = \frac{Pr(\theta = G) * Pr(s_0|\theta = G)}{Pr(\theta = G) * Pr(s_0|\theta = G) + Pr(\theta = B) * Pr(s_0|\theta = B)} \quad (1)$$

- Optimal Update to Forecast-Revision Problem:

$$\mathbf{E}(s_1|s_0) = Pr(\theta = G|s_0) * \mathbf{E}(s_1|\theta = G, s_0) + Pr(\theta = B|s_0) * \mathbf{E}(s_1|\theta = B, s_0) \quad (2)$$

- The "no inference-forecast gap" condition:

$$\underbrace{\mathbf{E}(s_1|s_0)}_{\text{Forecast}} = \underbrace{Pr(\theta = G|s_0)}_{\text{Inference}} * 100 \quad (3)$$

Experiment Design

- Baseline experiment was conducted with 279 participants through Prolific (online social science platform)
- Signals were framed as monthly revenue growth for 142 participants and as stock price growth for 137 participants
- Each participant went through 8 rounds of problems
- In each round, the DGP is fully specified by $Pr(\theta = G)$ and σ

Table 2: Parameter values for DGPs

Index	1	2	3	4	5	6	7	8
$\Pr(G)$	50%	50%	50%	50%	50%	50%	80%	20%
σ	50	60	70	80	90	100	100	100

Experiment Design

- Chance of receiving a USD 5 bonus payment
- Comprehension questions to rule out misunderstandings over the DGP

Answers are Then Compared to Rational Update

- Rational update is defined according to equatios (1) and (2)

$$\text{update} = \begin{cases} \text{answer} - \text{prior}, & \text{if } s_0 > 50 \\ \text{prior} - \text{answer}, & \text{if } s_0 < 50 \end{cases} .$$

- **Underreaction**: update is smaller than the rational update by more than 2.5
- **Overreaction**: update is larger than the rational update by more than 2.5
- **Near-ration**: update within 2.5 of rational update

Roadmap of Talk

Experiment

Results

Decision Procedures

Mechanisms

External Validity

Conclusion

Inference-Forecast Gap Under Baseline

Table 4: Aggregate patterns in *Baseline*

N=279, Obs.=2144	Classification			Update
	Underreaction	Near-rational	Overreaction	Mean (s.e.)
Inference	60.8%	15.2%	24.1%	14.3 (.7)
Forecast Revision	39.7%	6.4%	53.9%	32.7 (2)
Rational				23.3 (0.3)

Notes: The first three columns present the percentages of answers classified as Underreaction, Near-rational, and Overreaction. The last column shows the average belief movement from the (objective) prior to the posterior, as well as the rational benchmark. Observations with the signal equal to 50 are excluded. Standard errors are clustered by participant.

Results are Robust to Different Treatments

Table 3: Overview of additional treatments

Treatment	Section	Key differences from *Baseline*
Deterministic Outcome	3.2	Forecast outcome is a different variable (100 if $\theta = G$ and 0 if $\theta = B$)
Binary Signal	3.3	Signals are binary; forecast questions ask about full distributions
Nudge	4.1	Beliefs about states and forecasts are elicited on the same page
More Similar	5.1.2	State variable (profitability) = mean of signal or forecast outcome (profits); inference questions ask about the expectation of the state
Less Similar	5.1.3	Forecast outcome is a different variable (up if $\theta = G$ and down if $\theta = B$); forecast questions ask about full distributions

Similar Results with Deterministic Outcomes

Table 5: Aggregate patterns in *Deterministic Outcome*

N=100, Obs.=777	Classification			Update
	Underreaction	Near-rational	Overreaction	Mean (s.e.)
Inference	64.4%	14.8%	20.8%	13.4 (1.3)
Forecast Revision	39.9%	8.6%	51.5%	34.1 (3.5)
Rational				23.1 (.4)

Notes: The first three columns present the percentages of answers classified as Underreaction, Near-rational, and Overreaction. The last column shows average belief movements in the signal direction from the (objective) priors and their rational benchmark. Observations with the signal equal to 50 are excluded. Standard errors are clustered by participant.

Similar Results with Binary Signals

Table 7: Aggregate patterns in *Binary Signal*

N=140, Obs.=1120	Classification			Update
	Underreaction	Near-rational	Overreaction	Mean (s.e.)
Inference	61.0%	20.1%	18.9%	11.0 (0.9)
Forecast Revision	54.9%	6.7%	38.4%	14.2 (2.2)
Rational				18.7 (0.0)

Notes: The first three columns present the percentages of answers classified as Underreaction, Near-rational, and Overreaction. The last column shows average belief movements in the signal direction from the (objective) priors and their rational benchmark. The updates of forecast-revision answers are normalized by $Pr(\text{up}|G) - Pr(\text{up}|B)$ so that they are comparable to the inference updates. Standard errors are clustered by participant.

Roadmap of Talk

Experiment

Results

Decision Procedures

Mechanisms

External Validity

Conclusion

Are Participants Applying the Same Procedure to Both Problems?

- Inference-forecast gap should not arise if participants solved the forecast-revision problem by:
  - 1. First updating their beliefs about the states (thus answering the inference problem)
  - 2. Then correctly applying the LIE to form expectations about the forecast outcome
- Could participants be trying to apply this procedure but failing to do so?
  - Does not appear so
  - Nudges do not reduce the gap
  - Update biases on inference are weakly correlated with those in forecast-revision

Complexity-Reducing Nudges Have no Effect on the Inference-Forecast Gap

Table 7: Aggregate patterns in *Binary Signal*

N=140, Obs.=1120	Classification			Update
	Underreaction	Near-rational	Overreaction	Mean (s.e.)
Inference	61.0%	20.1%	18.9%	11.0 (0.9)
Forecast Revision	54.9%	6.7%	38.4%	14.2 (2.2)
Rational				18.7 (0.0)

Notes: The first three columns present the percentages of answers classified as Underreaction, Near-rational, and Overreaction. The last column shows average belief movements in the signal direction from the (objective) priors and their rational benchmark. The updates of forecast-revision answers are normalized by $Pr(\text{up}|G) - Pr(\text{up}|B)$ so that they are comparable to the inference updates. Standard errors are clustered by participant.

Problems Elicit Different Modal Responses

Table 10: Modes of behavior in *Baseline*

Mode	Criterion for answer	*Inference*	*Forecast Revision*
Non-update	$=$ prior	29.7%	21.9%
Exact representativeness	$=100$ if $s_0 > 50$, $=0$ if $s_0 < 50$	2.6%	20.3%
Naive extrapolation	$= s_0$	3.2%	11.9%
No inference-forecast gap (excluding the other modes)	inference $=$ forecast revision	3.3%	
Unclassified		61.8%	45.2%
Observations		2144	2144

Notes: The column "Criterion for answer" shows the criterion for an answer to be classified into a mode. Note that an answer may be classified into more than one mode. The percentages in the last two columns are the fractions of answers in each mode in *Inference* and *Forecast Revision*. Observations with the signal equal to 50 are excluded.

Roadmap of Talk

First Proposed Mechanism - Similarity

Hypothesis: The choice of heuristics is affected by the similarity between salient cues in the information environment and the variable to be updated

Table 11: Similarity between belief-updating questions and cues in *Baseline*

Cue	*Inference* $\Pr(\text{state}\vert\text{realized price})$	*Forecast Revision* $\mathbb{E}(\text{price}\vert\text{realized price})$	Behavior
$\mathbb{E}(\text{price}\vert\text{representative state})$	Not similar	Similar	Exact representativeness
Realized price	Not similar	Similar	Naive extrapolation
$\mathbb{E}(\text{price})$		Similar	Non-update
$\Pr(\text{state})$	Similar		Non-update

Alternative Treatment - More Similar

- The signal and the outcome variable are reframed as the firm's profit in the current month and in the next month, respectively
- The state variable is framed as the firm's profitability, defined as the long-run average of the firm's monthly profit, taking on values of either 0 or 100

Table 12: Similarity between belief-updating questions and cues in *More Similar*

Cue	*Inference* $\mathbb{E}$(profitability|realized profit)	*Forecast Revision* $\mathbb{E}$(profit|realized profit)	Behavior
$\mathbb{E}$(profit|representative state)	Similar	Similar	Exact representativeness
Realized profit	Similar	Similar	Naive extrapolation
$\mathbb{E}$(profit)		Similar	Non-update
$\mathbb{E}$(profitability)	Similar		Non-update

Increasing Similarity does Change Modal Procedures Applied to Inference

Table 13: Modes of behavior in *More Similar*

Mode	*Inference*	*Forecast Revision*
Non-update	36.3%	28.9%
Exact representativeness	15.3%	18.9%
Naive extrapolation	18.3%	26.9%
No inference-forecast Gap (excluding the other modes)	4.0%	
Unclassified	29.2%	25.5%
Observations	655	655

Notes: The criterion for an answer to be classified into a mode is the same as in Table 10. The percentages are the fractions of answers in each mode. Observations with the signal equal to 50 are excluded.

Second Proposed Mechanism - Timing

Hypothesis: the relative timing between the realization of states, signals, and outcomes may play a role in the higher prevalence of overreaction-inducing heuristics (such as naive extrapolation) in Forecast Revision than in Inference, thus contributing to the inference-forecast gap

Alternative Treatment - Timing

- Participants now follow a randomly chosen firm for two consecutive months, labeled the “first month” and the “second month.”
- Participants are randomized into two different conditions, the Future condition and the Past condition
- In the Past condition, participants observe the firm’s stock price growth in the second month as a signal and are asked to revise their expectation for the first month

Increasing Similarity does Change Modal Procedures Applied to Inference

Table 19: Modes of behavior in *Timing*

	Future Condition		*Past* Condition	
Mode	*Inference*	*Forecast Revision*	*Inference*	*"Precast" Revision*
Non-update	28.3%	14.0%	28.2%	14.4%
Exact representativeness	3.2%	24.5%	1.3%	21.2%
Naive extrapolation	3.4%	13.2%	3.3%	5.7%
No inference-forecast Gap (excluding the other modes)	2.6%		4.1%	
Unclassified	62.6%	48.9%	63.1%	57.0%
Observations	470	470	458	458

Notes: We separately report results from the *Future* condition and the *Past* condition. The criterion for an answer to be classified into a mode is the same as in Table 10. The percentages are the fractions of answers in each mode.

Roadmap of Talk

Experiment

Results

Decision Procedures

Mechanisms

External Validity

Conclusion

GDP Growth Rate Forecasts

- Some evidence of naive-extrapolation (overreaction inducing) and non-update (underreaction inducing) for GDP growth estimates by professional forecasters

$$w_{i,t} := \frac{f_{i,t}y_{t+1} - f_{i,t-1}y_{t+1}}{y_t - f_{i,t-1}y_{t+1}},$$

- Where y_t is GDP growth in time t and $f_{i,t}y_{t+n}$ is the forecast by individual i in time t for GDP growth in time $t + n$

- Naive-extrapolation implies $w_{i,t} = 1$ as $f_{i,t}y_{t+1} = y_t$

- Non-update implies $w_{i,t} = 0$ as $f_{i,t}y_{t+1} = f_{i,t-1}y_{t+1}$

Increasing Similarity does Change Modal Procedures Applied to Inference

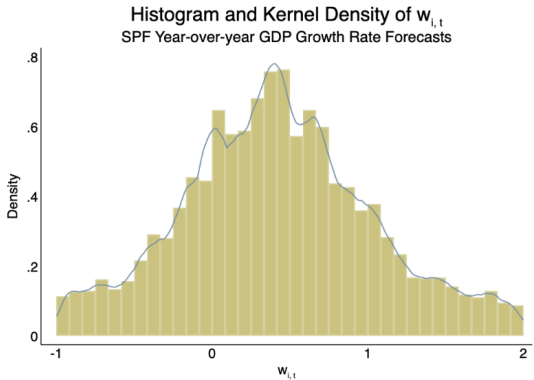

Figure 5: Histogram and kernel density for the weight on the most recent realization (relative to the prior forecast) in the *Survey of Professional Forecasters* (SPF) annual GDP growth rate forecasts, from 1968:Q4 to 2022:Q4. The width of both the bars and the kernel is $\frac{1}{12}$.

Roadmap of Talk

Experiment

Results

Decision Procedures

Mechanisms

External Validity

Conclusion

Conclusion

- Interesting experiment with all the standard caveats for this type of study
- Not clear if the similarity hypothesis is necessarily linked to the inference versus forecast-revision separation. Are inference problems always more similar to cues from the information environment in ways that elicit procedures that tend to underreact?

The End

Thanks!