Building Central Bank Credibility: The Role of Forecast Performance

Michael McMahon\(^1\) and Ryan Rholes\(^2\)

\(^1\)University of Oxford, CEPR and CfM
Irish Fiscal Advisory Council → Usual Disclaimer Applies

\(^2\)University of Oxford
Motivation

- Managing expectations is crucial for inflation-targeting framework
  - Effective communication requires credibility
  - Little is known in practice about the determinants, dynamics of central bank forecast credibility

**This Paper:** Use a large online experiment to study how features of forecast history impact the central bank’s forecast credibility

**We Consider:**

- Forecast Performance: How does overall forecast performance influence credibility?
- *Timing:* Does the timing of forecast errors matter for a central bank’s forecast credibility?
- *Communication:* Can central banks ‘talk their way out’ of a low-credibility position?
Preview of Results

Contributions

1. Relationship between performance and updating is flatter than theory predicts
   ▶ Under-punish consistently poor performance
   ▶ Under-reward excellent performance
   ▶ Over-precision/Under-precision

2. Timing of errors matters a lot - recent performance is key.
3. Communication can (sometimes) help offset poor recent performance.

Implications

Credibility evolves endogenously; rebuilding credibility could be harder if errors reduce capacity of central bank to influence expectations.
Bayesian Updating

- Participant $i$ prior belief about inflation given by:

$$
\pi_i \sim \mathcal{N}\left(\bar{\pi}_i, \frac{1}{\alpha_i}\right),
$$

1. $\bar{\pi}_i$ is $i$'s initial point forecast
2. $\alpha$ is a measure of $i$'s forecast precision.

- The central bank provides a potentially biased signal:

$$
\pi_{cb} = \pi + \tilde{\epsilon}, \quad \tilde{\epsilon} \sim \mathcal{N}\left(\gamma, \frac{1}{\beta}\right).
$$

1. $\beta$ is related to the precision of the central bank forecast
2. $\gamma$ is a possible systematic bias in the CB’s inflation forecast.
3. Assume $\gamma = 0$ for now.
The Role of Bias

- The optimal Bayesian inflation forecast:

\[ \mathbb{E}(\pi | \pi_{cb}) = \frac{\alpha \bar{\pi}_i + \beta \pi_{cb}}{\alpha + \beta} \]  \hspace{1cm} (3)

- Optimal update rate:

\[ u_i^* \equiv \frac{\mathbb{E}(\pi | \pi_{cb}) - \bar{\pi}_i}{(\pi_{cb} - \bar{\pi}_i)} = \frac{\beta}{\alpha + \beta} \]  \hspace{1cm} (4)

- If \( \gamma \neq 0 \) use adjusted signal \( (\pi_{cb} - \gamma) \)

1. If \( \beta \to \infty, \alpha \to 0 \) \( \Rightarrow u_i^* = 1 = 100\% \).

2. For any \( \beta \), \( \alpha \uparrow \) (prior precision \( \uparrow \)), update less \( (u_i^* \downarrow) \).

3. For any \( \alpha \), \( \beta \uparrow \) (signal precision \( \uparrow \)), update more \( (u_i^* \uparrow) \).
Participants’ Experience

Large online experiment (Prolific) with US users

1. Short survey
   - Economics knowledge
   - Understanding of and trust in various public institutions
   - Preferences for obtaining economic information
   - Familiarity with prevailing economic conditions

2. Instructions for inflation forecasting task (accessible later)

3. Comprehension quiz

4. Forecasting task

5. Informed which forecast had been selected for payment

6. Non-compulsory survey-of-decisions
Participants’ Experience

Large online experiment (Prolific) with US users

1. Short survey
2. Instructions for inflation forecasting task (accessible later)
   - Information they will get
   - How to interact with the available information
   - How to interact with our software
   - How we incentivized their forecasts
3. Comprehension quiz
4. Forecasting task
5. Informed which forecast had been selected for payment
6. Non-compulsory survey-of-decisions
Participants’ Experience

Large online experiment (Prolific) with US users

1. Short survey
2. Instructions for inflation forecasting task (accessible later)
3. Comprehension quiz
   ▶ 5 questions designed to test subjects’ understanding of our experimental instructions
   ▶ Must answer all five questions correctly to proceed
      ▶ More than 2 submissions with at least one wrong answer ⇒ Removed.
Participants’ Experience

Large online experiment (Prolific) with US users

1. Short survey
2. Instructions for inflation forecasting task (accessible later)
3. Comprehension quiz
4. Forecasting task
   ▶ 3 × decision periods
Participants’ Experience

Large online experiment (Prolific) with US users

1. Short survey
2. Instructions for inflation forecasting task (accessible later)
3. Comprehension quiz
4. Forecasting task
5. Informed which forecast had been selected for payment
6. Non-compulsory survey-of-decisions
Decision Periods

- Observe Past Central Bank Forecasts and Inflation Outcomes
- 1st set of Forecasts
  - Point Forecast
  - Range Forecast
- Initial Forecasts
- 2nd Set of Forecasts
  - Update Point Forecast?
  - Update Range Forecast?
- Updated Forecasts

Figure: Experimental Timeline: A single decision period

- Decision periods are independent
- Randomly select one forecast for bonus payment
Incentives

- Incentiving point forecasts:
  \[ F_{i,13} = 2^{-|E_{i,12}\{\pi_{13}\} - \pi_{13}|}. \]  

- Perfect forecast yields \( F_{i,13} = 1 \)
- \( F_{i,13} \) reduced by half each 1pp increase in forecasts error

- Range forecast:
  \[
  U_{i,t}(r_{i,t}) = \begin{cases}
  0 & \pi_{i,13} \not\in [u_{i,t}, u_{i,t}] \\
  \phi \left( \frac{1}{1+r_{i,t}} \right) & \pi_{i,13} \in [u_{i,t}, u_{i,t}] 
  \end{cases}
  \]  

- On average, participants
  - earned $3.75 for participation, $1.25 for bonus
  - equates to $13.20 per hour, on average
  - took 10-15 minutes to complete the experiment
3 Core Histories

![Graphs showing Core Histories for Early, Late, and Consistent Inflation Forecast Error.]

Legend:
- **Inflation**
- **Forecast**
- **Error**
## Summary of Forecast Performance by History (bps)

<table>
<thead>
<tr>
<th></th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Full Sample</th>
<th>HistAvg</th>
<th>LastYear</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calibration Data</td>
<td>110</td>
<td>95</td>
<td>34</td>
<td>80</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Forecast Performance</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consistent - Great</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>06</td>
<td>08</td>
</tr>
<tr>
<td>Consistent - Good</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>10</td>
<td>05</td>
</tr>
<tr>
<td>Consistent - Moderate</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>06</td>
<td>-07</td>
</tr>
<tr>
<td>Consistent - Bad</td>
<td>83</td>
<td>83</td>
<td>83</td>
<td>83</td>
<td>02</td>
<td>-19</td>
</tr>
<tr>
<td>Consistent - Terrible</td>
<td>171</td>
<td>171</td>
<td>171</td>
<td>171</td>
<td>-06</td>
<td>-42</td>
</tr>
<tr>
<td><strong>Timing &amp; Contextual Communication</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consistent - Bad</td>
<td>83</td>
<td>83</td>
<td>83</td>
<td>83</td>
<td>02</td>
<td>-19</td>
</tr>
<tr>
<td>Early</td>
<td>171</td>
<td>65</td>
<td>13</td>
<td>83</td>
<td>-51</td>
<td>12</td>
</tr>
<tr>
<td>Late</td>
<td>13</td>
<td>65</td>
<td>171</td>
<td>83</td>
<td>-52</td>
<td>-171</td>
</tr>
</tbody>
</table>

Numbers are average absolute forecast error in basis points.
Forecast Performance
Experimental Design - Forecast Performance

![Graphs showing forecast performance across different inflation scenarios]

- Great
- Good
- Moderate
- Bad
- Terrible

Inflation Forecast

- Blue dashed line: Inflation
- Red line: Forecast
## Treatment Summary: Forecast Performance

<table>
<thead>
<tr>
<th>History 1</th>
<th>History 2</th>
<th>History 3</th>
<th>Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1a</td>
<td>Early</td>
<td>Late</td>
<td>Great</td>
</tr>
<tr>
<td>T1b</td>
<td>Late</td>
<td>Early</td>
<td>Great</td>
</tr>
<tr>
<td>T2a</td>
<td>Early</td>
<td>Late</td>
<td>Good</td>
</tr>
<tr>
<td>T2b</td>
<td>Late</td>
<td>Early</td>
<td>Good</td>
</tr>
<tr>
<td>T3a</td>
<td>Early</td>
<td>Late</td>
<td>Moderate</td>
</tr>
<tr>
<td>T3b</td>
<td>Late</td>
<td>Early</td>
<td>Moderate</td>
</tr>
<tr>
<td>T4a</td>
<td>Early</td>
<td>Late</td>
<td>Bad</td>
</tr>
<tr>
<td>T4b</td>
<td>Late</td>
<td>Early</td>
<td>Bad</td>
</tr>
<tr>
<td>T5a</td>
<td>Early</td>
<td>Late</td>
<td>Terrible</td>
</tr>
<tr>
<td>T5b</td>
<td>Late</td>
<td>Early</td>
<td>Terrible</td>
</tr>
</tbody>
</table>
Results - Forecast Performance

![Graph showing forecast performance results with categories from Great to Terrible and performance metrics labeled as u', u'_{HistAvg}, and Optimal u'.]
Effect of Timing
What effect of time profile of errors?

- Use the full history to estimate $\beta$:

$$\beta^{-1} = \frac{\sum_{j=1}^{12} |E_{j-1}^{CB}(\pi_j) - \pi_j|}{12}.$$  

(7)

- Or, weight more heavily recent performance:

$$\beta^{-1} = \lambda \sum_{j=0}^{j=11} (1 - \lambda)^j |E_{t-2-j}^{CB}(\pi_{t-1-j}) - \pi_{t-1-j}|.$$  

(8)

![Weighting Function](image)
## Treatments - Timing

<table>
<thead>
<tr>
<th>History 1</th>
<th>History 2</th>
<th>History 3</th>
<th>Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early</td>
<td>Late</td>
<td>Consistent</td>
<td>97</td>
</tr>
<tr>
<td>Early</td>
<td>Consistent</td>
<td>Late</td>
<td>94</td>
</tr>
<tr>
<td>Late</td>
<td>Early</td>
<td>Consistent</td>
<td>80</td>
</tr>
<tr>
<td>Late</td>
<td>Consistent</td>
<td>Early</td>
<td>88</td>
</tr>
<tr>
<td>Consistent</td>
<td>Early</td>
<td>Late</td>
<td>79</td>
</tr>
<tr>
<td>Consistent</td>
<td>Late</td>
<td>Early</td>
<td>91</td>
</tr>
</tbody>
</table>
Results - Timing

-50
-25
0
25
50
75
Early
Consistent
Late
Deviation
Update
Optimal Update

U*

Medium
Reverse
Within
HistAvg

Deviation  Update  Optimal Update
Results - Timing

Weighting Function for Estimated $\lambda$

Quarters in Past

Early

Early$_{HistAvg}$

Late

Late$_{HistAvg}$
Contextualizing Communication
We introduce written comms. into *Late* in *E,C,L*
Consider 6 written reports:

- **Control**: General description of central banking
- **Control + Outlook**: Includes outlook on inflation that matches graphical forecast
- **Exogenous + Relative Performance**: Drop in forecast performance resulted from *exogenous shock* and bank has performed better or worse than counterparts
- **Endogenous + Relative Performance**: Drop in forecast performance resulted from *endogenous forces* and bank has performed better or worse than counterparts
<table>
<thead>
<tr>
<th>Name</th>
<th>Sample Size</th>
<th>Flesh-Kincaid Score</th>
<th>Reading Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>160</td>
<td>8</td>
<td>10th-12th</td>
</tr>
<tr>
<td>Control + Outlook</td>
<td>151</td>
<td>8.3</td>
<td>10th-12th</td>
</tr>
<tr>
<td>Exogenous + Better</td>
<td>131</td>
<td>8.5</td>
<td>10th-12th</td>
</tr>
<tr>
<td>Exogenous + Worse</td>
<td>152</td>
<td>8.5</td>
<td>10th-12th</td>
</tr>
<tr>
<td>Endogenous + Better</td>
<td>157</td>
<td>8.4</td>
<td>10th-12th</td>
</tr>
<tr>
<td>Endogenous + Worse</td>
<td>137</td>
<td>8.4</td>
<td>10th-12th</td>
</tr>
</tbody>
</table>
Results - Communication

![Diagram showing the results of communication]
Conclusion

What have we learned so far?

- Forecast performance matters but not as sharply as theory predicts
- Credibility is endogenous, dynamics are asymmetric:
  - Recency bias
  - Credibility takes longer to build than to lose
- MPRs, IRs, etc. are valuable as a way of rationalizing the past and reinforcing outlook

Implications

Credibility evolves endogenously; rebuilding credibility could be harder if errors reduce capacity of central bank to influence expectations.
Central Bank Announcement

The Fed uses interest rate policy to stabilize prices and keep employment high. We base monetary policy on how healthy the economy is now and how healthy we think it will be in the future. We use forecasts to guide our decisions. We do our best when making forecasts but the world is uncertain, and forecasts are never perfect.

Over the last year, our forecasts underpredicted inflation. This is because the pandemic lasted longer than initially expected and caused supply shortages. Our forecasts over this period were more accurate than private sector forecasts and other central banks. Our best guess is that inflation will decrease next quarter.
Longer-term Forecasts

Ustar

Early Consistent Late

Update Deviation Optimal Update

back
Changing the direction of forecast errors

U^*

Early Consistent Late

Update Deviation Optimal Update

- Back
Within-subject forecast credibility measure

- Early
- Consistent
- Late

Deviation, Update, Optimal Update
Within-subject forecast credibility measure

![Graph showing within-subject forecast credibility measure](image)

- Shading is 99%, 95%, 90% CIs
- Deviation
- Update
- Optimal Update