# Building Central Bank Credibility: The Role of Forecast Performance

Michael McMahon **Ryan Rholes**

*Department of Economics*
*University of Oxford*

# Motivation

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

# Motivation

- Managing expectations is crucial for the now dominate 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 historical forecast performance impacts a central bank's forecast credibility

# Motivation

- Managing expectations is crucial for the now dominate 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 historical forecast performance impacts a 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.

# 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:

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

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

- The central bank provides a potentially biased signal:

$$\pi_{cb} = \pi + \tilde{\epsilon}, \quad \tilde{\epsilon} \sim \mathcal{N}\left(\gamma, \frac{1}{\beta}\right). \tag{2}$$

  - $\beta$ is related to the precision of the central bank forecast
  - $\gamma$ is a possible systematic bias in the CB's inflation forecast.
    - Assume $\gamma = 0$ for now.

# The Role of Bias

- The optimal Bayesian inflation forecast:

$$\mathbb{E}(\pi|\pi_{cb}) = \frac{\alpha\bar{\pi}_i + \beta\pi_{cb}}{\alpha + \beta} \tag{3}$$

- Optimal update rate:

$$u_i^* \equiv \frac{\mathbb{E}(\pi|\pi_{cb}) - \bar{\pi}_i}{(\pi_{cb} - \bar{\pi}_i)} = \frac{\beta}{\alpha + \beta} \tag{4}$$

- 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

# 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

# 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 $\Rightarrow$ 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
   - $3 \times$ 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

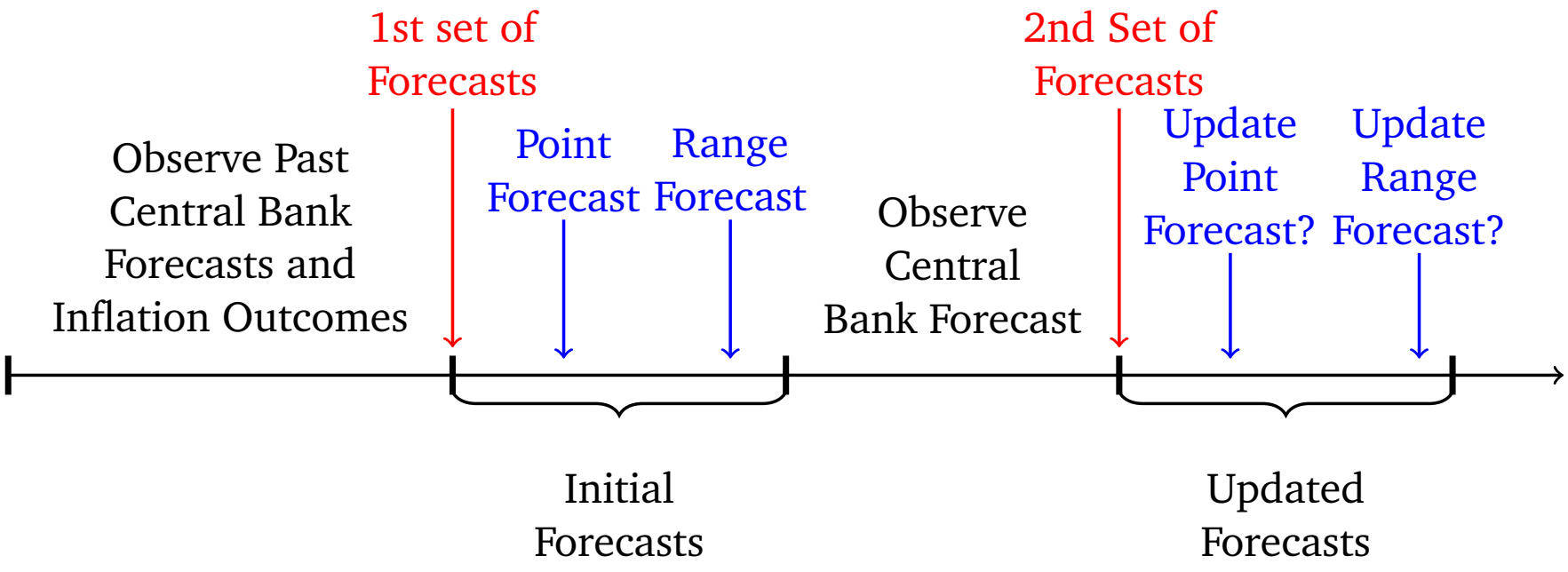



Figure: Experimental Timeline: A single decision period

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

# Incentives

- Incentiving point forecasts:

$$F_{i,13} = 2^{-|\mathbb{E}_{i,12}\{\pi_{13}\} - \pi_{13}|}. \tag{5}$$

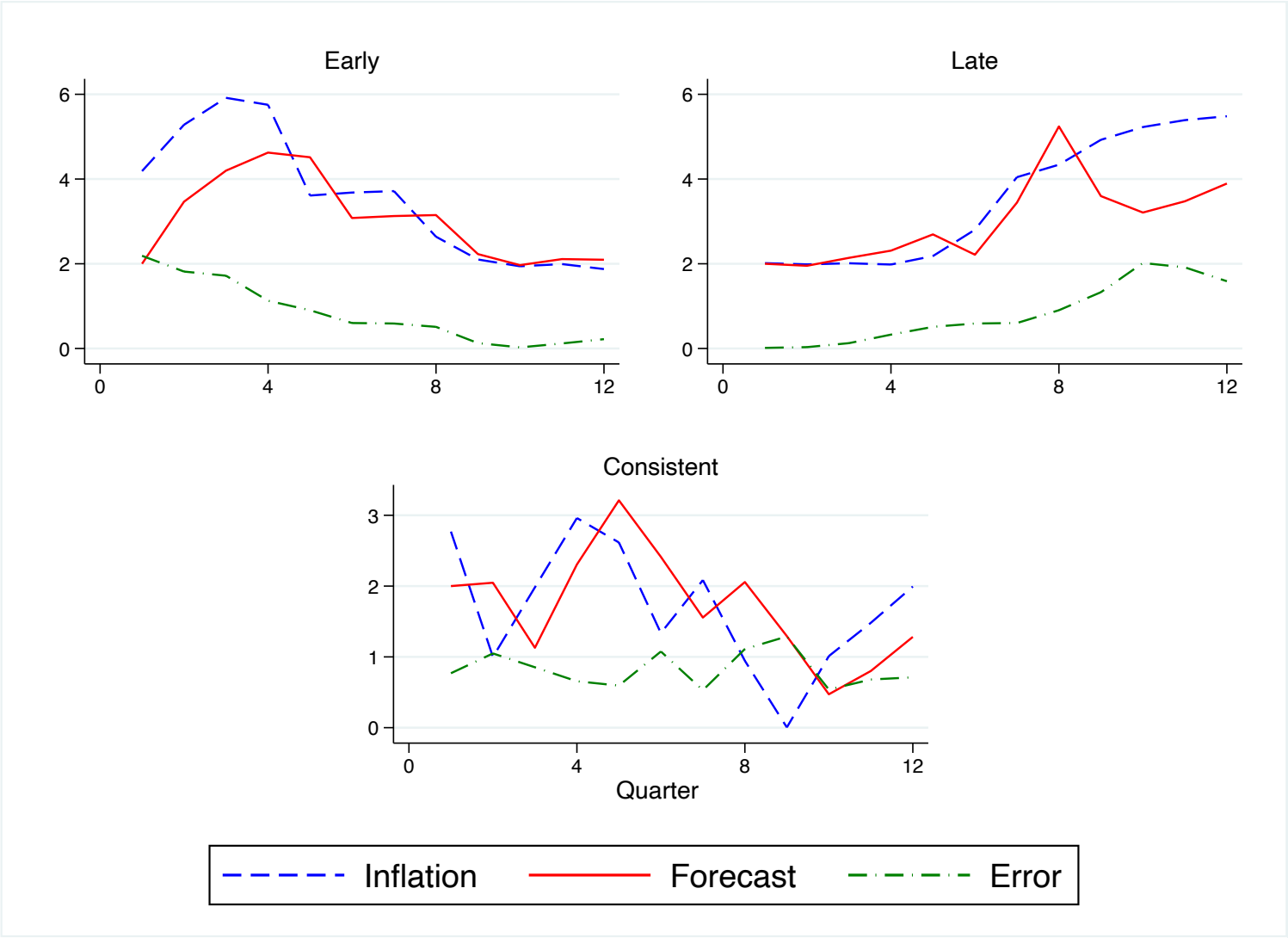
  - Perfect forecast yields $F_{i,13} = 1$
  - $F_{i,13}$ reduced by half each 1pp increase in forecasts error

- Range forecast:

$$U_{i,t}(r_{i,t}) = \begin{cases} 0 & \pi_{i,13} \notin [\underline{u_{i,t}}, \overline{u_{i,t}}] \\ \phi\left(\frac{1}{1+r_{i,t}}\right) & \pi_{i,13} \in [\underline{u_{i,t}}, \overline{u_{i,t}}]. \end{cases} \tag{6}$$

- 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



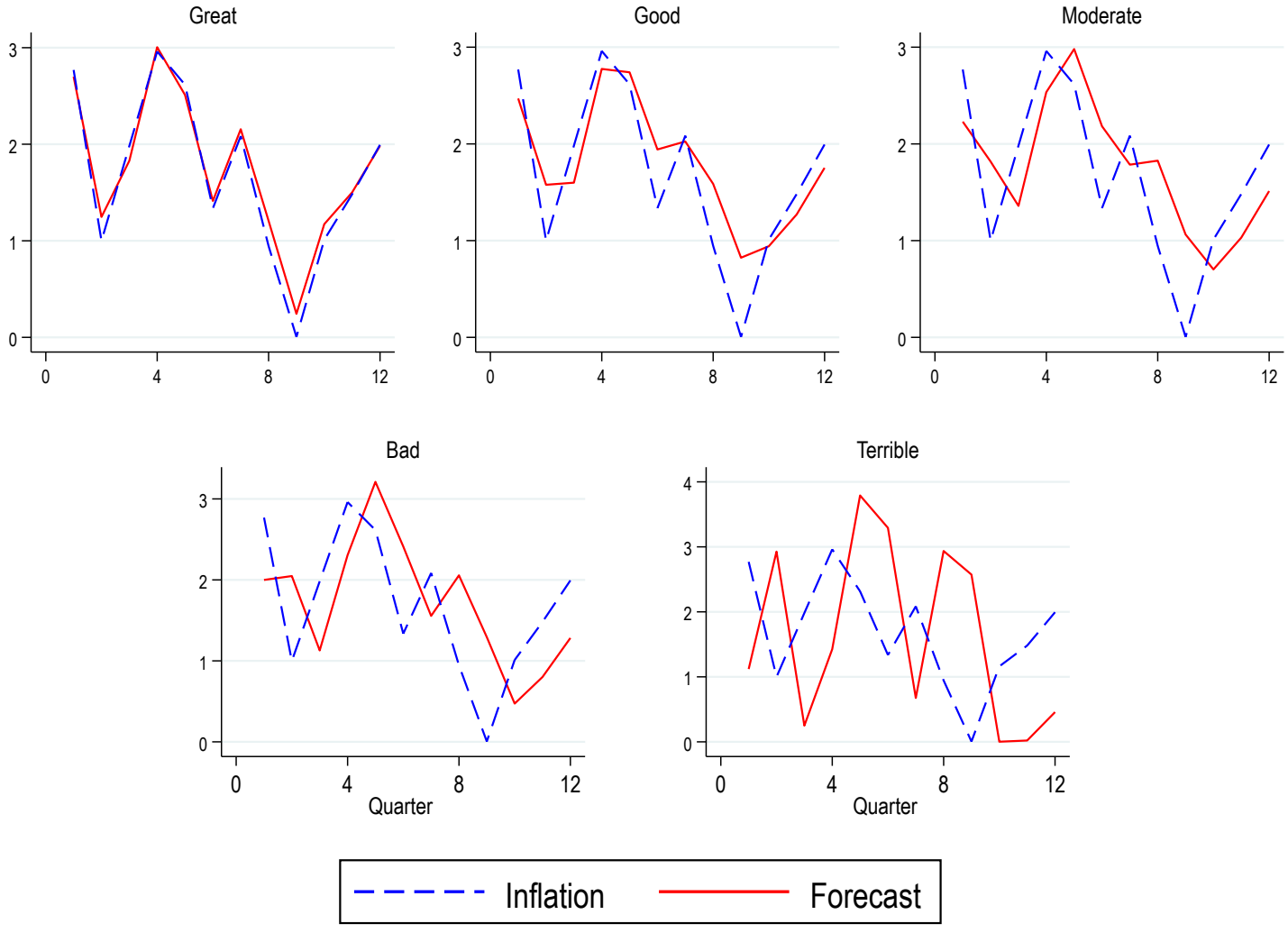
# All Histories

| | Summary of Forecast Performance by History (bps) | | | | | |
|---|---|---|---|---|---|---|
| | **Year 1** | **Year 2** | **Year 3** | **Full Sample** | $\gamma_{HistAvg}$ | $\gamma_{LastYear}$ |
| *Calibration Data* | 110 | 95 | 34 | 80 | | |
| *Consistent - Great* | 13 | 13 | 13 | 13 | 06 | 08 |
| *Consistent - Good* | 36 | 36 | 36 | 36 | 10 | 05 |
| *Consistent - Moderate* | 60 | 60 | 60 | 60 | 06 | -07 |
| ***Consistent - Bad*** | 83 | 83 | 83 | 83 | 02 | -19 |
| *Consistent - Terrible* | 171 | 171 | 171 | 171 | -06 | -42 |
| ***Consistent - Bad*** | 83 | 83 | 83 | 83 | 02 | -19 |
| *Early* | 171 | 65 | 13 | 83 | -51 | 12 |
| *Late* | 13 | 65 | 171 | 83 | -52 | -171 |

Numbers are average absolute forecast error in basis points.

# Forecast Performance

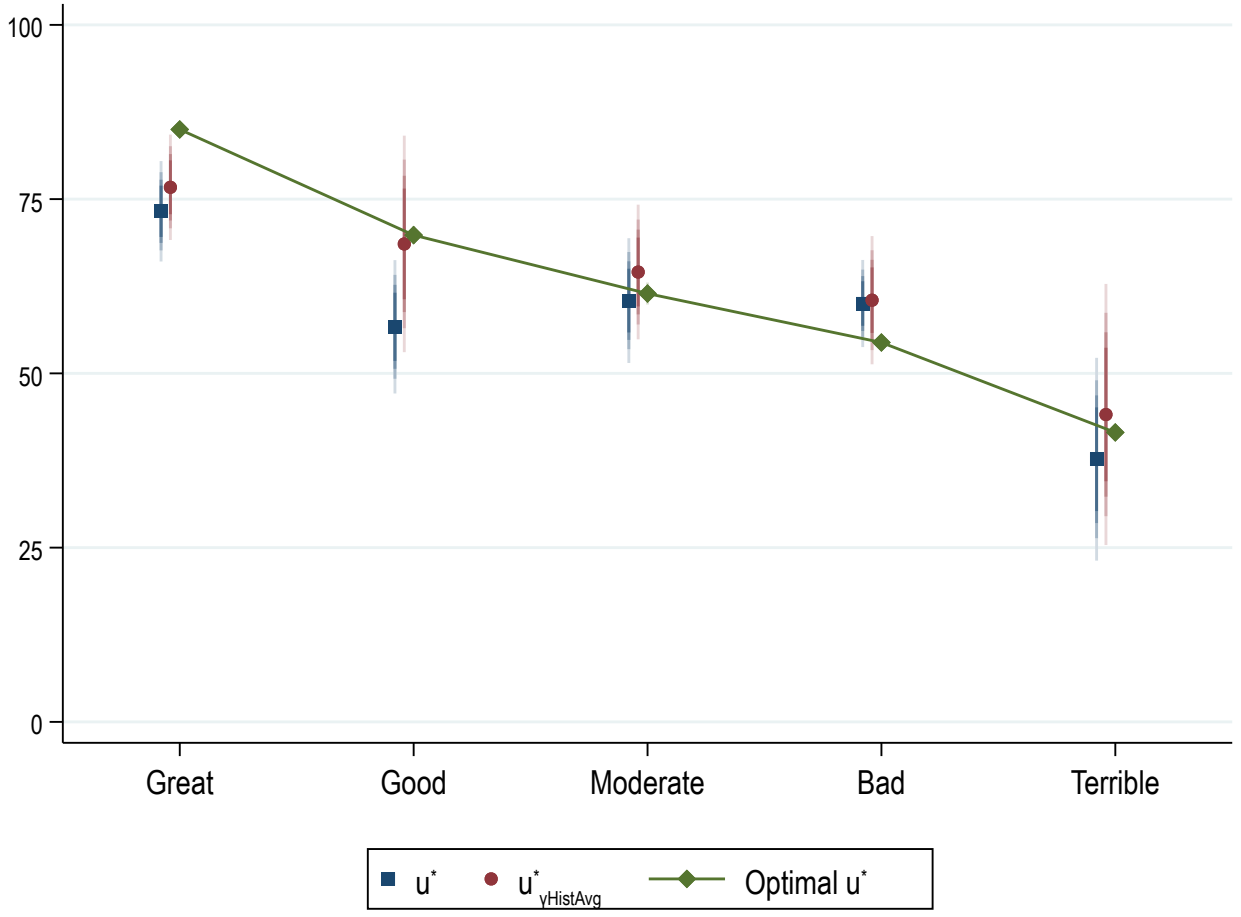
# Experimental Design - Forecast Performance



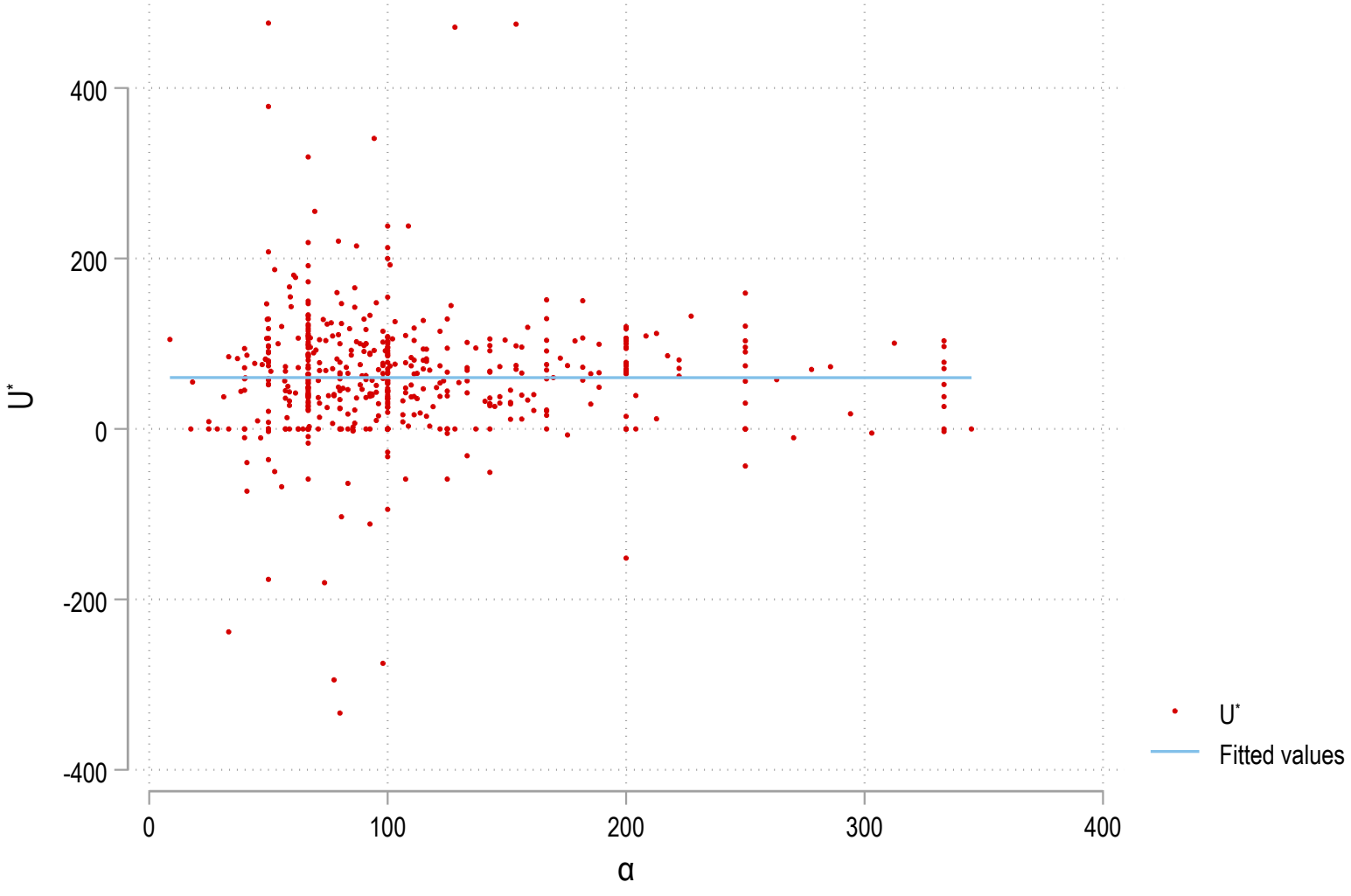
# Experimental Design - *ForecastPerformance*

| | Treatment Summary: Forecast Performance | | | |
|---|---|---|---|---|
| | **History 1** | **History 2** | **History 3** | **Sample Size** |
| *T1a* | *Early* | *Late* | *Great* | 46 |
| *T1b* | *Late* | *Early* | *Great* | 44 |
| *T2a* | *Early* | *Late* | *Good* | 44 |
| *T2b* | *Late* | *Early* | *Good* | 46 |
| *T3a* | *Early* | *Late* | *Moderate* | 33 |
| *T3b* | *Late* | *Early* | *Moderate* | 44 |
| *T4a* | *Early* | *Late* | *Bad* | 97 |
| *T4b* | *Late* | *Early* | *Bad* | 76 |
| *T5a* | *Early* | *Late* | *Terrible* | 46 |
| *T5b* | *Late* | *Early* | *Terrible* | 50 |

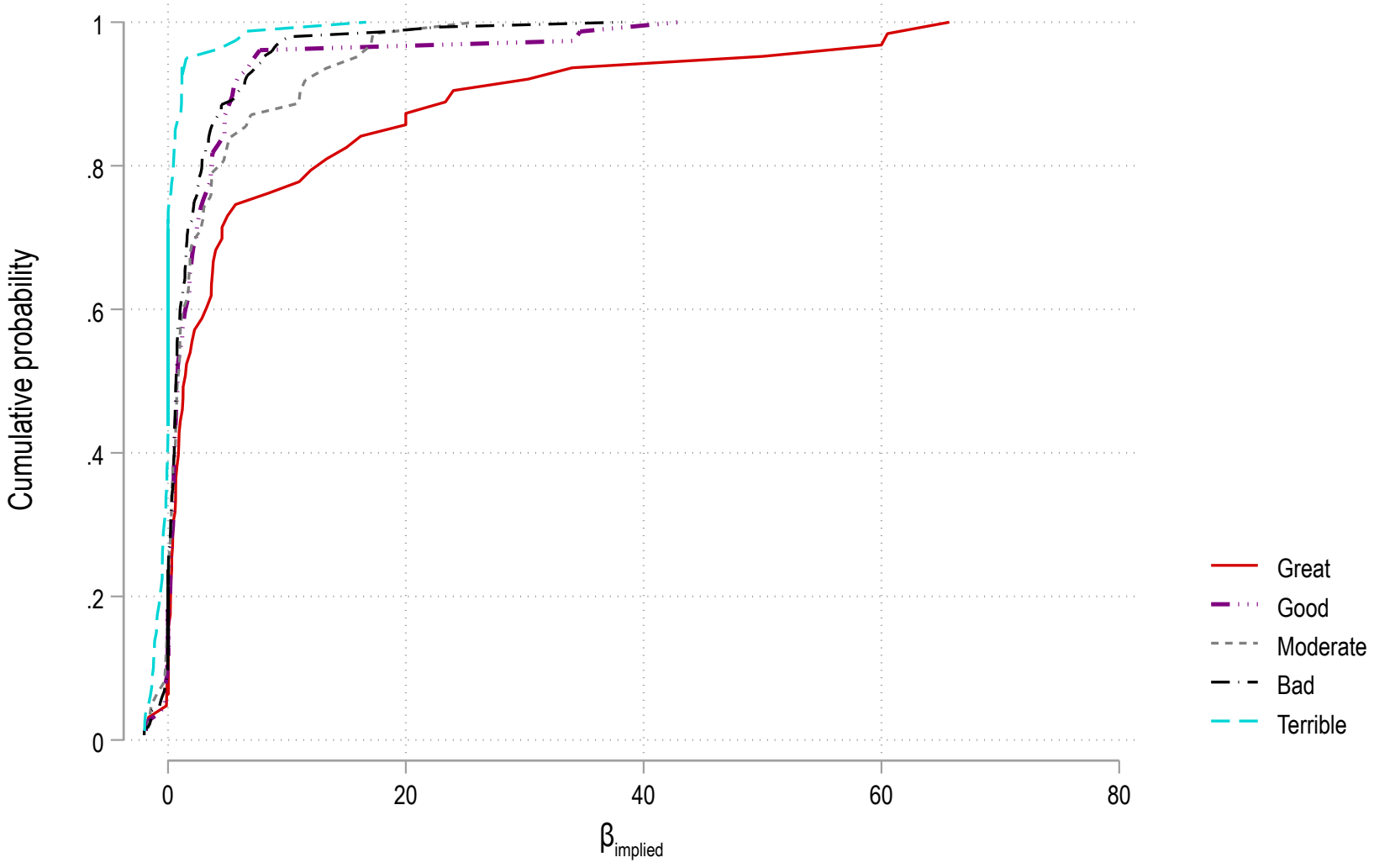
# Results - Forecast Performance



# Results - Forecast Performance



# Results - Forecast Performance



# Effect of Timing

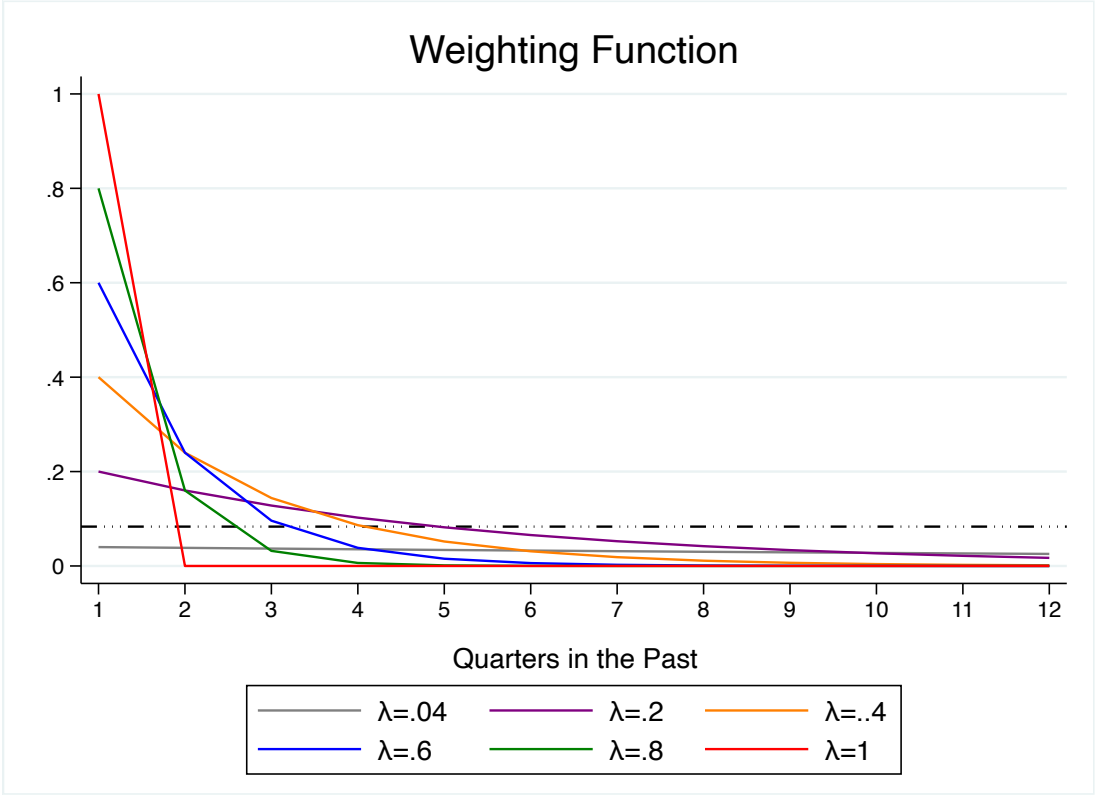
# What effect of time profile of errors?

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

$$\beta^{-1} = \frac{\sum_{j=1}^{j=12} |\mathbb{E}_{j-1}^{CB}(\pi_j) - \pi_j|}{12}. \tag{7}$$

- Or, weight more heavily recent performance:

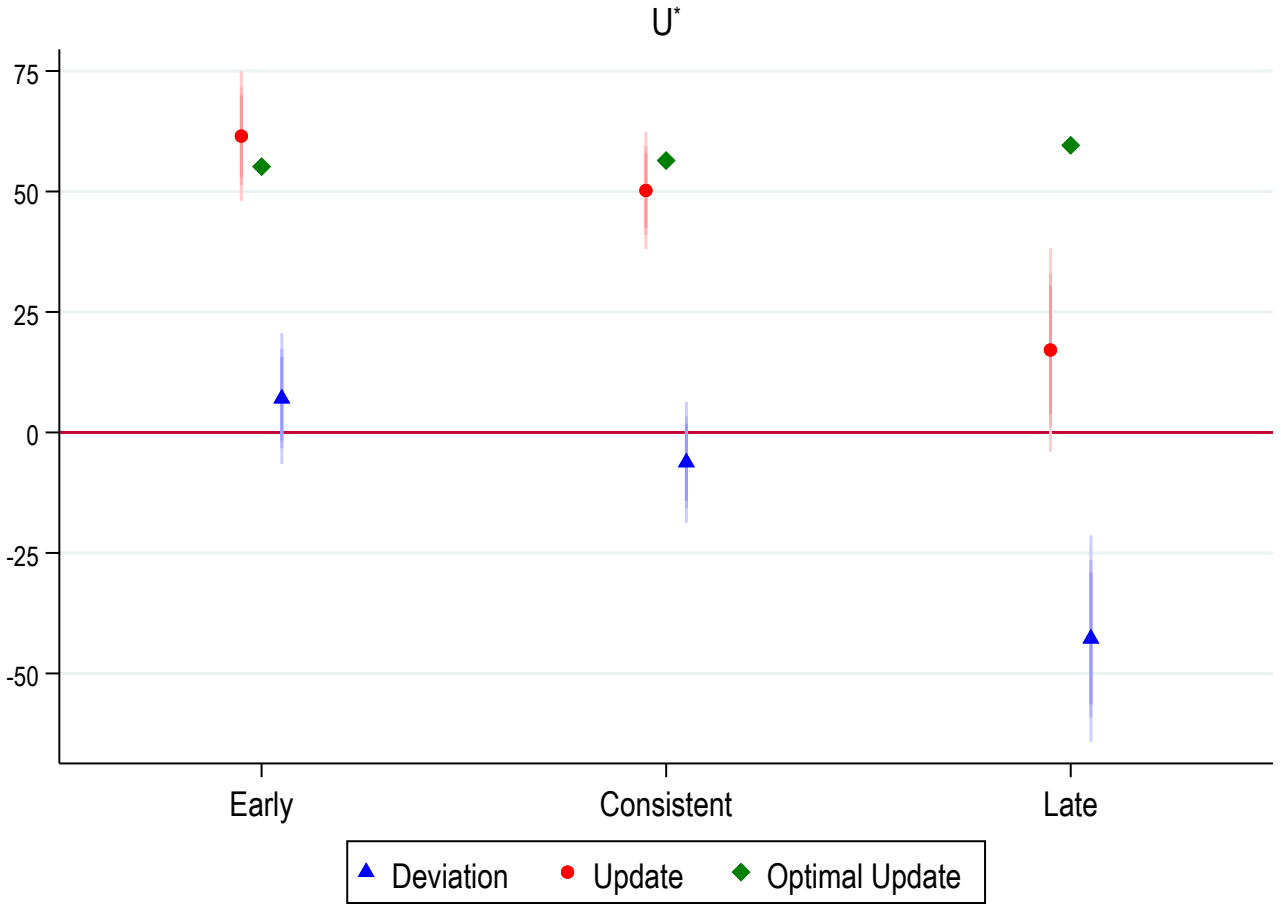
$$\beta^{-1} = \lambda \sum_{j=0}^{j=11} (1-\lambda)^j |\mathbb{E}_{t-2-j}^{CB}(\pi_{t-1-j}) - \pi_{t-1-j}| \tag{8}$$



# Treatments - Timing

| Treatment Summary - Timing | | | |
|---|---|---|---|
| **History 1** | **History 2** | **History 3** | **Sample Size** |
| *Early* | *Late* | *Consistent* | 97 |
| *Early* | *Consistent* | *Late* | 94 |
| *Late* | *Early* | *Consistent* | 80 |
| *Late* | *Consistent* | *Early* | 88 |
| *Consistent* | *Early* | *Late* | 79 |
| *Consistent* | *Late* | *Early* | 91 |

# Results - Timing



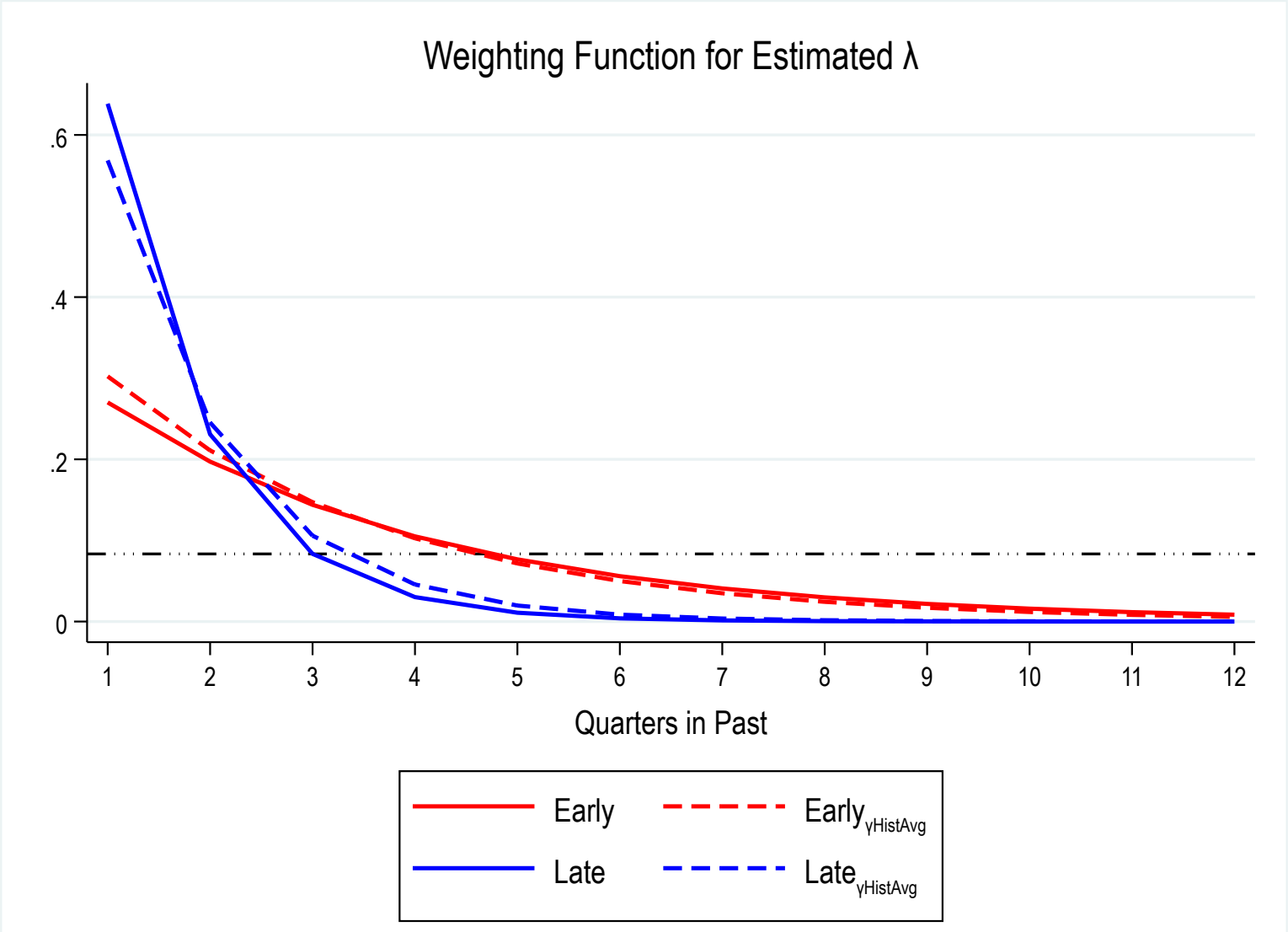

Medium | Reverse | Within

# Results - Timing - Estimated $\lambda$

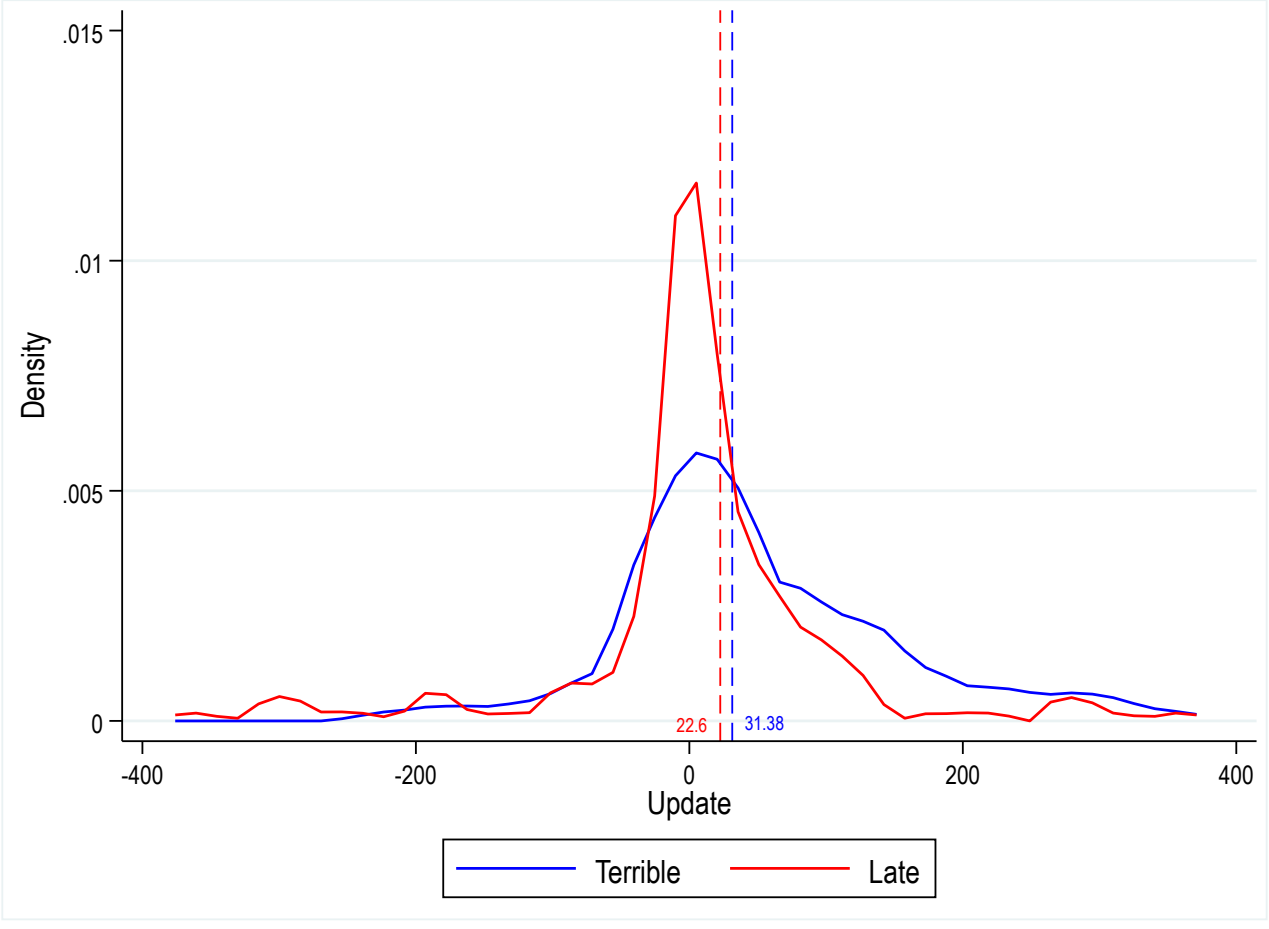
| | $\gamma_0$ | $\gamma_{HistAvg}$ |
|---|---|---|
| **Early** | 0.245 | 0.275 |
| | (0.0170) | (0.0160) |
| **Consistent** | 0.523 | 0.511 |
| | (0.022) | (0.022) |
| **Late** | 0.622 | 0.560 |
| | (0.0198) | (0.0222) |

Standard errors in parentheses

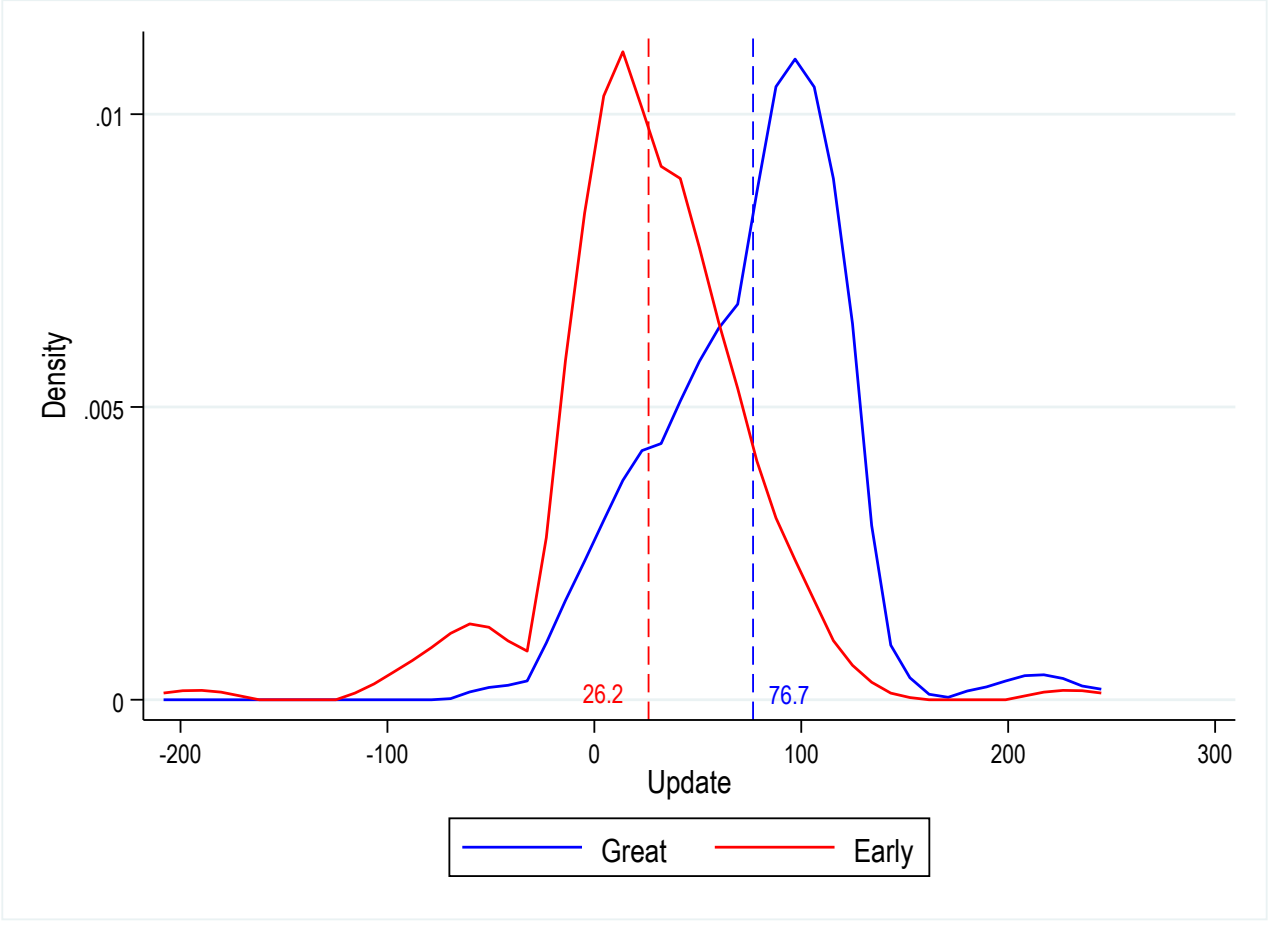
# Results - Timing



# Dynamics of Perceived Credibility



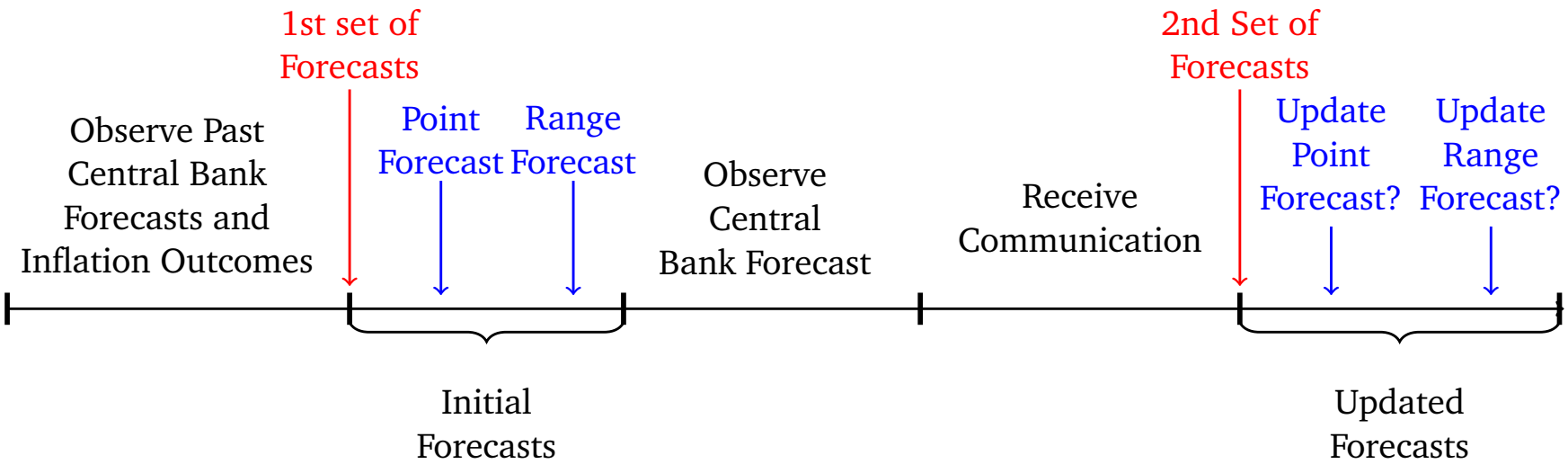
# Dynamics of Perceived Credibility



# Effect of Timing

# Experimental Design - Communication

We introduce written comms. into *Late* in *E,C,L*




Screen2

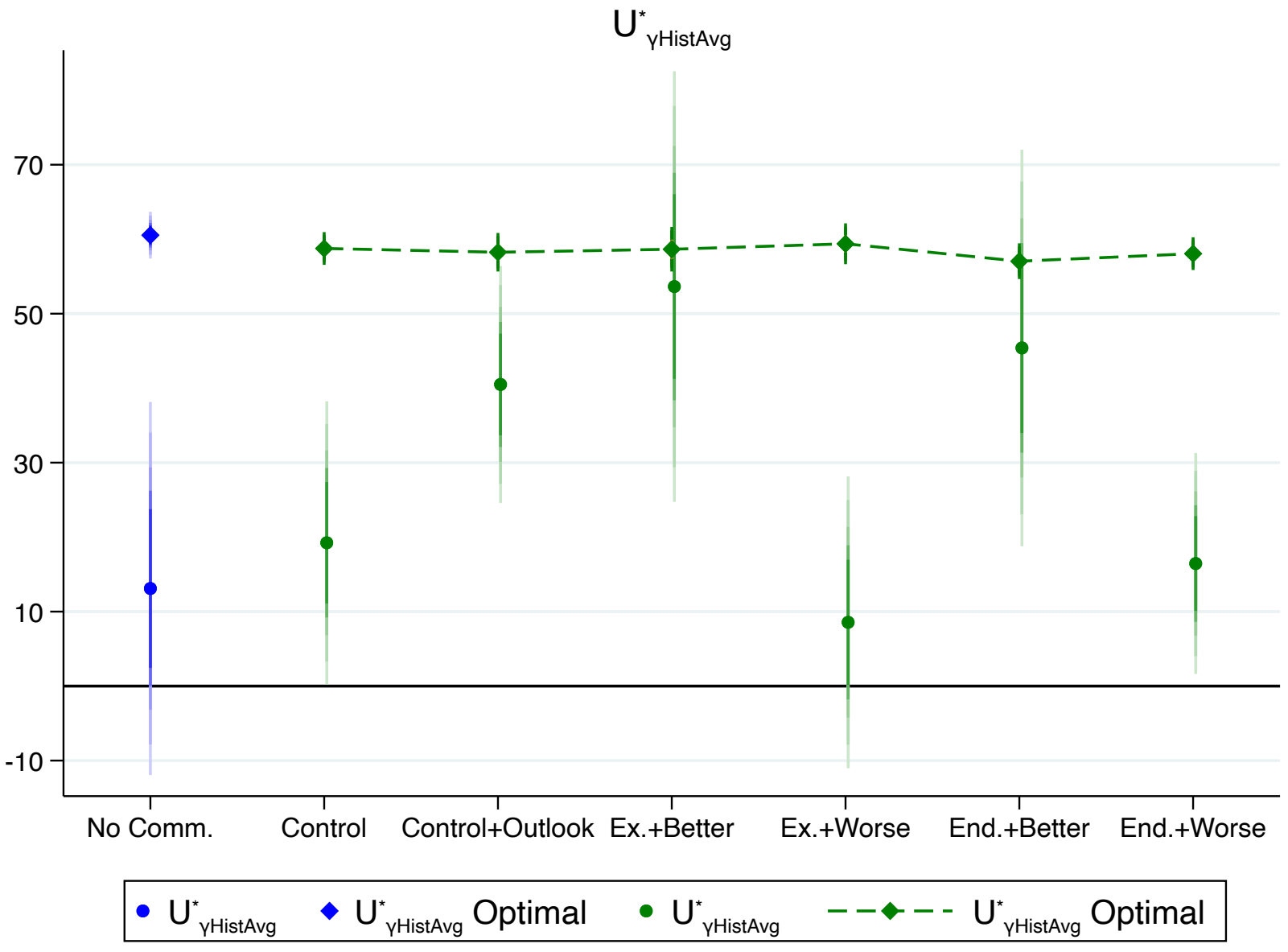
# Experimental Design - Communication

- 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

# Experimental Design - Communication

| Treatment Summary - Communication | | | |
|---|---|---|---|
| **Name** | **Sample Size** | **Flesh-Kincaid** | |
| | | *Score* | *Reading Level* |
| *Control* | 160 | 8 | 10th-12th |
| *Control + Outlook* | 151 | 8.3 | 10th-12th |
| *Exogenous + Better* | 131 | 8.5 | 10th-12th |
| *Exogenous + Worse* | 152 | 8.5 | 10th-12th |
| *Endogenous + Better* | 157 | 8.4 | 10th-12th |
| *Endogenous + Worse* | 137 | 8.4 | 10th-12th |

# Results - 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

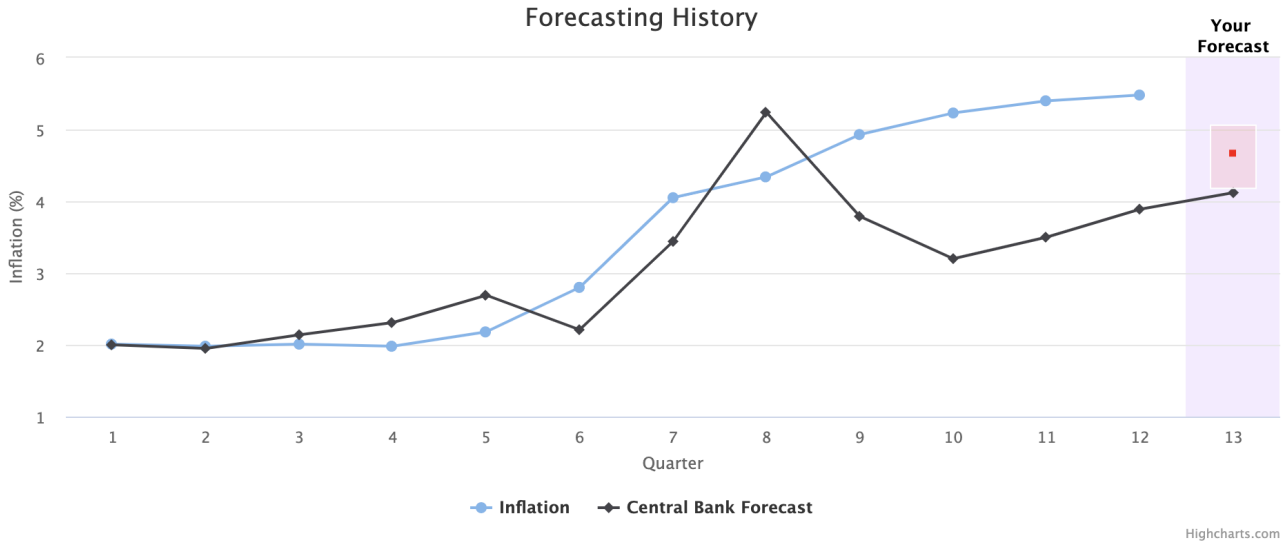
Credibility evolves endogenously; rebuilding credibility could be harder if errors reduce capacity of central bank to influence expectations.

# Screenshot for Comms

## Central Bank Announcement

**Avg. Forecast Error**

**First Year:** .13
**Second Year:**.65
**Third Year:** 1.71
**Full History:** .83




## 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.

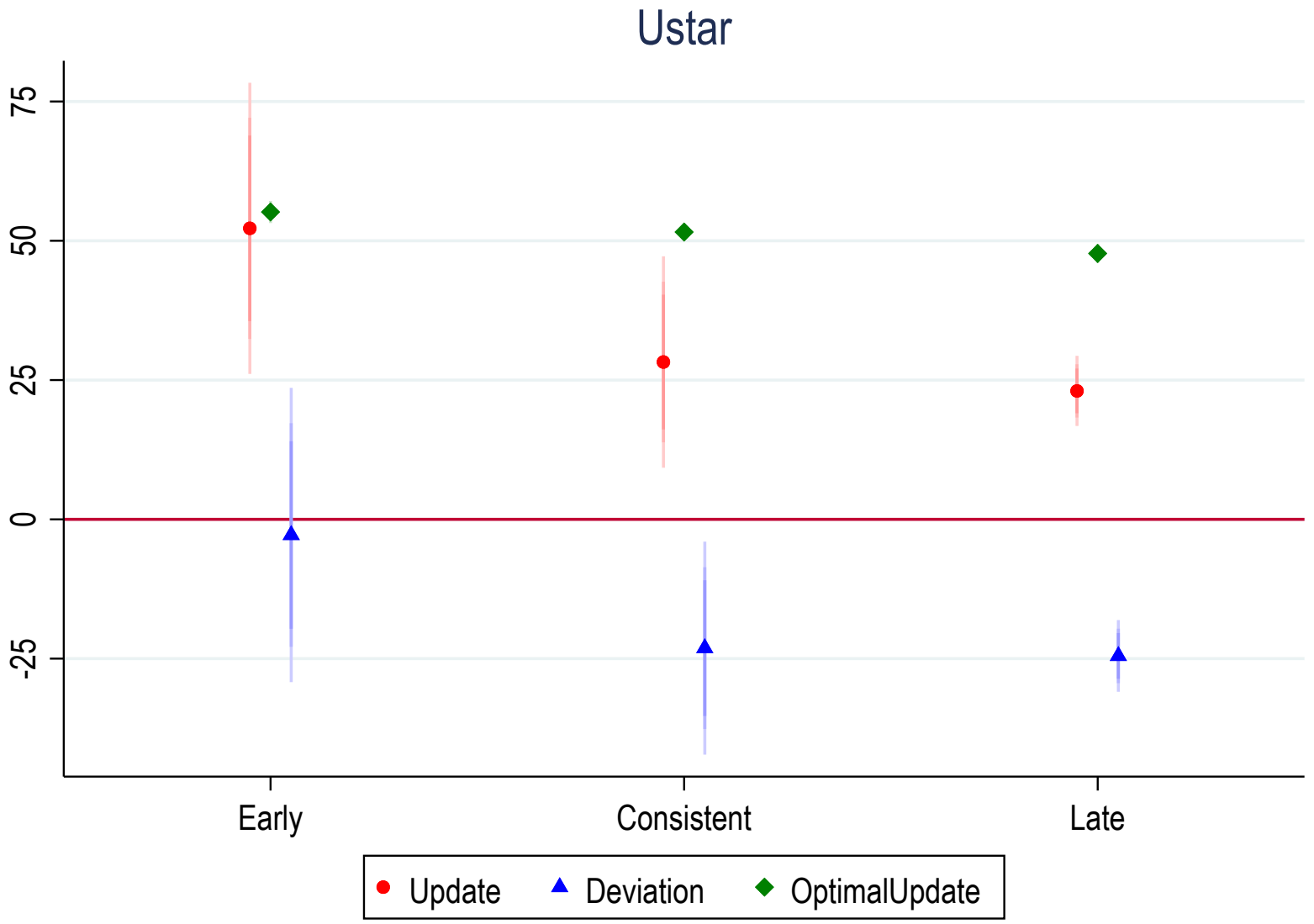
Over the last year, our forecasts underpredicted inflation. This is because the pandemic lasted longer than initially expected and caused supply shortages. Our forecasts over this period were more accurate than private sector forecasts and other central banks. Our best guess is that inflation will decrease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

Finished Reading

Full Instructions

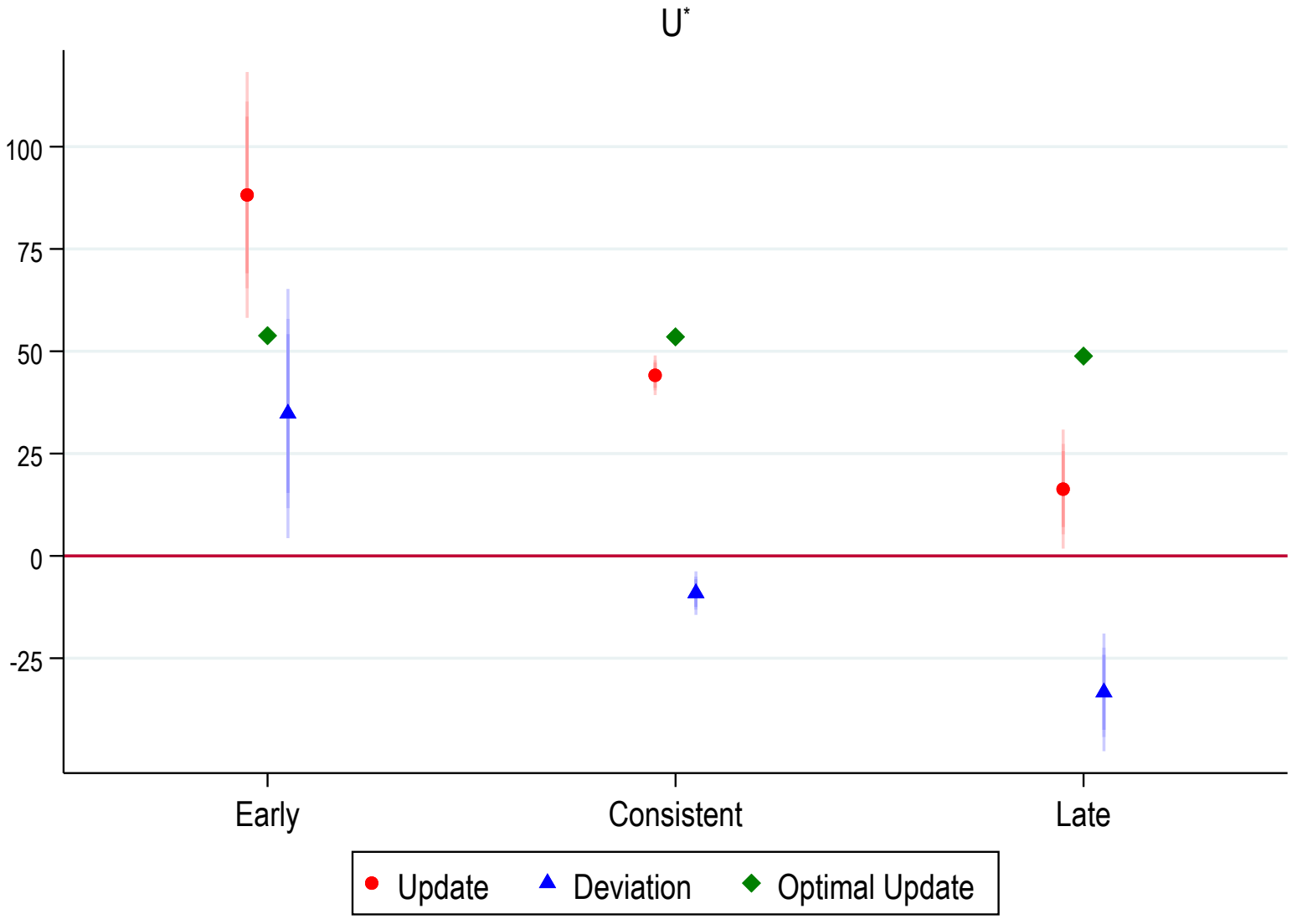
back

# Longer-term Forecasts



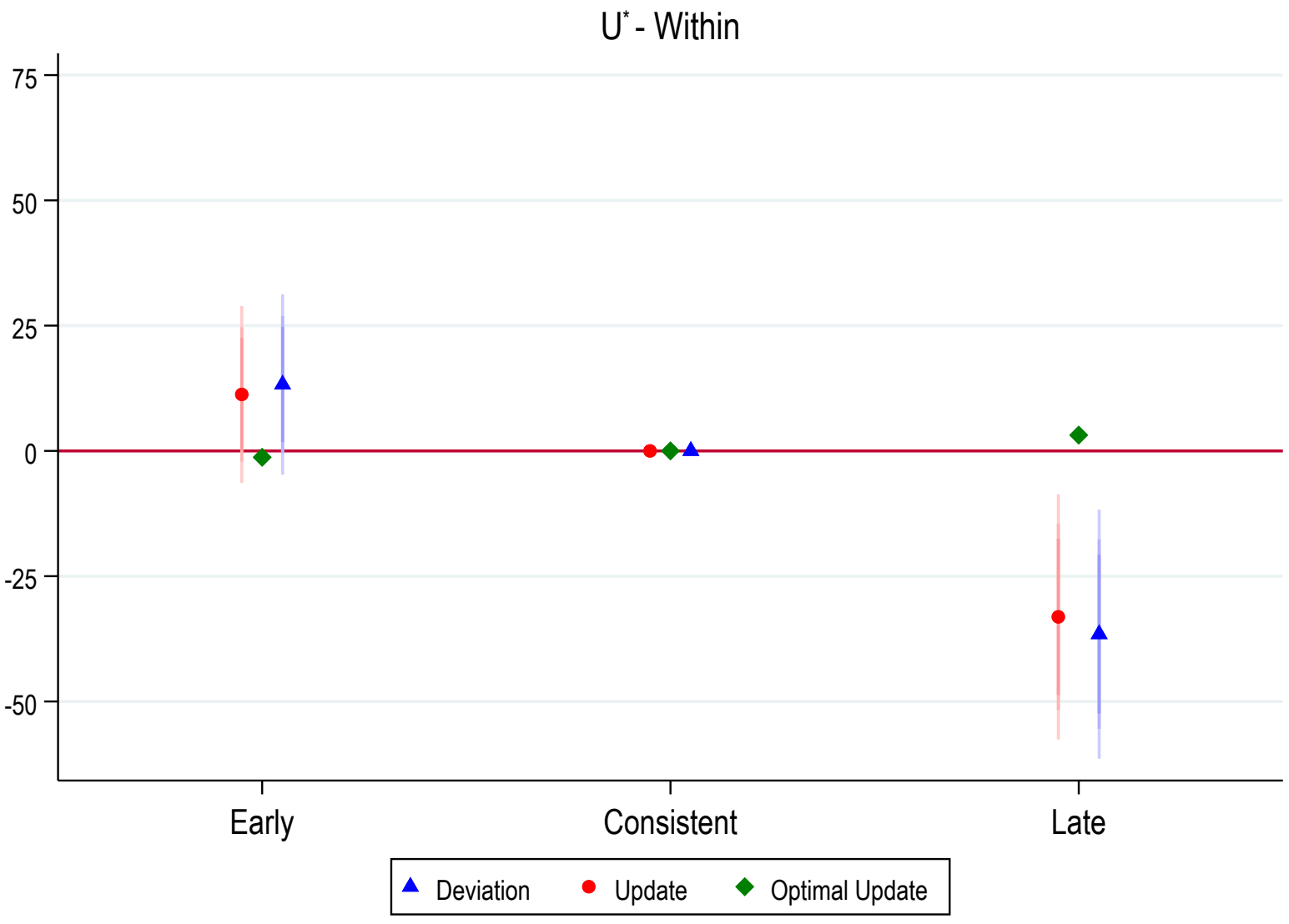

back

# Changing the direction of forecast errors




back

# Within-subject forecast credibility measure




back