BdE - VIII Annual Research Conference "Economics of Artificial Intelligence"

Simulating the Survey of Professional Forecasters – Discussion

Madrid

The views expressed herein belong to the authors only and do not necessarily reflect those of the NBS or the Eurosystem

Joan Paredes

14 November 2025

Summary of the paper

LLMs simulate SPF participants using real-time macro data and forecaster characteristics. Beautiful and intuitive equations!!

The problem (expectations)

$$y_{t+H} = f(x_t, z_t) + \varepsilon_{t+H}.$$

Humans

$$h_{i,t} = f(x_t, z_t) + \Delta_{i,t},$$

Machines

$$m_t = \mathbb{E}\Big[f(x_t, z_t) \mid x_t\Big].$$

Cooler machines (LLMs)

$$m_t^{\text{AI}} = \mathbb{E}^{\text{AI}}\Big[f(x_t, z_t) \mid x_t\Big].$$

Gap with Zt

$$\Delta_t^{\text{AI}} = m_t^{\text{AI}} - f(x_t, z_t).$$

Summary of the paper
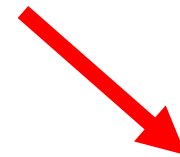
Augment info to the cooler machine (why not to other machines?) ;-(

$$m_t^{AI} = \mathbb{E}^{AI}\left[f(x_t, z_t) \mid x_t, \bar{h}_{t-1}\right], \quad \text{with human SPF info}$$

$$\bar{h}_{t-1} = \text{median}\left(h_{1,t-1}, h_{2,t-1}, \ldots, h_{N,t-1}\right)$$

$$= f(x_{t-1}, z_{t-1}) + \text{median}\left(\Delta_{1,t-1}, \Delta_{2,t-1}, \ldots, \Delta_{N,t-1}\right)$$

$$= f(x_{t-1}, z_{t-1}) + \bar{\Delta}_{t-1}.$$

$$\Delta_{i,t} = \gamma(w_{i,t}) + e_{i,t},$$

Our individual SPF agents $\quad m_{i,t}^{AI} = \mathbb{E}^{AI}\left[f(x_t, z_t) \mid x_t, f(x_{t-1}, z_{t-1}) + \bar{\Delta}_{t-1}, w_{i,t}\right]$

Summary of the paper

Results

- Median human and median AI are aligned. Most of the times...

- AI SPFs often lower forecasting error

- Best performance when LLMs are prompted with:
    Past SPF medians, **+53–757%**
    Real-time data, +8–37%
    Individual forecaster profiles. +4–15%

- Out-of-sample 2024: mixed

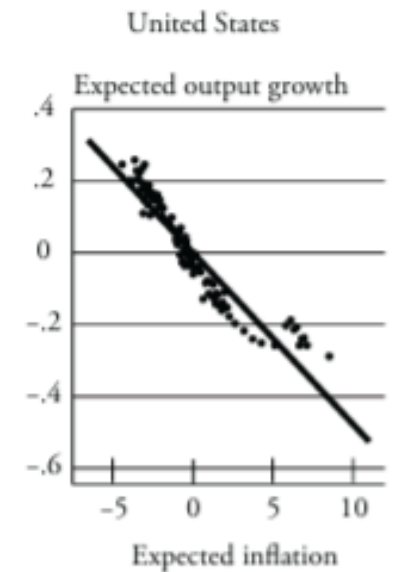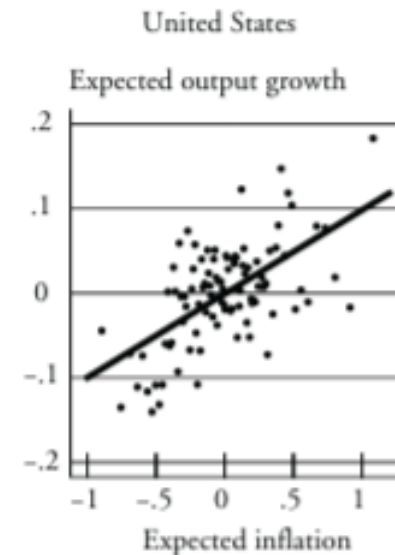# Comment 1 -> What is the target? 2 possible interpretations

A) Reproduce SPF expectations. $\Delta_t^{AI} = m_t^{AI} - f(x_t, z_t).$

- Cost-effective, high frequency

- It could be extended to households if micro characteristics known.

- Does AI inherit "demand vs supply"
 cognitive frames?

Communication and the Beliefs of Economic Agents (Candia et al. 2020)

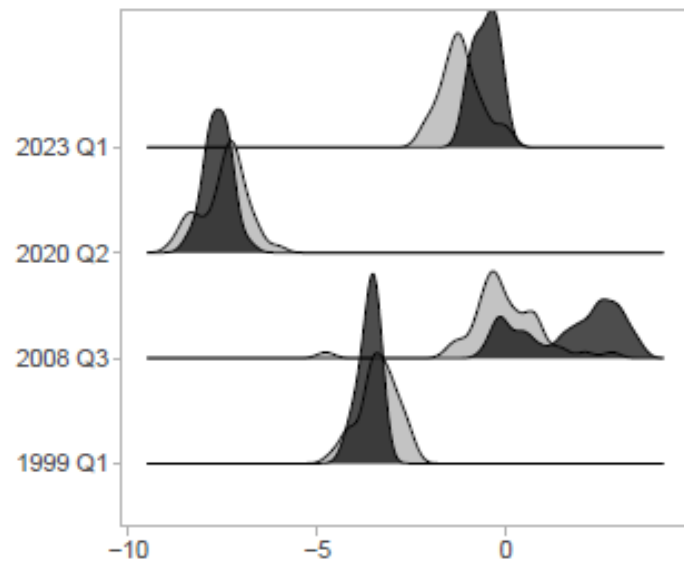# Comment 1 -> What is the target? 2 possible interpretations

- AI SPFs generated predictions are similar ...for some variables, GDP and unemployment, but ...

**(b)** CPI inflation rate, four quarters ahead

Comment 1 -> What is the target? 2 possible interpretations

- SPF disagreement is informative for policy (e.g., uncertainty, bias). Could AI dispersion be interpreted?

- How does *temperature* affect disagreement across simulated agents?
   → link to literature on bounded rationality and heterogeneous beliefs.

(b) CPI inflation rate, four quarters ahead

(f) Unemployment rate, four quarters ahead
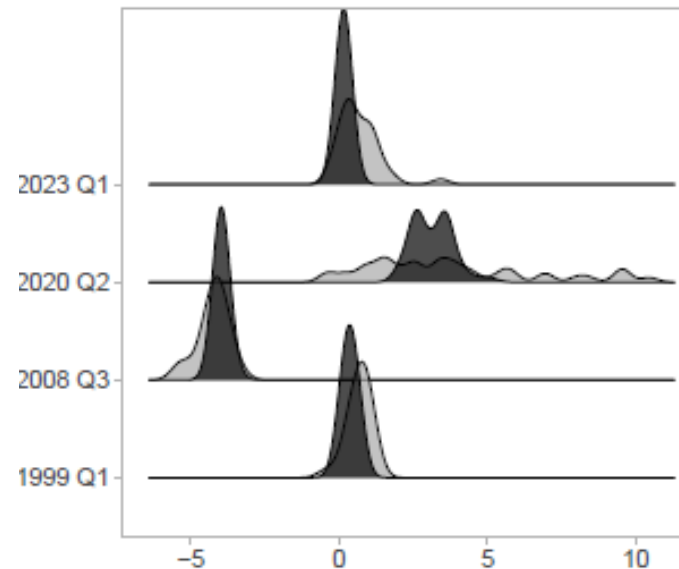
# Comment 1 -> What is the target? 2 possible interpretations

B) Achieve higher accuracy

- Then SPF might not be the right benchmark.

- Compare with VARs, ML models, or Fed projections.
  We can also inform VARs/MLs models about the past median SPF
  Can we improve the AIs SPFs providing to them their own error at t-1?
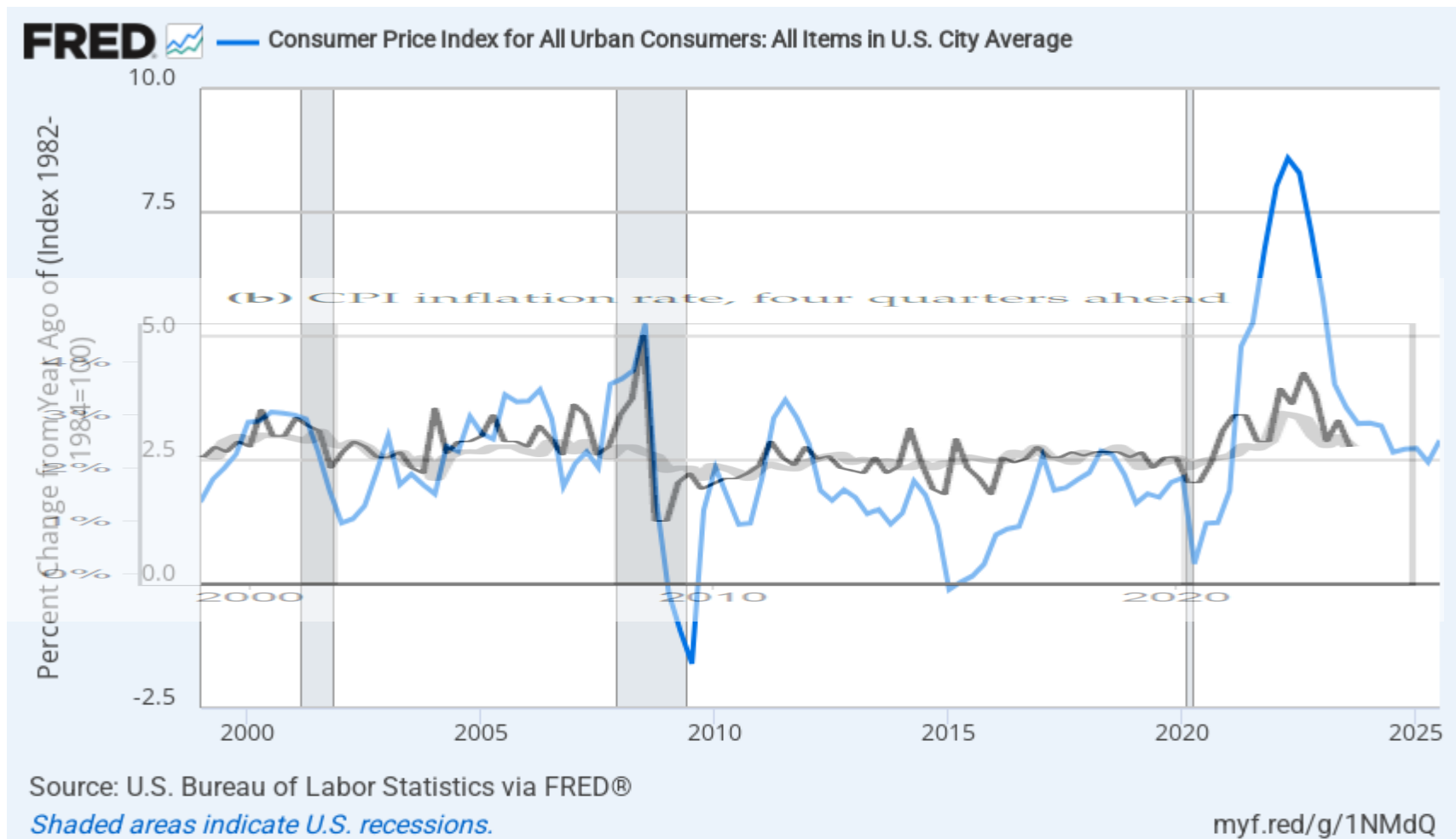
- Recursive MAEs over time could show model stability.

Good forecasters?
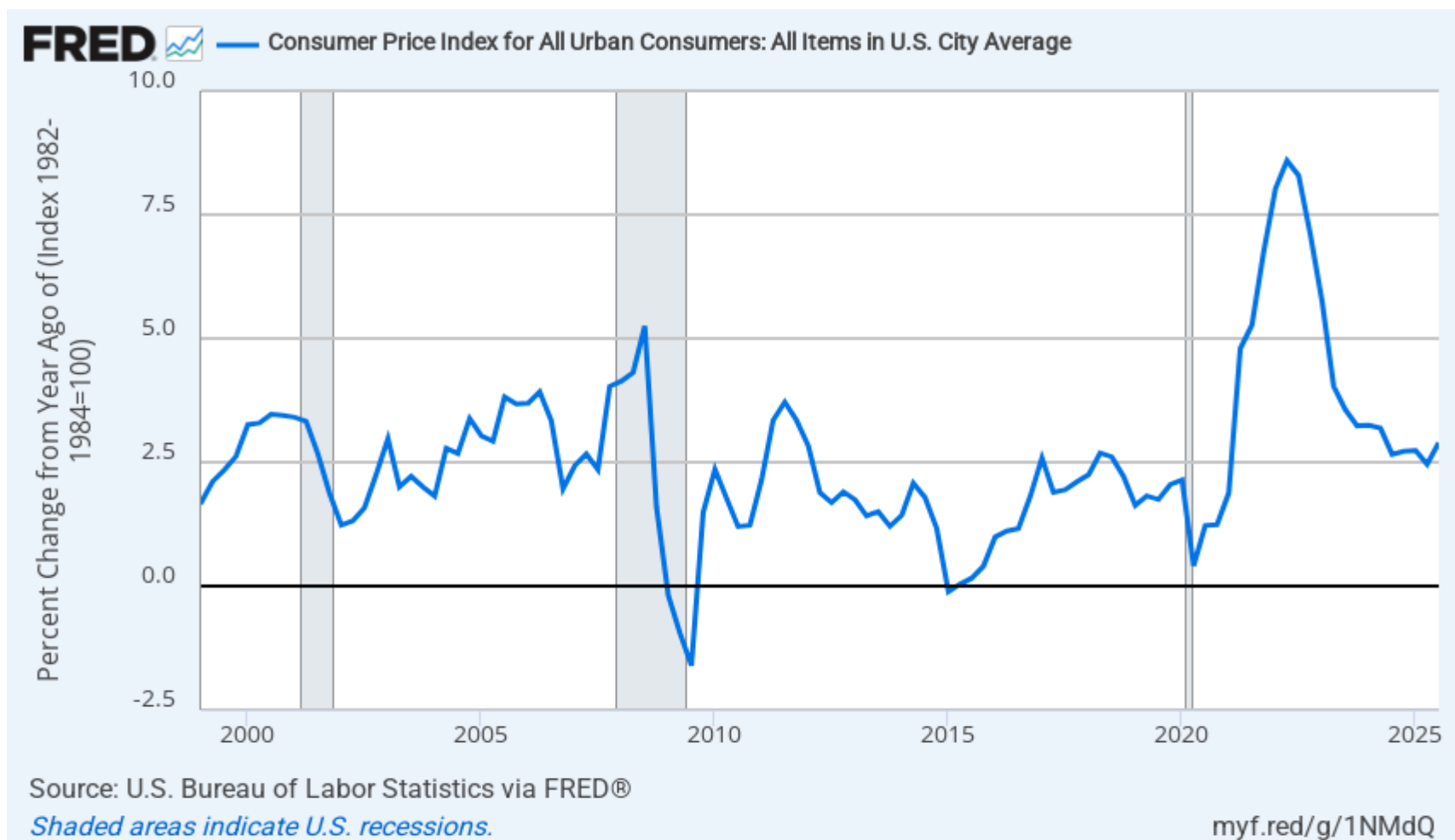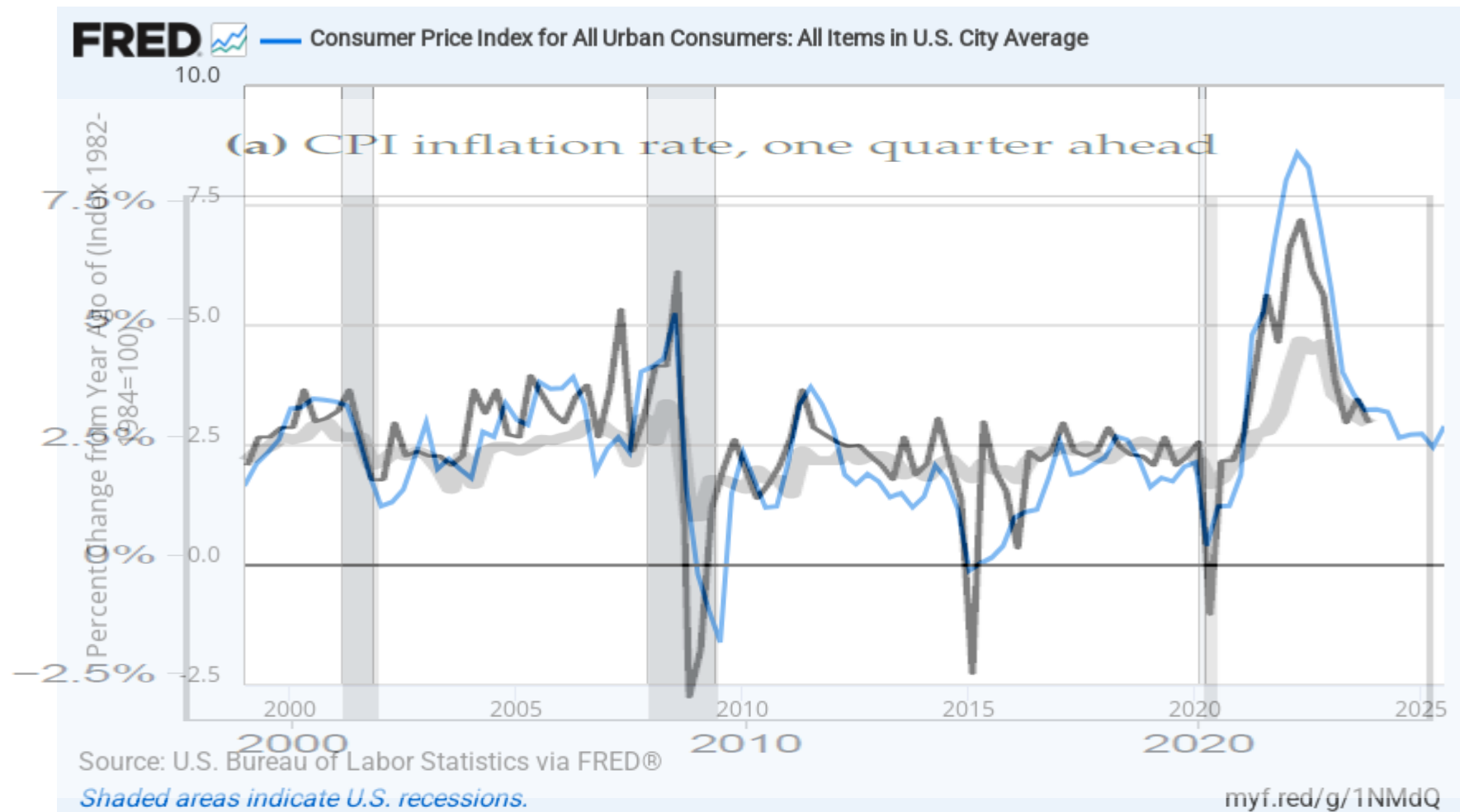
Good forecasters?

Good forecasters?

Comment 2 -> Individual Characteristics

- Human SPF: median underreacts to news (Bordalo et al., 2020); individuals overreact (Coibion & Gorodnichenko, 2015). Both violate the full-information rational expectations (FIRE) benchmark.
Do AI-SPFs replicate these asymmetries? The AI median seems to overreact. What about individual AI SPFs?

 - Who are the "best AI forecasters"?
Are education, institution type, or tenure driving accuracy?
Are characteristics capturing meaningful heterogeneity or adding structured bias?
Why does adding these characteristics outperform the median-only prompt?

Comment 3 -> Data leakage/Look ahead bias!

Authors carefully mitigate temporal leakage (real-time datasets, explicit prompt constraints).

Yet, any model estimated ex post (like VARs with full sample) embeds historical information.

Indeed, when you do a real real-time in 2024, then results in terms of accuracy are comparable. We need more evidence…

Comment 4 -> Policy Implications and Potential Extensions

Could AI-SPF serve as an early-warning tool for inflation expectations?

Simulations of communication strategies (FOMC-type experiments).

Hybrid "human + AI wisdom" ensembles (Schoenegger et al., 2024).

Concluding Remarks

Innovative and policy-relevant framework.

Opens the door to combining empirical forecast data with LLM-driven behavioral models.

Many promising extensions:
cross-country SPF replication, ECB SPF, expectation formation under alternative monetary regimes, FOMC interest rate simulations ;-)
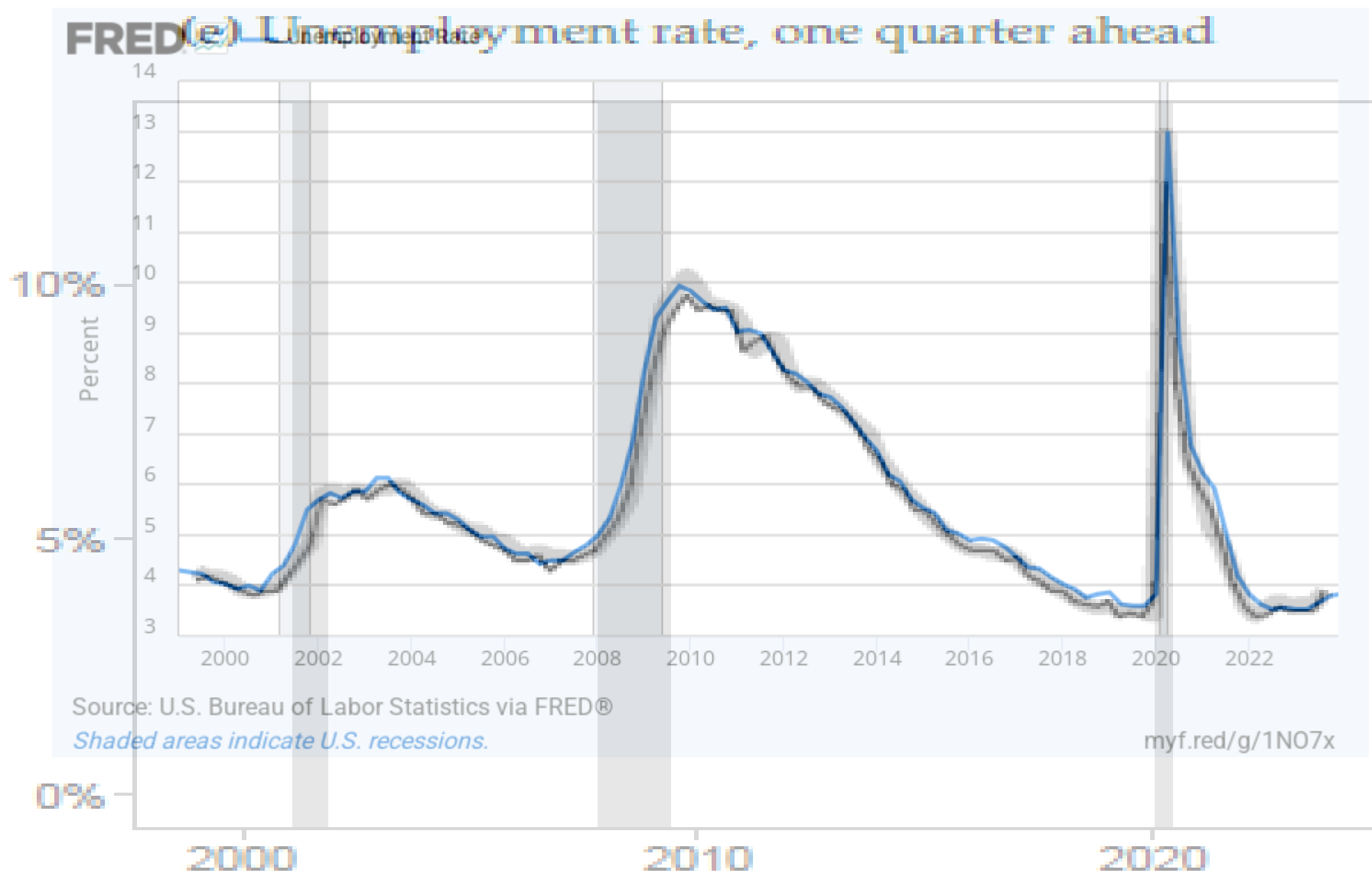
NBS is also represented by

Comment 4: Data leakage

Comment 4: Good forecasters?

Comment 4: Good forecasters?