Discussion of “Dual Interpretation of Machine Learning Forecasts.”

By Karin Klieber, Philippe Goulet Coulombe, Maximilian Goebel

13th ECB Conference on Forecasting Techniques, Frankfurt

Galina Potjagailo

Bank of England*

24 March 2026

*Disclaimer: The expressed views are my own and not necessarily those of the Bank of England (BoE).

What the paper does

- **Discusses an alternative intuitive way to interpret predictions**:
  - proximity: any prediction can be written as a **weighted sum of training observations** (data points) $\hat{y}_j = w_j y \quad \forall j \in$ Test Sample
  - correlation: rather than as a function of predictors/features, $\hat{y}_j = X_j \hat{\beta}$
- **“Data portfolio” weights** $w_j = K_j(\mathbf{K} + \lambda I)^{-1}$
  - reflect **proximity** between current and past economic conditions
  - $K_j = X_j X'$ measures **similarity between the new observation** j **and each training point** i - similar past economic states receive more weight
- For **linear models (e.g. Ridge)**:
  - portfolio weights can be **directly recovered - dual interpretation**
- **Paper shows that this is also the case for general ML models**:
  - requires additional steps or structure for some (e.g. linear final layer in NN)

Relevance for macro & empirical applications in the paper

- **Why the dual interpretation is useful in macro**:
    - time structure makes "similar past states" economically meaningful
    - predictions can be read as weighted averages of historical episodes
- **Empirical applications**:
    - FRED data, 1961-2024, quarterly and monthly
    - forecasts of:
        - unemployment and GDP growth during 2008
        - inflation in 2021–22
        - recessions as classification prediction exercise (monthly)
    - across models: Ridge, Random Forests, Neural Networks, HNN, Boosted Trees, Factor-augmented AR

What I like about the paper

- The **"historical nearest-neighbour" interpretation** is intuitive for policy and set out very clearly in the paper!
- It aligns well with how **policy makers think**:
  - reasoning in terms of comparable past episodes, rather than coefficients
  - easier to communicate to broader audiences
- **Relevant especially today**:
  - effects of ongoing oil price shocks due to war in Iran
  - which past episodes are most informative (1970s vs. 2022 vs. others)?
  - role of episodes that also featured economic slack
- More broadly, it **expands the set of interpretability tools**:
  - different representations are useful in different contexts
  - complements (rather than replaces) coefficients and Shapley values!

Comment I: Does the approach provide a distinct advantage for complex / opaque models?

- The interpretation identifies **which past episodes matter**, but not **why**:
  - why does the model place weight on a given period?
  - still requires additional tools (e.g. Shapley values or structure) to understand which features drive similarity?
- **Data constraints remain?**
  - large N, short T still requires shrinkage or factor structure
  - unclear how much the dual view alleviates this
- **Discuss broader scope of applicability**:
  - can dual interpretation also be useful in **small linear models** (e.g. hybrid Phillips curve regression)?
  - what about **VARs**? Goulet Coulombe and Klieber (2025) discusses this for local projections and past shocks - also interesting in terms of states, not only shocks?

Comment II: Which outputs are the most interesting for economic applications?

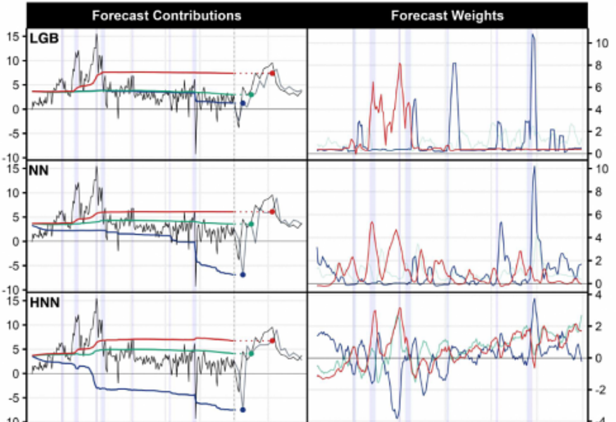
- **Range of outputs**
  1. **time series of w_j - directly, MA, or cumulative sum** - **main focus in the paper**
  2. contributions $c_{ij} = w_{ij} y_i$ - **of more interest, since varying with forecast?**
  3. historical importance of training observation i across OOS observations - **of even more interest, since varying with forecast and over time?**
  4. forecast concentration
  5. short position - also worth discussing more since it can inform on degree to which models read asymmetric information - advantage of non-linear models?
  6. leverage
  7. turnover

Output 1: time series and cumsum of proximity weights (focus in paper)

Cumulative sums not very intuitive

Weights series intuitive in some cases, erratic in others - and independent of the prediction (similar weights across horizons)

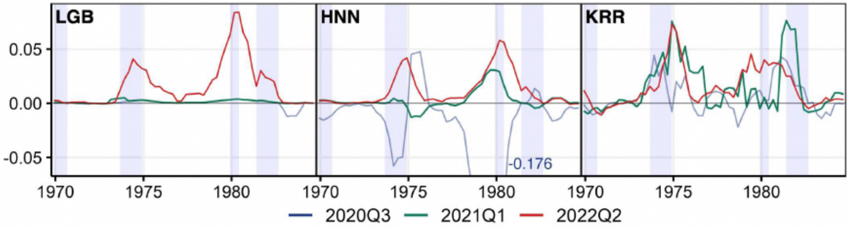
Figure 2: Dual Interpretation of Inflation Predictions ($h = 1$)



Output 2: MA forecast contributions

More telling - but only shown in one chart and for a sub-sample

Figure 3: An Alternative View: Moving Average Forecast Contributions for Inflation ($h = 1$)

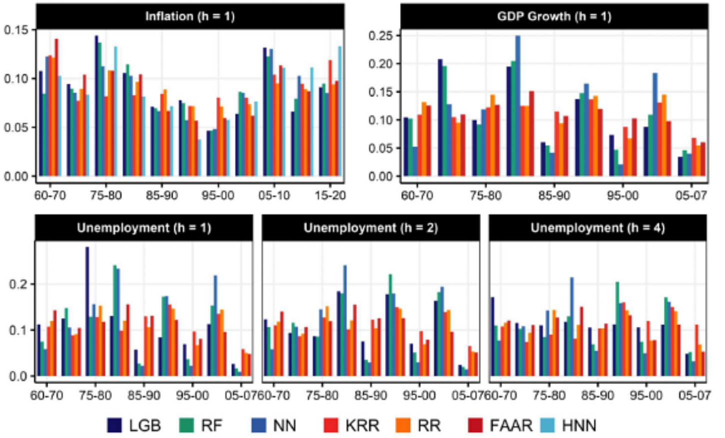


Notes: The figure presents c_{ji} as a moving average of four quarters. Lavender shading corresponds to NBER recessions.

Output 3: Historical importance of training observations for OOS observations

Even more telling - but only shown in appendix! Could be computed over an entire OOS sample (and sub-samples, to assess evolution over time) rather than only for single OOS observations

Figure 12: Overall Historical Importance



Comment III: Empirical applications

- **Empirical design feels fragmented**:
  - four applications, with some figures mixing them
  - difficult to follow and compare results across exercises
  - many models considered - are all are needed to make the point?
- **Limited forecast evaluation**:
  - forecasts shown for specific episodes (e.g. 2008, 2022)
  - no systematic out-of-sample evaluation over time
- **Link between interpretation and performance**:
  - unclear which model performs best
  - more intuition needed on how weighting certain periods relates to forecast accuracy

Suggestions for empirical exercise

- **Focus on a single application**:
  - e.g. inflation (timely, fewer challenges around revisions)
  - one other could maybe go in appendix
- **Step 1: Systematic forecast evaluation**:
  - run out-of-sample exercise over expanding or rolling window, e.g. since 2000
  - establish relative model performance over time
- **Step 2: Interpretation over time**:
  - analyse historical importance (Output type 3) across sub-periods
  - do models shift attention across training episodes (e.g. pre- vs post-2020)?
- **Step 3: Targeted case studies**:
  - zoom in on specific forecasts (e.g. 2022) to illustrate weight shifts - as already done

Great paper! Looking forward to final version

References i

Goulet Coulombe, P. and Klieber, K. (2025). Opening the black box of local projections. ECB Working Paper No. 3105.