Referee Report for

Weekly Inflation Forecasting: A Two-Step Machine Learning Methodology

General Response to Referee Comments

I would like to express my sincere appreciation to the referees for their careful reading of the manuscript and their constructive suggestions. In response to the valuable comments and critiques received, *I undertook a substantial revision of the research article*. The revised version incorporates several key enhancements and methodological refinements aimed at improving both the clarity and rigor of the analysis.

In particular, the following major changes were made:

1. Extension to High-Frequency Forecasting:

The original analysis, which focused solely on weekly inflation nowcasting, has been expanded to include both weekly and daily forecasts. This extension broadens the scope of the study and reinforces the practical relevance of the proposed methodology for real-time economic monitoring.

In fact, the title of the research article changed to *"High-frequency Inflation Forecasting: A Two-Step Machine Learning Methodology"*

2. Enhanced Two-Step Machine Learning Methodology:

The methodological framework was significantly restructured to improve clarity and robustness. The revised version now presents a more detailed and technically grounded explanation of the two-step approach, emphasizing its theoretical foundations, statistical validation through Kolmogorov–Smirnov tests, and practical implementation strategies for mixed-frequency data modeling.

3. Greater Conciseness and Coherence in Results Presentation:

The results section has been thoroughly revised to improve the presentation of key findings. Tables and figures have been updated with clearer annotations and supporting notes, and the structure of the empirical analysis has been streamlined to emphasize the most relevant results.

These revisions address the concerns raised by the referees regarding clarity, methodological soundness, and the interpretability of results. The updated version aims to offer a more compelling and technically rigorous contribution to the literature on high-frequency inflation forecasting in data-constrained environments.

I now provide a detailed, point-by-point response to the referee comments below.

Major comments

- Throughout the paper the distinction between out-of-sample and in-sample predictions seems to be blurry and difficult to follow. As a consequence, it is difficult to understand and interpret the paper's results. The paper would benefit from being much clearer on this:
    a) For instance, are the results in table 4 on page 16 based on in-sample or out-of-sample evaluations?
    b) How was figure 1, titled "Forecast interval…", constructed if only 20% of the sample are not used for estimation?! This can hardly be forecasts?!
    c) What about table 4? What is the evaluation sample here?!
    d) Etc…

I have taken care to clearly distinguish between in-sample and out-of-sample evaluations. Specifically:

a) Table 4 (now Table 1): The performance metrics reported in this table are based on out-of-sample evaluations. The monthly dataset is randomly split into 80% for training and 20% for validation, and all metrics (e.g., MSE, R^2) are computed on the 20% validation set. This methodological choice (random split for

training and validation sets) departs from the traditional temporal split, justified by the limited (monthly) sample size and the primary objective of this study: generating high-frequency inflation forecasts that effectively serve as interpolations within monthly intervals and real-time nowcasts. A random partitioning strategy, spanning the entire available sample period, ensures that the training set adequately captures diverse economic conditions and variability throughout the entire time series.

b) Figure 1 – "Forecast interval…": The forecast interval is not meant to represent formal statistical confidence bounds but instead serves as a reference for the range of variability across individual model predictions, highlighting the degree of consensus or divergence among models.

c) Evaluation Sample: To reiterate, the evaluation sample throughout the analysis refers exclusively to the 20% validation set, which is entirely held out during training. This design allows for a simulation of forecast performance on unseen data.

These distinctions and justifications are now presented in the revised manuscript.

- The paper would benefit from a more formal testing procedure of predictive performance, for instance, à la Giacomini and White (2006) "Tests of Conditional Predictive Ability".

I agree that formal predictive ability tests, such as the one proposed by Giacomini and White (2006), offer a rigorous framework for comparing competing forecast models. However, I did not incorporate this testing procedure in the revised version of the paper for two main reasons.

First, the primary goal of the study is to introduce and validate a high-frequency forecasting methodology for inflation in data-constrained settings—rather than to conduct formal hypothesis testing between forecast models. The paper focuses on demonstrating the feasibility and utility of transferring monthly-trained models to weekly and daily frequencies for real-time nowcasting, interpolation, and data augmentation.

Second, the application of Giacomini and White (2006) requires rolling or recursive out-of-sample forecasts over a sequential time horizon. This setup is not directly compatible with the 80/20 random split for training and validation sets in the study, which was selected to ensure representativeness and robust model validation across diverse economic conditions in the absence of abundant high-frequency inflation data.

That said, I acknowledge the value of incorporating conditional predictive ability tests in future extensions of this work, particularly when more extensive time-series data at higher frequencies become available. I now mention this in the revised manuscript as a potential avenue for further methodological refinement.

- The author could evaluate the "weekly" forecasts against the monthly forecasts to see whether the using of weekly data provides any additional information, i.e., whether there is an informational advantage when using more recent information; this is akin to a nowcasting exercise.

In the revised version, I directly address this point by assessing the informational advantage of using weekly (and daily) data for nowcasting purposes. Specifically, Sections 3.4 and 3.5 detail the application of the final monthly-trained model to generate daily and weekly year-on-year (y-o-y) inflation forecasts, using high-frequency feature inputs.

These sections provide visual demonstrations of the model's ability to capture intra-month inflation dynamics. The weekly and daily forecasts are shown to be statistically consistent with the monthly CPI inflation series, while also capturing meaningful within-month fluctuations. These higher-frequency estimates serve as early indicators of potential inflationary shifts, effectively anticipating changes before official monthly statistics are released.

This analysis demonstrates that weekly forecasts do indeed provide additional, timely information beyond what is available from monthly data alone. The model therefore supports real-time inflation nowcasting,

within-period interpolation, and functions as a data-augmentation strategy—bridging temporal gaps in official reporting and enhancing the granularity and responsiveness of inflation monitoring for economic analysis and policy design.

- Figure 3 would benefit from additionally plotting monthly forecasts, to see to what extend the weekly forecasts did better

The revised manuscript does not include that Figure.

- The equivalence between monthly and weekly values postulates in 2.1 and 2.2 hinges on the assumption of linearity of the model in the predictors. However, regression trees, as in the random forecasts or the Extra Tree Regressor case, are non-linear in the predictors because they are based on interaction effects of variables. Therefore, the linearity assumption fails in that case – that should be discussed.

In the revised manuscript, Section 2.1 provide a theoretical justification for applying monthly-trained models to weekly/daily data based on the linearity of expectations and aggregation operators. However, this justification is intended to support the logic of transferring feature construction across time frequencies—not to assume or impose linearity in the forecasting models themselves.

As clarified in Section 2.1, the assumption pertains to the aggregation of features and target variables (i.e., the use of means from daily observations to construct weekly and monthly variables), ensuring their distributions are comparable under the assumption of consistent daily data. The Kolmogorov–Smirnov tests, also introduced in the revised version, empirically validate this assumption by confirming distributional equivalence between daily/weekly and monthly aggregated variables.

Importantly, the revised manuscript acknowledges and embraces non-linear relationships in the modeling phase. As described in Section 2.5, the forecasting framework deliberately includes non-linear algorithms such as Random Forests and Extra Trees. These are not expected to preserve linearity assumptions, but are selected precisely to capture interaction effects and complex dynamics in the high-frequency inflation data.

Thus, the linear aggregation assumption underpins the data preparation logic, while the machine learning stage remains flexible and model-agnostic, accommodating both linear and non-linear specifications. This separation of assumptions and stages is now made clearer in the revised text.

- On page 12, the comparison of MSEs reported in other papers for forecasting inflation series of other countries is completely redundant since these are different inflation series with potentially other DGPs.

In the revised manuscript, that comparison is excluded.

- Are the expectations in 2.1 and 2.2 conditional or unconditional?

Unconditional.

- What about seasonal effects within a month or week?

In the revised manuscript, intra-period seasonality is partially addressed through the inclusion of lagged values of the target variable. As described in Section 2.3 (Features), I incorporate multiple lags of the CPI inflation—specifically at 1, 2, 3, 6, 9, and 12 months—which help the model capture monthly and annual seasonal patterns and the persistence inherent in inflation dynamics.

While these lagged variables contribute to modeling broader seasonal cycles, I acknowledge that the current framework does not explicitly incorporate intra-month or intra-week seasonal effects (e.g., paydays or

weekend price shifts). This is noted as a limitation in the revised version, and I identify it as a promising direction for future research, where calendar-based features or dummy variables could be used to more directly capture such short-term seasonal behaviors.

This clarification is now reflected in Sections 2.3 and 4 of the manuscript.

- What is the unit of the left-hand-side variable? Normalized CPI in levels? That isn't clear from the text.

In the revised manuscript, I clarify that the target variable is the year-on-year (y-o-y) inflation rate, calculated from the Consumer Price Index (CPI) and expressed in percentage points. This is explicitly described in Section 2.2 (Target Variable).

- All tables and figures need notes that make them self-comprehensive, i.e., that the reader can understand the tables/figures without having to look for additional information in the text

Noted.