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

- The paper compares forecast models using weekly data, but a key concern is the missed opportunity to utilize daily data directly in the forecasting models. By leveraging daily data, the comparisons might have been more insightful, potentially yielding even more accurate weekly forecasts. This aspect warrants further exploration to enhance the methodology.

I have extended the analysis to include both weekly and daily inflation nowcasts, thereby addressing the methodological gap identified in the original submission. The revised version now explicitly integrates daily feature data into the high-frequency forecasting pipeline, using the same two-step machine learning framework developed for weekly predictions.

This enhancement is detailed in Section 2 (Methodology) and further implemented in Section 3 (Empirical Application). Daily forecasts are generated using the final monthly-trained model applied to daily features, and their temporal dynamics are evaluated alongside the weekly and observed monthly inflation rates.

This extension not only increases the model's granularity and real-time applicability but also allows for a more comprehensive comparison across different frequency levels. The results indicate that daily forecasts capture intra-month inflation dynamics, adding significant informational value, especially in periods of heightened volatility. This change aligns directly with the referee's suggestion and substantially strengthens the empirical and methodological contribution of the paper.

- The paper's explanation of the two-step methodology, particularly regarding estimation and projection windows, is somewhat unclear, making it challenging to follow the precise steps of the forecasting process. More clarity in the presentation of these technical details would help readers better understand and evaluate the robustness of the approach.

I have substantially clarified and reorganized Section 2.1, which now presents the two-step methodology in a structured manner under the titles: (1) Mixed-Frequency Data Preparation and (2) Machine Learning Prediction Pipeline.

Step 1: Mixed-Frequency Data Preparation

This step describes how high-frequency features (daily and weekly) are aggregated into monthly equivalents to align with the target variable (monthly y-o-y CPI inflation). This allows for a unified training dataset despite initial frequency mismatches. The section formalizes the aggregation process and introduces Kolmogorov–Smirnov (K–S) tests to validate the assumption of distributional equivalence between high-frequency and monthly-aggregated data. This validation is essential for ensuring that a model trained on monthly data can generalize appropriately when applied to weekly or daily inputs.

Step 2: Machine Learning Prediction Pipeline

This step outlines the process of training machine learning models on the monthly dataset and using the final model to produce high-frequency inflation forecasts. It details the procedures for feature selection, hyperparameter tuning, model validation, and ultimately, the generation of weekly and daily nowcasts. The forecasting model, trained exclusively on monthly data, is applied to high-frequency (weekly and daily) features to obtain temporally disaggregated predictions.

To address the concern about estimation and projection windows, I now provide a clear explanation of how training, validation, and prediction samples are defined and used. The training set includes 80% of monthly observations, the validation set comprises the remaining 20% (used for out-of-sample evaluation); justification is provided for using random sampling instead of a purely temporal split due to the interpolation objective. And prediction sets involve applying the final model to weekly and daily inputs, aligned with their respective dates.

Additionally, the revised version emphasizes that high-frequency forecasting in this study is understood as nowcasting or interpolation—real-time prediction ahead of official releases or within-period estimation using mixed-frequency indicators—while also functioning as a data-augmentation strategy.

Therefore, beyond traditional projection windows, the central goal is to demonstrate the model's capacity to:

- Support real-time inflation nowcasting in data-scarce environments;
- Perform intra-period interpolation to reconstruct plausible inflation paths within each month;
- Act as a data-augmentation tool to enhance temporal granularity and fill gaps in official statistics.

- The methodology holds promise for weekly inflation forecasting, but the paper could benefit from clearer exposition and more comprehensive demonstrations. Improving the statistical evaluation of model performance and considering daily data comparisons could significantly bolster the paper's impact.

The revised version includes several enhancements to strengthen both the clarity and depth of the statistical evaluation. First, the exposition of the methodology has been entirely restructured and expanded in Section 2, now distinguishing between the Mixed-Frequency Data Preparation and the Machine Learning Prediction Pipeline. Each component is presented with step-by-step detail, including feature aggregation, validation of distributional equivalence (via Kolmogorov–Smirnov tests), model training, cross-validation, and prediction. The revised Figure 1 provides a concise visual representation of the entire forecasting pipeline.

Second, in terms of model performance evaluation, I have extended the empirical analysis:

- ***Comparative Evaluation Against Econometric Benchmarks:*** The revised version includes forecast accuracy comparisons against Bridge Equation and MIDAS regression models, two widely used econometric approaches for mixed-frequency forecasting. These benchmarks are introduced and evaluated in Section 3.3, and results show that the proposed two-step machine learning methodology outperforms both benchmark models in out-of-sample accuracy metrics (MSE and R^2).
- ***Robustness Across Multiple Feature Selection Strategies:*** To enhance parsimony and avoid overfitting, I implement four distinct feature selection methods: (1) correlation filtering, (2) principal component analysis (PCA), (3) L1 regularization (Lasso), and (4) random forest importance scores. Each selection method generates a different subset of predictors, and model performance is assessed across these subsets to confirm robustness. These are detailed in Section 2.4 and evaluated comparatively in Table 1 (Section 3.2).
- ***Algorithm Diversity and Model Tuning:*** Eight forecasting approaches are tested, including Ridge, Lasso, Random Forest, Extra Trees, Gradient Boosting, AdaBoost, and two ensemble combinations. All algorithms are fine-tuned using 5-fold cross-validation, and results are benchmarked on a (out-of-sample) validation set. The final model—Ridge regression trained on Lasso-selected features—emerges as the top-performing single model, with additional gains obtained through ensemble forecasts (e.g., WAM-Best).
- ***Daily Forecasting Demonstration:*** As per the referee's suggestion, the analysis has been extended to include daily nowcasts in addition to weekly forecasts. Section 3.4 presents the daily forecasting results, which illustrate the model's capacity to detect intra-month inflation dynamics with high temporal resolution.

- The paper could discuss endogeneity issues and consider incorporating a more diverse set of economic and financial indicators to enhance the model's predictive power.

While the current version of the paper focuses on high-frequency predictors available in the Bolivian context—such as wholesale prices, commodity prices, Google Trends data, and selected financial variables—I recognize the importance of expanding the set of indicators to further enhance predictive accuracy.

In the Conclusion (Section 4), I explicitly acknowledge this limitation and identify future research directions that could address it. In particular, I propose incorporating richer and more granular data sources, including high-frequency financial transactions, payment system data, and additional macroeconomic indicators as they become available.

- It is not clear if the analysis is an in-sample or out-of-sample of the models at a monthly frequency, if the latter, what is the out-of-sample forecast period? One-month-ahead forecast as mentioned for only one model? Please clarify before introducing empirical analysis.

I now provide a clearer explanation of how training, validation, and prediction samples are defined and used. Specifically, the monthly dataset is partitioned into an 80/20 split: 80% of the observations are used for training (in-sample), and the remaining 20% serve as the validation set (out-of-sample). The out-of-sample evaluation mimics a situation where one-month-ahead (or multi-month) nowcasts are estimated using features available in months without official inflation reports.

Additionally, the final model is applied to aggregated weekly and daily inputs—each aligned with their respective dates—to generate high-frequency forecasts. In this study, high-frequency forecasting is understood as nowcasting or interpolation, while also functioning as a data-augmentation strategy. The rationale for using random sampling instead of a purely temporal split is discussed in the manuscript, as it better accommodates the interpolation objective by ensuring that the training set captures a wide range of economic conditions over the entire sample period. These clarifications are detailed in Sections 2.5 and 3.1 of the revised version.

- The analysis overlooks intra-month seasonal patterns, such as paydays and weekends.

In the current version, the methodology aggregates daily observations into weekly and monthly measures, which smooths out finer intra-month seasonal effects like paydays and weekends. However, I acknowledge that explicitly accounting for these intra-month patterns could potentially enhance the model's predictive power. This aspect is not incorporated in the present analysis, but noted as an interesting avenue for future research.

- The document assumes that (i) the monthly average of daily observations equals the inflation accumulated over 30 days, and (ii) the weekly average of daily observations matches the inflation accumulated over 7 days. While not necessarily incorrect, this assumption raises concerns related to the previous point. Using sums of daily observations might be more appropriate than relying on averages.

I clarify that the choice of using averages instead of sums is fully consistent with the construction of Bolivia's Consumer Price Index (CPI). In Bolivia, the CPI is computed as the average of weekly-compiled prices over the month, not as a sum of daily price changes. Therefore, aggregating daily data through averages better reflects the official methodology and the underlying price dynamics captured in the CPI.

Using averages ensures that the aggregated values are directly comparable to the published CPI figures and maintain the same scale and interpretation. Conversely, summing daily observations would not align with the official computation and could introduce biases or misrepresent the inflation measure. This methodological decision is now discussed in the revised version (see Section 2.1), and I emphasize that our approach is designed to mirror the official process for calculating inflation in Bolivia, thereby enhancing the validity of our forecasts.

- There exists a large literature on interpolation that could offer crucial insights and enhance this study. For example, methodologies exist that impose constraints to ensure that the sum of weekly inflations (unobserved) aligns with the monthly inflation (observed).

I recognize that the interpolation literature offers many methodologies that impose constraints to align disaggregated forecasts with aggregated observations. In our study, however, the approach is grounded in the specific construction of Bolivia's CPI, which is calculated as an average of weekly-compiled prices over the month. Thus, using averages directly mirrors the official methodology and preserves the interpretability of our forecasts.

Beyond focusing solely on the interpolation literature, our analysis also accounts for established mixed-frequency techniques. In particular, I compare the performance of the two-step machine learning methodology against Bridge Equation and MIDAS regression models. These benchmarks serve to demonstrate the advantages of this approach in capturing high-frequency dynamics and providing real-time nowcasts.

- It's important to exercise caution when using variables in levels, particularly when normalizing (z-score) variables with a unit root. This could explain why the autoregressive term in Lasso is so significant.

While caution is warranted when normalizing variables with potential unit roots, this concern is less critical in the machine learning framework.

- The document does not specify how cross-validation was conducted. In a time series context, independent k-fold validation isn't feasible. The document seems to neglect this aspect and omits an adequate description.

In the revised version, I clarify that we use standard 5-fold cross-validation based on a random split of the monthly dataset. This approach is justified by the interpolation objective of the study, which focuses on generating nowcasts within the month rather than long-term forecasts. The random split ensures that the training set captures diverse economic conditions, thereby enhancing model robustness. This methodology and its rationale are now explicitly described in Section 2.5 of the manuscript.

- Dimensionality reduction (from 500 to 86 variables) is performed using simple correlations. Why not allow the algorithms to handle this process, either as an initial step before forecasting or by jointly selecting and forecasting

The revised version, beyond correlation-based feature selection, also considers alternative feature selection methods—such as PCA, Lasso regularization, and random forest importance—which allow for algorithmic selection and jointly evaluate forecasting performance. These complementary approaches provide robustness checks and demonstrate that the final results are not overly sensitive to the initial filtering method.

Minor comments

- How does the model performance compare when applied to different periods or under varying economic conditions, such as high inflation or economic shocks? Additional sensitivity analyses could strengthen the model's robustness.
- The paper could have a roadmap at the end of the introduction.
- Testing the proposed model with data from other developing countries could help validate its applicability and robustness. Exploring the impact of unexpected global events, such as the pandemic, on inflation rates could add another dimension to the study.

In the revised manuscript, I have added a brief roadmap at the end of the introduction to guide readers through the subsequent sections. Additionally, I acknowledge that assessing model performance under varying economic conditions, such as periods of high inflation or during economic shocks, would further strengthen the analysis. I have now expanded the discussion on limitations and noted that future research should test the proposed model using data from other developing countries and explore the impact of structural breaks. These points are discussed in the Conclusion sections.