Box Article 1:

E-Commerce price index prediction with time series mining and automated machine learning[1]

The BSP uses the Consumer Price Index (CPI) as its primary indicator of inflation. The CPI is generated by the Philippine Statistics Authority (PSA) using data from its Retail Prices Survey (RPS). The conduct of RPS requires substantial financial resources, extensive human labor, and may be susceptible to unforeseen circumstances. As a complement to the RPS, this study uses big data analytics and machine learning (ML) techniques to create an e-commerce-based CPI. This provides the BSP with a supplementary tool for monitoring inflation trends.

Using new data sources to examine CPI food price patterns

This study utilizes data from a major e-commerce company in the Philippines, covering the period from April 2021 to October 2024. It focuses on the food and non-alcoholic beverages group, which makes up 37.75 percent of the CPI basket. The dataset includes over 19 million price quotes for over 659,000 unique items across 16,000 product categories from 10,960 online merchants. The study highlights 480 food products that are used to forecast CPI trends.

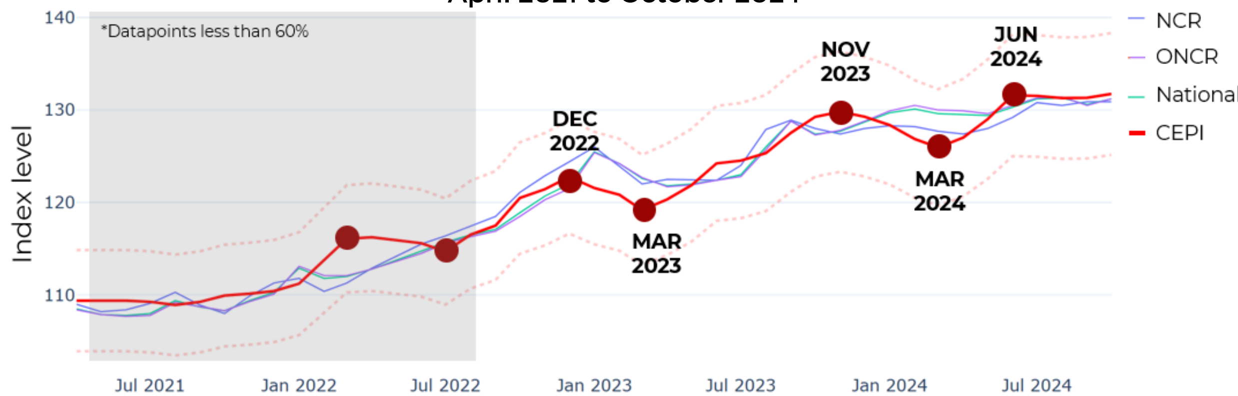
Analyzing food price trends using Machine Learning in three stages

This study follows a three-step process to create the food Composite E-Commerce Price Index (CEPI) and forecast the food CPI. The first step is data preprocessing, which ensures dataset accuracy and consistency. This includes data ingestion and validation, normalizing prices based on product sizes, standardizing merchant-level prices, detecting and removing outliers, and harmonizing category names. The second step focuses on building the CEPI using the Dynamic Time Warping (DTW) algorithm to find products with price patterns similar to CPI movements. The composite index is created by calculating the mean of the top 10 percent of products most similar to the CPI, then using linear regression to align it with CPI scales. The final step involves testing different regression-based machine learning models, such as Linear, Decision Tree, Random Forest, Ridge Regression, Huber Regression, AdaBoost, K-Nearest Neighbors, and Gradient Boosting Regression, to forecast the CPI.

CEPI strongly correlates with the official food price index, showing regional variations

The results show a strong correlation between the food CEPI and the official food CPI. Over 24-months, the correlation remains consistently high at around 88 percent across all regions. However, when extended to 36 months, the correlation varies. In the National Capital Region (NCR), the food CEPI correlates strongly with the food CPI at 81.96 percent. At the national level, this correlation drops to 69.82 percent, and further declines to 51.61 percent in areas outside NCR (AONCR). This decrease is due likely to limited online shopping data in earlier periods and the uneven adoption of e-commerce across the Philippines.

Figure 1
Official Food CPIs vs. (CEPI)
April 2021 to October 2024



Source: Bangko Sentral ng Pilipinas staff calculations

Table 1
Correlations of CEPI with Official Food CPIs

Timeframe	National	NCR	AONCR
24 months	88.15%	88.22%	88.51%
36 months	69.82%	81.96%	51.61%

* All correlations have a p-value < 0.05 (statistically significant relationship).
Source: Bangko Sentral ng Pilipinas staff calculations

The food CEPI shows noticeable fluctuations in its deviation from the official food CPI. As seen in Table 2, while the food CEPI generally follows the trends of the official food CPI, there are significant differences, particularly in areas outside the AONCR. These variations suggest a need for further investigation into the factors causing these discrepancies and their impact on inflation analysis and forecasting.

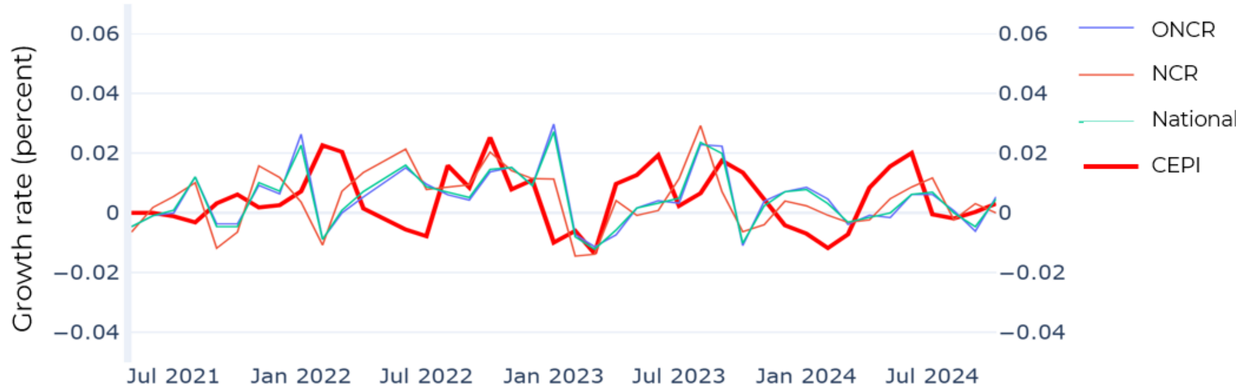
Table 2
CEPI Deviations from the Official Food CPIs

Food Indices	Standard Deviation	Average
National	2.31	4.08
NCR	1.92	3.45
ONCR	2.40	4.21

Source: Bangko Sentral ng Pilipinas staff calculations

The food CEPI also shows a slightly higher average growth rate of 0.0072 and a larger standard deviation of 0.0192, suggesting more volatility in e-commerce food prices. These findings emphasize the importance of including e-commerce data alongside the official food CPI to better understand food price trends.

Figure 2
Official Food CPI vs. CEPI (Monthly Growth Rates)
April 2021 to October 2024



Source: Bangko Sentral ng Pilipinas staff calculations

Table 3
Descriptive Statistics of Official Food CPIs and the CEPI Growth Rates

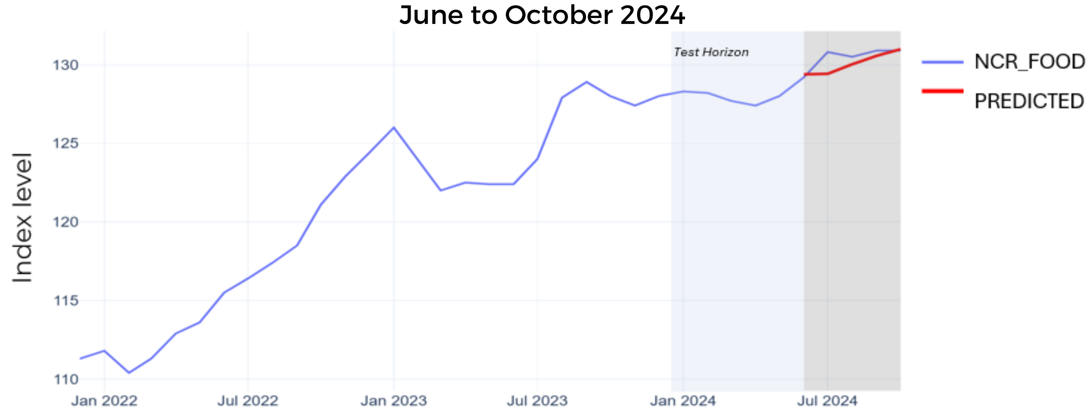
Measure	National	NCR	AONCR	CEPI
Mean	0.0060	0.0058	0.0060	0.0072
Min	−0.0121	−0.0144	−0.0113	−0.0301
Max	0.0272	0.0293	0.0298	0.0578
Std. Dev.	0.0104	0.0109	0.0108	0.0192

Source: Bangko Sentral ng Pilipinas staff calculations

Machine Learning effectively predicts food price trends with Ridge Regression

The forecasting results show that ridge regression outperforms other algorithms, achieving the highest R-squared value of 0.74 and the lowest Mean Absolute Error (MAE) of 1.269. As seen in Figure 3, ridge regression delivers reliable predictions with minimal error, as confirmed by out-of-sample validation.

Figure 3
Forecasting Results, Official Food CPI, and CEPI
June to October 2024



Source: Bangko Sentral ng Pilipinas staff calculations

Using the best-performing model (ridge regression), the study found that certain products play a key role in predicting overall food price changes. Coffee prices were the most important indicator, contributing about 8.5 percent to the prediction accuracy. Milk prices followed at 5.5 percent, and chocolate at 4.5 percent, with other common food items also contributing to the predictions. This is significant because it shows how modern data analysis can pinpoint which everyday products signal upcoming price changes. By combining different analysis tools, this approach offers a promising way to track prices alongside traditional methods. Expanding the model's data sources, such as incorporating the Department of Trade and Industry's e-Presyo, is expected to enhance the model's accuracy and potentially improve its forecasting reliability.

ENDNOTE:

1/ The authors of this box article are Carmelita G. Esclanda-Lo, Chelsea Anne S. Ong, and Gabriel A. Masangkay. The study has been accepted for publication in the Statistical Journal of the International Association of Official Statistics and will be published by SAGE Journals in the first half of 2025.