

Analysis of India's CPI over the years

Table of contents

Introduction	2
Motivation	2
Objective	3
Data Overview	3
Source	3
Contents	3
Preprocessing	4
Exploratory Data Analysis (EDA)	5
General CPI time series	5
Year-over-Year Percentage Change	5
General CPI YoY % Change	6
Vegetables YoY % Change	7
Fuel and Light YoY % Change	8
Transport and Communication YoY % Change	9
STL Decomposition	9
General CPI STL Decomposition	10
Milk and Products STL Decomposition	11
Vegetables STL Decomposition	12
Clothing STL Decomposition	13
Volatility & Risk Analysis	13
General CPI - Volatility Analysis	14
Vegetables Volatility Analysis	15
Transport and Communication Volatility Analysis	16
Milk and Products Volatility Analysis	17

Structural Breaks and Anomalies	17
Structural Breakpoints	18
Outliers and Anomalies	18
General CPI	18
Fuel and Light	20
Vegetables	21
Milk and Products	23
Forecasting	24
Why Forecast CPI?	24
ARIMA Models	24
Prophet	24
Results	25
Comparison	26
Rural vs. Urban Comparison (Case Study)	27
Conclusion & Insights	28
Overall CPI Trends	28
Food vs. Non-Food Dynamics	29
Inflation Risk & Volatility	29
Structural Shifts & Anomalies	29
Rural vs. Urban Case Study	29
Policy Implications	29
References & Citations	30

Introduction

Motivation

The **Consumer Price Index (CPI)**, which measures **inflation**, has a direct impact on monetary policy decisions, corporate pricing strategies, and household purchasing power. Even minor changes in the CPI can have a significant impact on India's economic stability and an average household, where a significant portion of spending goes towards essentials like healthcare, food, and fuel.

Over the **past decade**, price dynamics have become even more volatile and uncertain due to the COVID-19 pandemic, global supply-chain disruptions, and shifting fiscal policies. Policy-makers, analysts, and businesses can make **better plans** if they have a clear, **data-driven understanding** of these trends and can differentiate between predictable seasonal cycles and real shocks.

Objective

This project conducts a comprehensive time-series analysis of India's monthly CPI (2013-2024) with four key aims:

1. **Explore and decompose CPI** by examining YoY % changes and applying STL to extract seasonal, trend, and residual components, revealing underlying patterns and anomalies.
2. **Quantify volatility** by computing rolling standard deviations of month-over-month percentage changes, highlighting “hot” (volatile) and “calm” periods.
3. **Detect structural breaks and outliers** in the trend and remainder series to pinpoint significant inflation shifts and one-off shocks (e.g., pandemic lockdowns).
4. **Forecast future CPI** with ARIMA and Facebook Prophet models, backtesting on the most recent 12-month window to compare predictive accuracy and inform short-term inflation expectations.

Together, these analyses will transform raw CPI data into actionable insights about the drivers, risks, and likely trajectories of inflation in India.

Data Overview

Source

The primary dataset was obtained from the Kaggle repository:

All India Index Upto Feb24 (available at <https://www.kaggle.com/datasets/sumit08/all-india-index-upto-feb24>), which provides monthly CPI values for various categories from **January 2013 to February 2024**.

Contents

The dataset includes:

- A **Date** split into **Year** and **Month** columns
- A **Sector** column indicating **Rural**, **Urban**, or **Rural and Urban** CPI
- CPI indices for multiple categories such as:
 - Food-related (e.g., *Fruits, Vegetables, Cereals and products, Meat and fish*, etc.)
 - Non-food (e.g., *Transport and Communication, Clothing and footwear, Health, Education*, etc.)
- A **General index** column representing the overall CPI

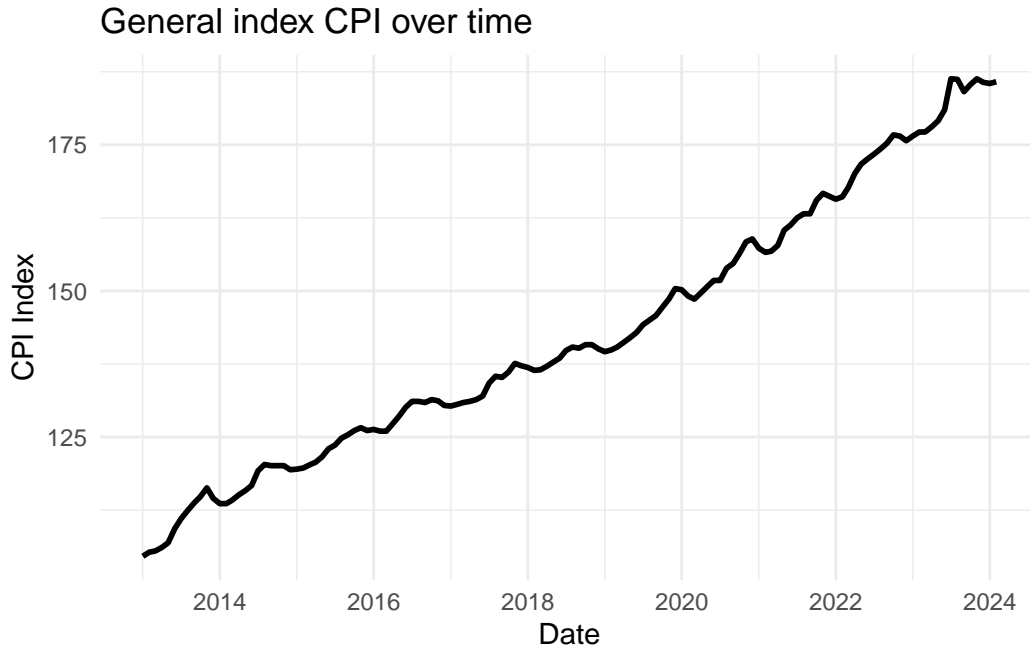
Preprocessing

- Fixed a data entry error in the **Month** column (**Marcrh** to **March** in row 45)
- Renamed category columns to replace dots with spaces (e.g., **Meat.and.Fish** to **Meat and fish**)
- Combined the **Year** and **Month** into a proper **Date** column (YYYY-MM-01) for time-series analysis
- Split the data into three separate data frames:
 - **rural_df**
 - **urban_df**
 - **combined_df** based on the **Sector** column for targeted analysis
- **Linear interpolation** was used to fill **missing CPI values** for April 2019 and April-May 2020, ensuring continuity for time series analysis. These gaps may have arisen due to reporting delays or disruptions in data collection.

Note: For brevity and clarity, the core analysis in this report focuses primarily on the *Rural and Urban* (combined) CPI data. Selected insights from the *Rural* and *Urban* segments are included in a later section to highlight key differences where relevant.

Exploratory Data Analysis (EDA)

General CPI time series



While the time series plot of the General CPI provides a broad view of how overall price levels have evolved in India, it does not fully capture the dynamics of inflation. In particular, it does not convey the *rate* at which prices are rising or falling over time. To gain deeper insight into inflationary trends and volatility, we turn to derivative measures such as the **Year-over-Year (YoY) percentage change**, which allow for clearer interpretation of price momentum and turning points.

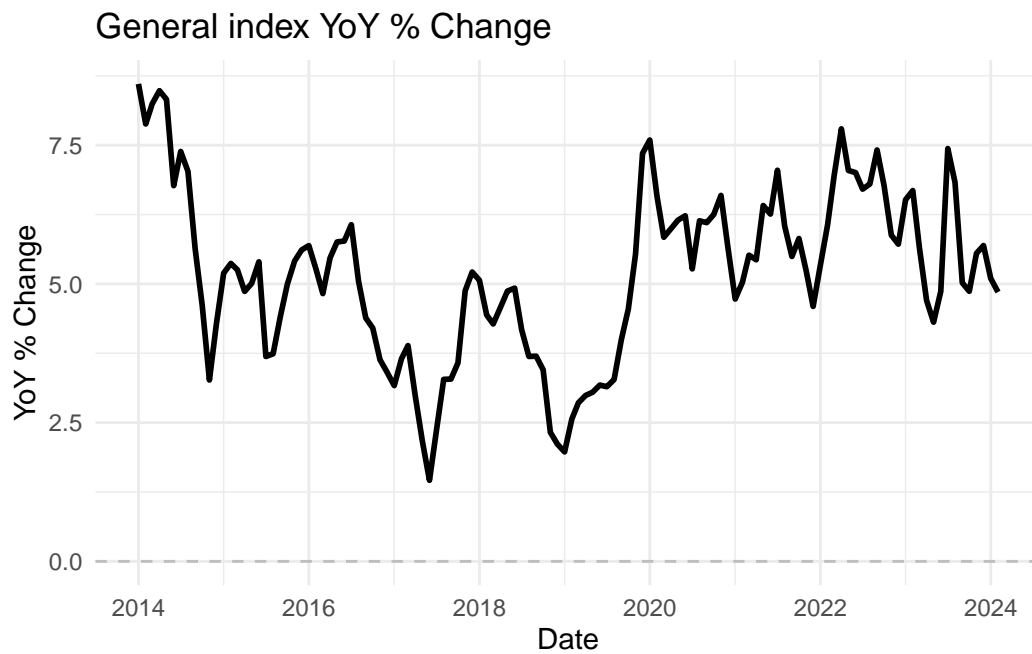
Year-over-Year Percentage Change

The **year-over-year (YoY) percentage change** measures the relative difference between the current CPI value and the CPI value 12 months earlier:

$$\text{YoY } \%_t = 100 \times \frac{\text{CPI}_t - \text{CPI}_{t-12}}{\text{CPI}_{t-12}}$$

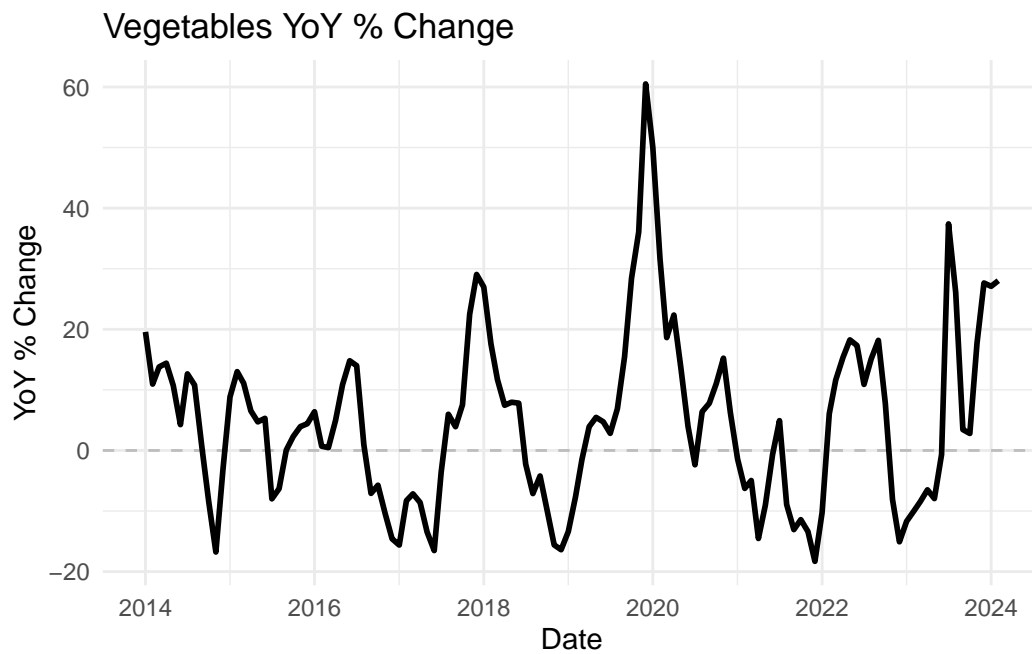
This metric highlights how the underlying inflation rate accelerates or decelerates over entire years, smoothing out short-term monthly noise.

General CPI YoY % Change



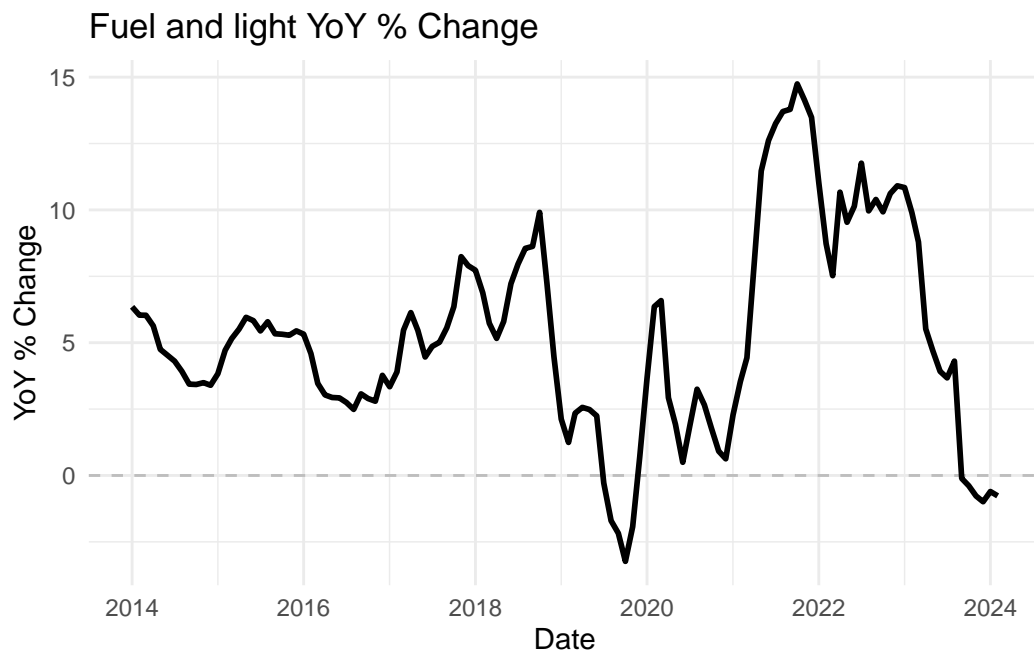
- General CPI inflation oscillates between approximately 2 % and 8 %, with notable spikes above 7 % in 2020 corresponding to COVID-19 supply disruptions and food price surges during nationwide lockdowns.
- The mid-2022 peak, coinciding with global commodity price shocks following Russia's invasion of Ukraine, highlights India's exposure to external supply-chain disruptions despite underlying domestic resilience.

Vegetables YoY % Change



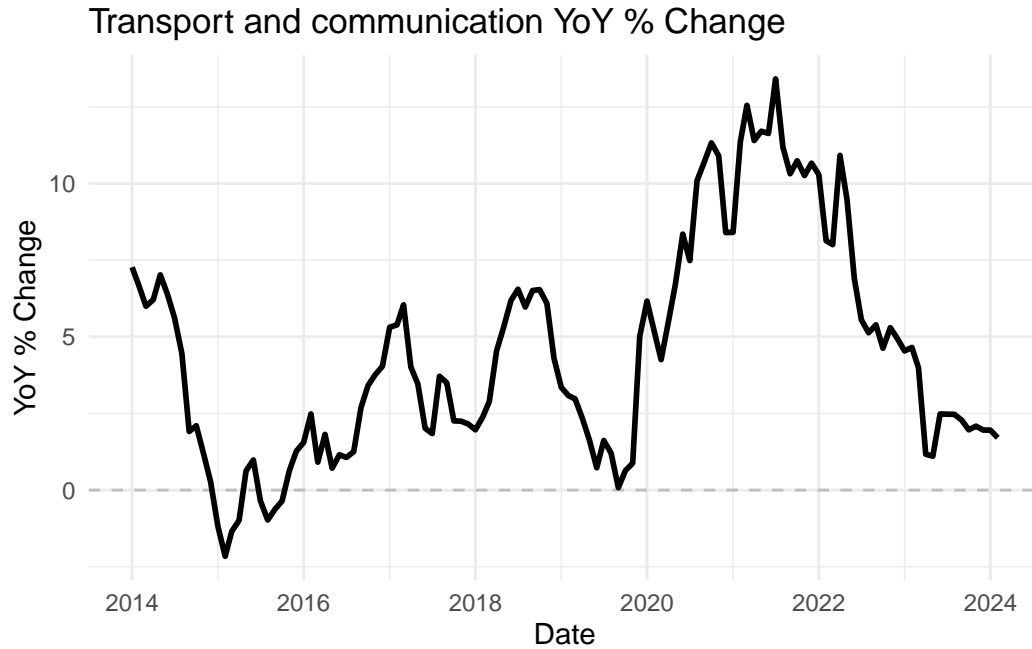
- The YoY % change in the Vegetables CPI exhibits pronounced volatility, oscillating between deflation below -15 % and inflation spikes above 40 %, highlighting its high sensitivity to supply disruptions.
- Significant surges during 2019-2020 correspond with pandemic-related supply chain interruptions and lockdown measures, followed by recurring peaks driven by erratic monsoon patterns and weather-induced crop damage.

Fuel and Light YoY % Change



- Fuel and Light CPI experienced dramatic inflation spikes above 10 % in 2021-22, reflecting the global energy price shocks that followed Russia's February 2022 invasion of Ukraine.
- A steep decline to near-zero inflation by 2023-24 illustrates how rapidly energy prices normalized as global commodity markets stabilized and supply-chain pressures eased.

Transport and Communication YoY % Change



- The Transport and Communication CPI surged from 2020 to 2022, peaking above 12 %, driven by pandemic-induced shifts toward private transportation amid COVID-19 safety concerns.
- The rapid post-2022 decline reflects the normalization of mobility patterns and stabilization of communication service pricing as pandemic restrictions were lifted.

STL Decomposition

STL (Seasonal and Trend decomposition using Loess) is a widely used technique for analyzing time series data by breaking it down into three additive components:

- **Trend (T_t):** The long-term progression or direction in the data.
- **Seasonal (S_t):** Recurring patterns or cycles that repeat over fixed periods (e.g., months or quarters).
- **Remainder (R_t):** Irregular fluctuations and residual noise not captured by trend or seasonality.

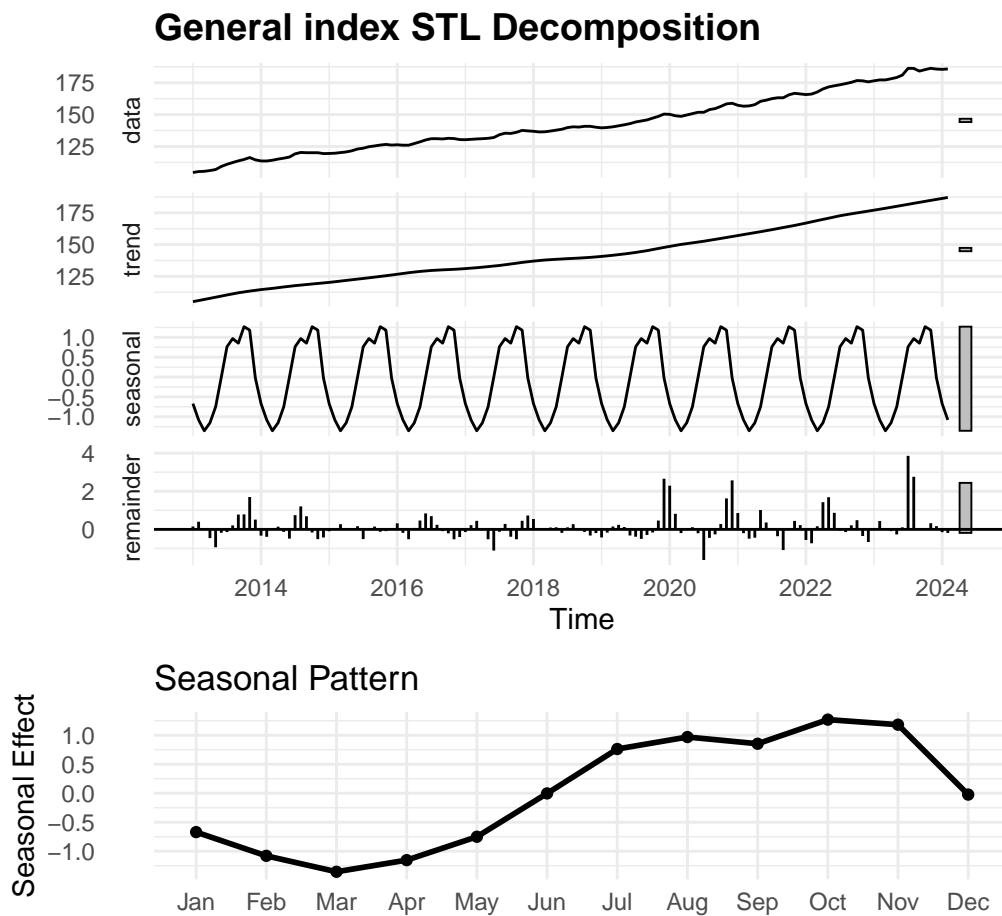
Mathematically, it can be expressed as:

$$Y_t = T_t + S_t + R_t$$

STL is particularly powerful because: - It handles both short-term and long-term seasonal patterns. - It uses **LOESS smoothing** (Locally Estimated Scatterplot Smoothing), allowing flexibility in modeling complex non-linear trends. - It can handle missing values and works well with real-world, noisy data.

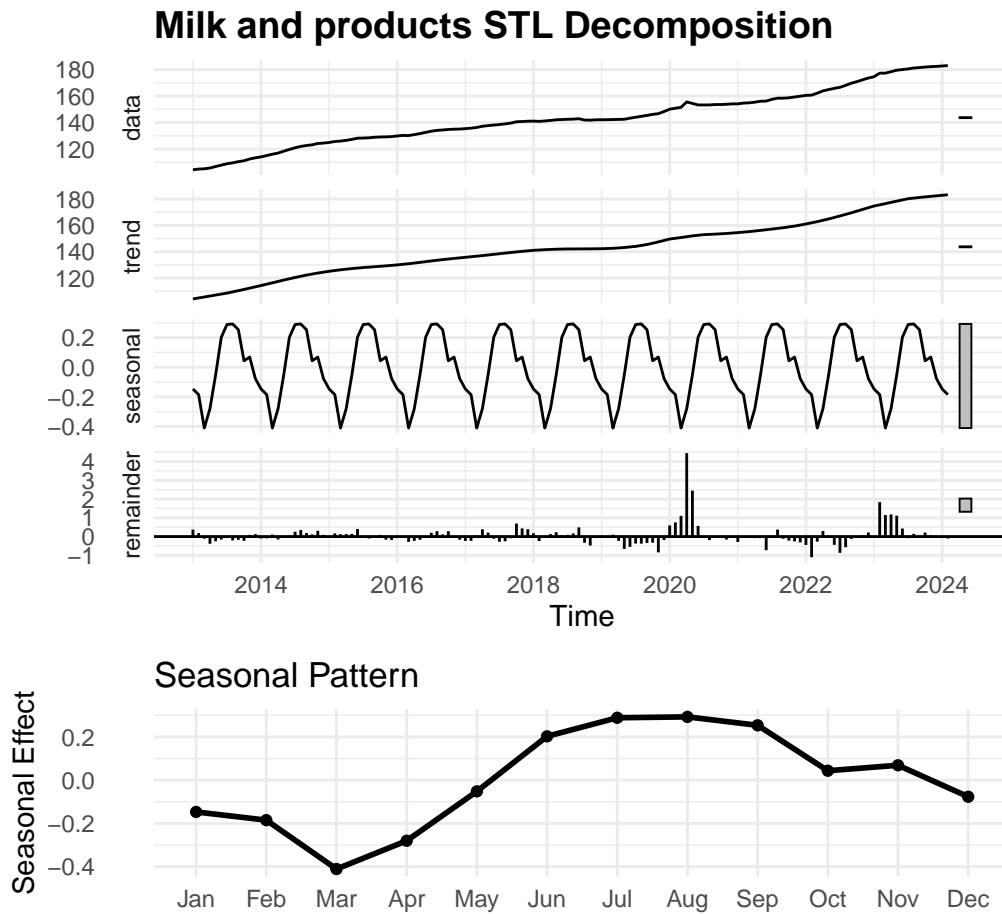
In this analysis, STL decomposition is applied to selected CPI categories to uncover underlying seasonal cycles, long-term inflation trends, and unusual anomalies in price behavior.

General CPI STL Decomposition



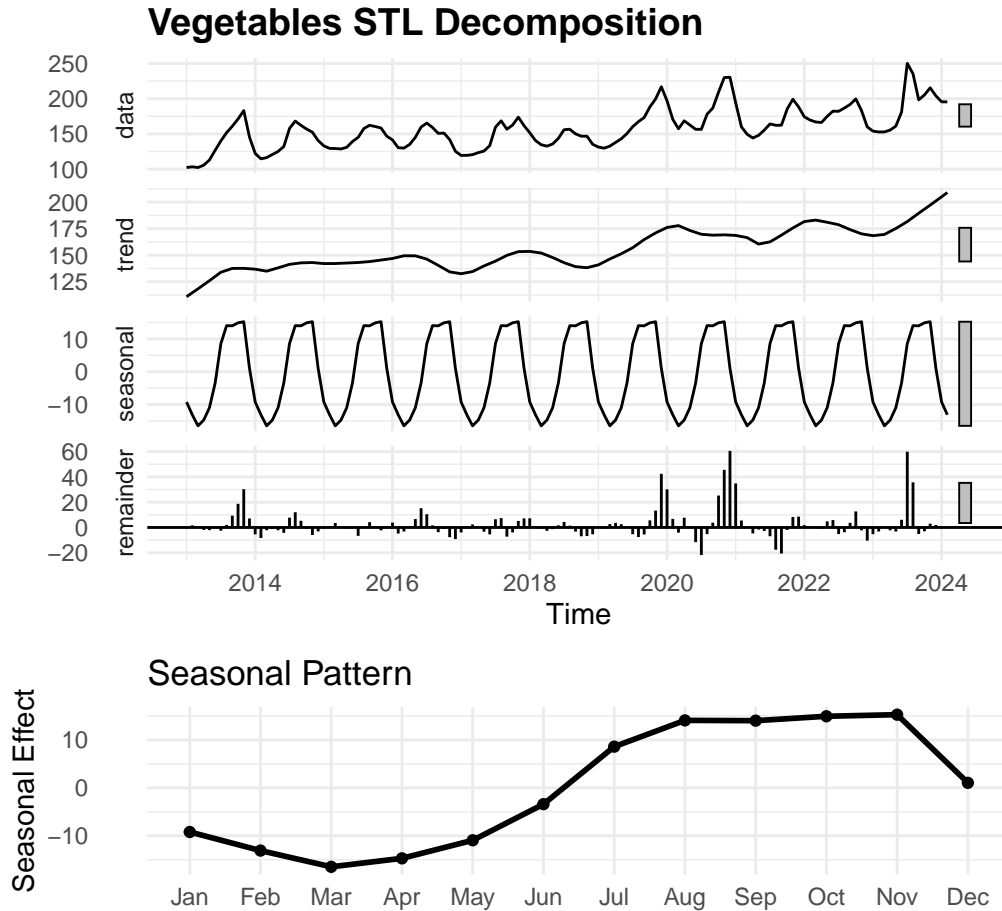
- The seasonal component peaks during the festive months (October-November) at around and 1.0 point and dips in late winter (February-March) at around -1.0 point, indicating mild seasonality.
- Noticeable remainder spikes in 2020 and 2023-24 reflect external shocks such as the COVID-19 pandemic and global commodity volatility.

Milk and Products STL Decomposition



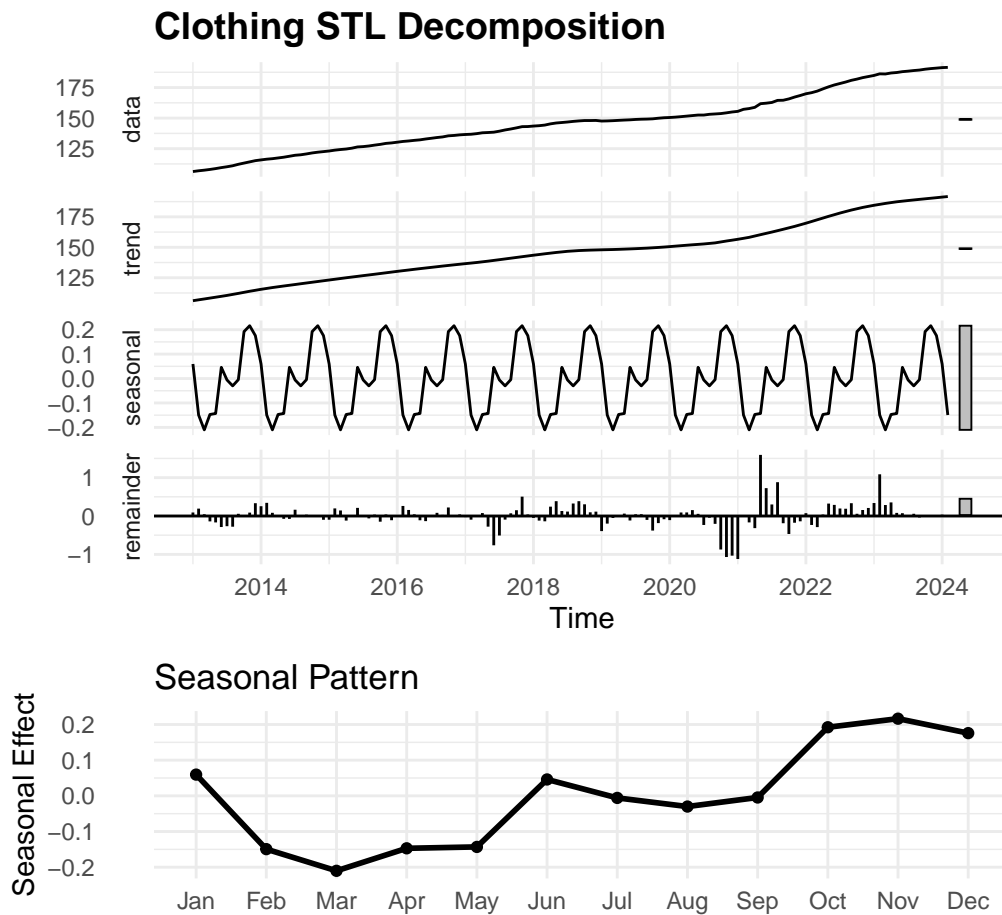
- Seasonal component reaches its lowest point in March at -0.4 points and peaks between June and September above +0.2 points, reflecting reduced milk supply and higher fodder costs during the summer and early monsoon months.
- Trend shifts with significant remainder spikes in 2020 and 2022-23 reflect pandemic-related supply disruptions and government interventions in milk procurement and pricing policies.

Vegetables STL Decomposition



- Exhibits strong seasonal swings, oscillating between +10 and -10 points, and prices often rise during monsoons despite higher harvests, due to poor monsoons, weak storage infrastructure, or festival-season demand surges, making vegetable prices highly sensitive to seasonal stress.
- Trend breaks in 2020 and 2023-24 align with major remainder peaks during the pandemic lockdowns and the 2023 tomato-onion price crisis, showing how vegetables are highly sensitive to supply shocks and weather disruptions.

Clothing STL Decomposition



- Seasonal component dips to around -0.2 from February to May, stays flat through monsoon months, and peaks at +0.2 during October-December, likely reflecting festive-season shopping.
- The sharp trough in late 2020 reflects COVID-19 lockdown-related demand collapse, while the 2021 remainder spike aligns with post-pandemic recovery, global cotton price surges, and GST-driven increases in input costs, all of which contributed to higher apparel prices.

Volatility & Risk Analysis

To understand not just the level of inflation but also its stability, we examine two related metrics:

1. **Month-over-Month (MoM) Percentage Change** The MoM percentage change measures the relative difference between one month's CPI and the previous month's CPI :

$$\text{MoM } \%_t = 100 \times \frac{\text{CPI}_t - \text{CPI}_{t-1}}{\text{CPI}_{t-1}}$$

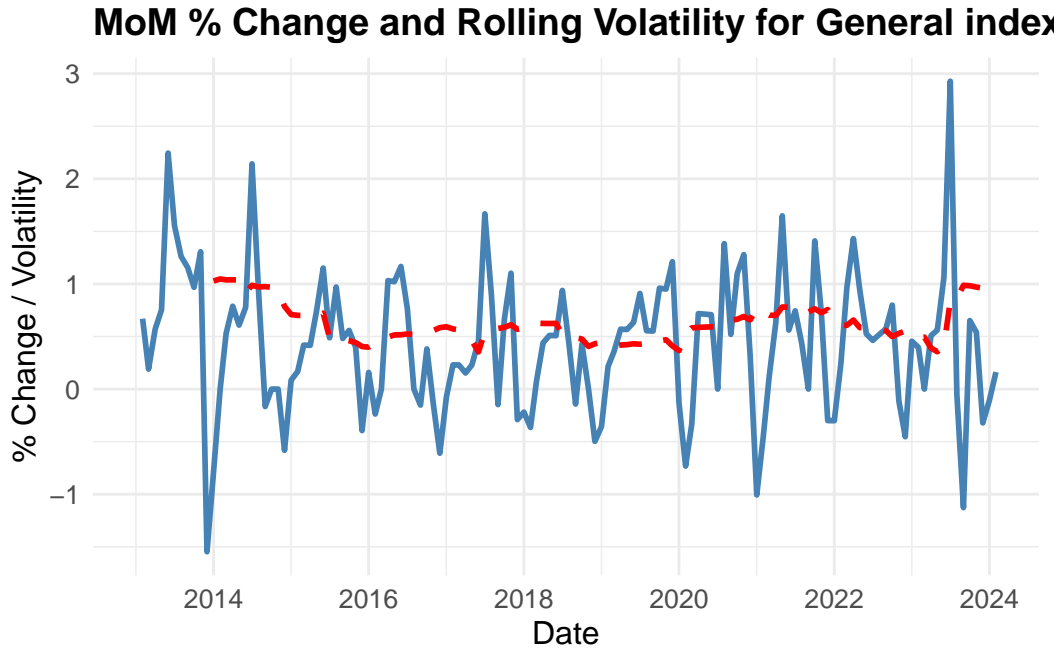
This highlights short-term price dynamics and shows when inflation accelerates or decelerates on a month-to-month basis.

2. **Rolling Volatility (12-Month Rolling Standard Deviation)** Rolling volatility captures how consistent or erratic those monthly changes have been over a longer window. We compute the standard deviation of the last 12 MoM percentage changes at each point in time:

$$\sigma_t = \sqrt{\frac{1}{11} \sum_{i=t-11}^t (\text{MoM } \%_i - \overline{\text{MoM } \%})^2}$$

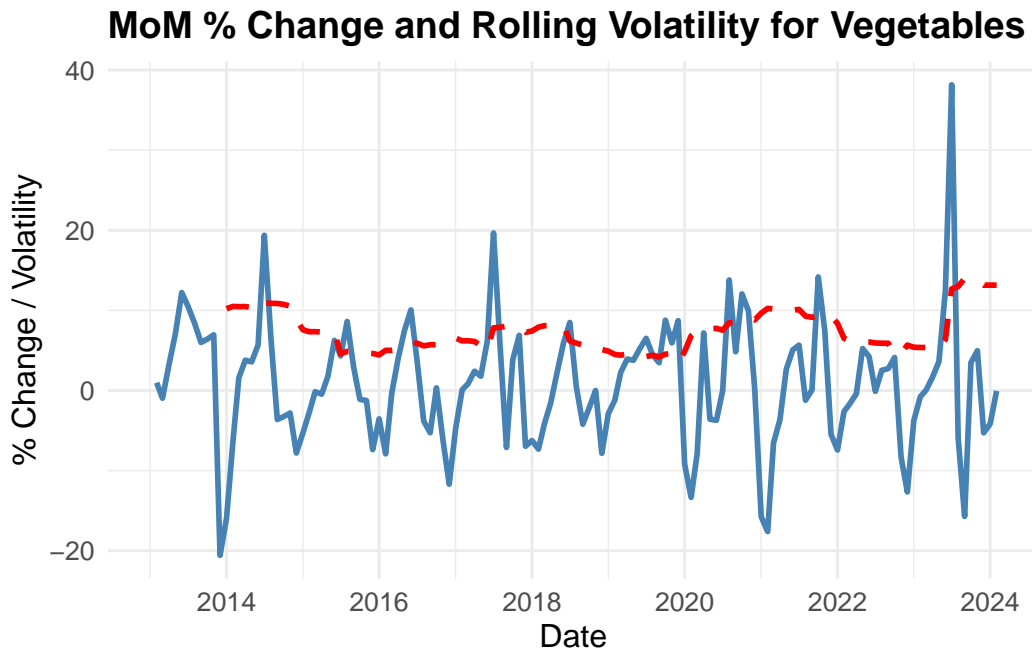
A higher rolling SD indicates more volatile month-to-month price swings (“hot” periods), while a lower rolling SD signals greater stability (“calm” periods).

General CPI - Volatility Analysis



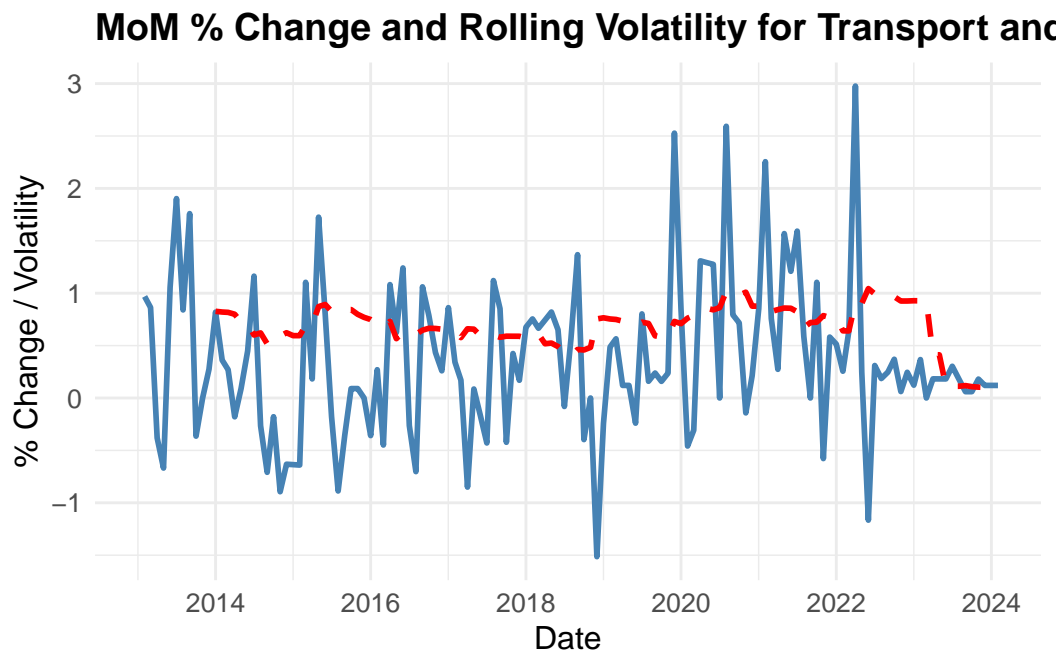
- Despite overall price stability, monthly inflation exhibits frequent oscillations, with MoM % changes swinging from nearly -1.5% to +3%, especially during early 2020 and late 2023 indicating short-term economic shocks or rapid corrections.
- The rolling 12-month volatility (red line) remains relatively stable around 0.5-1% but rises notably post-2022, likely due to post-pandemic demand recovery, global commodity price instability, and lingering supply chain pressures.

Vegetables Volatility Analysis



- Shows extreme short-term volatility, with MoM % spike reaching nearly +40% in May 2023, likely due to the tomato-onion price crisis. This alone causes rolling volatility to surge from ~5% to over 13-15%, reflecting acute supply shocks.
- Sustained high volatility throughout the decade, especially during mid-2020 and 2023, which shows the category's heightened sensitivity to monsoon irregularities, crop failures, and disruptions in perishable goods logistics.

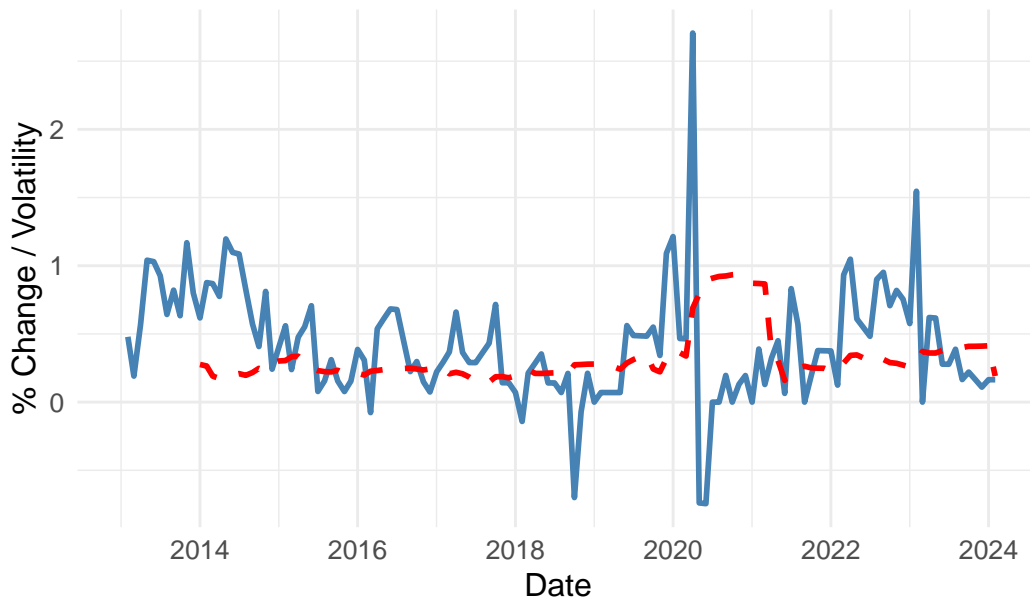
Transport and Communication Volatility Analysis



- MoM % changes swung dramatically (-1 % to +3 %) in early 2022, then stabilized to near zero by mid-2022, mirrored by rolling SD collapsing from ~1 % to almost 0 % at the start of 2023, highlighting the rapid normalization of mobility and data-service demand post-pandemic.
- Low pre-2020 volatility (~0.5 %) followed by a brief spike in 2020-21 illustrates how sudden policy-driven behavioral shifts (lockdowns) can generate short-lived bursts of price instability.

Milk and Products Volatility Analysis

MoM % Change and Rolling Volatility for Milk and products



- Volatility sits at ~0.25 % for most of the decade but jumps to nearly 1 % by early 2021, driven by a +2.5 % MoM spike in April 2020 and subsequent dip below -0.5 %, capturing both panic-buying and supply-chain disruptions during the initial COVID-19 wave.
- The rapid rise and fall in volatility around 2020-21 underscores how perishable-goods markets can experience sharp, transient shocks, followed by equally swift recoveries as supply chains adapt.

Structural Breaks and Anomalies

To truly understand the evolution of CPI over time, it is not enough to examine trends or seasonality alone. Major economic shocks, policy changes, or supply chain disruptions can lead to **structural shifts** in inflation dynamics. Additionally, sudden, short-term price shocks often leave distinct signatures in the data that are not explained by regular patterns.

Hence, we apply **structural break and anomaly detection techniques** to:

- Identify **persistent regime changes** in inflation patterns, such as a shift in trend growth rates or volatility.
- Detect **one-time disruptions or anomalies**, often caused by external shocks like pandemics or geopolitical events.

This enables a deeper understanding of when, how, and why the CPI behavior changed significantly.

Structural Breakpoints

Structural breakpoints refer to points in time where the underlying data generation process changes significantly. In the context of CPI, this could mean a shift in the long-term inflation trend due to **economic reforms**, **tax changes**, or large-scale disruptions like a **pandemic**.

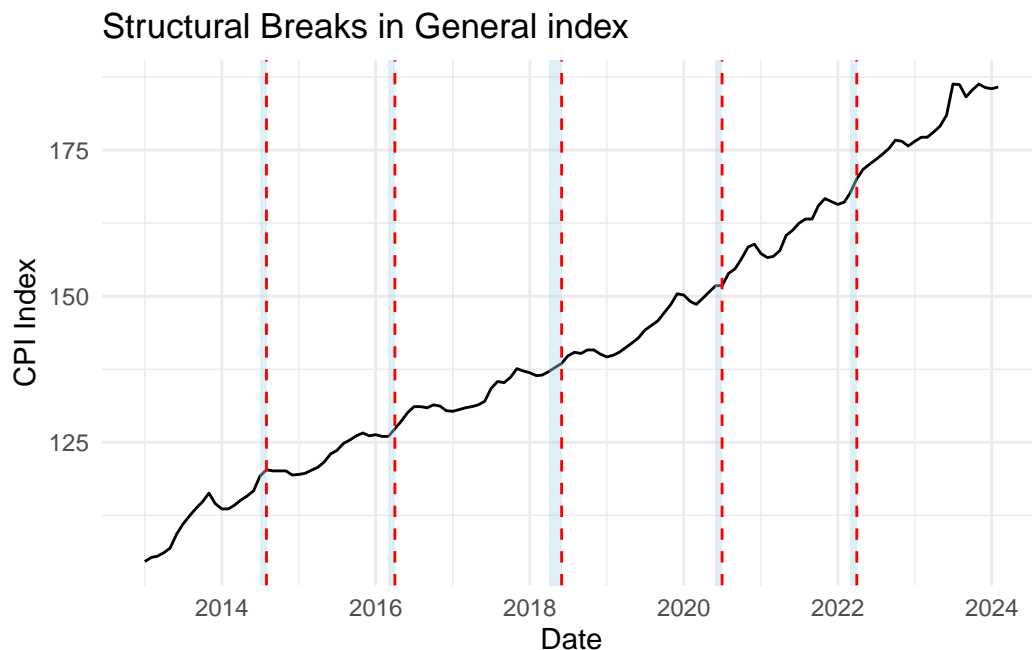
We use methods like the `breakpoints()` function from the `strucchange` package to detect such breaks by analyzing changes in regression relationships over time.

Outliers and Anomalies

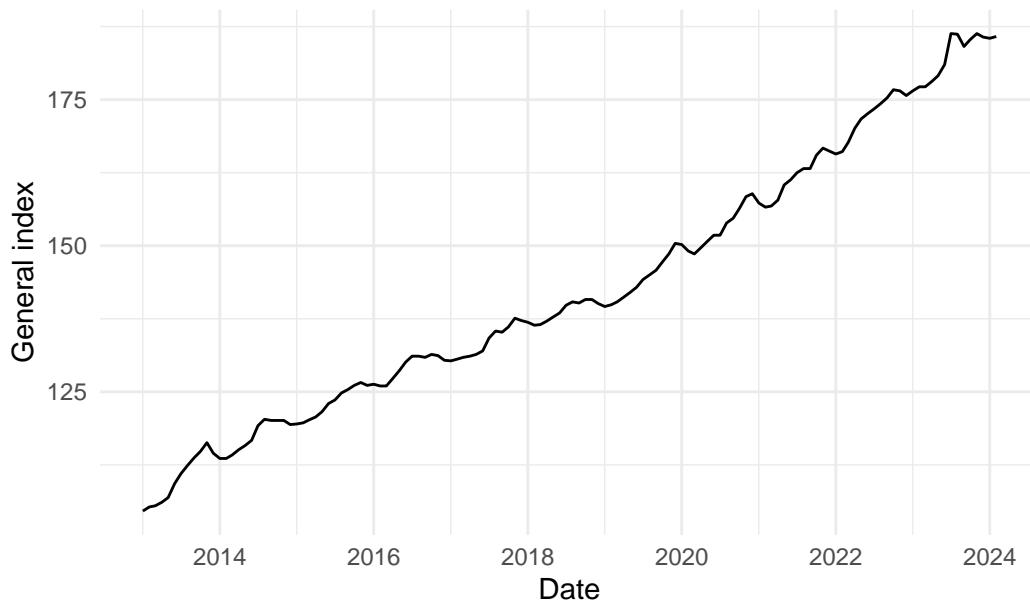
Outliers are sudden, short-term deviations from expected behavior that are not part of the regular trend or seasonality. These may result from **temporary supply shocks**, **extreme weather events**, or **abrupt policy interventions**.

We apply statistical tools like `tsoutliers()` from the `forecast` package to identify and classify these anomalies.

General CPI

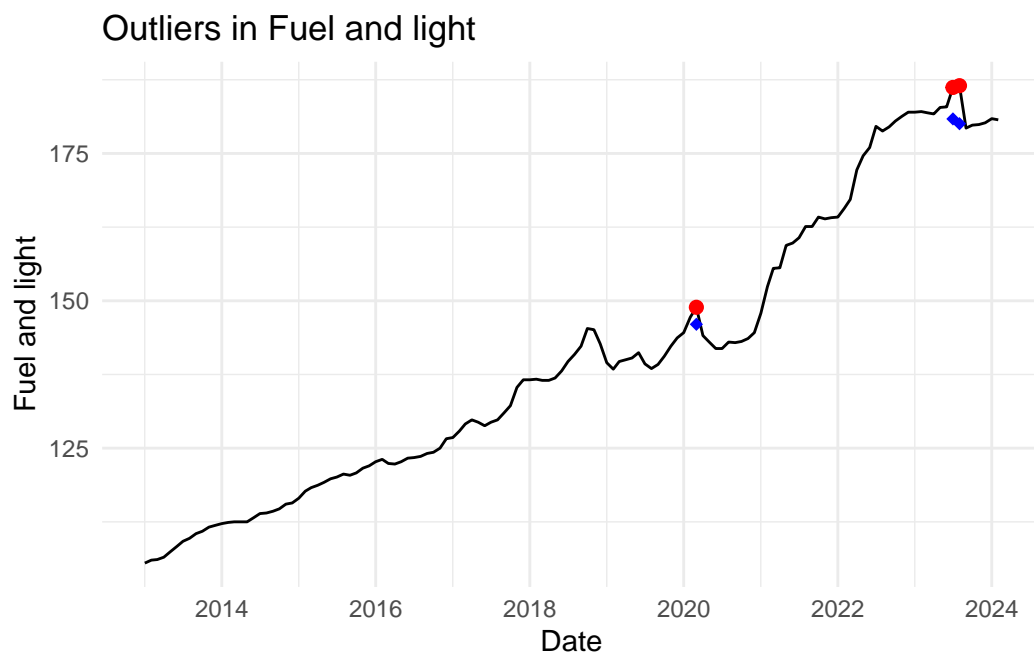
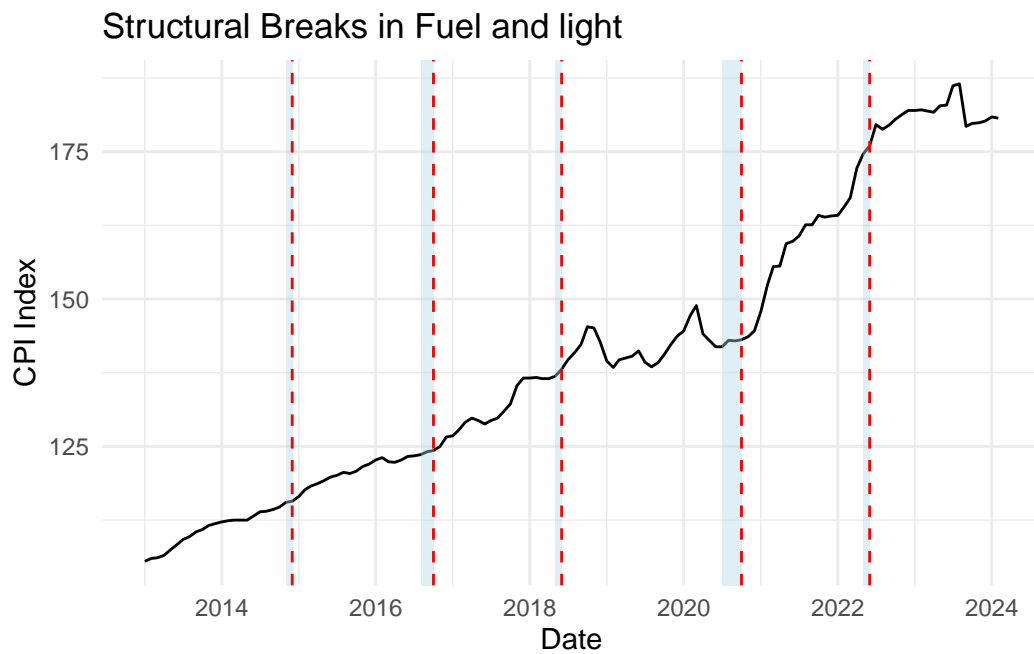


Outliers in General index



- Five structural breaks are detected in the General CPI trend: mid-2014 (global oil price collapse), early-2016 (demonetization shock), spring-2018 (GST implementation), mid-2020 (COVID-19 lockdowns), and early-2022 (post-pandemic commodity price surge).
- No significant outliers are identified in the residual series, indicating that most price movements are captured by the identified trend shifts and regular seasonal patterns.

Fuel and Light

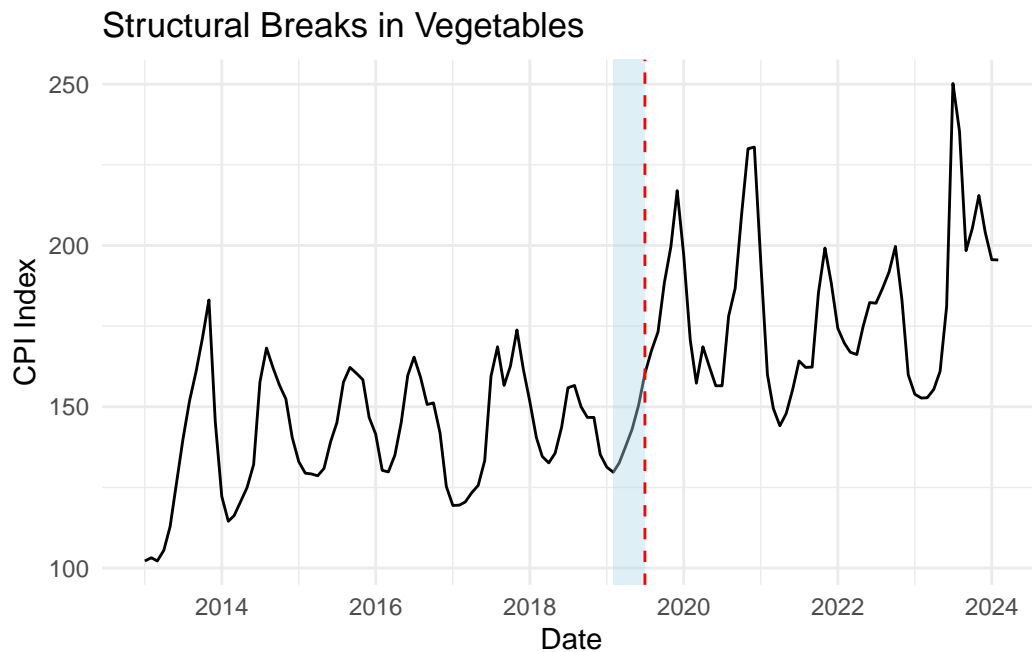


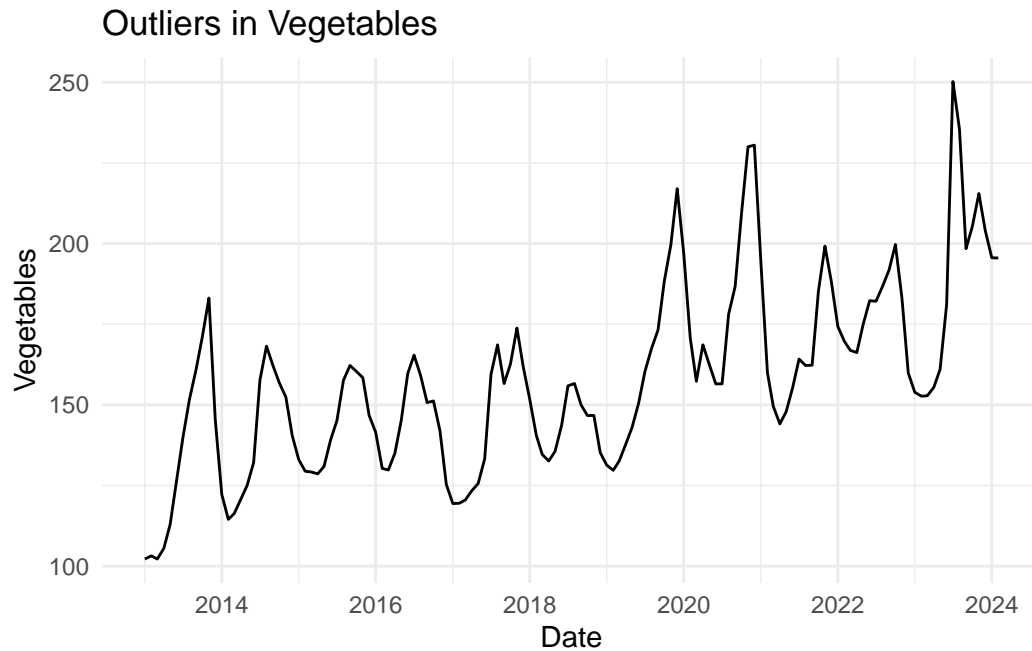
- Five distinct structural breaks are detected in the Fuel and Light CPI trend: at end-2014 (global oil price crash), late-2016 (OPEC production cuts), spring-2018 (renewed

demand growth), mid-2020 (COVID-19 lockdown-induced price collapse), and mid-2022 (Russia-Ukraine conflict driving energy costs).

- Significant one-off outliers occur in March 2020 (sharp price drop during nationwide lockdown) and around July-August 2023 (large deviations likely reflecting temporary subsidy adjustments and rapid global market corrections).

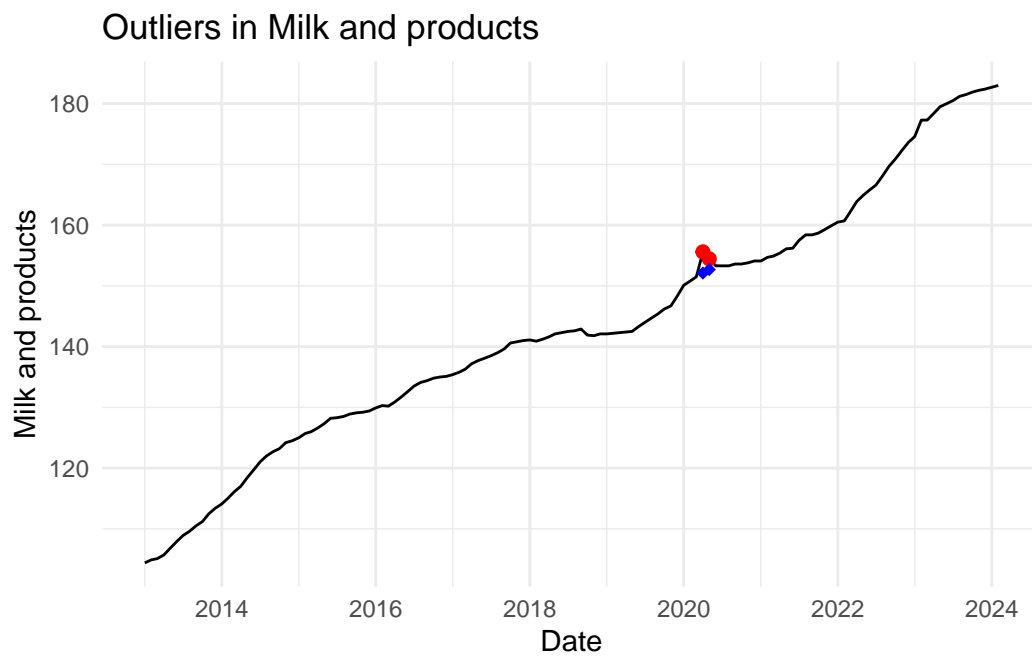
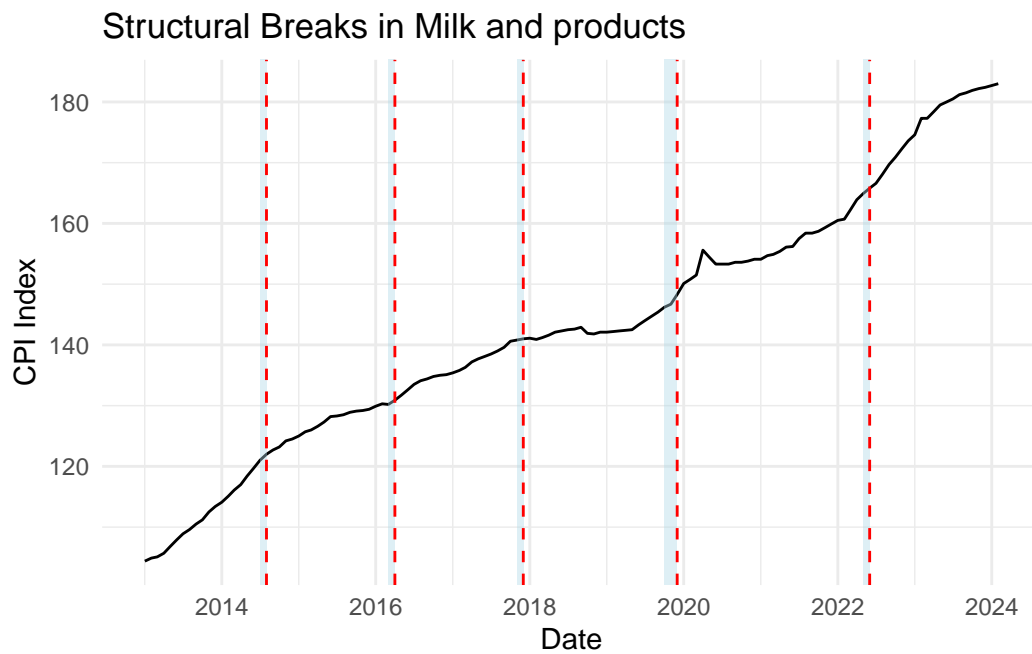
Vegetables





- A single structural break is detected in mid-2019, marking a distinct shift in the underlying price trend. This aligns with a period of significant food inflation and supply disruptions linked to erratic monsoons and pre-pandemic logistical stresses.
- No statistical outliers are detected in the residuals, reinforcing the idea that extreme price fluctuations in vegetables are not anomalous but rather part of their typical seasonal volatility.

Milk and Products



- Five distinct shifts are detected in the Milk and Products CPI trend: mid-2014 (lean-season supply constraints), early-2016 (demonetization-induced demand disruptions),

late-2017 (GST implementation), late-2019 (rising food inflation from erratic weather and input cost pressures), and mid-2022 (global dairy price surges).

- Two significant anomalies occur in April and May 2020, with replacement values of ~152.1 and ~152.7, capturing the pronounced supply shocks and panic buying during the initial COVID-19 lockdowns.

Forecasting

Why Forecast CPI?

Forecasting CPI helps policymakers, businesses, and investors anticipate inflationary pressures, make budgeting decisions, set pricing strategies, and inform interest-rate policy.

ARIMA Models

- **AutoRegressive Integrated Moving Average (ARIMA)** models are classical linear time-series models that combine:
 - **AR(p)**: regression on its own lagged values,
 - **I(d)**: differencing to induce stationarity,
 - **MA(q)**: modeling the error term as a moving average of past errors.
- **Pros**: solid theoretical footing; interpretable parameters; handles a wide range of linear patterns.
- **Cons**: assumes linear relationships; struggles with complex seasonality or non-linear trends without seasonal ARIMA extensions.

Prophet

- Developed by Facebook, **Prophet** is a **semi-parametric** model that decomposes a series into: trend (piecewise linear or logistic), seasonality (yearly, weekly, daily via Fourier series), and holidays.
- **Pros**: automated handling of multiple seasonalities; robust to missing data and outliers; intuitive hyperparameters.
- **Cons**: less interpretable coefficients; can overfit if holidays/seasonalities are over-specified.

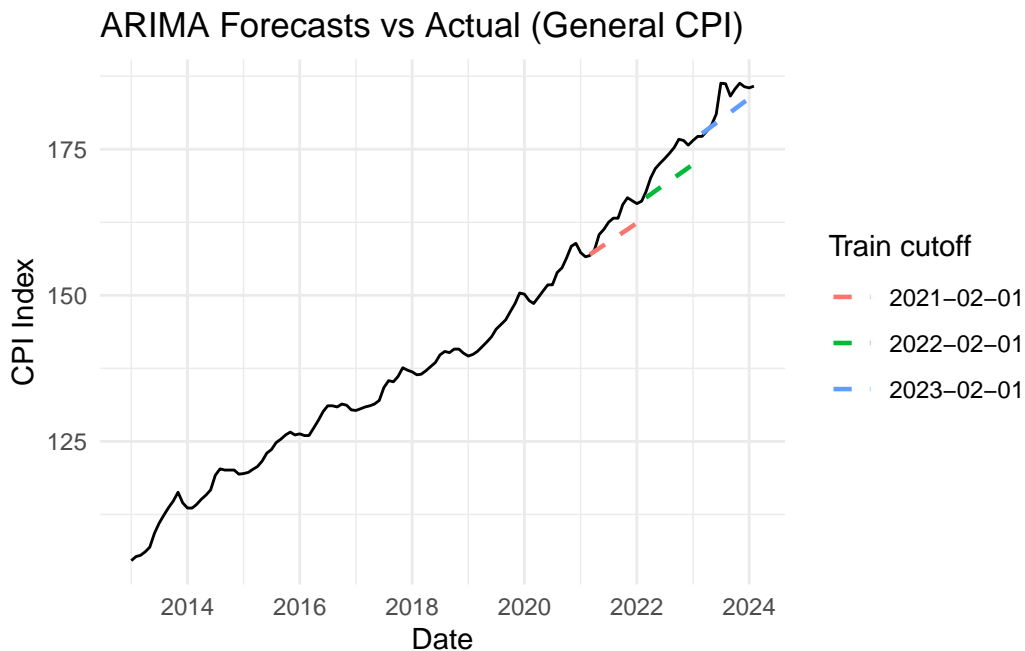
Now, we employ a **rolling-origin evaluation** to assess forecast accuracy over multiple one-year horizons. We repeatedly split the data at three cutoffs:

- **Cutoff = 2021-02-01**: Train on January 2013 to February 2021, then forecast March 2021 to February 2022
- **Cutoff = 2022-02-01**: Train on January 2013 to February 2022, then forecast March 2022 to February 2023
- **Cutoff = 2023-02-01**: Train on January 2013 to February 2023, then forecast March 2023 to February 2024

For each fold, we fit both ARIMA and Prophet to the training period, generate 12-month forecasts, and compare them to the held-out observations. The resulting **cutoff** column in the results table indicates the end of each training window, and the accompanying error metrics RMSE (Root Mean Squared Error) and MAPE (Mean Absolute Percentage Error) measure how closely each model's forecasts matched the actual CPI values over the subsequent year.

Results

When using **ARIMA** :



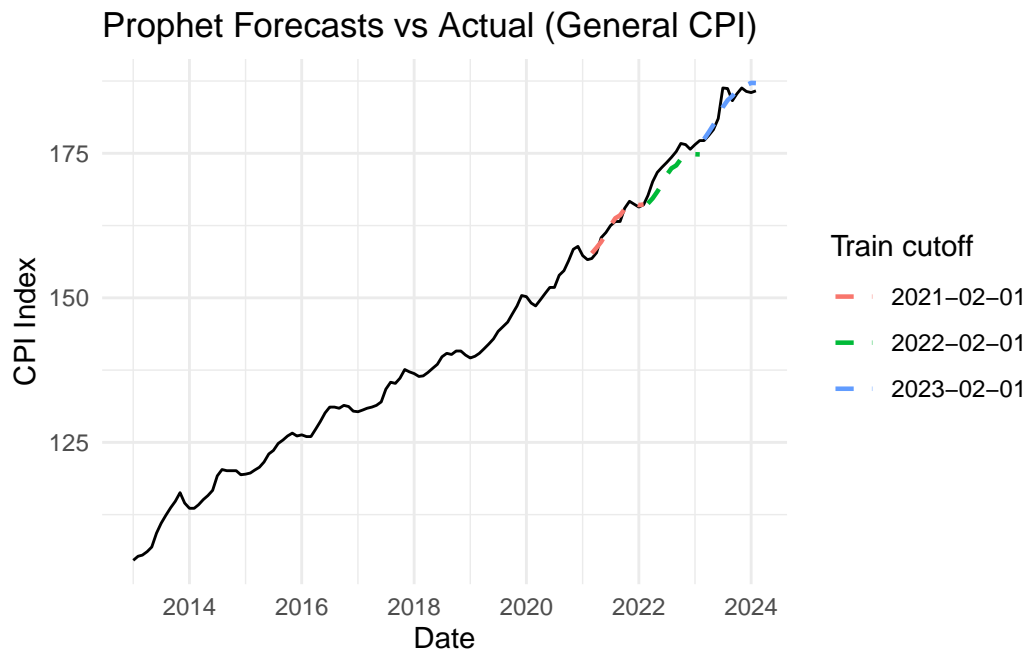
```
# A tibble: 3 x 3
  cutoff    RMSE  MAPE
<date>    <dbl> <dbl>
```

```

1 2021-02-01 11.4 0.0573
2 2022-02-01 9.45 0.0483
3 2023-02-01 15.8 0.0716

```

When using **Prophet** :



```

# A tibble: 3 x 3
  cutoff      RMSE  MAPE
  <date>    <dbl> <dbl>
1 2021-02-01 13.0  0.0612
2 2022-02-01  8.48 0.0433
3 2023-02-01 13.5  0.0597

```

Comparison

```

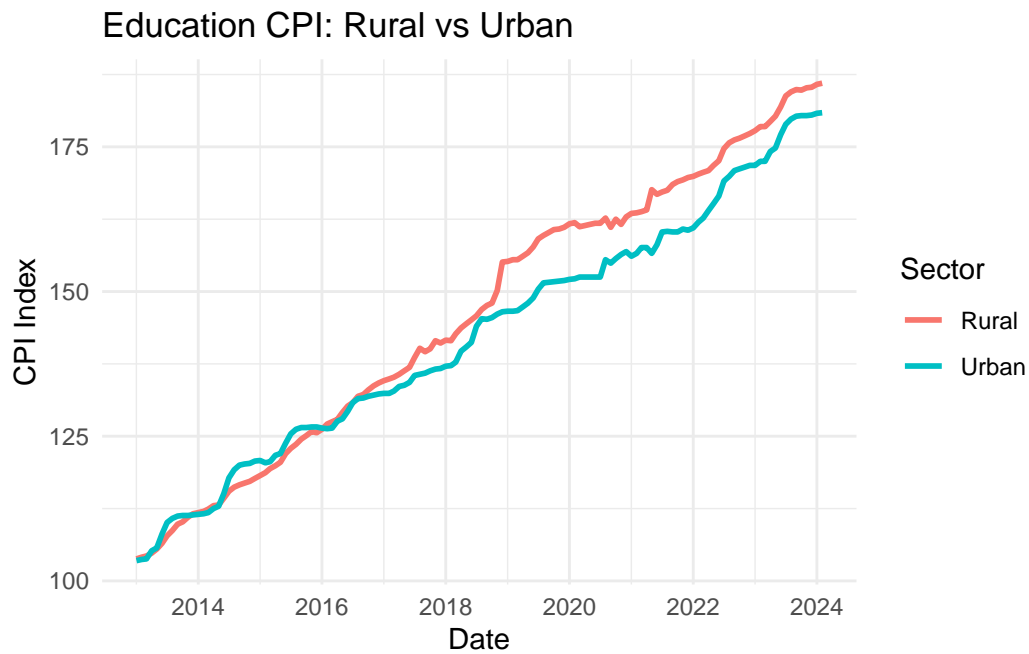
# A tibble: 2 x 3
  model  `mean MAPE` `mean RMSE`
  <chr>    <dbl>    <dbl>
1 ARIMA    0.0591    12.2
2 Prophet  0.0548    11.7

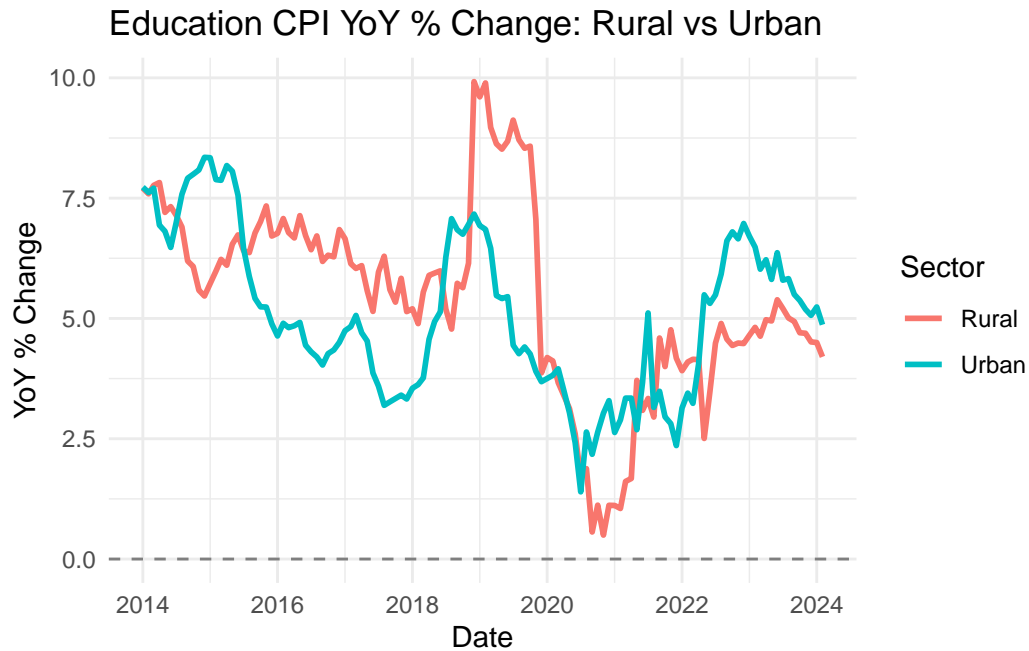
```

Based on the rolling-origin cross-validation results, **Prophet consistently outperforms ARIMA** in both MAPE and RMSE, indicating better predictive accuracy for modeling CPI trends in our dataset.

Rural vs. Urban Comparison (Case Study)

Before diving into combined-data results, it is instructive to apply the same analysis to the rural and urban CPI series to uncover regional differences. Below, we compare the **Education** index for the two sectors.





- **Divergence since 2018:** Until 2018, the CPI for rural and urban education remained closely aligned. Post-2018, rural education CPI rose more sharply, reflecting faster fee growth or cost increases in rural areas, possibly driven by expanding private schooling or reduced subsidies.
- **YoY % change shows sharper rural inflation spikes:** In 2019-2020, rural education CPI inflation peaked near 10%, while urban rates remained relatively lower, possibly due to the surge in online education and coaching which made education relatively more affordable. The post-pandemic trend shows convergence, but rural inflation still remains slightly elevated.

Conclusion & Insights

Overall CPI Trends

- The General CPI rose steadily from 2013-2024, with mild seasonality peaking around October-November and dipping in February-March. Forecasting tests showed **Prophet** outperformed ARIMA in 12-month accuracy.

Food vs. Non-Food Dynamics

- **Vegetables, Milk & Products:** pronounced seasonal cycles (e.g. +40% MoM spike for vegetables in May 2023 or lean-season peak in milk), and high rolling volatility during 2020-21 and 2023 shocks.
- **Clothing & Others:** minimal seasonality but clear lockdown troughs in early 2020 and rebounds in 2021, tied to input-cost jumps.

Inflation Risk & Volatility

- **Rolling SD** rose from ~0.5% pre-2020 to ~1% post-2022, highlighting periods of elevated price instability around the COVID lockdowns and global commodity surges.

Structural Shifts & Anomalies

- **Breakpoints** align with major events like 2014 oil collapse, 2016 demonetization, 2017 GST rollout, 2020 lockdowns or 2022 energy-price shock.
- **Outliers** in April 2020 (for Milk and Products and Fuel and Light) captured one-off lockdown panics, but extreme swings for Vegetables data is part of its regular seasonal volatility.

Rural vs. Urban Case Study

- **Education CPI** diverged post-2018: rural fees climbed faster (YoY peaks near 10% in 2019-20), while urban costs were moderated by the rapid rise of online learning platforms during and after COVID.

Policy Implications

- **Strengthen cold-chain and storage infrastructure for perishables** Extreme vegetable volatility underscores the need for improved logistics, warehousing, and market linkages to buffer supply shocks and stabilize food prices.
- **Expand digital education infrastructure in rural areas** The post-2018 divergence in Education CPI, driven by uneven access to online learning, suggests a targeted push for broadband, devices, and teacher training to make digital education affordable and dampen fee inflation.

- **Adopt adaptive monetary policy frameworks** The succession of structural breaks, from oil-price collapses to GST rollout to pandemic shocks means central banks should integrate real-time break-and-anomaly detection into their inflation models, enabling more responsive interest-rate decisions.

References & Citations

1. **Bureau of Labor Statistics**, *The 2014 plunge in import petroleum prices: What happened?* – documents the 60%–70% global oil price collapse during 2014–2015. Available at: <https://www.bls.gov/mxp/publications/presentations/import-petroleum-price-decline-2014-presentation.pdf>
2. **World Bank**, *What triggered the oil price plunge of 2014–2016...* – explains the global oversupply and demand slowdown behind that crash. Available at: <https://blogs.worldbank.org/en/developmenttalk/what-triggered-oil-price-plunge-2014-2016-and-why-it-failed-deliver-economic-impetus-eight-charts>
3. **Brookings**, *Early Lessons from India's Demonetization Experiment* – analyzes macroeconomic disruptions and cash shortages from the 2016 policy action. Available at: <https://www.brookings.edu/articles/early-lessons-from-indias-demonetization-experiment/>
4. **AP News**, *India's top court upholds legality of the 2016 currency ban* – discusses the Supreme Court's ruling and its economic implications. Available at: <https://apnews.com/article/397ca5ebd3084cd57c27f94fbeb8420>