Forecast of Polish Inflation Rate

Erste Group CEE Research Interview Task

Aleksandar Gradev

28.05.2025

ERSTE Group

Task Overview

- The task is to analyze the historical dynamics of Polish inflation and **forecast the HICP for the first four months of 2025**

- My approach begins with **a diagnosis of the time series** and observation of the key characteristics of time series **like stationarity, seasonality, trend and autocorrelation**

- I then build and compare **3 models**:
  - A **basic AR(1)** model
  - A **univariate Auto ARIMA** model selected by AIC minimizing
  - A **multivariate ARIMAX** model with **external macroeconomic variables**

- **Goals** of this task:
  - Create **a both technically and economically meaningful forecast**
  - Show that **my AIC** with Erste's CEE Research team is **approximately 0 :)**

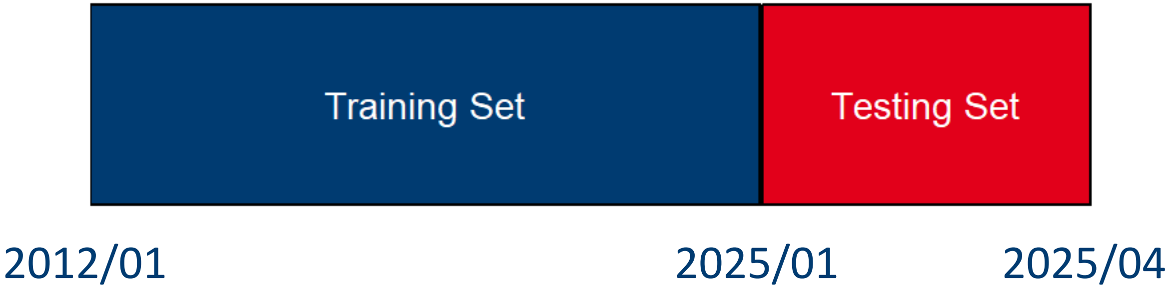
ERSTE Group

Data Overview & Libraries Used

Data description:

- Original data set is downloaded from Eurostat
- Represents percentage change in Polish HICP
- Monthly frequency
- Spans between January 1997 and April 2025
- Last accessed on the 21st of May

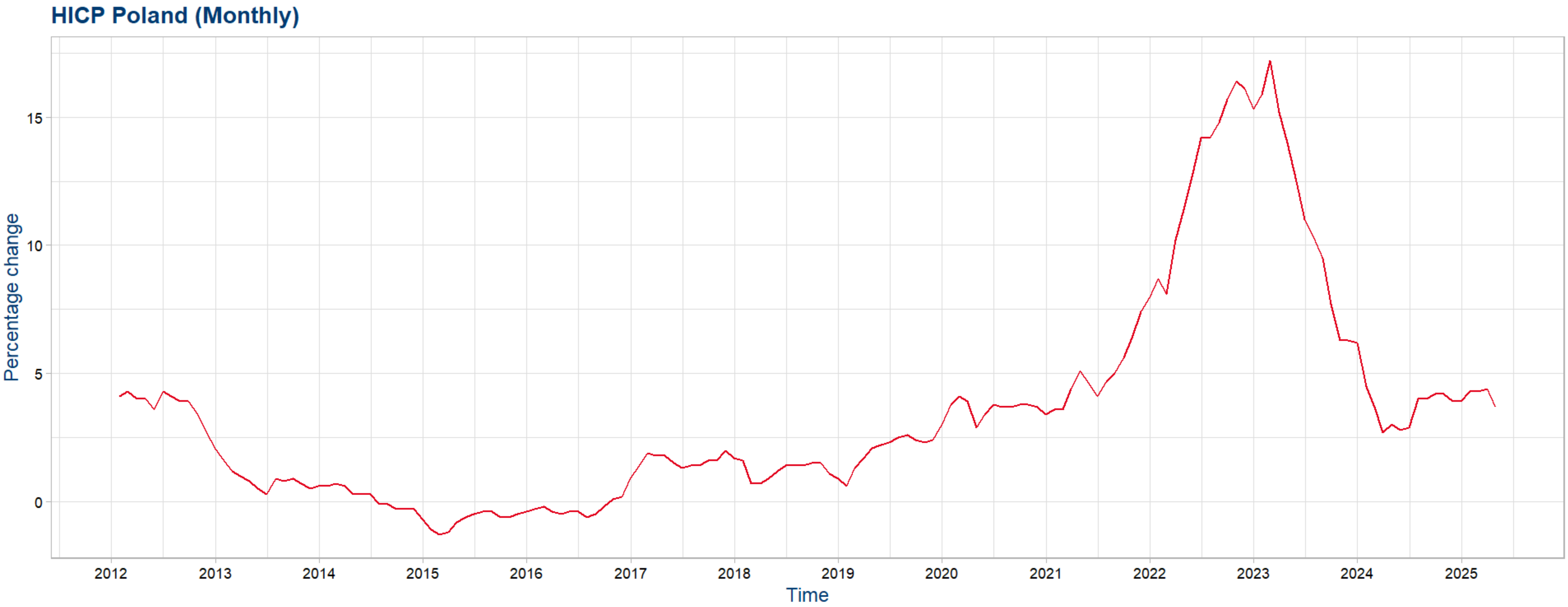
For the sake of this task I have split the data in the following 2 data subsets:



During the project I have used the following libraries:

- lubridate
- ggplot2
- tseries
- forecast
- quantmod
- dplyr
- tidyr

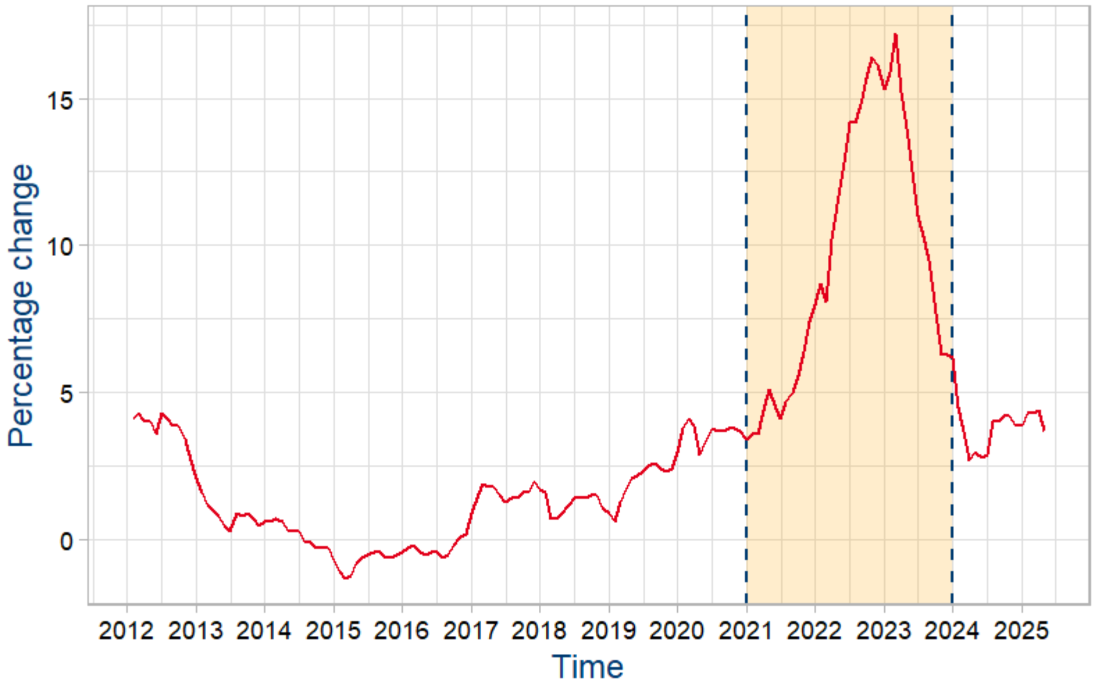
Data Overview & Libraries Used



TS Diagnostics (Stationarity)

- Stationarity implies **constant mean, variance and autocovariance** over time
- Most time series models (like ARIMA) **require stationarity** to perform well
- Due to the COVID spike in inflation one can't visually decide if the time series is stationary
- Use **Augmented Dickey-Fuller** Test:
  - At $\alpha = 0.05$
  - H_0: there is unit root (non-stationary)
  - H_A: there is no unit root (stationary)
- **Conclusion**: Reject H_0 => the series **is stationary** at 5% level

HICP Poland (Monthly)




```
##
##  Augmented Dickey-Fuller Test
##
## data:  train_ts
## Dickey-Fuller = -3.5707, Lag order = 5, p-value = 0.03827
## alternative hypothesis: stationary
```

ERSTE Group

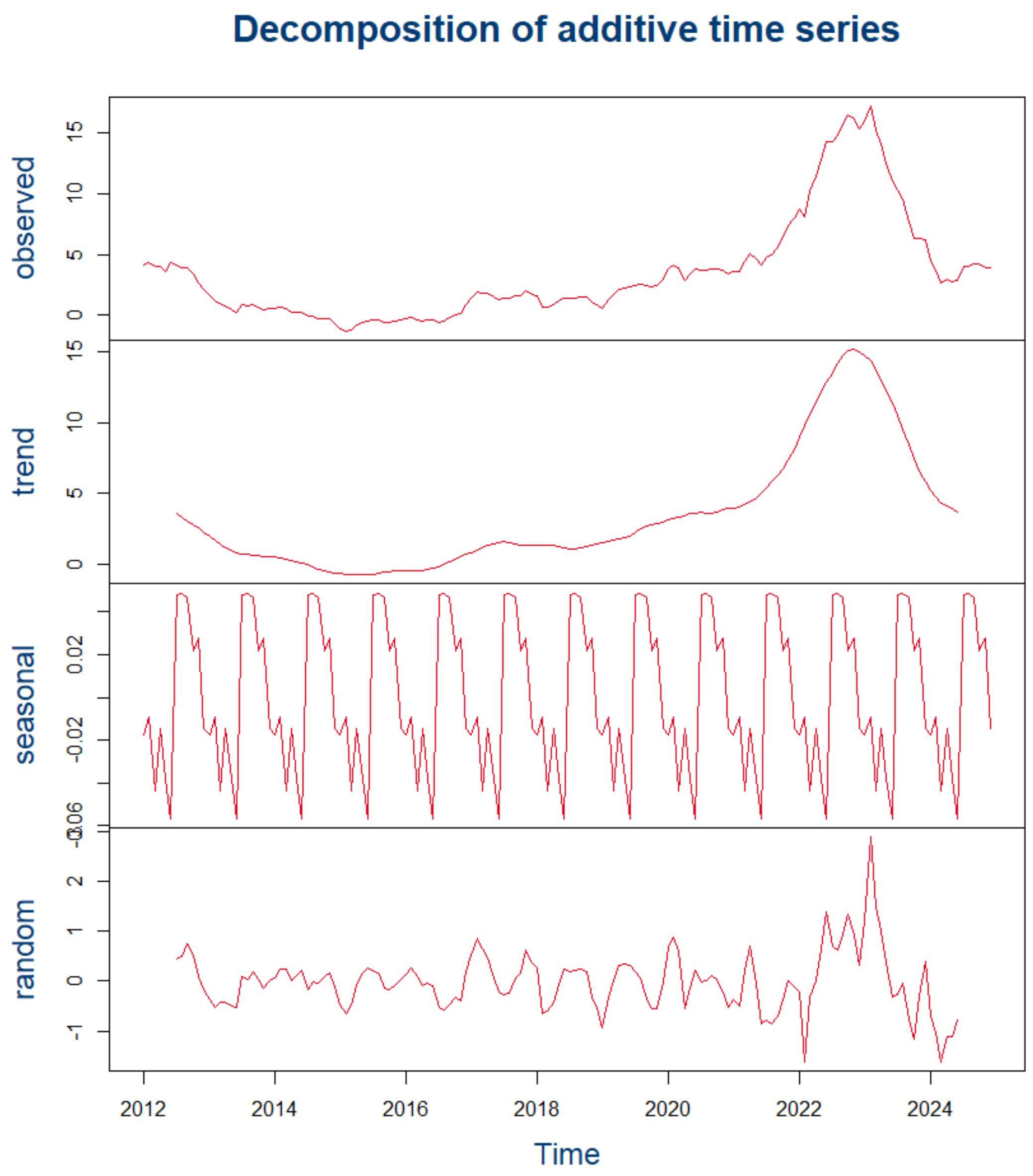
TS Diagnostics (Decomposition)

A time series can be decomposed into **Trend**, **Seasonality** and **Noise:**

- **Trend** - not permanent
  - No clear trend before 2021
  - Strong upward movement between 2021 and 2023 due to COVID
  - Fast decline after 2023

- **Seasonality -** regular pattern detected, but variation is small (≈ 0.08)
  - Seasonality is statistically visible but economically minor
  - If seasonality is not impactful, SARIMA model will drop it

- **Noise -** what's left after removing trend and seasonality
  - Stable before 2020 and more volatile around 2021–2023
  - Indicates that external factors likely play a role



ERSTE Group

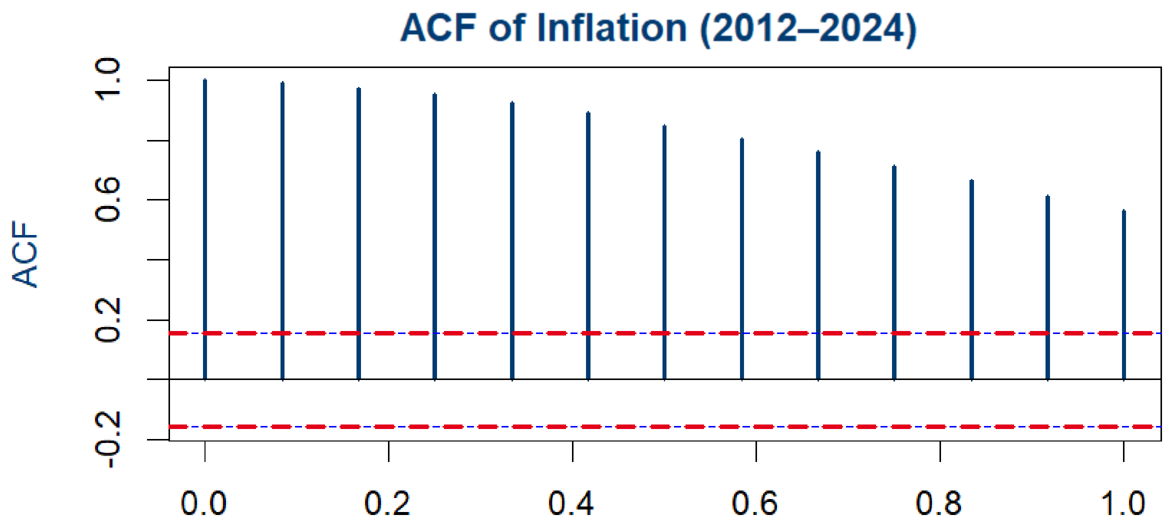
TS Diagnostics (Autocorrelation)

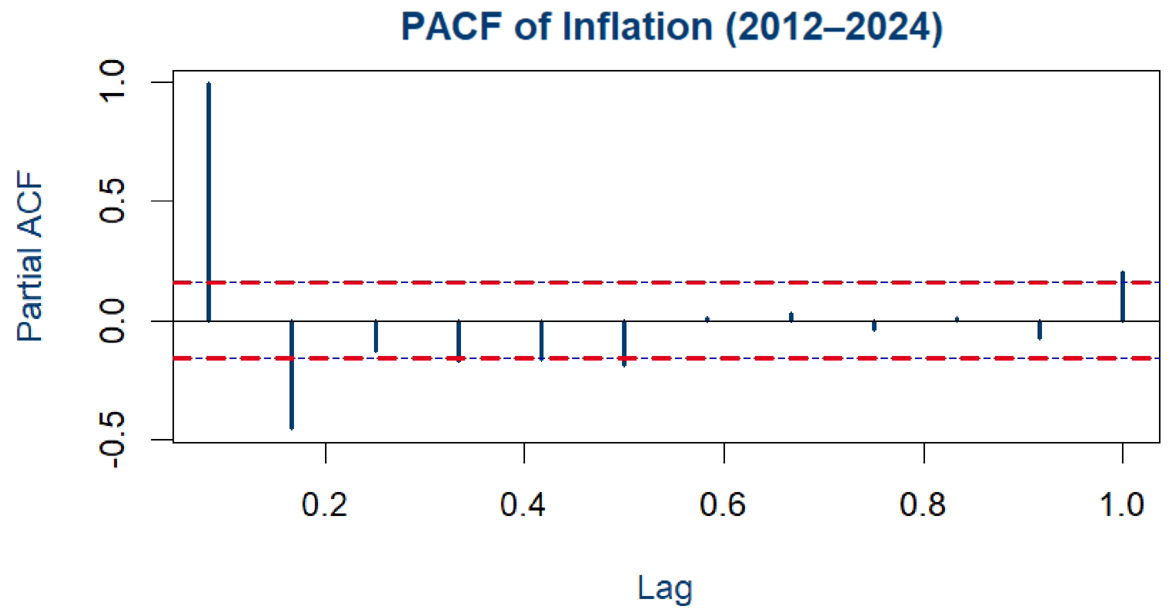
Autocorrelation measures how current values are related to past values

- **ACF (Autocorrelation Function)**
  - Captures correlation between current and past values (past errors in MA process)
  - **Strong correlation at lag 1** (≈ 1)
  - No clear cut-off

- **PACF (Partial ACF)**
  - shows the pure correlation between a value and its lag (isolates true AR influence)
  - **Huge value at lag 1, then drops significantly**

- This pattern suggests the time series is driven by an **AR(1) process**





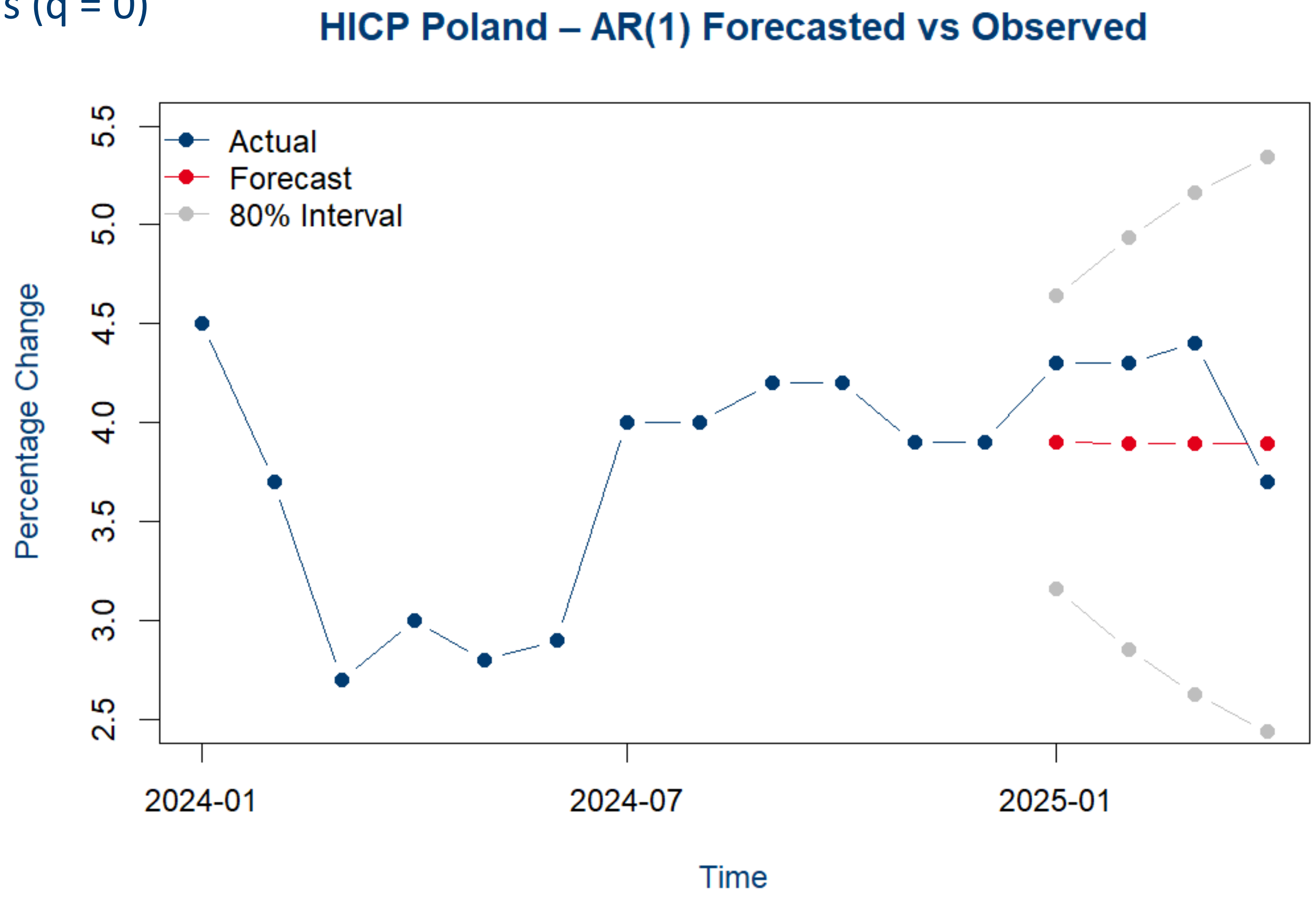
Univariate Model - AR(1)

- **ARIMA(1,0,0) based on diagnostics:**
  - PACF showed a strong lag-1 spike => use p = 1
  - ADF test confirmed stationarity => use d = 0
  - ACF decays gradually => no need for MA terms (q = 0)

- **Model output:**
  - AR(1) coefficient ≈ 0.99 => strong persistence
  - Residual variance $\sigma^2 \approx 0.34$
  - AIC ≈ 281.8

- **Forecast:**

```
##          Observed Forecast
## 2025 Jan      4.3 3.897829
## 2025 Feb      4.3 3.895688
## 2025 Mar      4.4 3.893575
## 2025 Apr      3.7 3.891490
```



Group

Univariate Model - Auto ARIMA

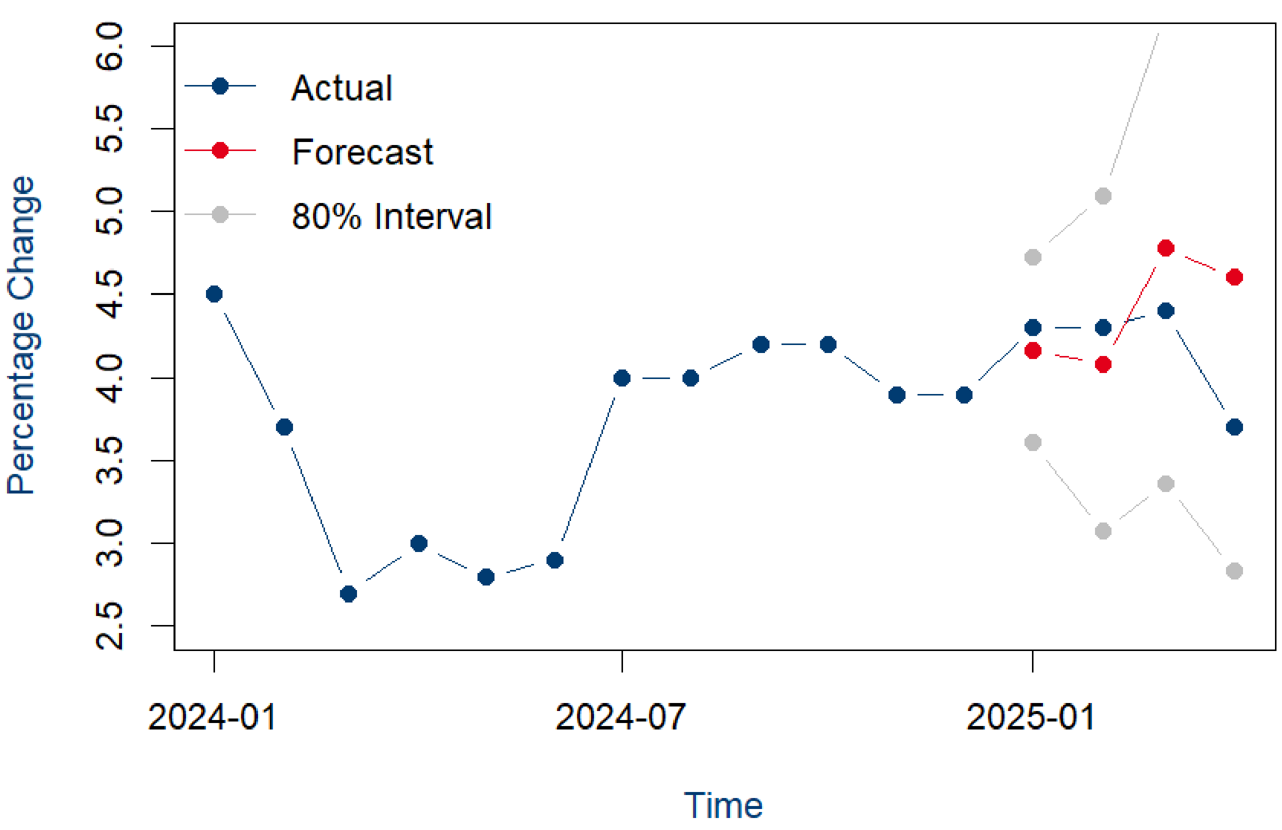
- ***auto.arima()*** to let R choose the best (p,d,q) based on AIC minimization
- Allows seasonality detection (if present uses SARIMA)

- **Selected Model - ARIMA(1,1,0)(0,0,1)[12]:**
  - 1 AR term, 1 order of differencing, 1 seasonal MA term
  - Residual variance $\sigma^2 \approx 0.19$ (better than AR(1))
  - AIC ≈ 198.2 (better than AR(1))

- **Forecast:**

```
##          Observed Forecast
## 2025 Jan      4.3 4.165496
## 2025 Feb      4.3 4.084557
## 2025 Mar      4.4 4.777414
## 2025 Apr      3.7 4.605298
```

HICP Poland – Auto ARIMA Forecasted vs Observed



Expanding the Model I

Unemployment Rate

- Captures domestic labour market pressure
- According to the Phillips Curve there is inverse relationship between inflation and unemployment
- Lag of 1 month performs best (inflation reacts with a short delay to labour shifts)
- Data from Eurostat, monthly % of population

Oil Price (in PLN)

- Oil is a key input for transportation, heating and production costs
- Oil shocks are historically linked to inflation spikes
- Measured in PLN to reflect true domestic cost (USD price * USD/PLN rate)
- Lag of 9 months best captures delayed cost pass-through
- Data from FRED and Yahoo Finance *(quantmod)*

EUR/PLN Exchange Rate (YoY log diff)

- Poland's largest trade partners (Germany, France) operate in EUR
- EUR/PLN affects prices of cars, industrial equipment and other euro-denominated imports
- Used log return over 12 months + 1-month lag to reflect pricing contract delays
- Data from Yahoo Finance *(quantmod)*

USD/PLN Exchange Rate (YoY log diff)

- Many key commodities (especially oil and food) are priced in USD
- Depreciation of the PLN makes imports more expensive (inflationary pressure)
- Used 12-month log return to smooth volatility and capture trend
- Data from Yahoo Finance *(quantmod)*

ERSTE Group

Expanding the Model II

Global Food Price Index

- Food makes up a large share of Poland's consumer basket
- Index reflects global agricultural prices (wheat, oils, dairy, meat)
- Captures import-driven food inflation (e.g. droughts, trade restrictions, supply chain distortions, wars)
- Data from FAO

Average Nominal Gross Wage

- Wages reflect domestic demand strength and cost-push inflation
- As wages increase, consumption increases and firms pass labour costs to prices
- Lag of 2 months chosen based on test performance
- Data from Statistics Poland (GUS)

10-Year Government Bond Yield

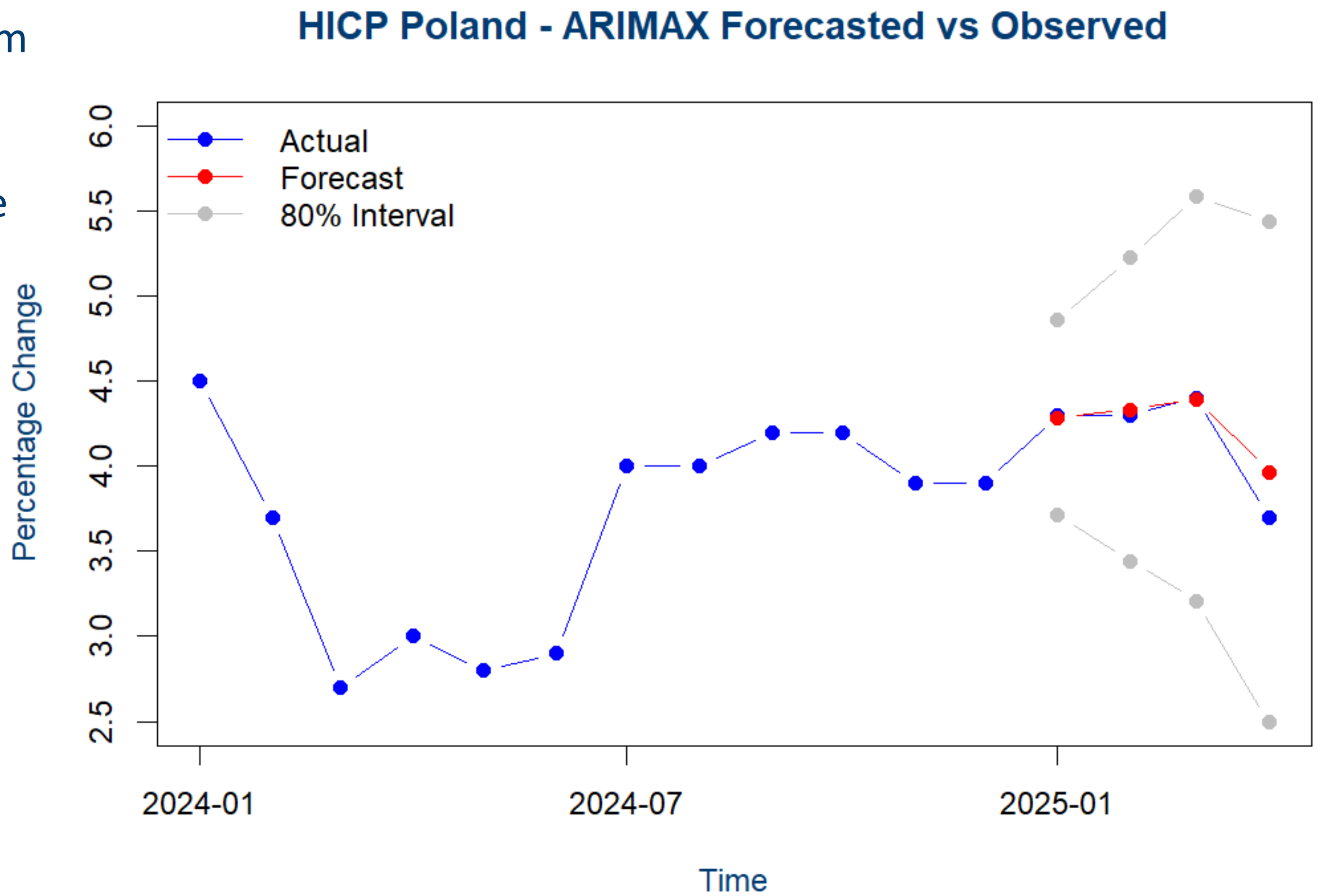
- Long-term yield reflects market expectations of future inflation and monetary policy
- Yields increase when a higher inflation or tighter monetary stance is expected
- Used 6-month lag based on test performance
- Data from FRED

ERSTE Group

Multivariate Model - ARIMAX

- Trained using ***auto.arima()*** with exogenous variables
- **Selected Model - ARIMA(2,0,1)(1,0,0)[12] + 11 regressors:**
  - 2 AR terms, 1 MA term and 1 seasonal AR term (12 months)
  - Residual variance $\sigma^2 \approx 0.199$
  - AIC ≈ 211.1 (slightly worse than the univariate model)
- **Forecast:**

```
##          Observed Forecast
## 2025 Jan      4.3 4.284120
## 2025 Feb      4.3 4.332600
## 2025 Mar      4.4 4.394824
## 2025 Apr      3.7 3.965498
```



ERSTE Group

Multivariate Model - ARIMAX

- **Forecast:**

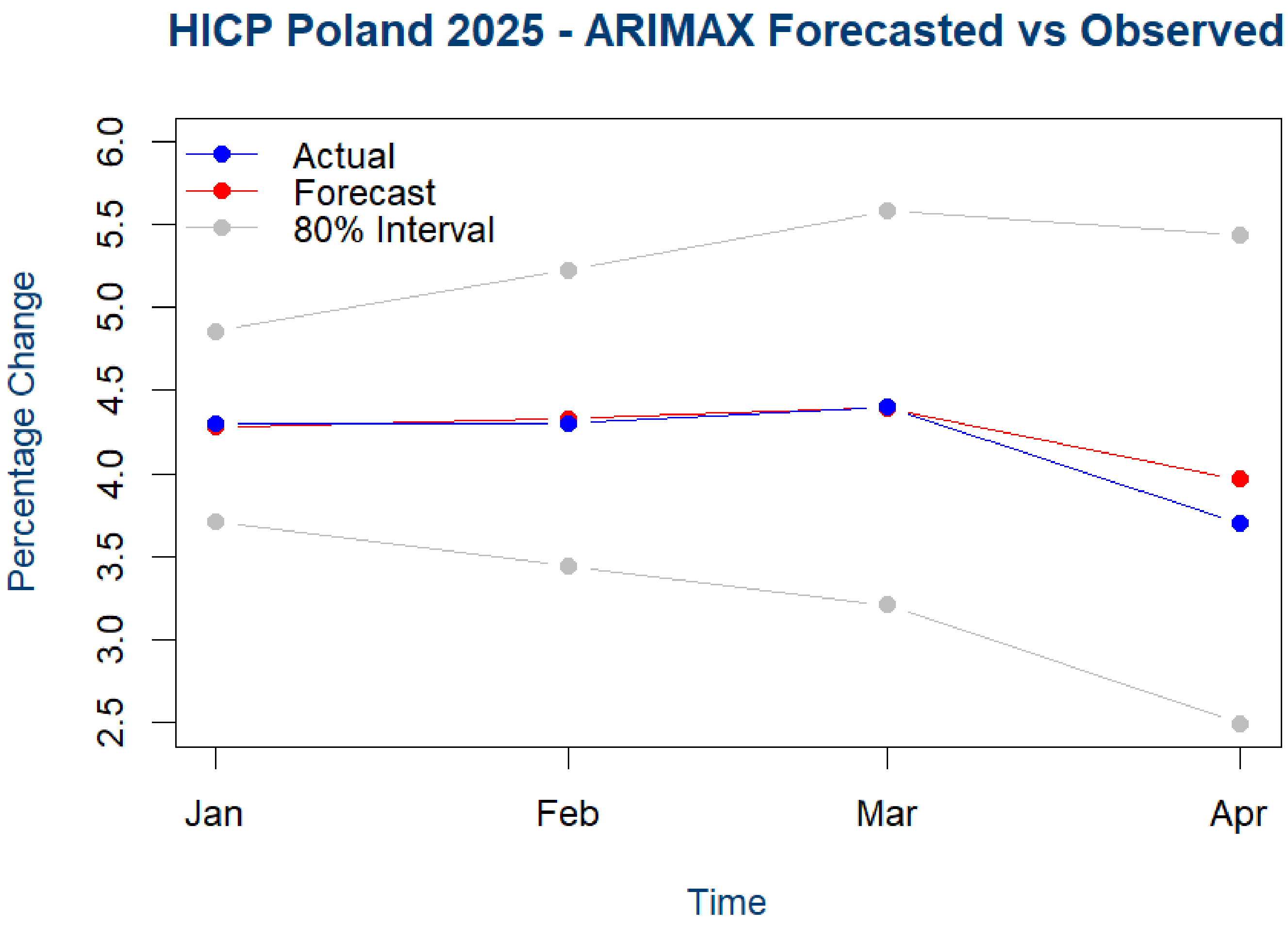
```
##           Observed Forecast
## 2025 Jan       4.3 4.284120
## 2025 Feb       4.3 4.332600
## 2025 Mar       4.4 4.394824
## 2025 Apr       3.7 3.965498
```

- AIC measures in-sample fit and penalizes model complexity => Slightly higher AIC in ARIMAX is due to more parameters (external variables)

- ARIMAX achieved better forecast accuracy for the forecasting period (2025/01 – 2025/04)

- External variables helped capture real-world shocks

- This improved out-of-sample generalization, even with a slight AIC trade-off



ERSTE Group

Thank you!

ERSTE Group