TÉCNICO LISBOA

DS 2025

Data Science Lab 5

Team nr: 2	**Student 1:** Åmund Grimstad	**IST nr:** 1116675
	Student 2: Arthur de Arruda Chau	**IST nr:** 1116090
	Student 3: Benjamin Raymond Kuhn	**IST nr:** 1115778
	Student 4: João Rafael Freitas Lourenço	**IST nr:** 425699

TIME SERIES ANALYSIS

Traffic Time Series

DATA PROFILING

Data Dimensionality and Granularity

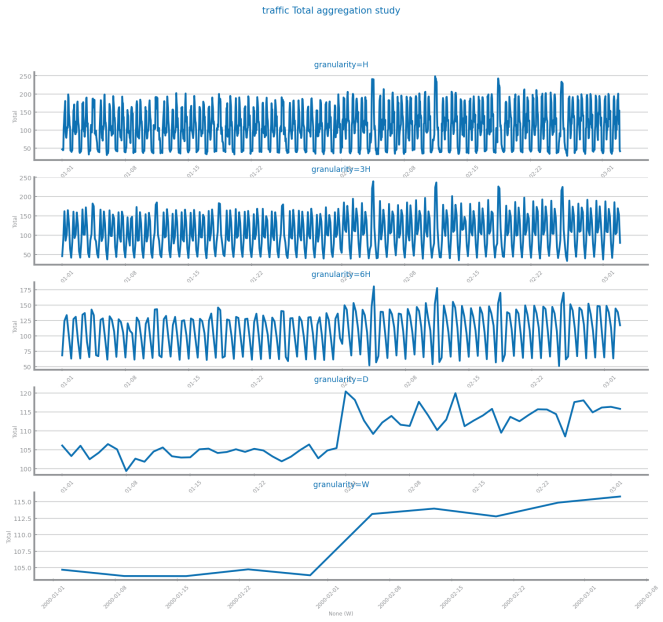


Figure 1: Traffic Time Series at five different granularities

Data Distribution

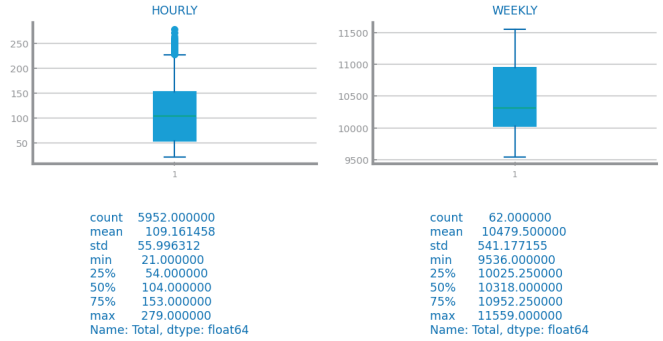


Figure 2: Boxplot(s) for Traffic Time Series

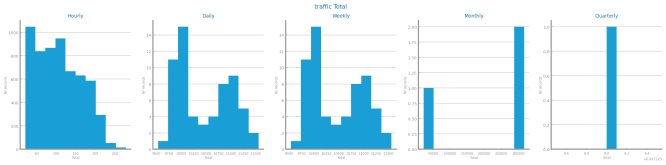


Figure 3: Histogram(s) for Traffic Time Series

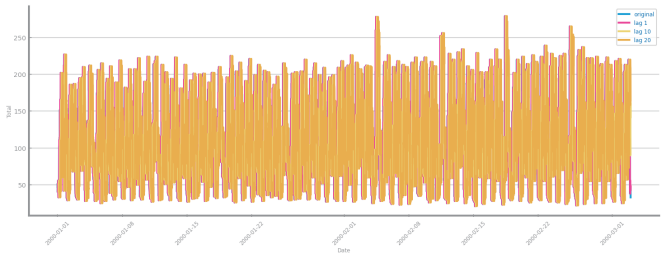


Figure 4: Autocorrelation lag-plots for original Traffic Time Series

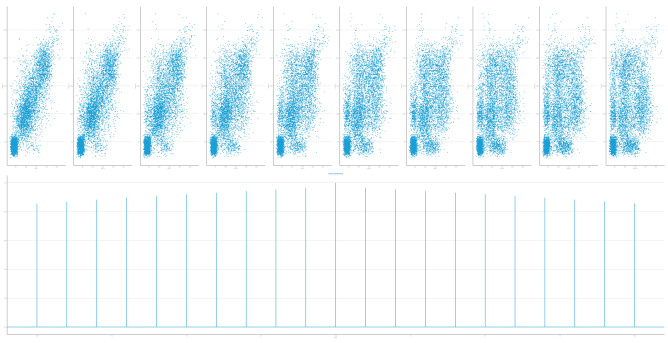


Figure 5: Autocorrelation correlogram for original Traffic Time Series

Data Stationarity

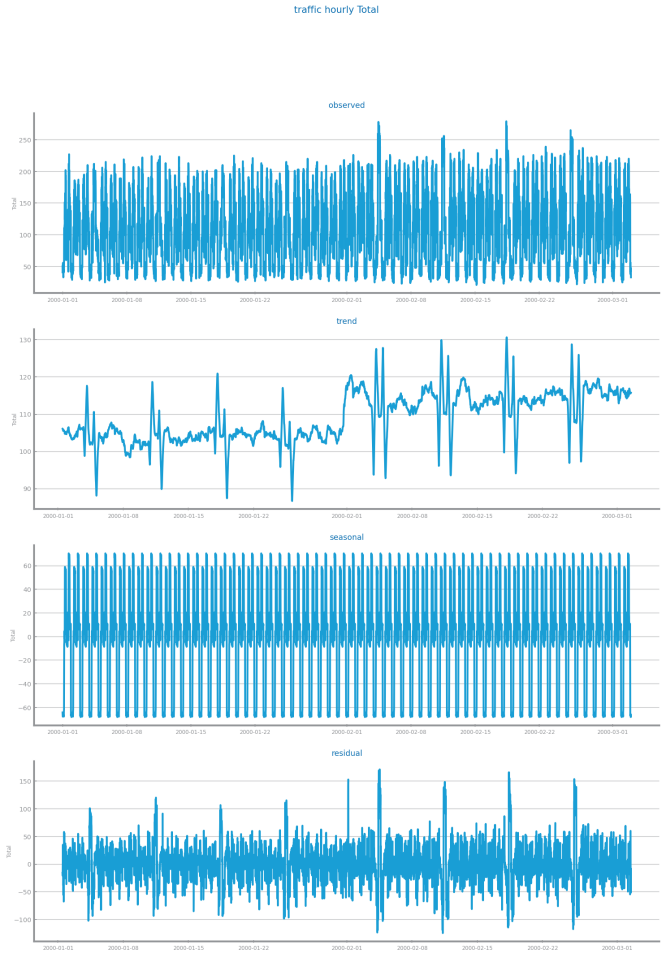


Figure 6: Components study for Traffic Time Series

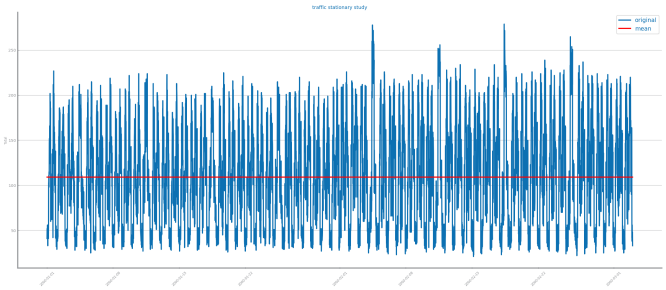


Figure 7: Stationarity study for Traffic Time Series

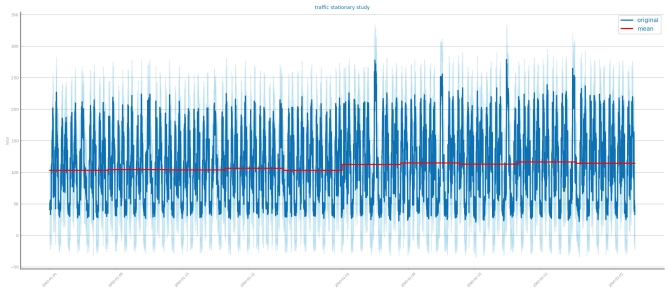


Figure 8: Stationarity study with bins for Traffic Time Series

Figure 9: Augmented Dickey-Fuller test results for Traffic Time Series

DATA TRANSFORMATION

Aggregation

Figure 10: Forecasting predictions from Persistence Realist and Linear Regression on **hourly Traffic Aggregation**

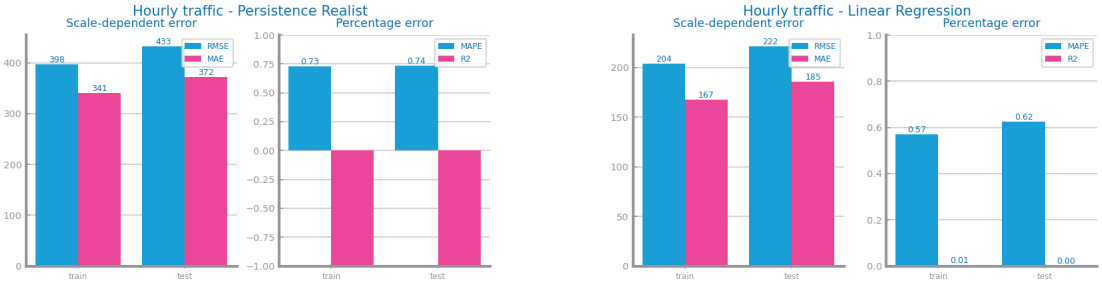


Figure 11: Performance of Persistence Realist and Linear Regression on **hourly Traffic Aggregation**

Figure 12: Forecasting predictions from Persistence Realist and Linear Regression on **3hourly Traffic Aggregation**

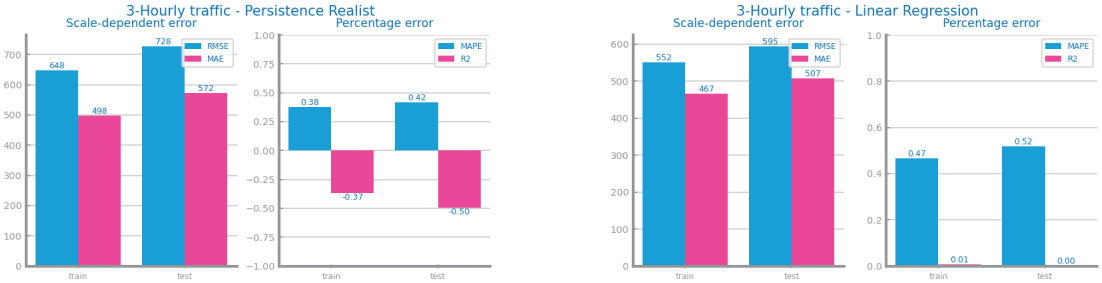


Figure 13: Performance of Persistence Realist and Linear Regression on **3hourly Traffic Aggregation**

Figure 14: Forecasting predictions from Persistence Realist and Linear Regression on **6hourly Traffic Aggregation**

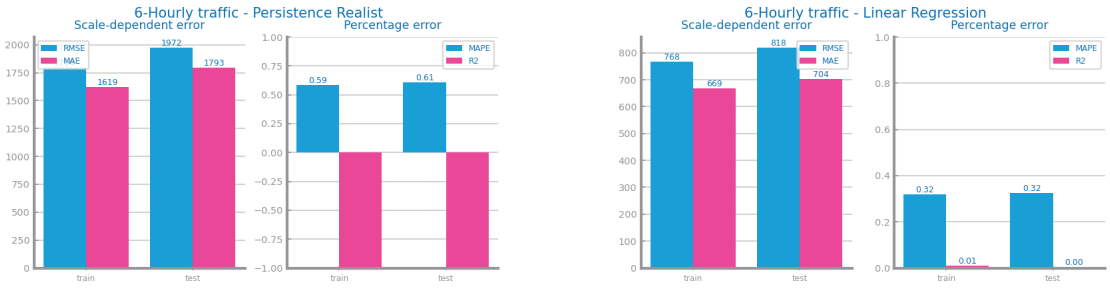


Figure 15: Performance of Persistence Realist and Linear Regression on **6hourly Traffic Aggregation**

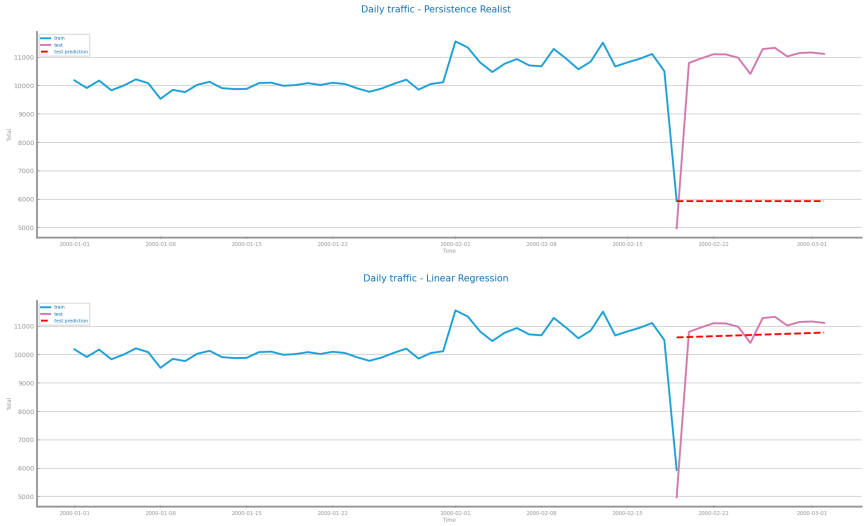


Figure 16: Forecasting predictions from Persistence Realist and Linear Regression on **daily Traffic Aggregation**

Figure 17: Performance of Persistence Realist and Linear Regression on **daily Traffic Aggregation**

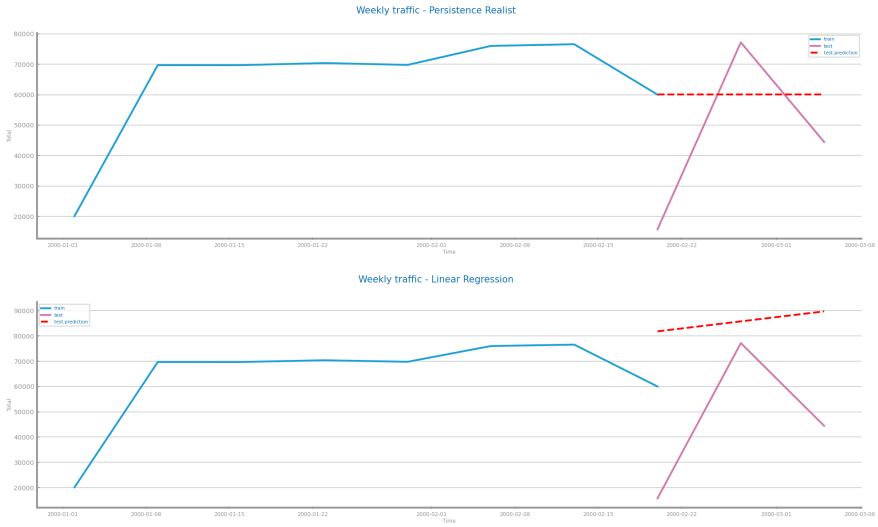


Figure 18: Forecasting predictions from Persistence Realist and Linear Regression on **weekly Traffic Aggregation**

Figure 19: Performance of Persistence Realist and Linear Regression on **weekly Traffic Aggregation**

Differentiation

Figure 20: Forecasting plots after first and second differentiation of Traffic Time Series

Figure 21: Forecasting plots after first and second differentiation of Traffic Time Series

Smoothing

Figure 22: Forecasting plots after different smoothing parameterisations on Traffic Time Series

Figure 23: Forecasting results after different smoothing parameterisations on Traffic Time Series

Scaling

Figure 24: Forecasting plots after different scaling parameterisations on Traffic Time Series

Figure 25: Forecasting results after different scaling parameterisations on Traffic Time Series

Inflation Rate Time Series

DATA PROFILING

Data Dimensionality and Granularity

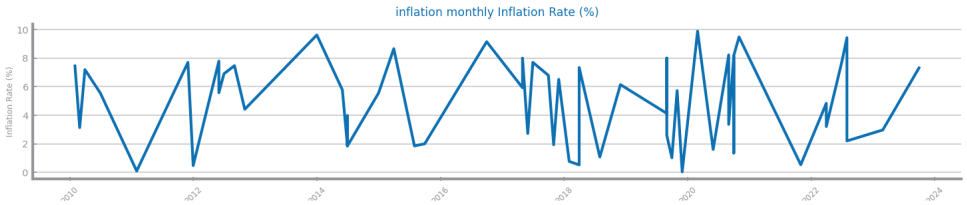


Figure 26: U.S. Inflation Rate dimensionality over time.

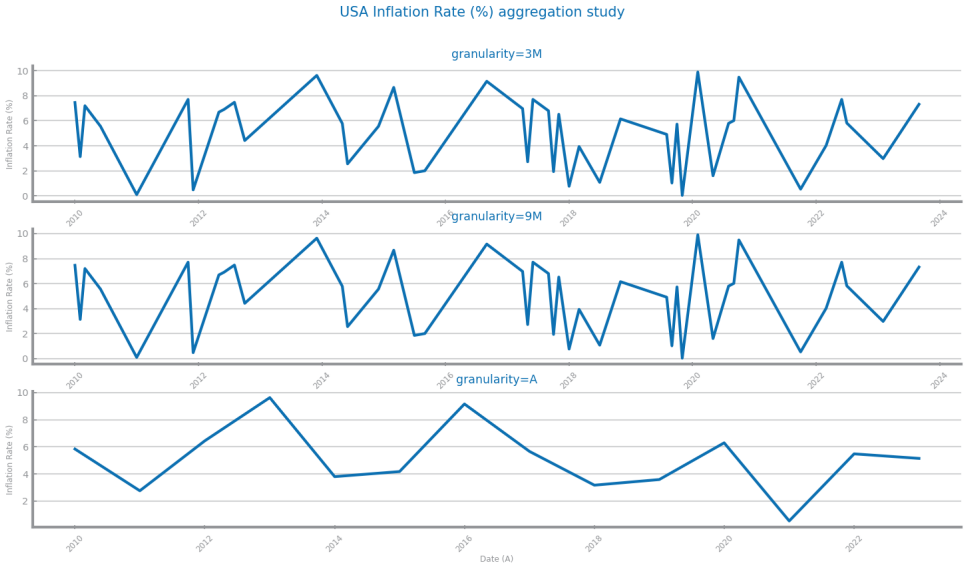
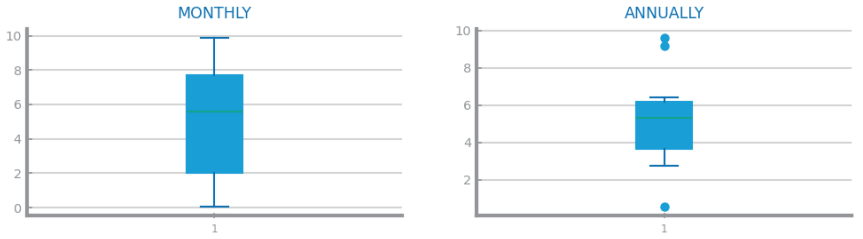


Figure 27: Inflation Rate Time Series at three different granularities

Data Distribution



count	57.000000
mean	5.016842
std	3.000707
min	0.030000
25%	2.000000
50%	5.580000
75%	7.710000
max	9.900000

Name: Inflation Rate (%), dtype: float64

count	14.000000
mean	5.118488
std	2.420472
min	0.530000
25%	3.638750
50%	5.312500
75%	6.179375
max	9.630000

Name: Inflation Rate (%), dtype: float64

Figure 28: Boxplot(s) for Inflation Rate Time Series

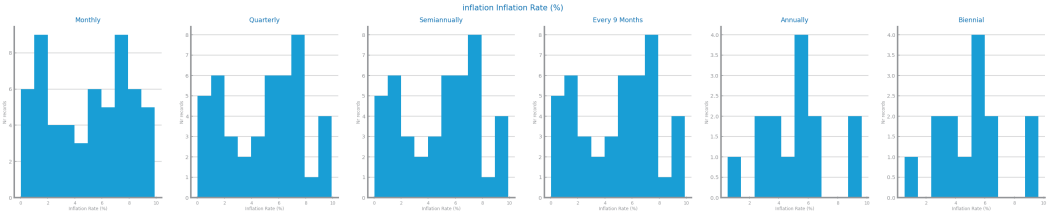


Figure 29: Histogram(s) for Inflation Rate Time Series

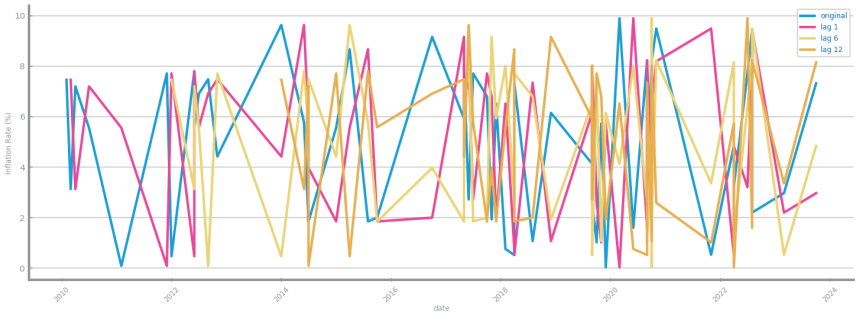


Figure 30: Autocorrelation lag-plots for original Inflation Rate Time Series

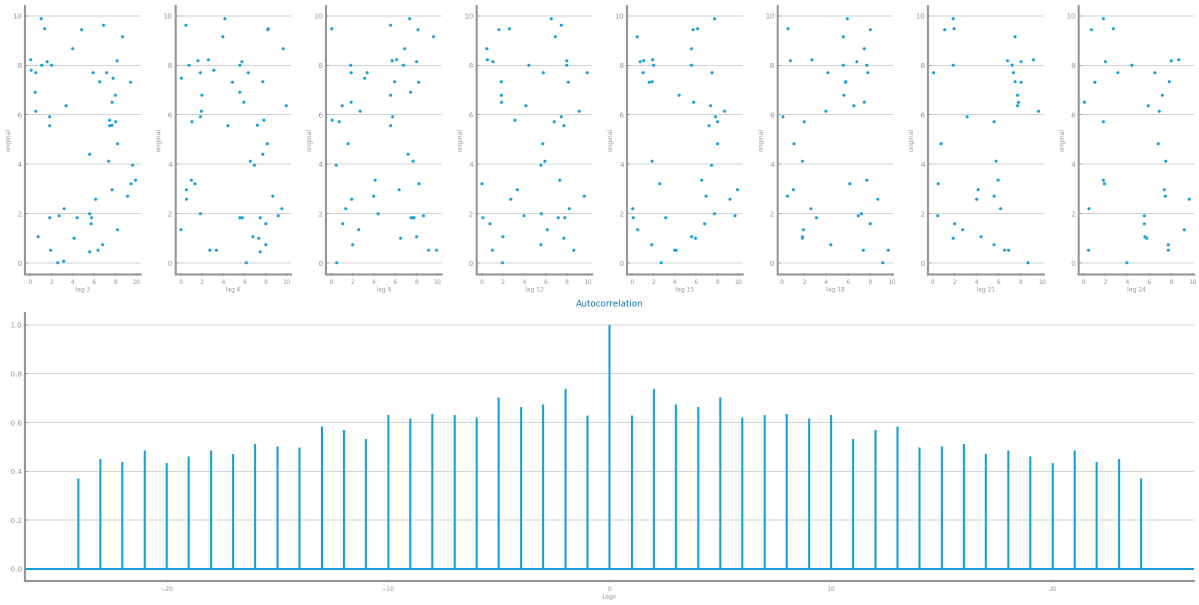


Figure 31: Autocorrelation correlogram for original Inflation Rate Time Series

Data Stationarity

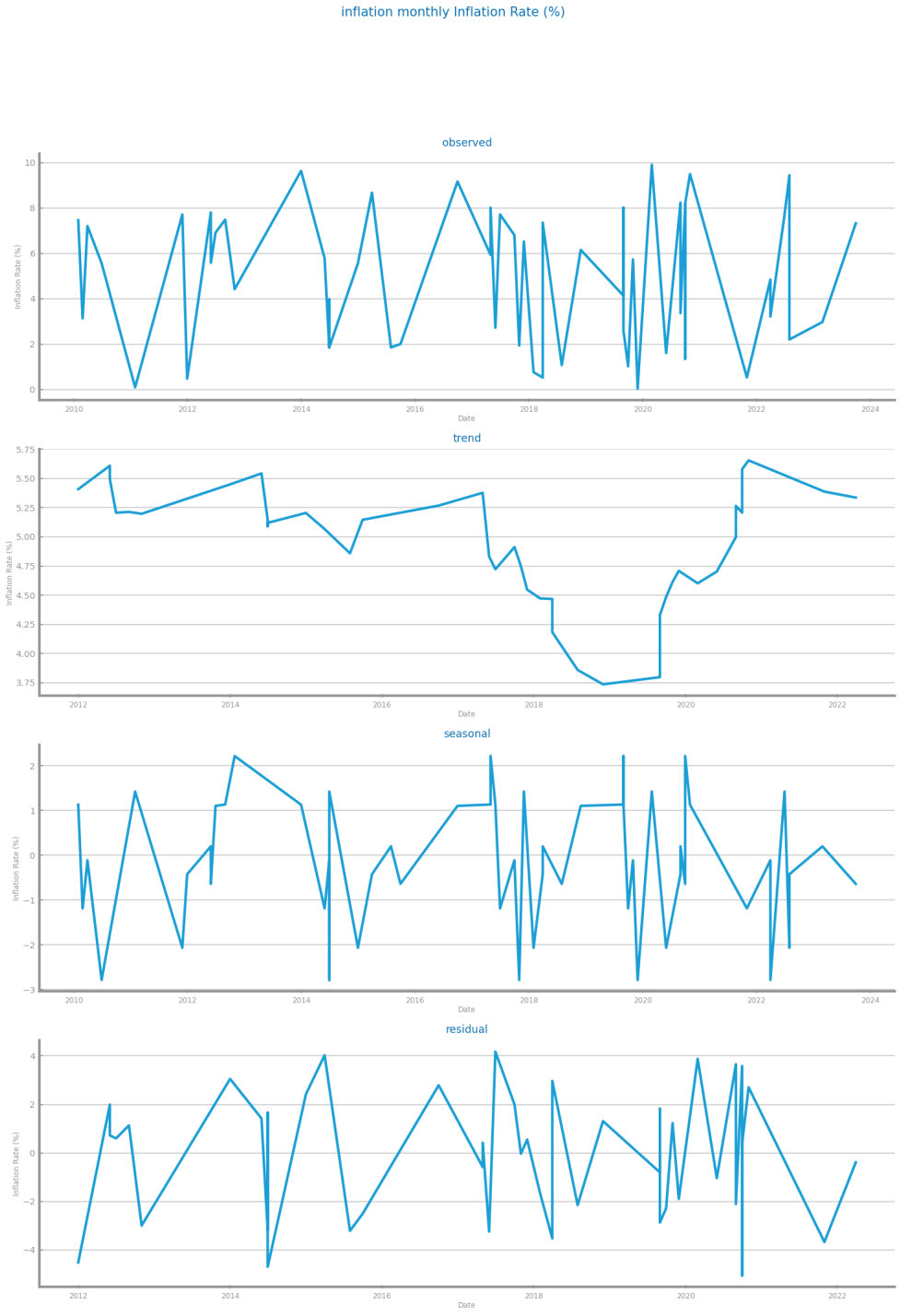


Figure 32: Components study for Inflation Rate Time Series

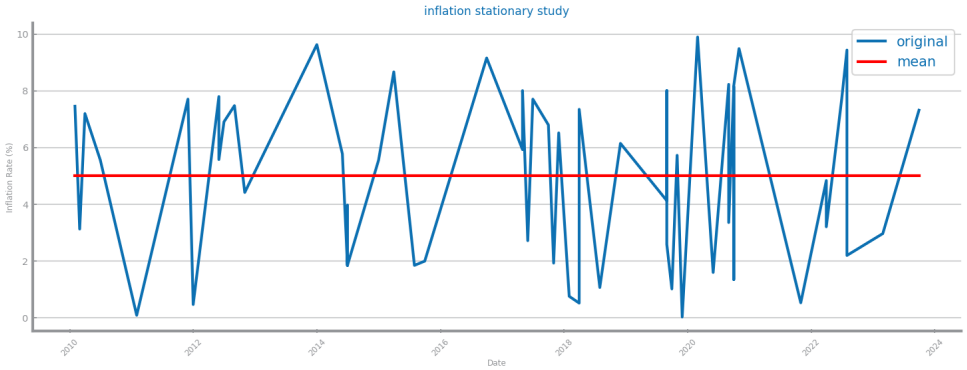


Figure 33: Stationarity study for Inflation Rate Time Series

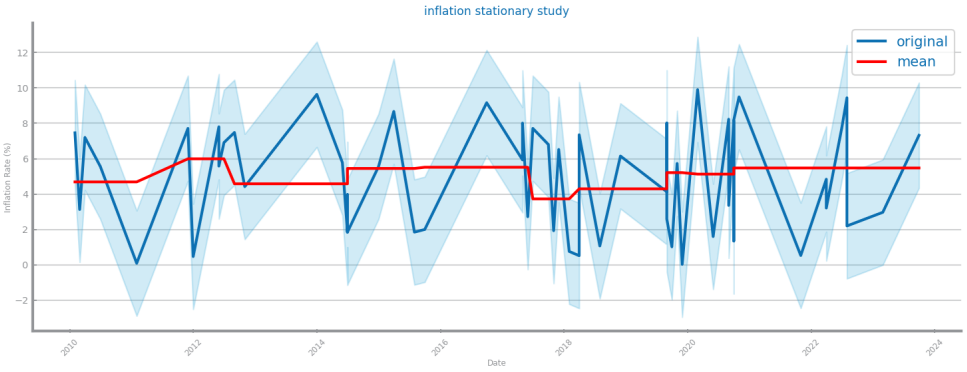


Figure 34: Stationarity study with bins for Inflation Rate Time Series

```
ADF Statistic: -10.410
p-value: 0.000
Critical Values:
        1%: -3.553
        5%: -2.915
        10%: -2.595
The series is stationary
```

Figure 35: Augmented Dickey-Fuller test results for Inflation Rate Time Series

DATA TRANSFORMATION

Aggregation

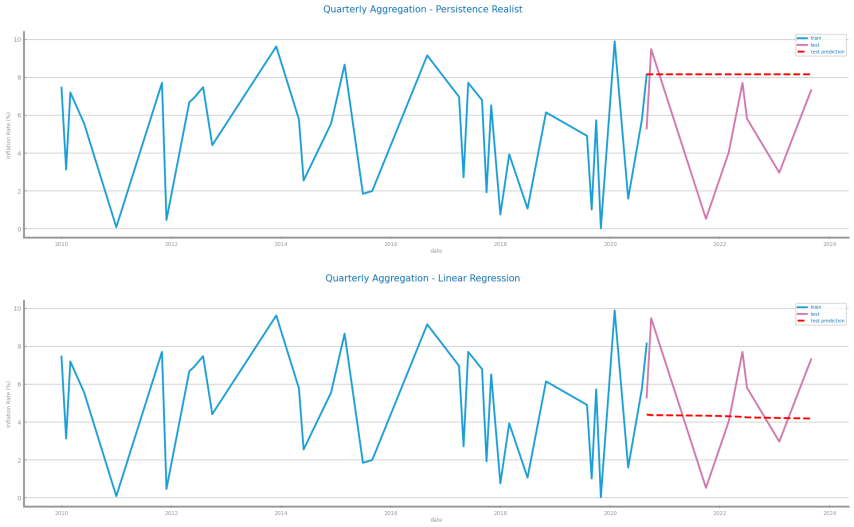


Figure 36: Forecasting predictions from Persistence Realist and Linear Regression on **Quarterly Inflation Aggregation**

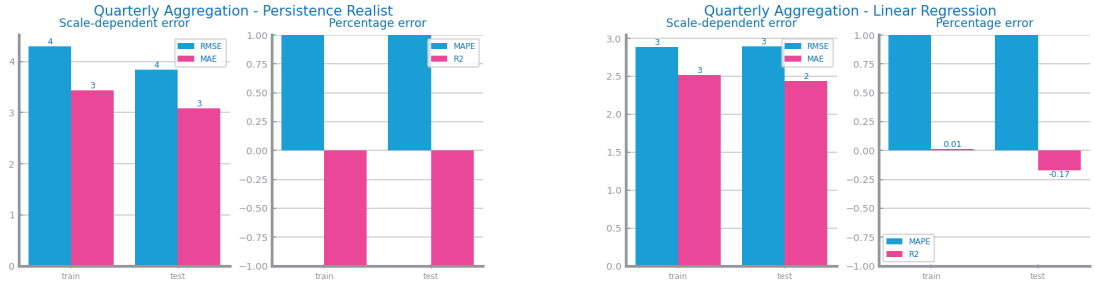


Figure 37: Performance of Persistence Realist and Linear Regression on **Quarterly Inflation Aggregation**

Figure 38: Forecasting predictions from Persistence Realist and Linear Regression on **Semi-Annual Inflation Aggregation**

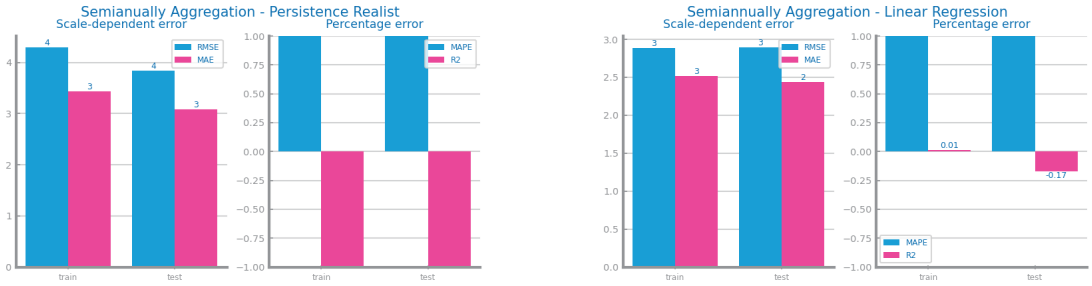


Figure 39: Performance of Persistence Realist and Linear Regression on **Semi-Annual Inflation Aggregation**

Figure 40: Forecasting predictions from Persistence Realist and Linear Regression on **Annual Inflation Aggregation**

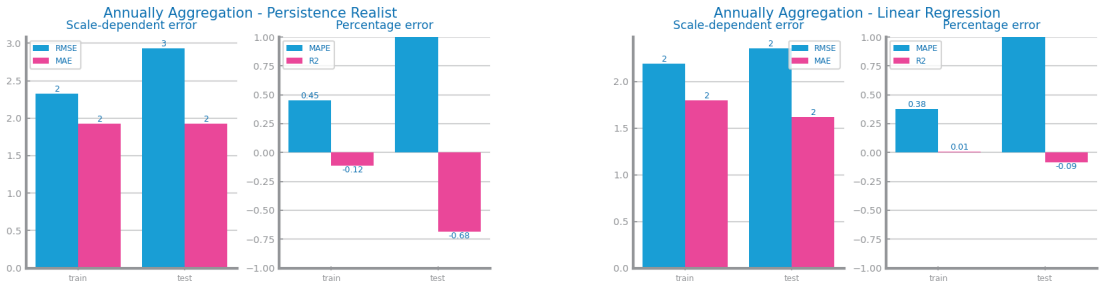


Figure 41: Performance of Persistence Realist and Linear Regression on **Annual Inflation Aggregation**

For all aggregation techniques, the Linear Regression model yields a lower MAE and RMSE than the Persisent Realist Model for the test set. Among the Linear Regression models, the Annual aggregation yields the lowest MAE for the test set (1.6).

Differentiation

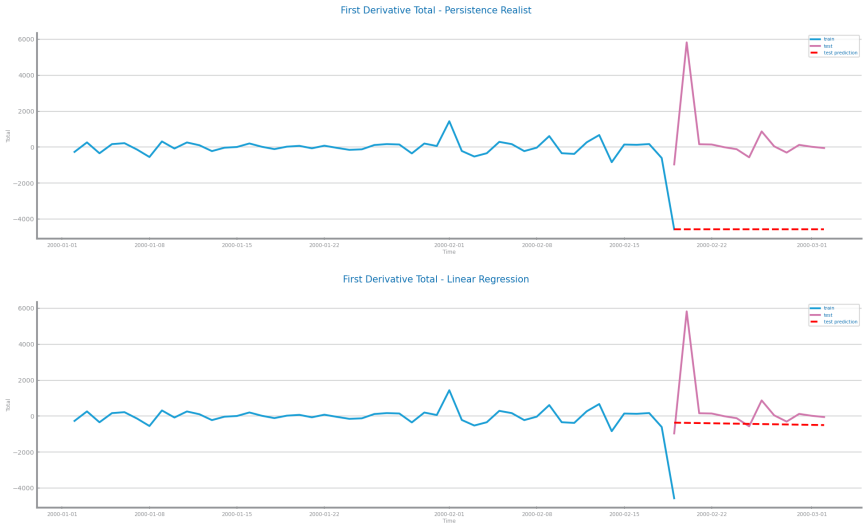


Figure 42: Forecasting predictions from Persistence Realist and Linear Regression on **Annual Inflation Differentiation (1st Order)**

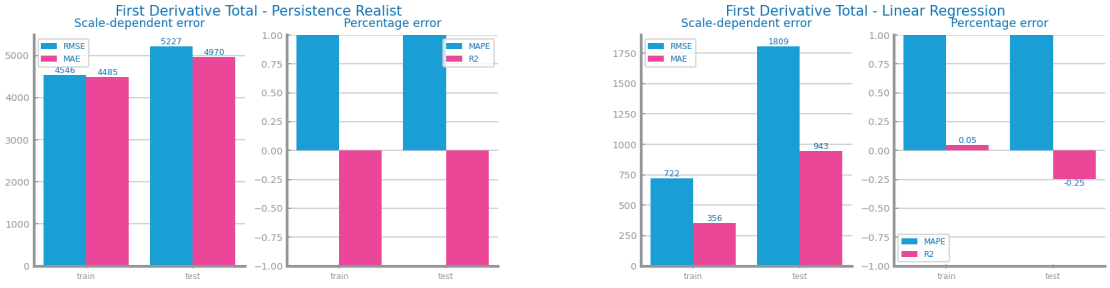


Figure 43: Performance of Persistence Realist and Linear Regression on **Annual Inflation Differentiation (1st Order)**

Figure 44: Forecasting predictions from Persistence Realist and Linear Regression on **Annual Inflation Differentiation (2nd Order)**

Figure 45: Performance of Persistence Realist and Linear Regression on **Annual Inflation Differentiation (2nd Order)**

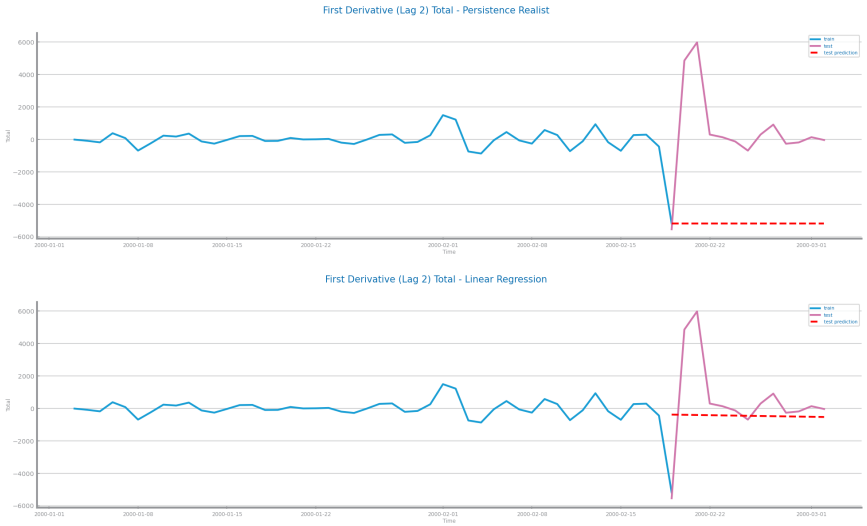


Figure 46: Forecasting predictions from Persistence Realist and Linear Regression on **Annual Inflation Differentiation with Lag 2**

Figure 47: Performance of Persistence Realist and Linear Regression on **Annual Inflation Differentiation with Lag 2**

for diff_1 and diff_2, the Linear Regression Model yields a lower MAE and RMSE for the test set than the Persistent Realist models. For the diff_lag2 treatment, the Persistent Realist has a lower MAE (but higher RMSE) than the Linear Regression Model. But overall, the best model, a Linear Regression Model trained on the diff_1 treatment, is worse than the best model without differential treatment. So, we do not differentiate.

Smoothing

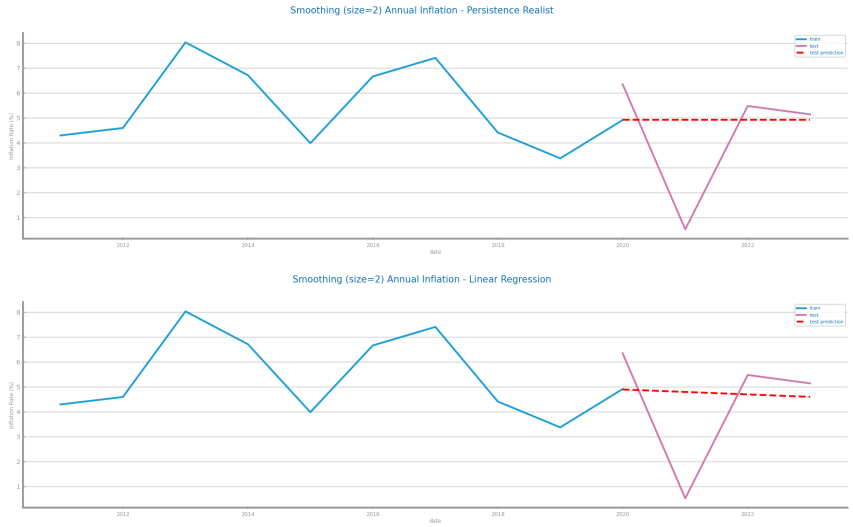


Figure 48: Forecasting predictions from Persistence Realist and Linear Regression on **Annual Inflation Smoothing (windowSize=2)**

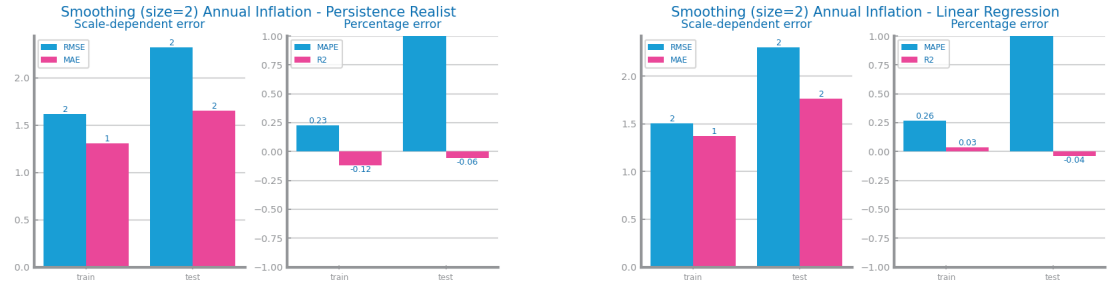


Figure 49: Performance of Persistence Realist and Linear Regression on **Annual Inflation Smoothing (windowSize=2)**

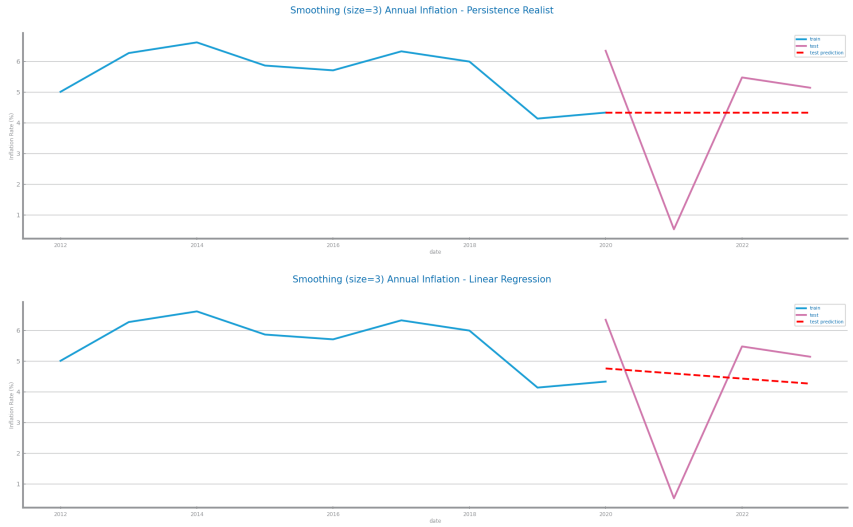


Figure 50: Forecasting predictions from Persistence Realist and Linear Regression on **Annual Inflation Smoothing (windowSize=3)**

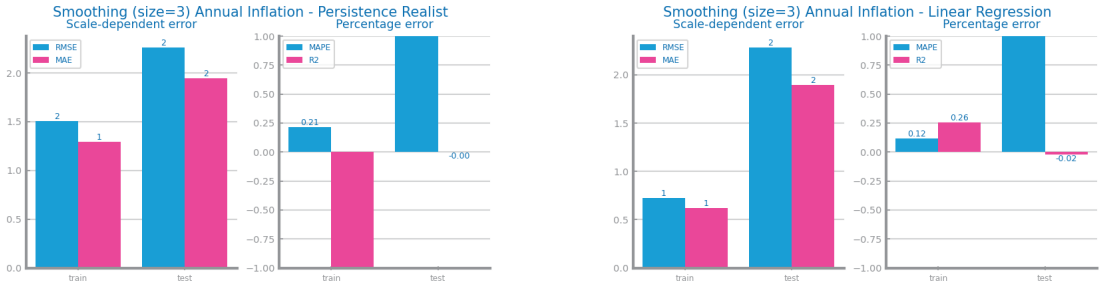


Figure 51: Performance of Persistence Realist and Linear Regression on **Annual Inflation Smoothing (windowSize=3)**

Figure 52: Forecasting predictions from Persistence Realist and Linear Regression on **Annual Inflation Smoothing (windowSize=4)**

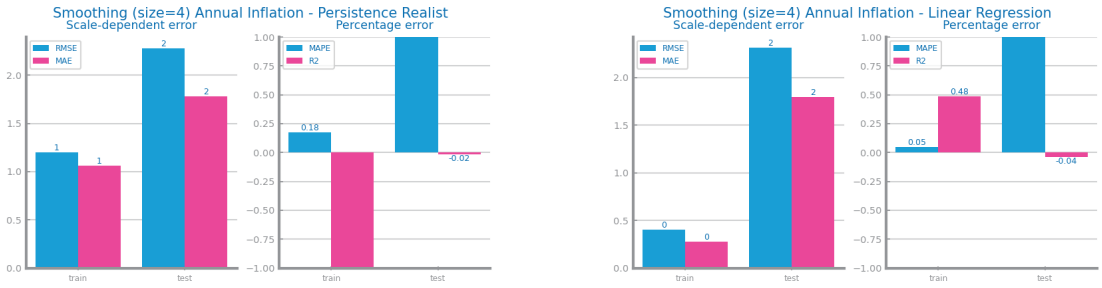


Figure 53: Performance of Persistence Realist and Linear Regression on **Annual Inflation Smoothing (windowSize=4)**

Choosing the smoothing with windowSize=2 since the Persistence Scored better than the best after aggregation and differentiation.

Scaling

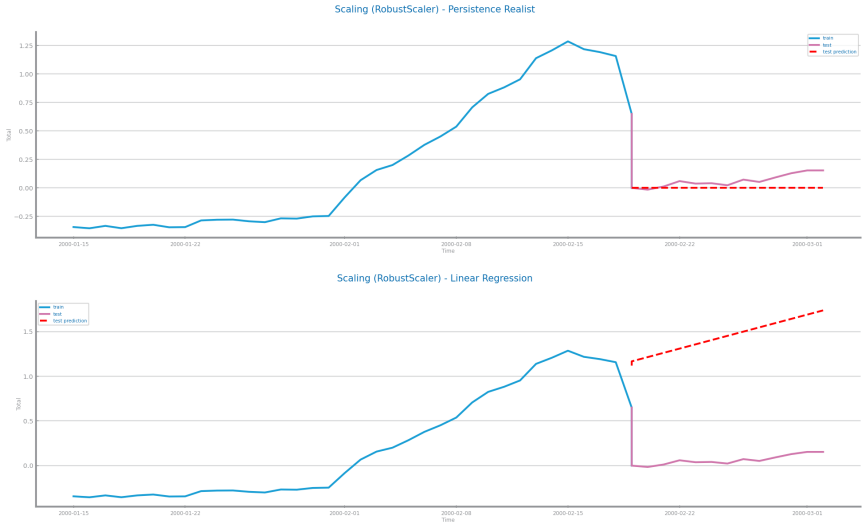


Figure 54: Forecasting predictions from Persistence Realist and Linear Regression on **Annual Inflation Scaling - RobustScaler**

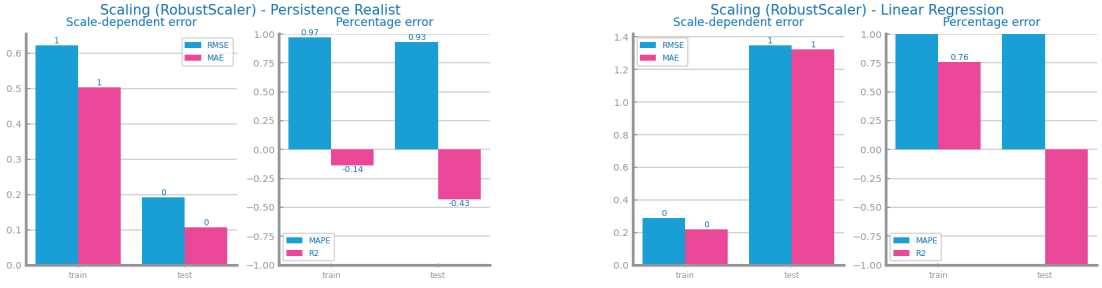


Figure 55: Performance of Persistence Realist and Linear Regression on **Annual Inflation Scaling - RobustScaler**

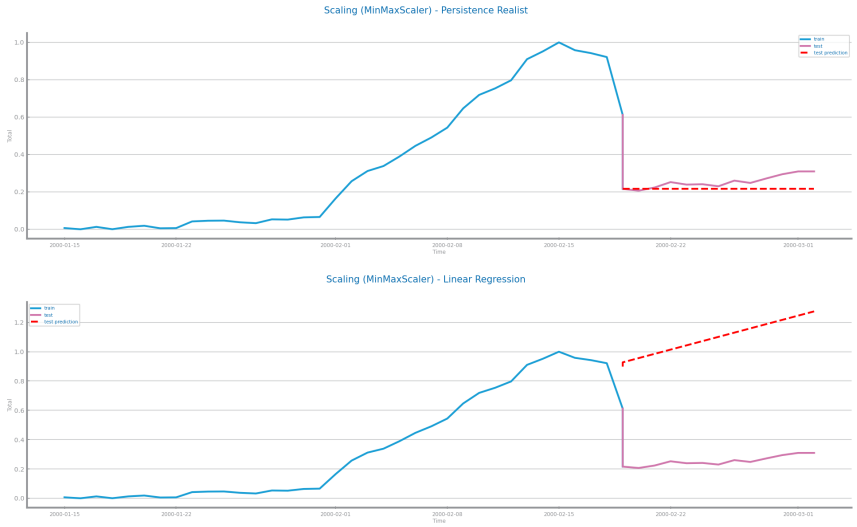


Figure 56: Forecasting predictions from Persistence Realist and Linear Regression on **Annual Inflation Scaling - MinMaxScaler**

Figure 57: Performance of Persistence Realist and Linear Regression on **Annual Inflation Scaling - MinMaxScaler**

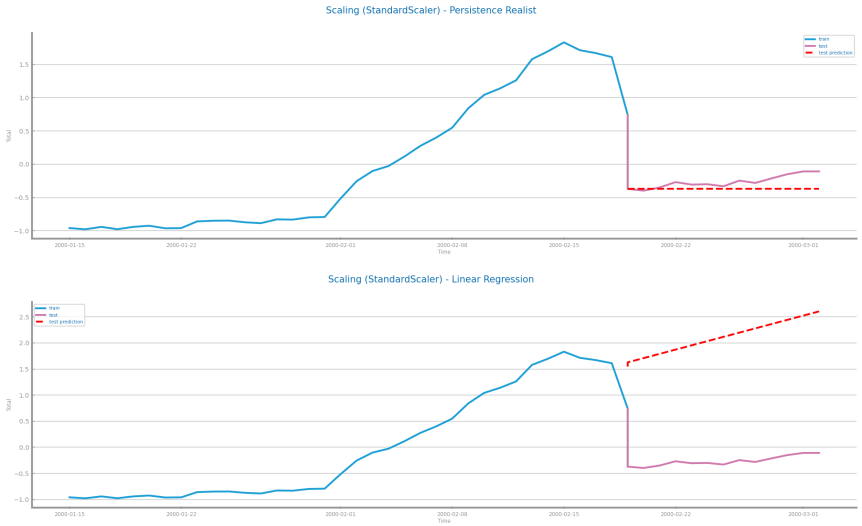


Figure 58: Forecasting predictions from Persistence Realist and Linear Regression on **Annual Inflation Scaling - StandardScaler**

Figure 59: Performance of Persistence Realist and Linear Regression on **Annual Inflation Scaling - StandardScaler**

The best results are achieved with MinMaxScaler combined with the Persistence Realist model, due to its lowest Test MAE (0.35) and significantly better Test MAPE (0.61).