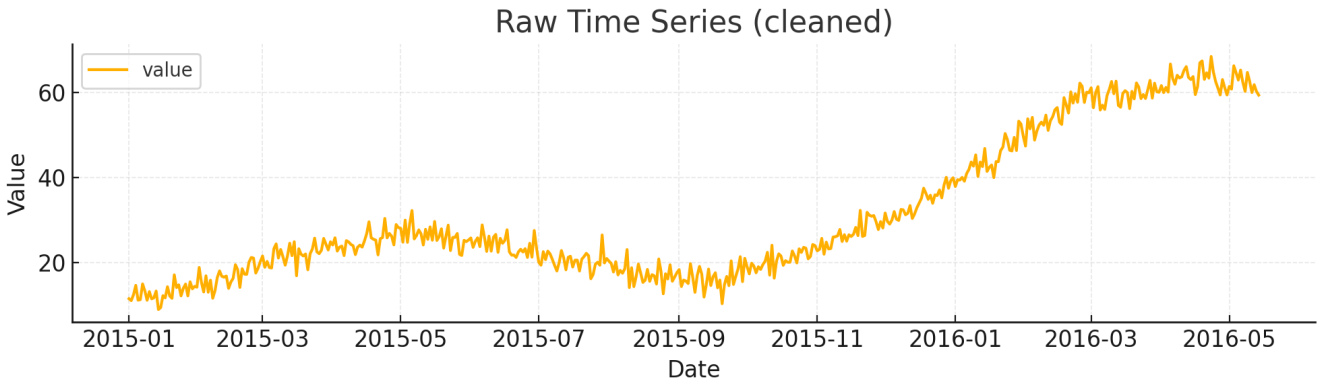
Full Project Report: Time Series Analysis & Models

Dataset path in workspace:
/mnt/data/time_series_dataset.csv

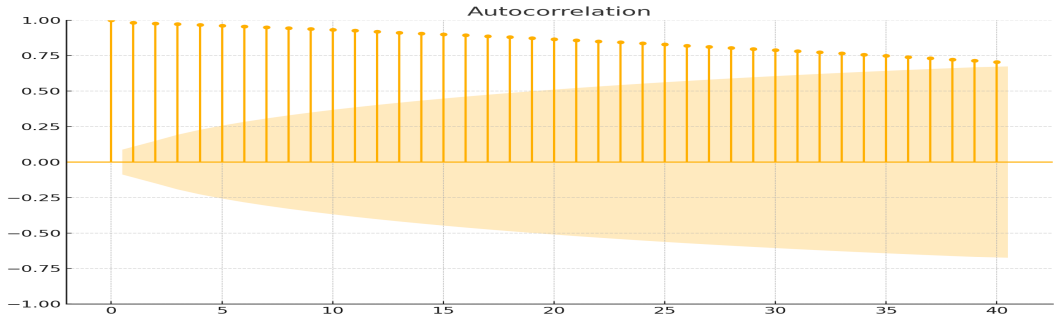
1) Dataset summary

Total rows: 500
Columns: ['value']
Train rows: 400 (as used in experiments)
Test rows: 100
No missing values detected.

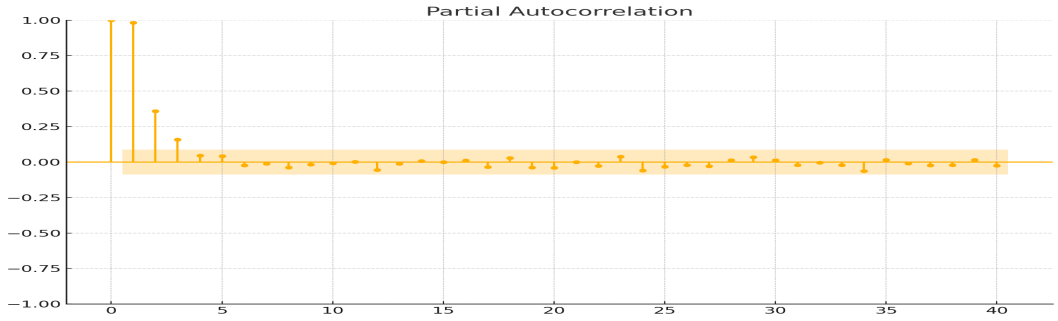
Raw time series plot



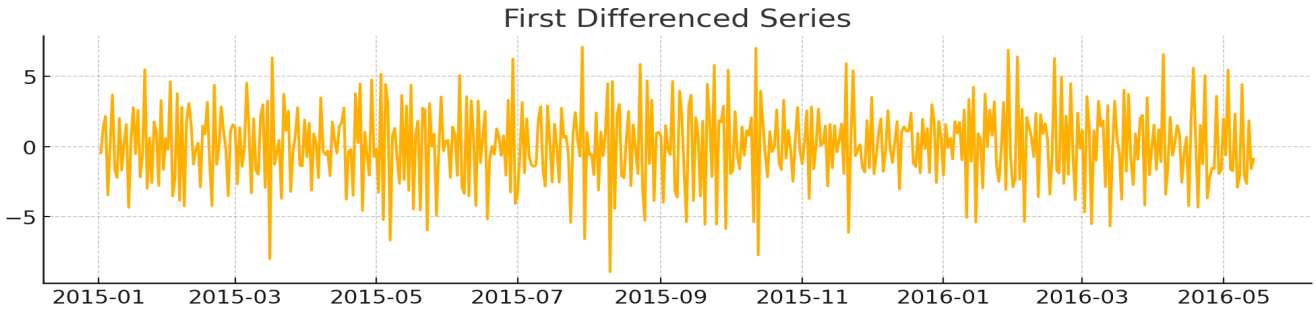
ACF plot



PACF plot



First differenced series (plot)



ADF Test Result

ADF statistic: 0.718863
p-value: 0.990207
Critical value (1%): -3.4436840273842058
Critical value (5%): -2.8674204319994674
Critical value (10%): -2.5699020441557052

Model performance summary

Model performance recorded in experiments:
- Custom KF (fixed Q/R): MAE = 21.18998, RMSE = 25.13353, MAPE = 34.59983%
- UnobservedComponents (MLE baseline): MAE = 13.78422, RMSE = 16.98045, MAPE = 22.48700%
- EM-enhanced KF (EM run): MAE = 156.20694, RMSE = 180.35882, MAPE = 254.26001%
- SARIMA (1,1,1)(1,1,1)[7]: MAE = 7.53170, RMSE = 9.54703, MAPE = 12.34873%

Python code used (snippet):

```
# Example: ACF, PACF, ADF in Python
import pandas as pd
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
from statsmodels.tsa.stattools import adfuller

df = pd.read_csv("/mnt/data/time_series_dataset.csv", parse_dates=["date"]).set_index("date")
ts = df['value']

# ACF and PACF plots
plot_acf(ts, lags=40)
plot_pacf(ts, lags=40, method='ywm')

# ADF test
adf_res = adfuller(ts)
print("ADF statistic:", adf_res[0])
print("p-value:", adf_res[1])
```

Saved train/test CSVs in workspace:
data_train.csv, data_test.csv (saved in current working directory)