TUNIS BUSINESS SCHOOL
UNIVERSITY OF TUNIS

MODELING AND FORECASTING BANK CREDIT USING THE BOX–JENKINS METHODOLOGY

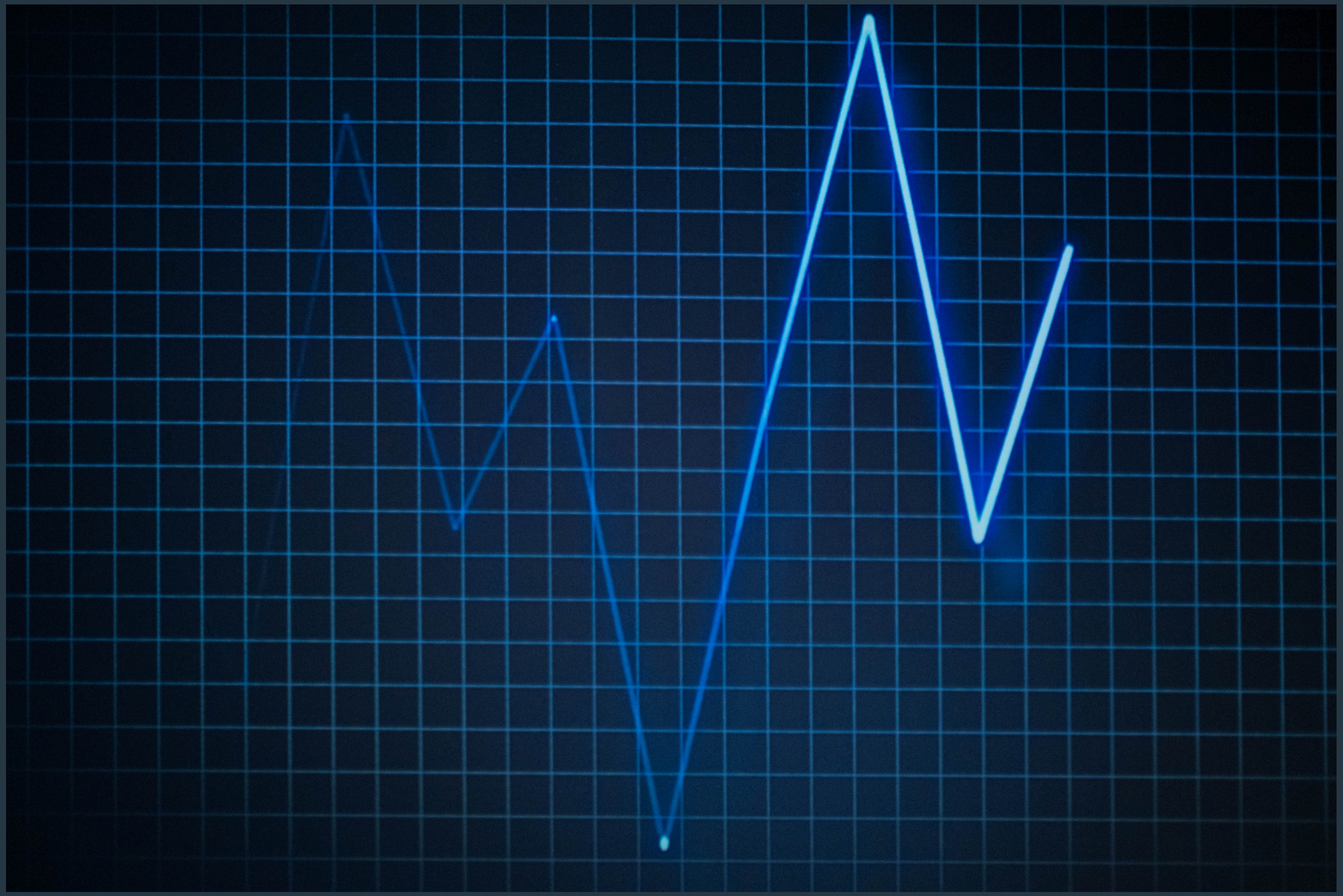
TIME SERIES ANALYSIS OF “SECURITIES IN BANK CREDIT – ALL COMMERCIAL BANKS”

Time Series Analysis Project — 2025

Prepared by: Linda Zaibi, Tassnime Raddaoui & Yasmine Yazidi

OUTLINE

1. Introduction
2. Problem Statement
3. Box-Jenkins Methodology
4. The Data: Securities Invested In Bank Credit.
5. Testing for Stationarity
6. Choosing the Model & Parameters
7. Forecasting
8. Conclusion



INTRODUCTION

Why this data?
This dataset shows how much banks invest in securities (like Treasury bonds). It’s a core part of the banking system’s balance sheet.

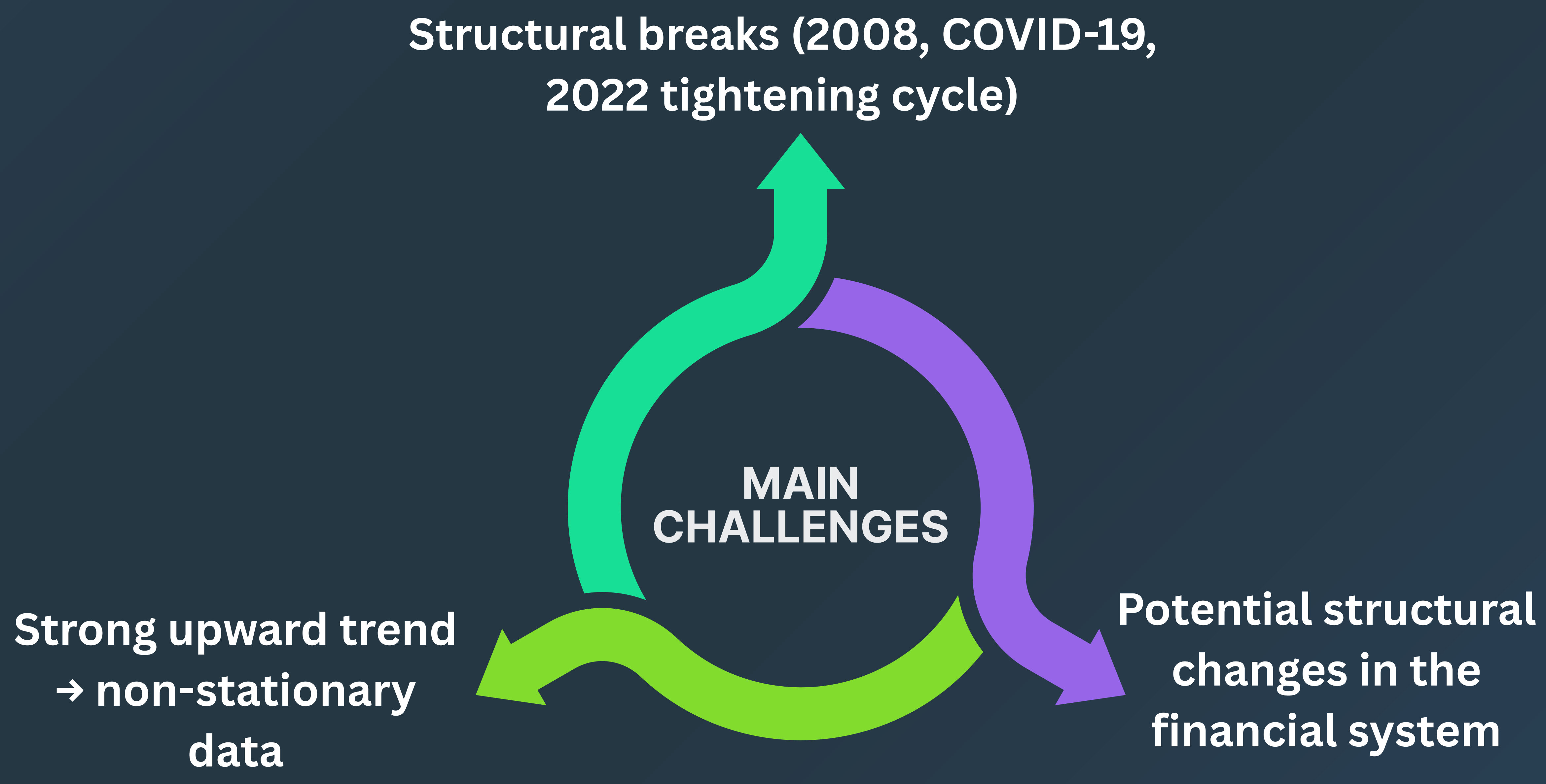
- Bank credit reflects economic activity, lending conditions, and financial stability.
- Movements in credit influence businesses, households, and monetary policy.
- The dataset is long, monthly, and reliable → ideal for time-series modeling.

We wanted to understand how banks adjust their investment positions over time — and whether we can predict that behavior for the next 30 days.

Our goal: Understand its behavior + generate a short-term forecast.

PROBELM STATEMENT

How can we model the dynamics of **U.S. bank credit** using ARIMA/SARIMA to produce reliable, interpretable short-term forecasts?



BOX-JENKINS METHODOLOGY

Step 01

Identification

stationarity, transformations, ACF/PACF

Step 02

Estimation

selecting ARIMA(p,d,q)

Step 03

Diagnostic Checking

residual tests, AIC/BIC

Step 04

Forecasting

generating multi-step forecasts

Step 05

Extensions

structural breaks, volatility insights

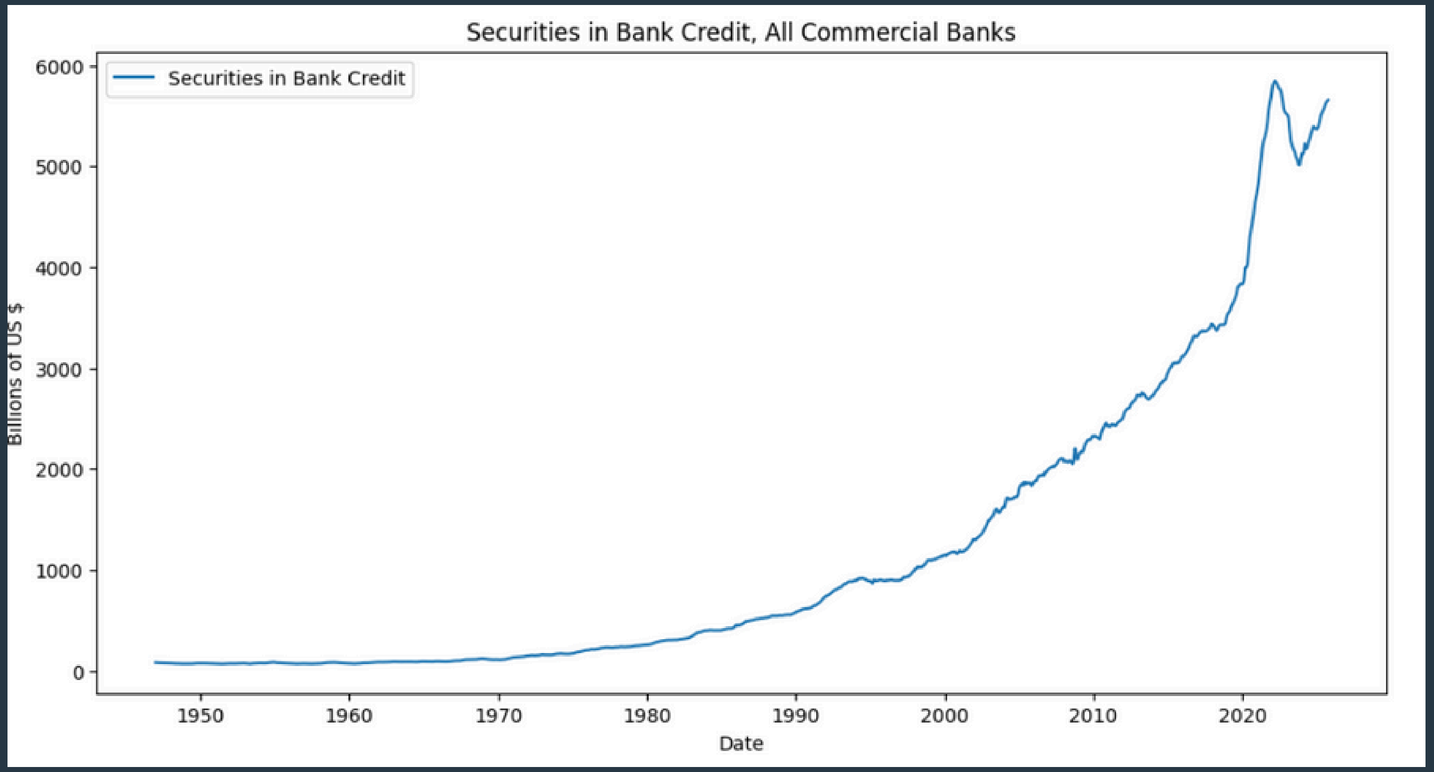
STEP 1: VISUAL & TREND DIAGNOSIS (WHAT WE DID & FOUND)

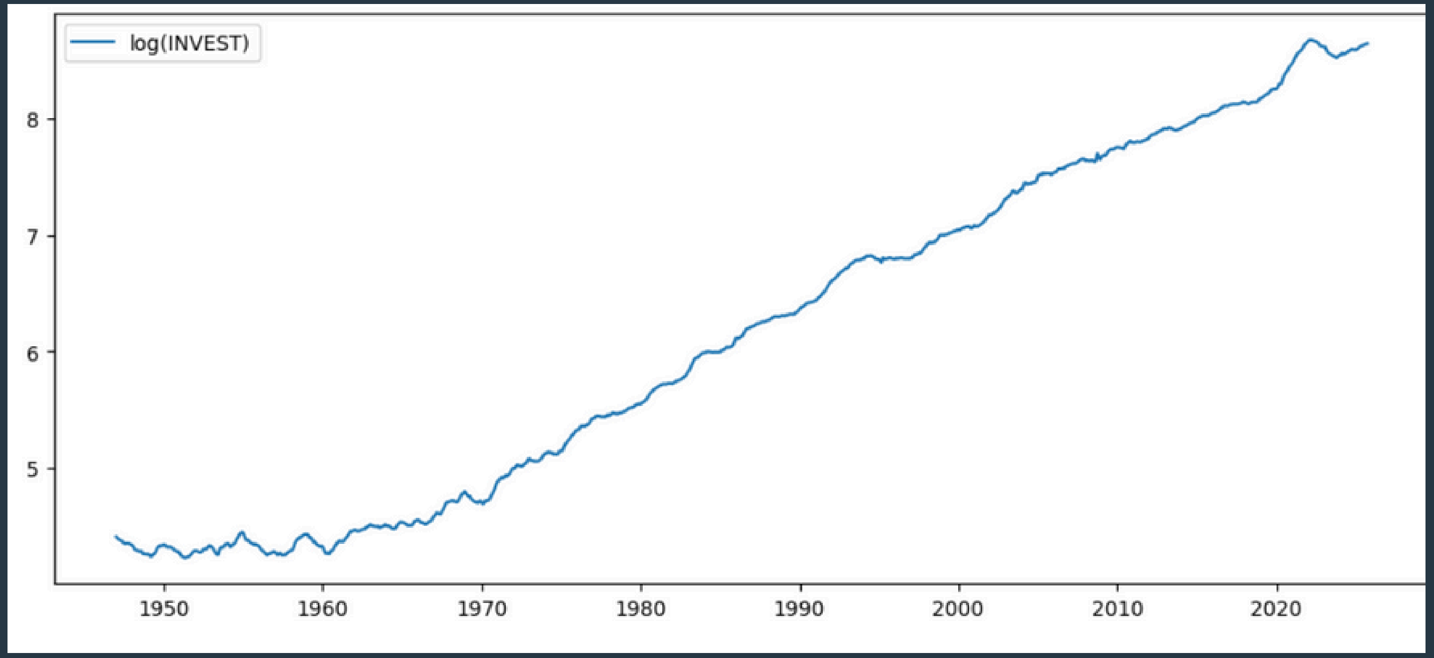
- We first plotted the original series and noticed that it was growing very quickly, almost like an exponential curve.
- Then we calculated a 60-month rolling average to see the long-term trend.

→ This rolling curve confirmed that the growth wasn’t random it had a clear, smooth upward pattern.

- Because the growth looked too steep and curved, we applied a log transformation to make the data more stable and easier to model.

- After taking the log, the series became much ***straighter, smoother, and more linear.***
- This confirmed that using a log transformation was the right choice to reduce the effect of rapid growth and ***prepare the data for ARIMA modeling.***





STEP 2: CHECKING STATIONARITY

What We Did:

- Tested if the series moves randomly or has a trend using ADF and KPSS.
- Checked raw, logged, and % change series.
- Applied first difference of log to stabilize fluctuations.

What We Found:

- Raw and log series → clearly trending → not stable.
- Percent change → more stable but still a bit unpredictable.
- Decision: Needed differencing → log + difference ready for modeling.

Is the Data Stable Over Time?

```
--- ADF Test: INVEST pct change ---
ADF stat -6.6830, p-value 0.0000
Used lags: 22 nobs: 922
crit values: {'1%': np.float64(-3.437462363899248), '5%': np.float64(-2.8646798473884134),
7076)}

--- KPSS Test (c): INVEST pct change ---
KPSS stat 0.5967, p-value 0.0229
Used lags: 15
crit values: {'10%': 0.347, '5%': 0.463, '2.5%': 0.574, '1%': 0.739}
```

```
#After comfirming that the time series we have is not stationary we move to the detrending by differencing
data["log_INVEST"] = np.log(data["INVEST"])
data["d_log_INVEST"] = data["log_INVEST"].diff()
data = data.dropna()
```

Step 3 & 4: Identifying Patterns & Selecting the Models

Best model: ARIMA(1,2,1) the one with the lower AIC

What We Did:

Looked for Patterns:

- Plotted ACF & PACF of the differenced log series.
- Checked for cutoff vs decay patterns to suggest candidate AR (autoregressive) and MA (moving average) orders.
- Selected the Best Model:
- Ran auto_arima to automatically pick the model that balances fit vs simplicity (AIC).

What We Learned:

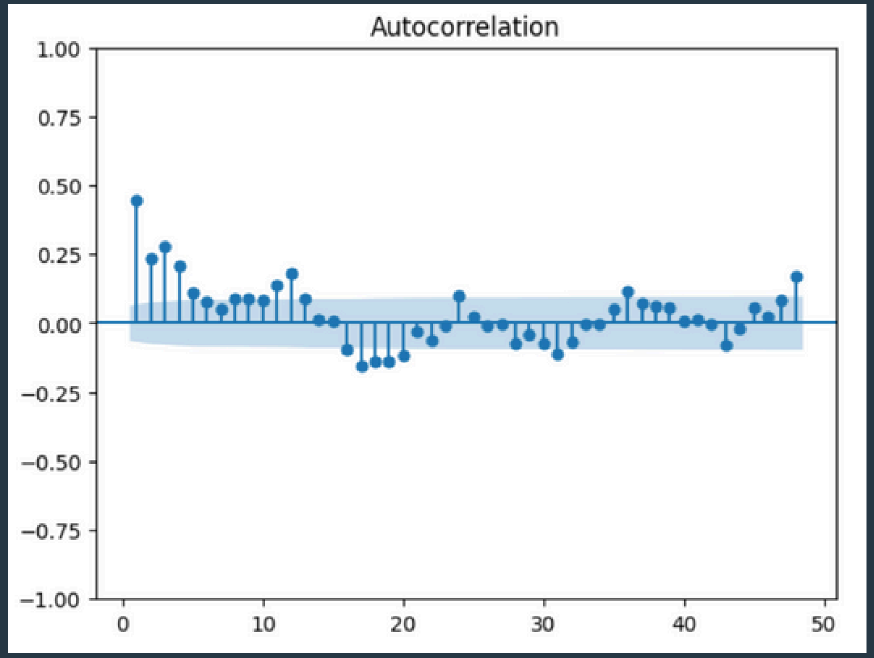
ACF & PACF suggested low-order AR & MA, but not decisive visually.

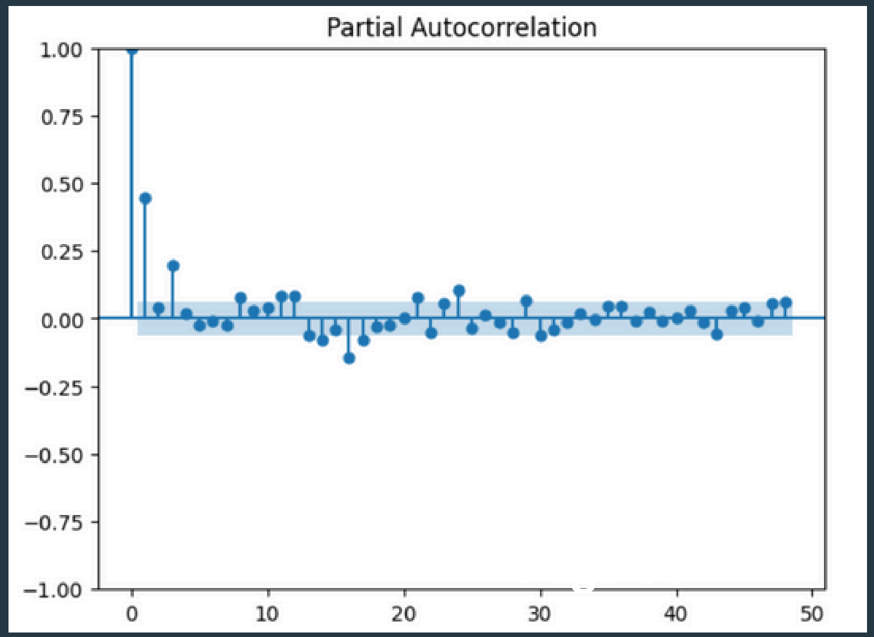
Auto_arima selected ARIMA(1,2,1) on log(INVEST) → needed 2 differences to capture trend acceleration

```
Performing stepwise search to minimize aic
 ARIMA(0,2,0)(0,0,0)[0] intercept   : AIC=-58
 ARIMA(1,2,0)(0,0,0)[0] intercept   : AIC=-59
 ARIMA(0,2,1)(0,0,0)[0] intercept   : AIC=-60
 ARIMA(0,2,0)(0,0,0)[0]             : AIC=-58
 ARIMA(1,2,1)(0,0,0)[0] intercept   : AIC=-61
 ARIMA(2,2,1)(0,0,0)[0] intercept   : AIC=-60
 ARIMA(1,2,2)(0,0,0)[0] intercept   : AIC=-61
 ARIMA(0,2,2)(0,0,0)[0] intercept   : AIC=-61
 ARIMA(2,2,0)(0,0,0)[0] intercept   : AIC=-60
 ARIMA(2,2,2)(0,0,0)[0] intercept   : AIC=-61
 ARIMA(1,2,1)(0,0,0)[0]             : AIC=-61
 ARIMA(0,2,1)(0,0,0)[0]             : AIC=-60
 ARIMA(1,2,0)(0,0,0)[0]             : AIC=-59
 ARIMA(2,2,1)(0,0,0)[0]             : AIC=-60
 ARIMA(1,2,2)(0,0,0)[0]             : AIC=-61
 ARIMA(0,2,2)(0,0,0)[0]             : AIC=-61
 ARIMA(2,2,0)(0,0,0)[0]             : AIC=-60
 ARIMA(2,2,2)(0,0,0)[0]             : AIC=-61

Best model:  ARIMA(1,2,1)(0,0,0)[0]
Total fit time: 13.362 seconds
```

BEST MODEL IS THE ONE WITH THE LOWER: AIC: A





Step 5: Fitting the Model

Estimating ARIMA Parameters

What We Did:

Fitted SARIMAX with order (1,2,1).

Key Findings:

AR(1) ≈ 0.42 → captures trend momentum.
MA(1) ≈ -0.99 → adjusts for recent shocks.
Model captures main trend, small innovation variance.

```
Fitting SARIMAX (1, 2, 1) (0, 0, 0, 0)
                               SARIMAX Results
==============================================================================
Dep. Variable:             log_INVEST   No. Observations:                  945
Model:               SARIMAX(1, 2, 1)   Log Likelihood                3070.561
Date:                Sun, 23 Nov 2025   AIC                          -6135.121
Time:                        23:47:04   BIC                          -6120.580
Sample:                    02-01-1947   HQIC                         -6129.578
                         - 10-01-2025
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.4206      0.021     19.749      0.000       0.379       0.462
ma.L1         -0.9883      0.006   -164.334      0.000      -1.000      -0.976
sigma2      8.537e-05   2.62e-06     32.598      0.000    8.02e-05    9.05e-05
===================================================================================
Ljung-Box (L1) (Q):                   0.13   Jarque-Bera (JB):               372.48
Prob(Q):                              0.72   Prob(JB):                         0.00
Heteroskedasticity (H):               0.76   Skew:                             0.21
Prob(H) (two-sided):                  0.01   Kurtosis:                         6.05
===================================================================================
```

STEP 6: CHECKING THE MODEL

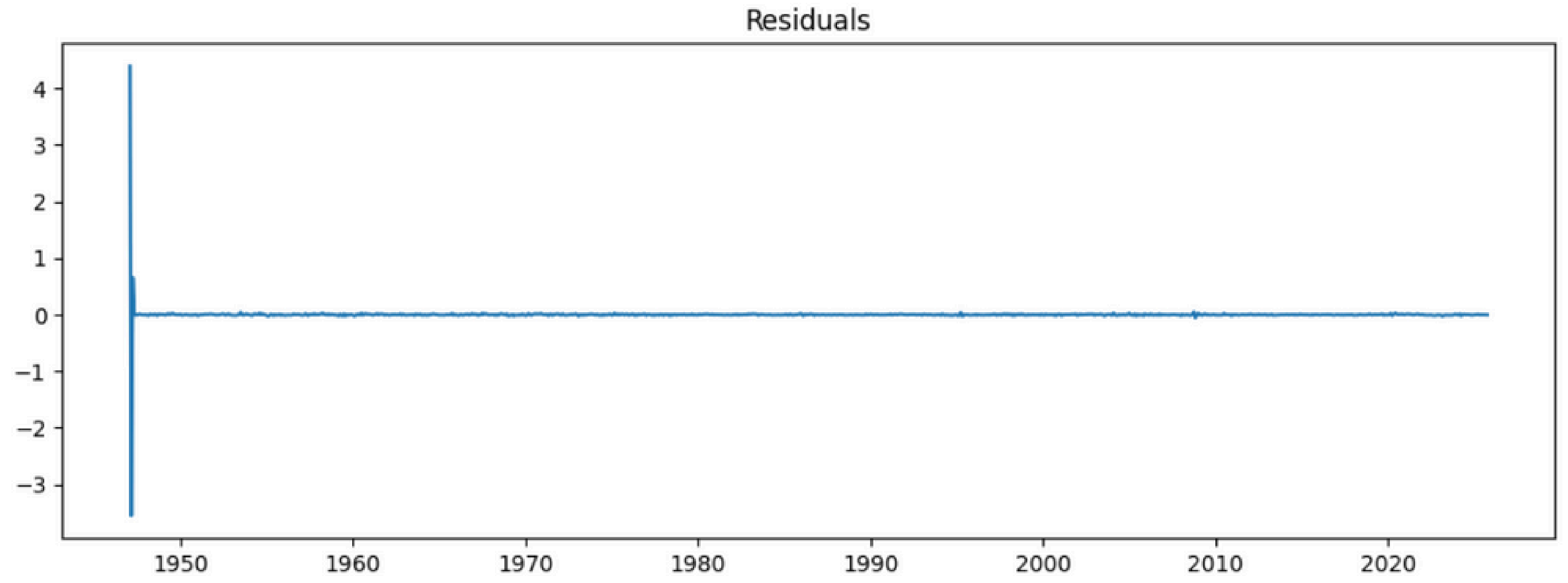
Residual Tests:

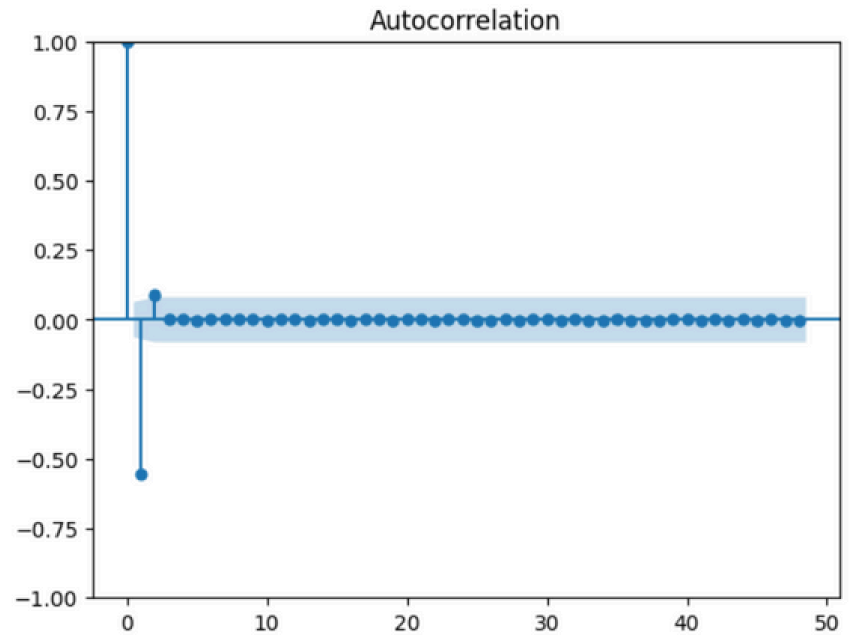
- Plot residuals → any pattern left?
- Ljung-Box → serial correlation?
- Jarque-Bera → normality?
- ARCH → volatility clustering?

Findings:

- Residuals → non-normal, volatility clustering → model captures trend but not extremes.
- Suggests ARIMA-GARCH for volatility if needed.

Are The Residuals Healthy?






```
         lb_stat      lb_pvalue
12  297.545867  1.537862e-56
24  297.587831  5.128757e-49
36  297.622791  6.427535e-43
Jarque-Bera stat, p-value: SignificanceResult(statistic=np.float64(8881091.00971726), pva
ARCH test: (np.float64(76.71483637149242), np.float64(2.2040298019212862e-12), 8.2588528
```

STEP 7: FORECASTING & EVALUATION

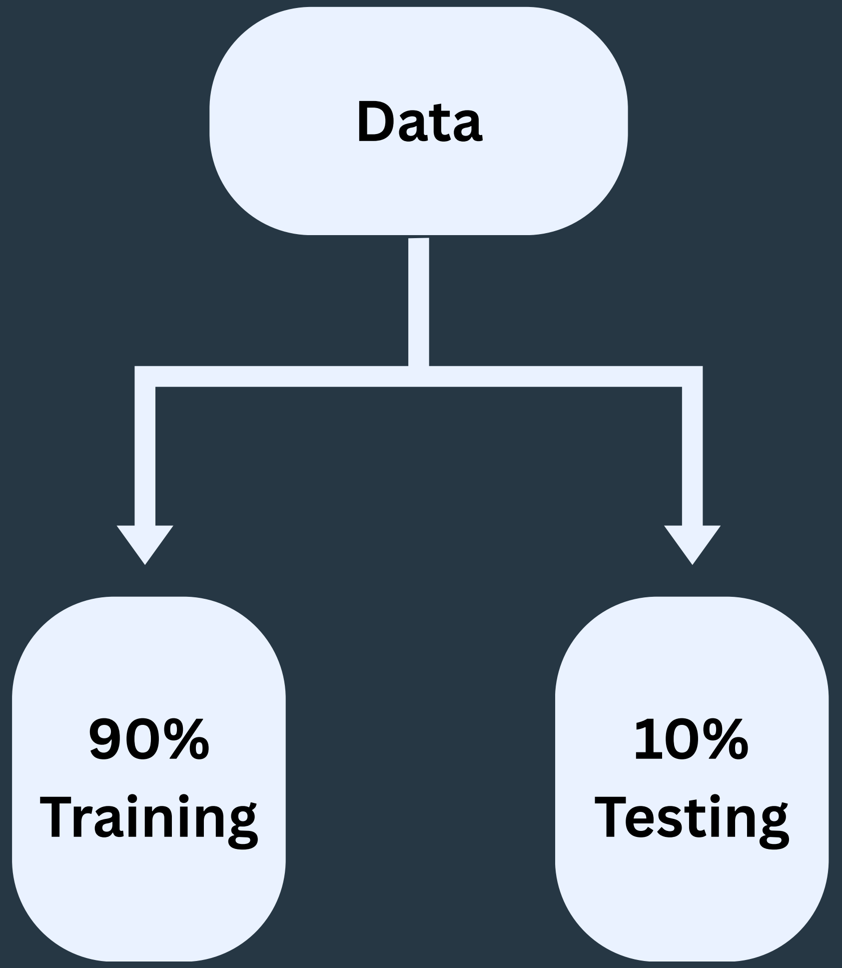
We kept the first 90% of the data as training, then forecasted each day in the last 10% one by one.

We measured forecast accuracy using three key metrics:

- RMSE (Root Mean Squared Error) → penalizes large mistakes
- MAE (Mean Absolute Error) → average error magnitude
- MAPE (Mean Absolute Percentage Error) → error in percentage, easy to interpret

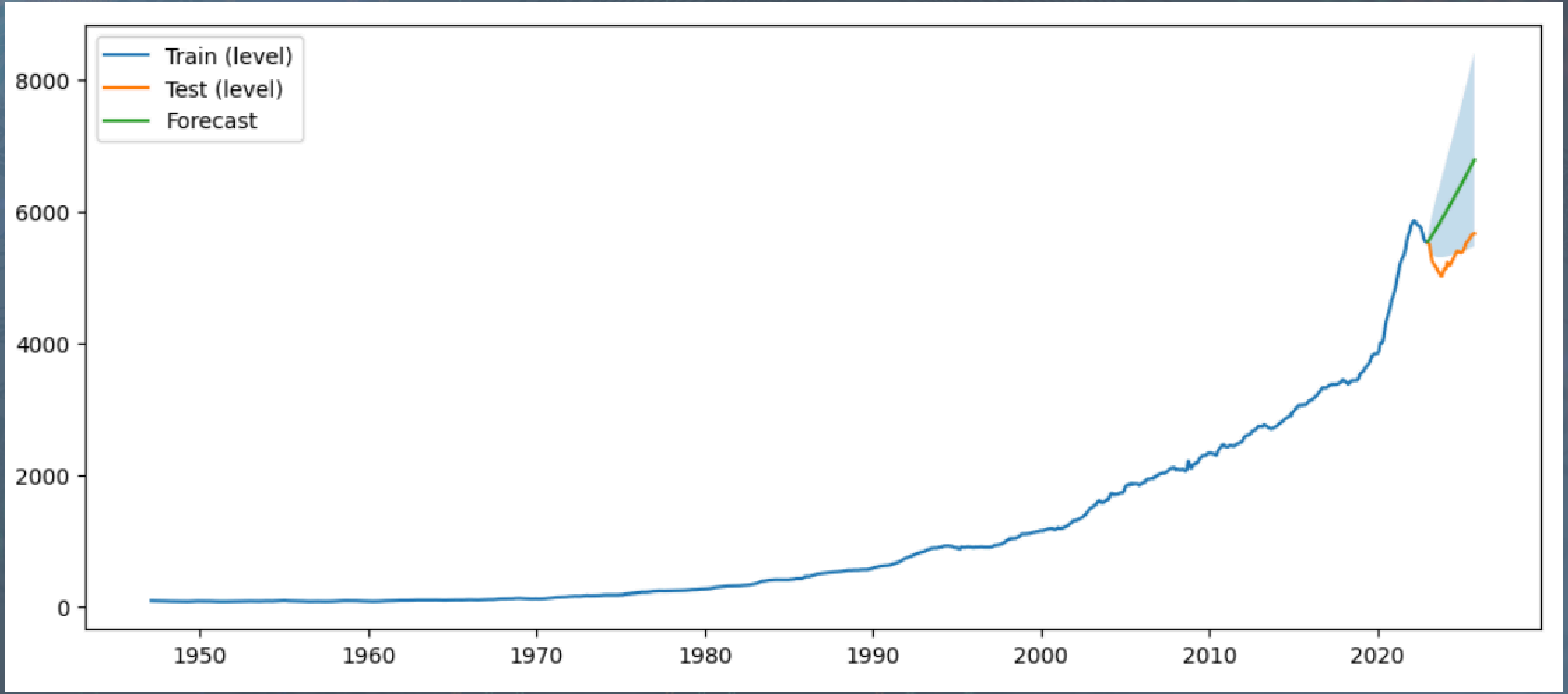
Results:

- Accuracy ≈ 85%
- Example: 2030 forecast = $7,116 Billion (95% CI: 5,335–9,492)



Did you know?
Securities held by banks grew ~70x since 1947!

MODEL PERFORMANCE: IN-SAMPLE FIT AND OUT-OF-SAMPLE FORECAST



The Model Successfully Captures the Trend: The forecast correctly predicts the continued upward trajectory of bank securities. However, we can see a gap between the forecast and the actual test data. For tactical, month-to-month decision-making, we would need to incorporate additional leading indicators that can explain this short-term volatility.

KEY TAKEAWAYS: WHAT WE DID SO FAR

What We Did:

- Plotted the data → saw strong exponential growth
- Applied log transform & differencing → stabilized trend
- Checked patterns with ACF & PACF → suggested AR/MA orders
- Auto_arima selected ARIMA(1,2,1) → best model by AIC
- Fitted SARIMAX, ran residual diagnostics → captured main trend, some volatility remains
- Forecasted future values → 2030: ~7,116 Billion USD


THANK YOU