TIME SERIES FORECASTING

AUTOREGRESSIVE MODELS FOR TIME SERIES

Alternative Approach: ARIMA

Exponential smoothing and ARIMA

Most widely-used approaches to time series forecasting Provide complementary approaches to the problem

Exponential smoothing

Description of trend and seasonality in the data

ARIMA models :

Autocorrelations in the data

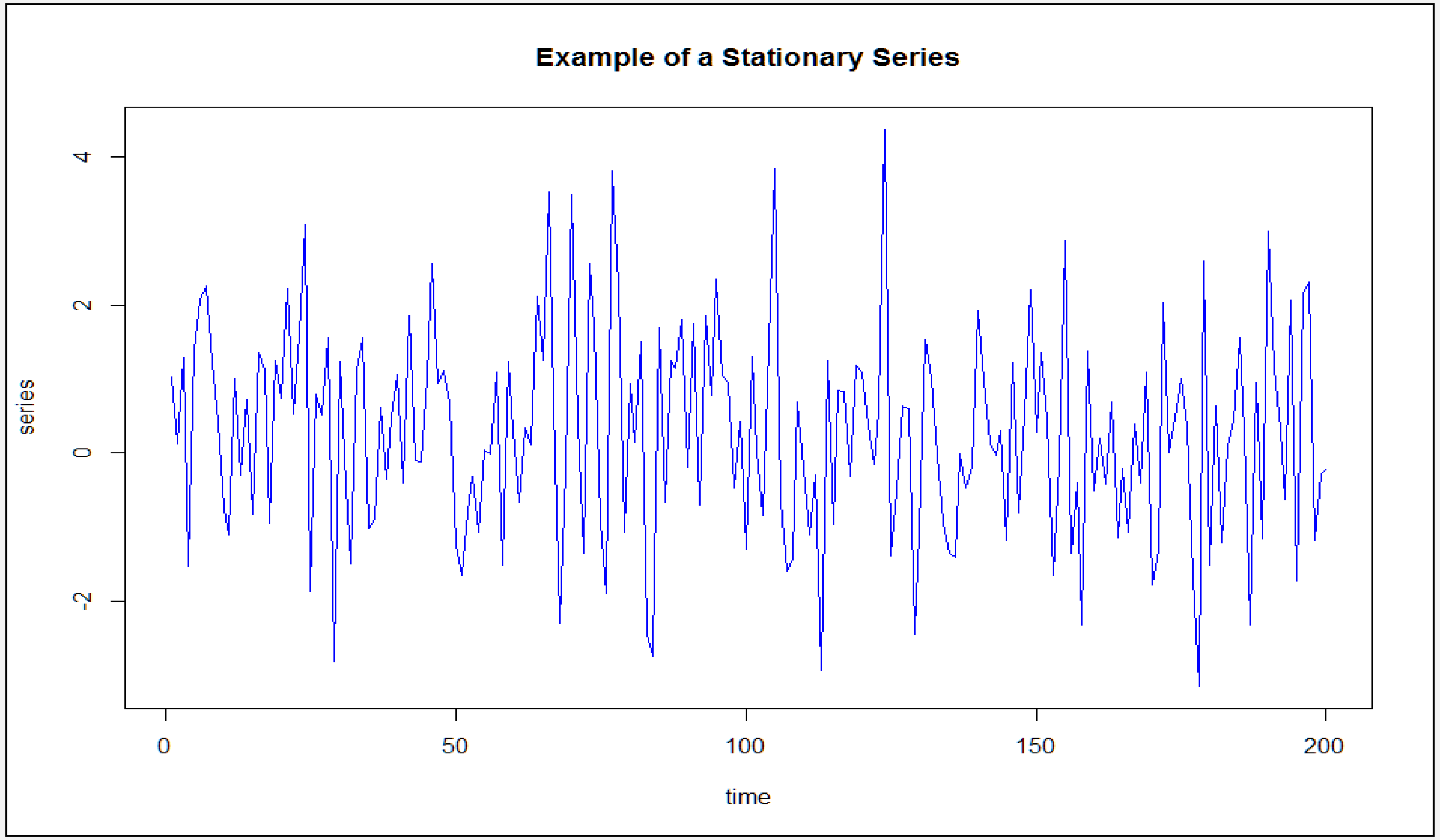
Stationary Time Series

- Only Stationary Series can be forecasted!!
- If Stationarity condition is violated, the first step is to stationarize the series

A stationary time series is one whose properties do not depend on the time at which the series is observed. The series will not have any predictable pattern

Another name: White Noise

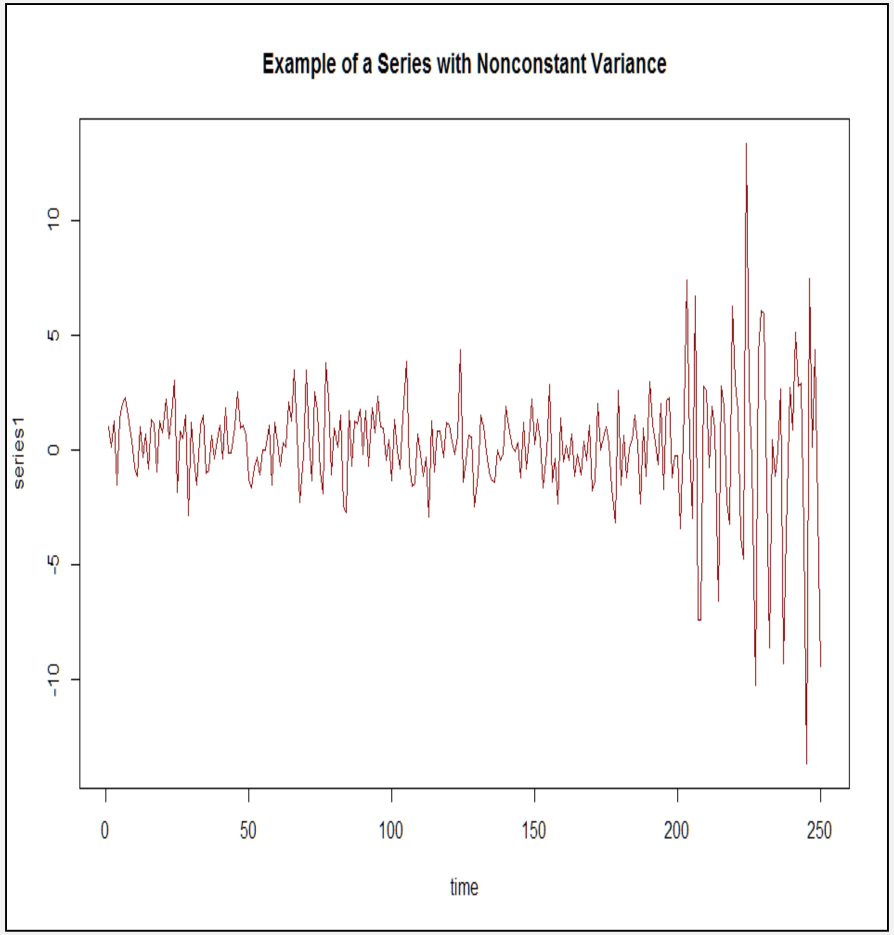
Stationary Time Series

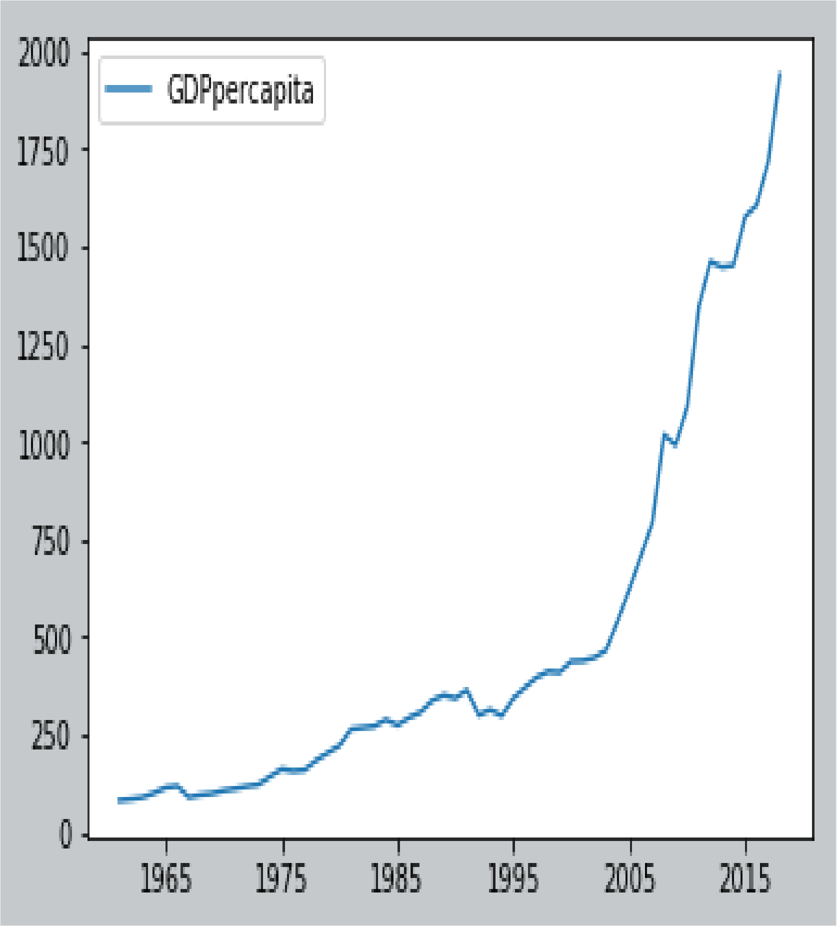


Stationary Time Series

- Mean of the time series will be a constant.
  - Time series with trends are not stationary
  - Variance of the time series will be a constant
  - The correlation between the t-th term in the series and the t+m-th term in the series is constant for all time periods and for all m

Non-Stationary Time Series





Steps for Analysis

1. Visualization
2. Stationarization
   - Do a formal test of hypothesis
   - If series non-stationary, stationarize
3. Explore Autocorrelations and Partial Autocorrelations
4. Build ARIMA Model
   - Identify training and test periods
   - Decide on model parameters
   - Compare models using accuracy measures
   - Make prediction

Simple Test for Stationarity

Augmented Dickey-Fuller Test

Tests whether a time series is NON-STATIONARY

Null hypothesis H_0: Time series non-stationary
Alternative hypothesis H_a: Time series is stationary

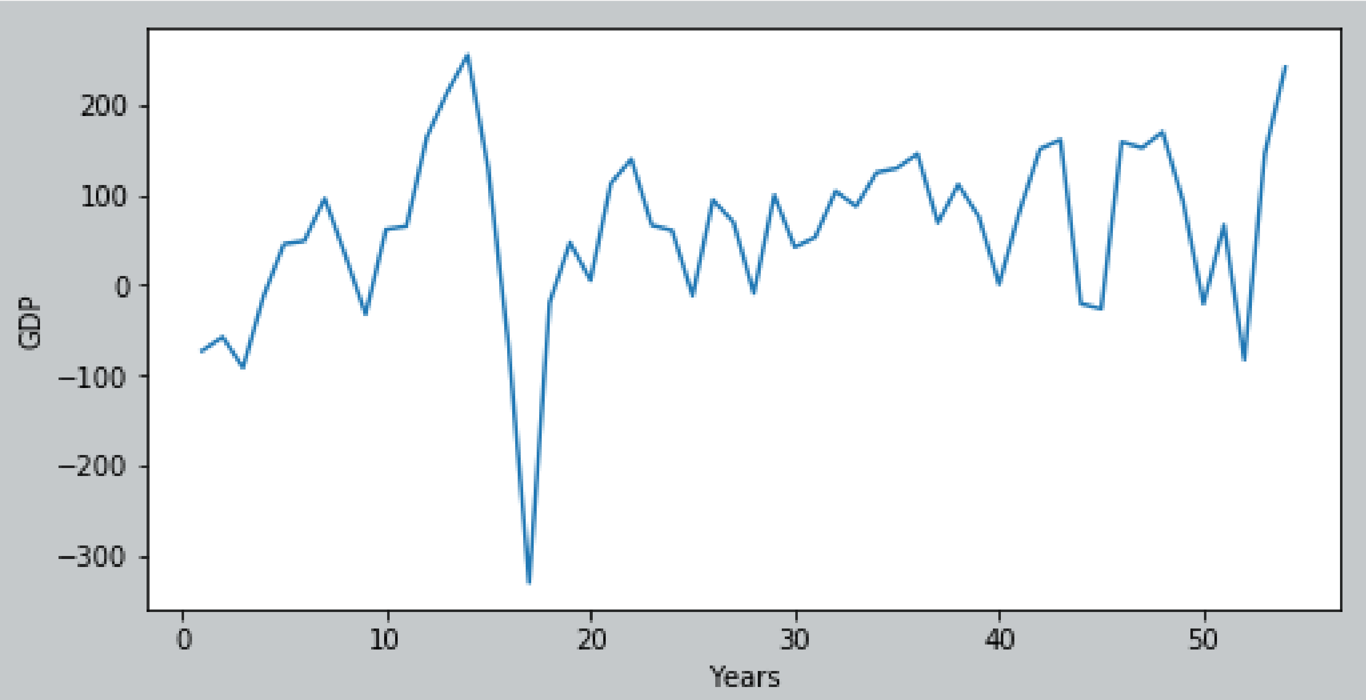
Rejection of null hypothesis implies that the series is stationary

Stationary Series

It is possible to make a non-stationary series stationary by taking differences between consecutive observations

Difference Series

- Simple but effective method to stationarize a time series
- Take difference of consecutive terms in a series
- Known as a Difference Series of Order 1



Autoregression

- When value of a time series depends on its value at the previous time point
- Various economic and financial series are autoregressive
  - GDP of a country
  - Stock prices
  - Consumption expenditure

Autoregression

- Current value of the series may depend on only one past observation
  - Current GDP depends only on the past year's GDP
  - AR(1) process

$$Y(t) = \beta_1 Y(t-1) + \varepsilon(t)$$

Autoregression

- It can depend on several past observations
  - AR(p) process
  - p: parameter (to be determined from data)

$$Y(t) = \beta_1 Y(t\text{-}1) + \beta_2 Y(t\text{-}2) + \beta_3 Y(t\text{-}3) + \ldots + \beta_p Y(t\text{-}p) + \varepsilon(t)$$

$\beta_{1,}\ \beta_{2,\ \ldots}\ \beta_p$: autoregressive parameters of various orders

$\varepsilon(t)$: White noise, iid r.v. with mean 0, variance σ^2

Explore Autocorrelation

- Autocorrelation: Correlation with self
- Autocorrelation of different orders gives *inside information* regarding time series
- Determines order p of the series

$$-1 \le \mathbf{ACF} \le 1$$

$$\mathrm{ACF}(0) = 1$$

ACF makes sense only if the series is stationary

Explore Autocorrelation

- Correlation between Original series and Lag(1) series = ACF(1)
- Correlation between Original series and Lag(2) series = ACF(2)
- Correlation between Original series and Lag(3) series = ACF(3)
- And so on ….

acf(BASF, lag = 50)

Explore Autocorrelation

- Autocorrelations decreasing as lag increases
- Autocorrelations significant till high order
- Significant autocorrelations imply observations of long past influences current observation

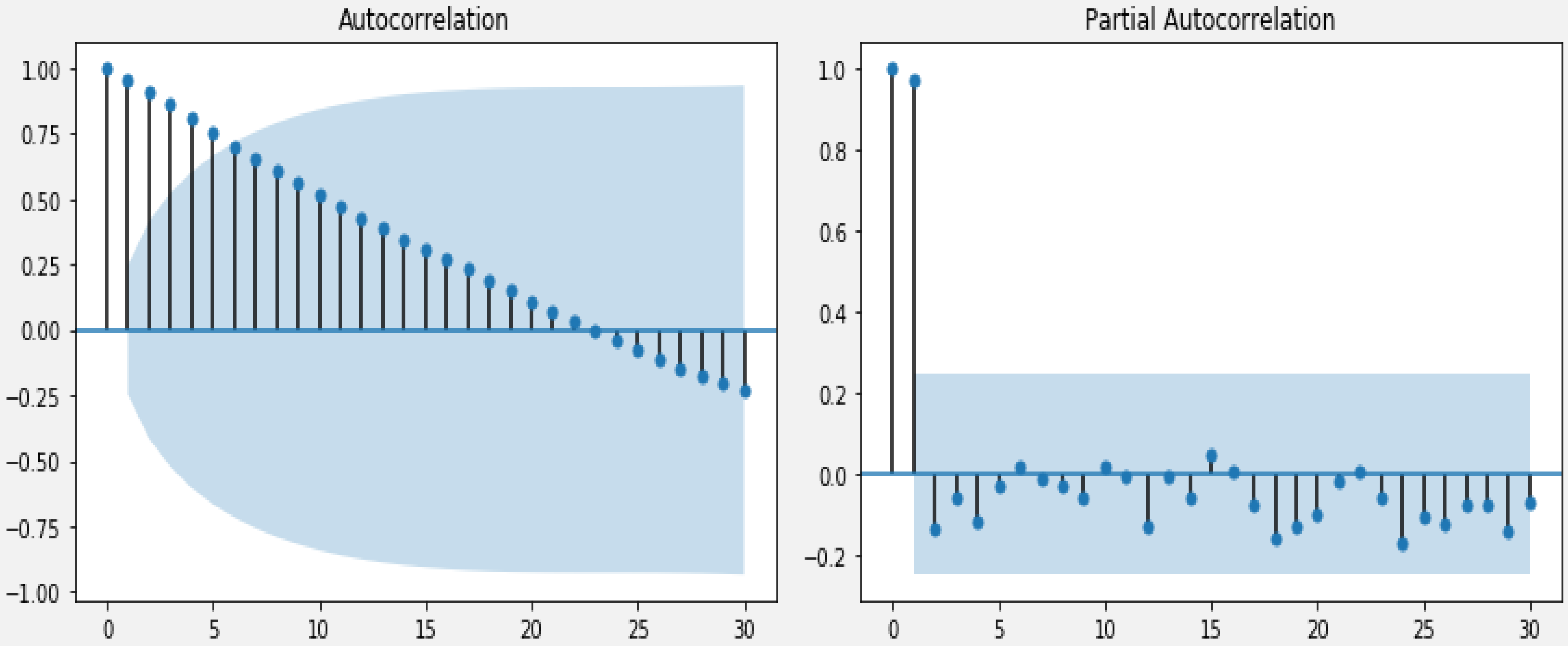
Is there really such a dependency or the dependency is only apparent?

Explore Partial Autocorrelations

- Partial autocorrelation adjusts for the intervening periods
- PACF(1) = ACF(1)
- PACF(2) is the correlation between Original and Lag(2) series AFTER the influence of Lag(1) series has been eliminated
- PACF(3) is the correlation between Original and Lag(3) series AFTER the influence of Lag(1) and Lag(2) series has been eliminated
- PACF(50) is the correlation between Original and Lag(50) series AFTER the influence of Lag(1) through Lag(49) series has been eliminated

Explore Partial Autocorrelation

Dramatic reduction in PACF beyond order 1



Tools for Identification

- ACF and PACF together to be considered for identification of order of Autoregression
- Seasonal ACF show significant correlation at seasonal points

Generalized Model: ARIMA

AR is a special and simpler form of a general class of models: ARIMA(p, d, q)

AR: Current observation is regressed on past observations

$Y(t) = \beta_1 Y(t-1) + \beta_2 Y(t-2) + \beta_3 Y(t-3) + \ldots + \beta_p Y(t-p) + \varepsilon(t)$

MA (Moving Average): Current observation is regressed on past forecast errors

$Y(t) = \varepsilon(t) + \alpha_1\, \varepsilon(t-1) + \alpha_2\, \varepsilon(t-2) + \ldots + \alpha_q\, \varepsilon(t-q) \qquad |\, \alpha_1 \,| < 1$

$\alpha_{1,}\, \alpha_2$, ..., α_q : Moving average parameters

Generalized Model: ARIMA

ARMA: When Current observation is a linear combination of past observations and past white (random) noises

$$Y(t) = \beta_1 Y(t-1) + \beta_2 Y(t-2) + \beta_3 Y(t-3) + \dots + \beta_p Y(t-p) + \varepsilon(t) + \alpha_1 \varepsilon(t-1) + \alpha_2 \varepsilon(t-2) + \dots + \alpha_q \varepsilon(t-q)$$

Theoretically (p, q) may take any value but usually values higher than 2 not preferred in practical situation

Generalized Model: ARIMA

- If the original series is not stationary, differencing is necessary
- Most often differencing of order 1 makes the series stationary
- But higher order differencing may be needed
- Order = d

ARIMA(p, d, q) identifies a non-seasonal model which needs to be differenced d times to make it stationary and contains p AR terms and q MA terms

Seasonal ARIMA

ARIMA(p, d, q)x(P, D, Q)[freq]

Seasonal difference = D

Appropriate for seasonal series

Final Model

- Based on accuracy choose the model which works best
- Must have proper interpretability
- Often a simple model works better