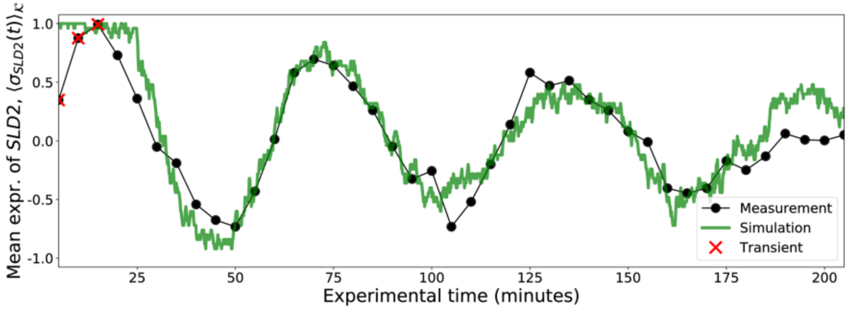
TIME SERIES PREDICTION

2143488 BIG DATA AND ARTIFICIAL INTELLIGENCE

DR. JING TANG

TIME SERIES



- Sequence of observations (at regular intervals) $s_t \in R$ ordered in time
- Examples:
  - Meteorology: temperature…
  - Economy and Finance: GDP, spread…
  - Marketing: sales…
  - Industry: power consumption…
  - Web: clicks
  - Genomics: gene expression during cell cycle..

WHY TIME SERIES?

- **Prediction** of the future based on the past
- **Control** of the process producing the series
- **Understanding** of the mechanism generating the series
- **Description** of the salient features of the series.

CONTINUOUS VS. DISCRETE TIME SERIES

- **Continuous time series:** if **observations** are made **continuously through time**, even when the measured variable can only take a discrete set of values.
  - E.g., a binary process at continuous time is a continuous time series.
- **Discrete time series:** if **observations** are taken only **at specific times**, usually equally spaced, even the measured variable is a continuous variable.
  - A data-reporting interval that is infrequent or irregular (sampled vs aggregated)
  - Gaps where values are missing due to reporting interruptions (e.g., intermittent server or network downtime)

UNIVARIATE VS. MULTIVARIATE TIME SERIES

- **Univariate time series**: where one type of measurement is made repeatedly on the same object or individual.
- **Multivariate time series**: when observations are taken on two or more variables, it may be possible to use the variation in one time series to explain the variation in another series.
  - E.g., it is interesting to see how sea level is affected by temperature and pressure, and to see how sales are affected by price and economic conditions.

TIME SERIES VS. RANDOM SAMPLES OF INDEPENDENT OBSERVATIONS

- When **Successive observations** of a time series are **dependent**, future values may be predicted from past observations.
  - **Deterministic**: If a time series can be predicted exactly, it is said to be deterministic
    - E.g., $x_t = 2x_{t-1}$
  - **Stochastic**: Most time series are stochastic in that the future is only partly determined by past values, so that exact predictions are impossible and must be replaced by the idea that future values have a probability distribution, which is conditioned by a knowledge of past values
    - E.g., $x_t = 2x_{t-1} + \epsilon, \epsilon \sim N(0, \sigma^2)$

A GENERAL MODEL OF TIME SERIES

- A general model of times series:

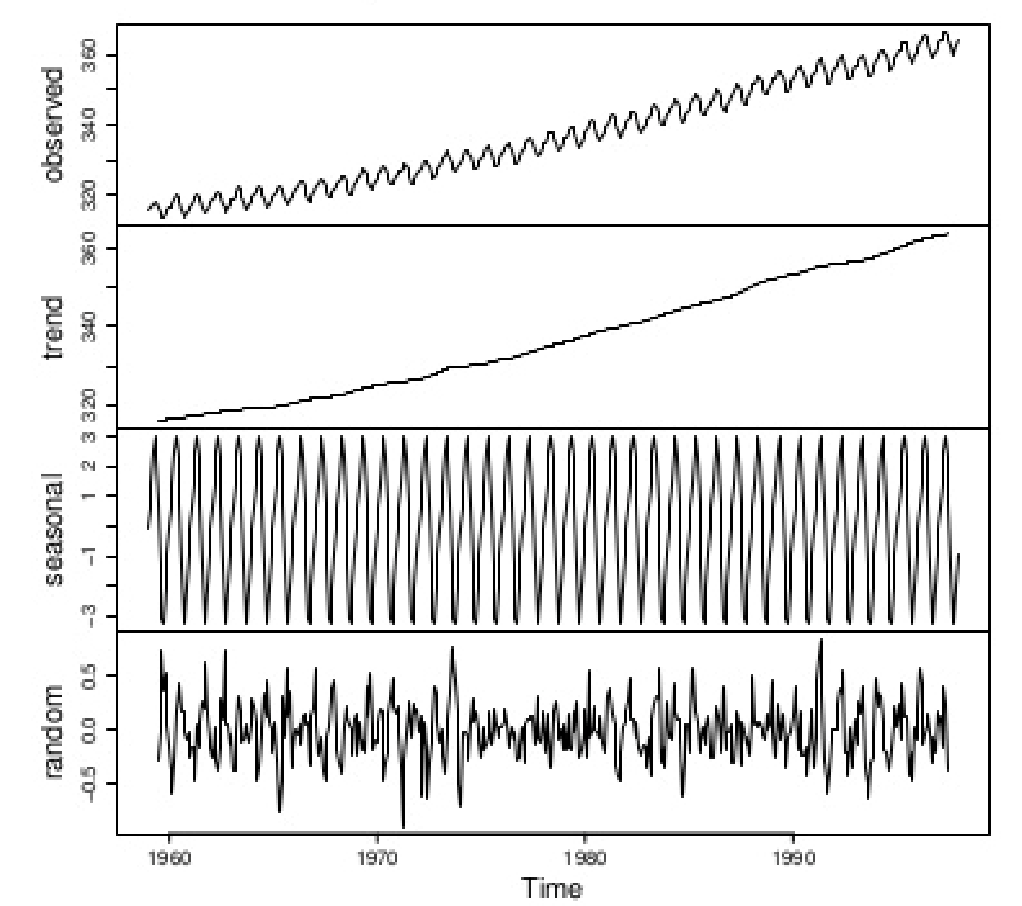
$$x_t = g(t) + \epsilon_t \quad t = 1, \dots, T$$

- Two components:
  - **Deterministic / Systematic part**: $g(t)$, also called **signal or trend**, which is a function of time
  - **Stochastic part**: a residual term ϵ_t, also called **noise**, which follows a probability law.

TYPES OF VARIATION

- **Trend**: stabilized long-term increase or decrease of mean
  - After remove the trend variation, called **detrended**
- **Seasonal**: in daily, weekly, or monthly fluctuations (<1 year)
  - After remove the seasonal variation, called **deseasonalized**
  - **Additive** (constant) **vs. Multiplicative** (% of mean)
- **Cyclic**: Rises/falls that not a fixed period (>1 year)
  - E.g., economic cycles
- **Irregular**: may or may not be completely random

DECOMPOSITION OF TIME SERIES



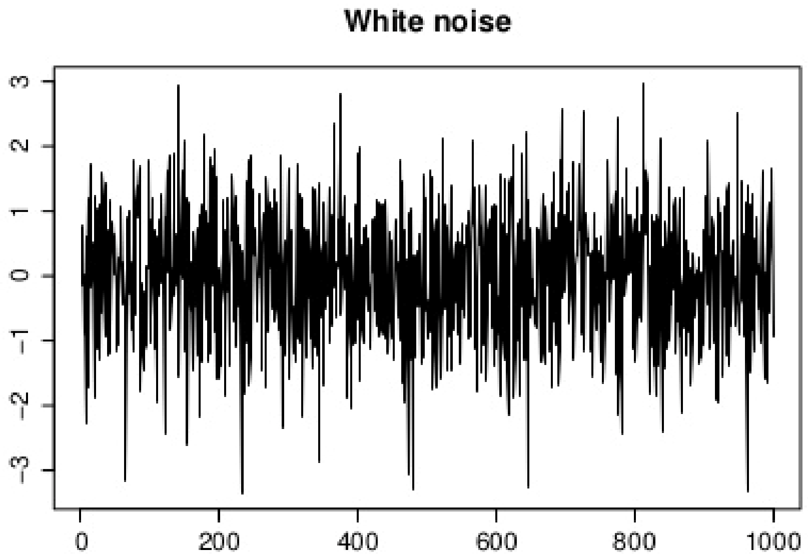
$e.g., x_t = \alpha + \beta t + \epsilon_t$

Differencing to remove the trend

$e.g., x_t = m_t + s_t + \epsilon_t$

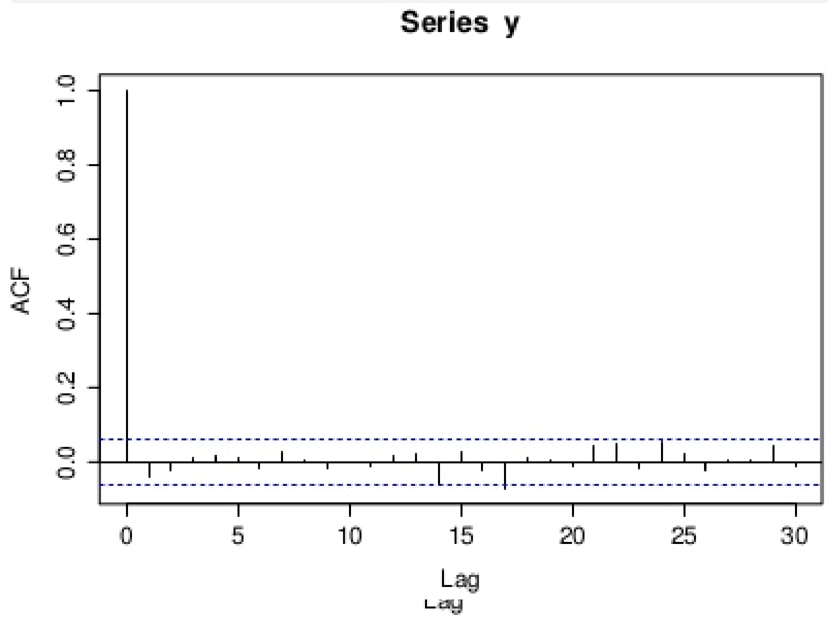
Differencing to remove the seasonal impact

GAUSSIAN PURELY RANDOM ϵ_t



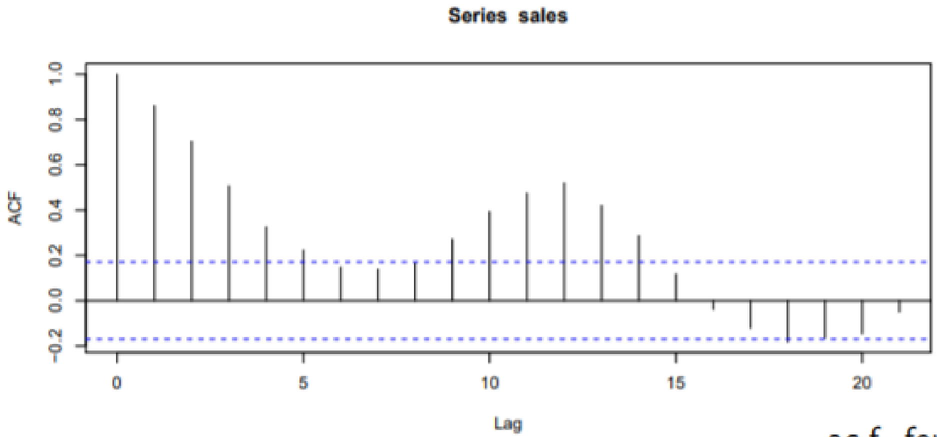
Autocorrelation Function

$$Corr(\epsilon_t, \epsilon_{t+k}) = \frac{\sum_{t=1}^{N-k}(\epsilon_t - \bar{\epsilon})(\epsilon_{t+k} - \bar{\epsilon})}{\sum_{t=1}^{N-k}(\epsilon_t - \bar{\epsilon})^2}$$

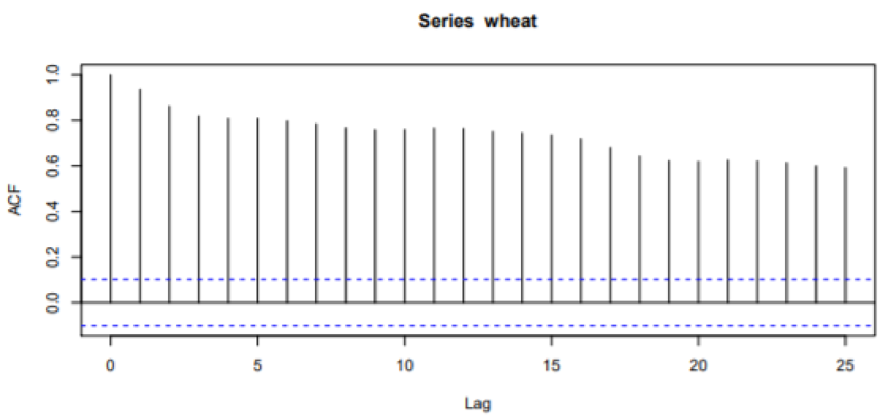


ACF EXAMPLES

ac.f. for house sales



ac.f. for wheat price index



AUTOREGRESSION PROCESS

- A process x_t is said to be an **autoregressive process** of order n, $AR(n)$, if

$$x_t = a_1 x_{t-1} + \cdots + a_n x_{t-n} + \epsilon_t$$

- **Auto**: like a linear regression, but not on independent varies but on its **past values**
- Properties of **stationarity** depends on the coefficiencies $a_i, 1 \ldots n$

MARKOV PROCESS: AR(1)

- Markov process, AR(1):

$$x_t = ax_{t-1} + \epsilon_t$$
$$= a(ax_{t-2} + \epsilon_{t-1}) + \epsilon_t$$
$$= \epsilon_t + a\epsilon_{t-1} + a^2\epsilon_{t-2} + \cdots$$

- Then:
  - $E(x_t) = 0$
  - $Var(x_t) = \sigma_\epsilon^2(1 + a^2 + a^4 + \cdots)$
  - $|a| < 1 \Rightarrow Var(x_t) = \frac{\sigma_\epsilon^2}{1-a^2}$ $\Rightarrow$autocorrelation $\rho(k) = a^k$

GENERAL ORDER PROCESS: AR(N)

- It has been shown that condition necessary and sufficient for the stationarity is that the complex roots of the equation

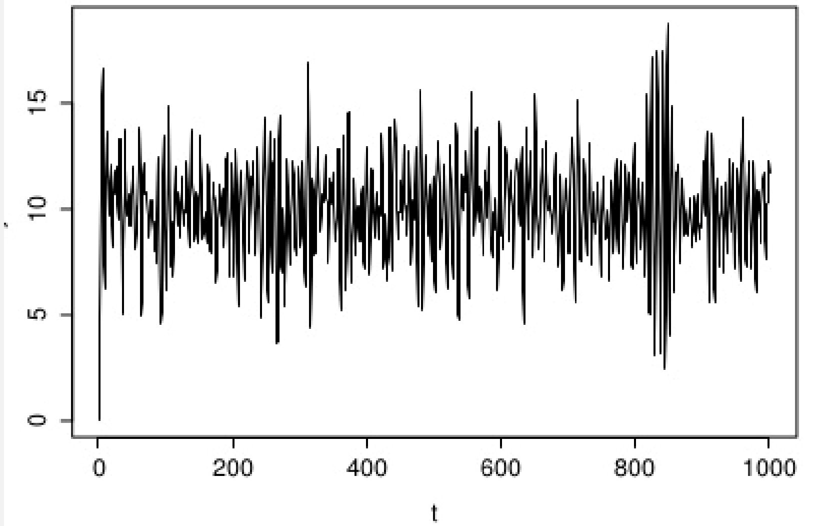
$$x(z) = 1 - \alpha_1 z - \cdots - \alpha_n z^n = 0$$

  lie outside the unit circle.

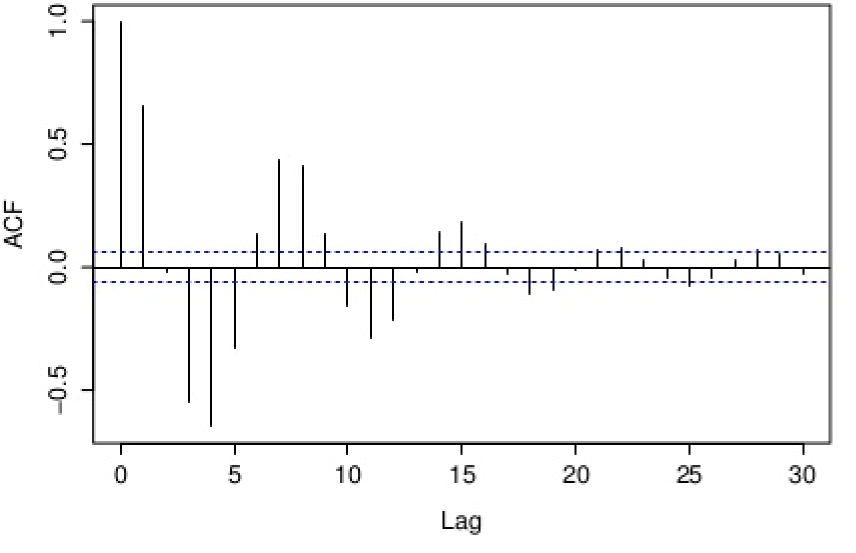
- The autocorrelation function of an AR(n) attenuates slowly with the $lag\ k$ (exponential decay or damped sine wave pattern).

- On the contrary the partial autocorrelation function cuts off at $k > n$, i.e. it is not significantly different from zero beyond the $lag\ n$.

AR(2)

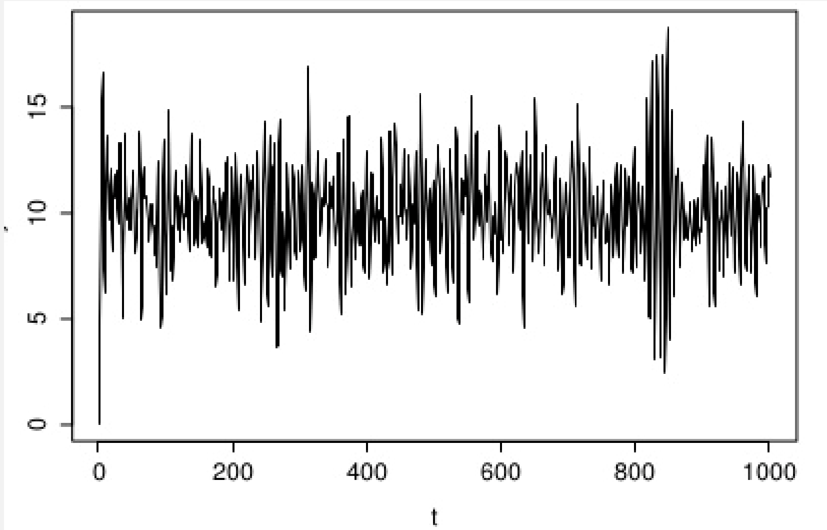


15
10
5
0
0
200
400
600
800
1000
t

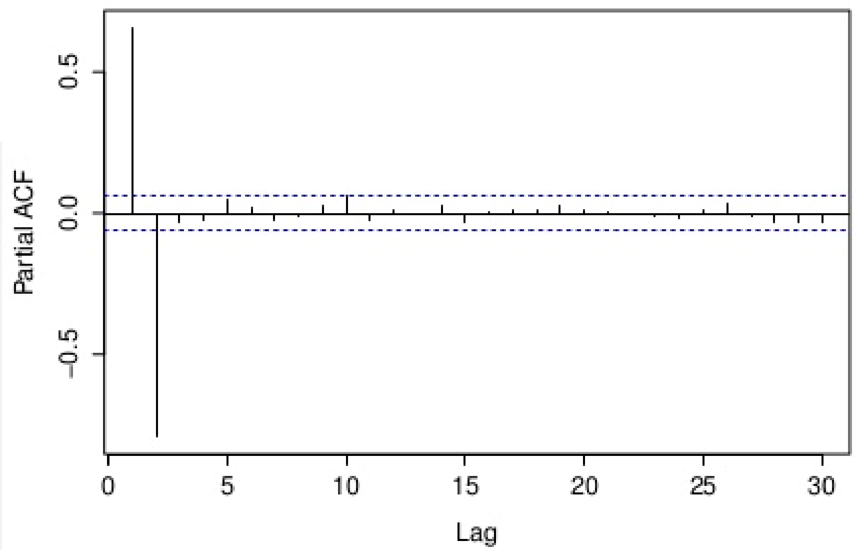


1.0
0.5
0.0
−0.5
ACF
0
5
10
15
20
25
30
Lag

AR(2)



Partial Autocorrelation Function



ONE STEP-AHEAD VS. ITERATED PREDICTION

ONE-STEP-AHEAD PREDICTION

- n previous values of the series are available and the forecasting problem can be cast in the form of a **generic regression problem**
- In literature a number of supervised learning approaches have been used with success to perform one-step-ahead forecasting on the basis of historical data.

ITERATED (OR RECURSIVE) PREDICTION

- **Predicted output is fed back** as input for the next prediction.
- As the feedback values are typically distorted by the errors made by the predictor in previous steps, the iterative procedure may produce undesired effects of accumulation of the error.
- **Low performance is expected in long horizon tasks**. This is due to the fact that they are essentially models tuned with a one-step-ahead criterion which is not capable of taking temporal behavior into account.

4 METHODS FOR TIME SERIES FORECASTING

- **Classical Time Series Model**
  - Get start with it
  - Use if explicit modeling is good
- General Machine Learning
  - If you are unsure about modeling assumptions
  - Use proper validation to ensure good performance
- Feature Engineering
  - For more complex data
  - If you have a priori knowledge about the domain
- Deep Learning
  - If you have a lot of data
  - If you frequently want to iterate & experiment
  - If explicit modeling & feature engineering is too costly

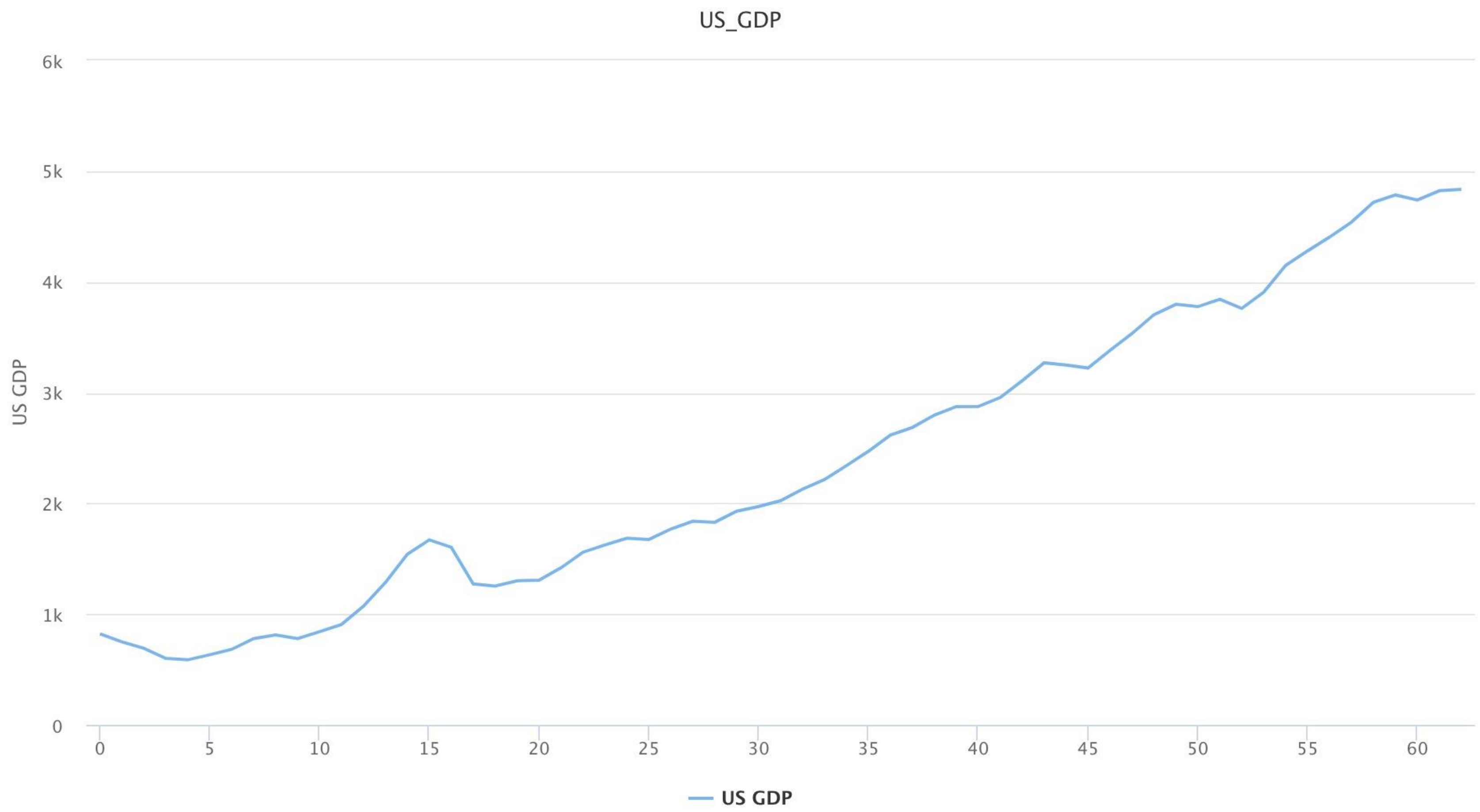
PROCEDURE OF CLASSICAL TIME SERIES FORECASTING

1. Visulize the time series
2. Stationarize the series manually
3. Plot ACF/PACF charts and find optimal parameters
4. Build the ARIMA model
5. Make Predictions (smallest AIC)

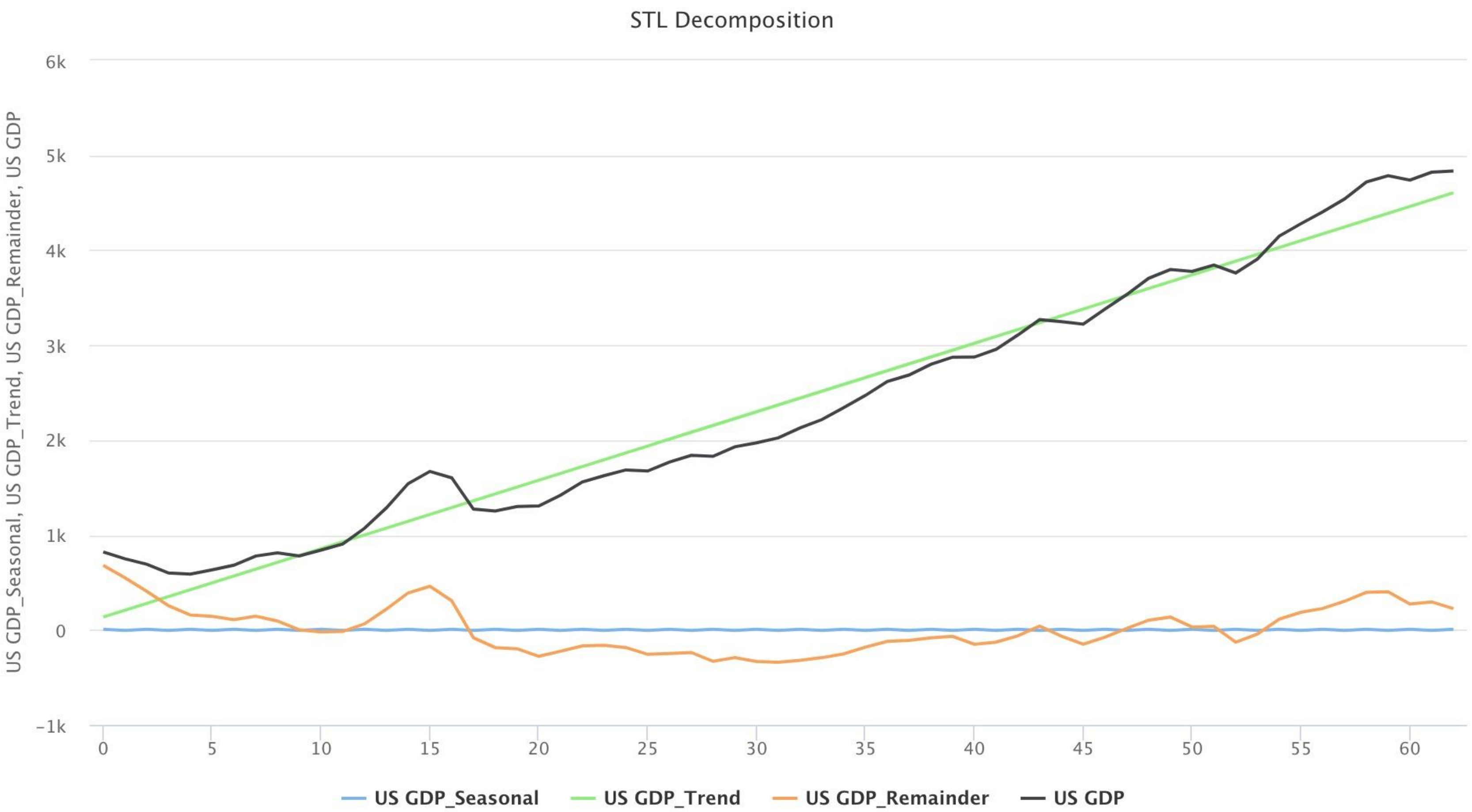
PROCEDURE OF CLASSICAL TIME SERIES FORECASTING

1. Visulize the time series
2. Stationarize the series manually
3. Plot ACF/PACF charts and find optimal parameters
4. Build the ARIMA model
5. Make Predictions (smallest AIC)

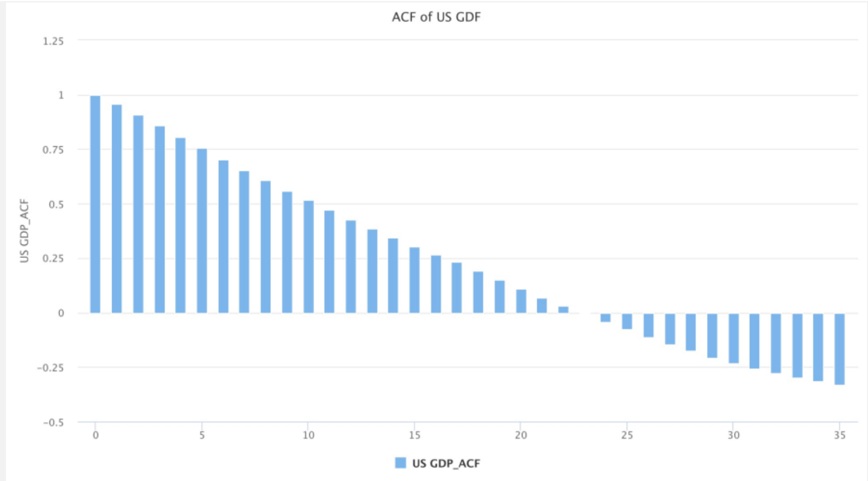
EXAMPLE: FORECAST US_ GDP
1. VISUALIZE ORIGINAL DATA

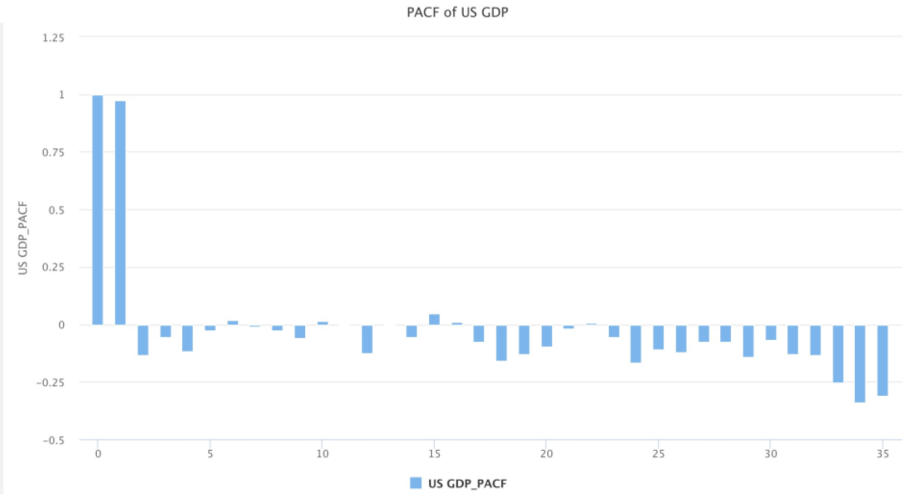


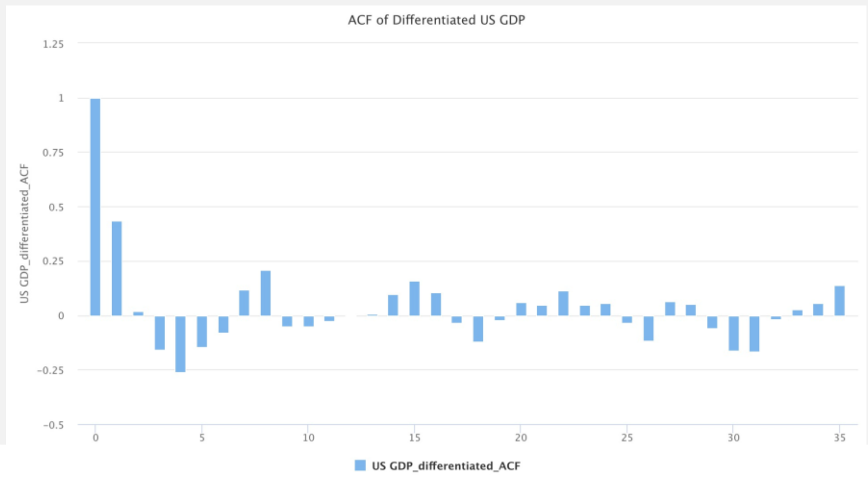
EXAMPLE: FORECAST US_ GDP
2. STL DECOMPOSITION

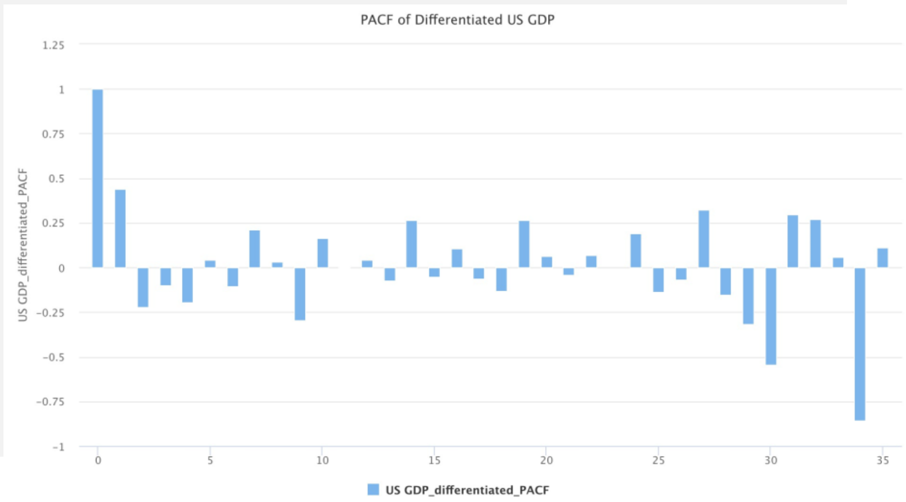


EXAMPLE: FORECAST US_ GDP
3. CHECK ACF/PACF





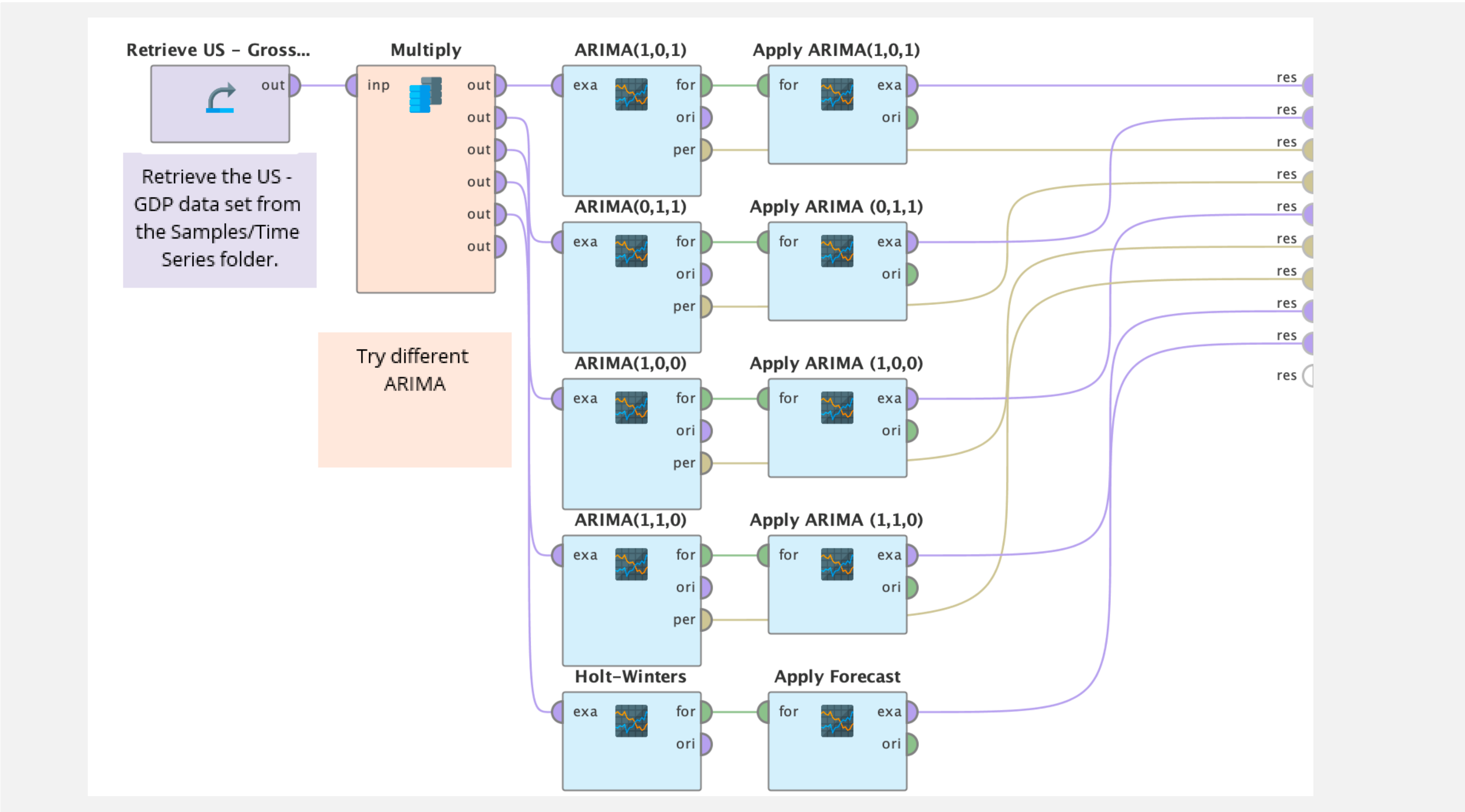




ARIMA(p,d,q)

- Auto Regressive Integrated Moving Average
- p: Auto Regressive, lags of variable itself (Refer PACF)
- d: Integrated, or Trend, differencing steps required to make stationary
- q: Moving Average, lags of previous information shocks (Refer ACF)

EXAMPLE: FORECAST US_ GDP ARIMA FORECAST



HW6: FORECAST DAILY AVERAGE PRESSURE (PRES) IN TIANTAN, BEIJING IN 2017 MARCH

1. Select attributes (univariable model)
2. Group data (aggregate pressure daily by average)
3. Visualize data (the trend, seasonal pattern of pres)
4. Calculate ACF and PACF
5. Apply ARIMA
6. Evaluate the result is ok or not? How to improve it?