Kherson State Maritime Academy

Херсонська державна морська академія

Speech for graduate students of specialty
275 – ’’Transport technologies’’

Виступ для аспірантів спеціальності
275 – «Транспортні технології»

Dr.Sc., Prof.,
professor of Software Engineering Department TNTU

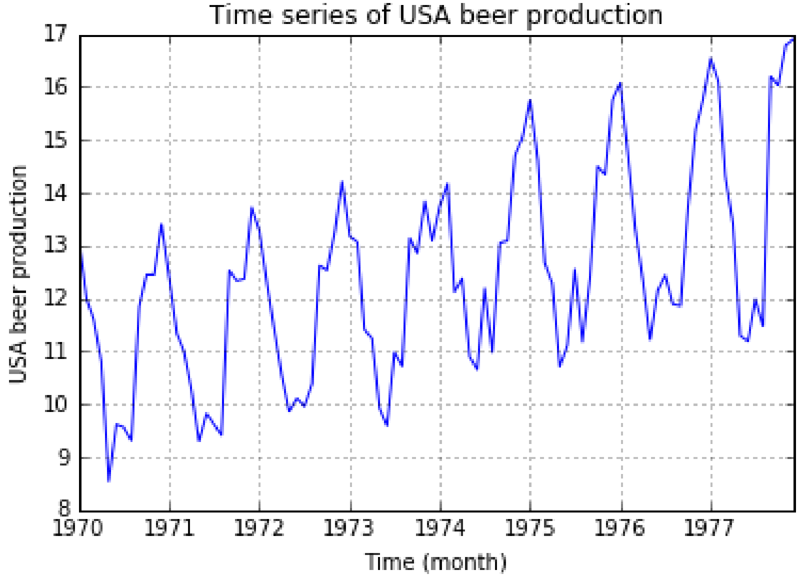
докт. техн. наук., проф., професор кафедри програмної інженерії
ТНТУ ім.І.Пулюя

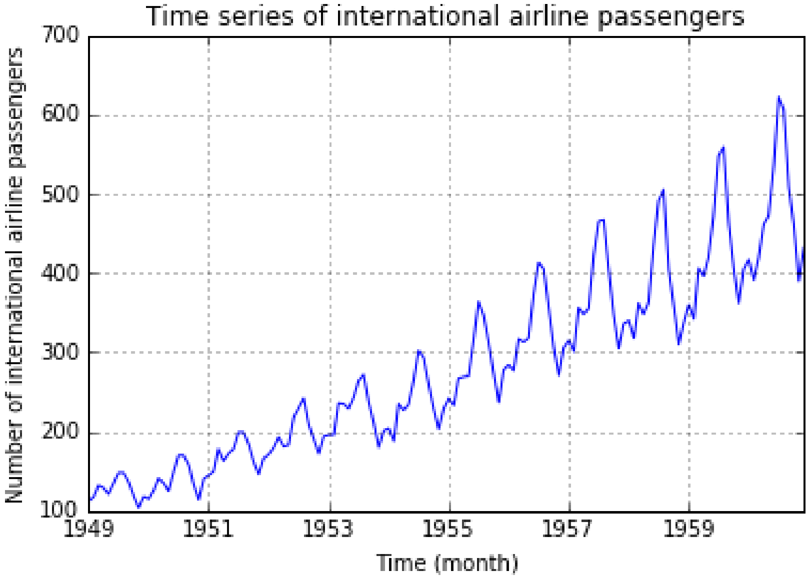
ANALYSIS AND FORECAST OF TIME SERIES FOR TRANSPORT TECHNOLOGIES

АНАЛІЗ І ПРОГНОЗ ЧАСОВИХ РЯДІВ ДЛЯ ТРАНСПОРТНИХ ТЕХНОЛОГІЙ

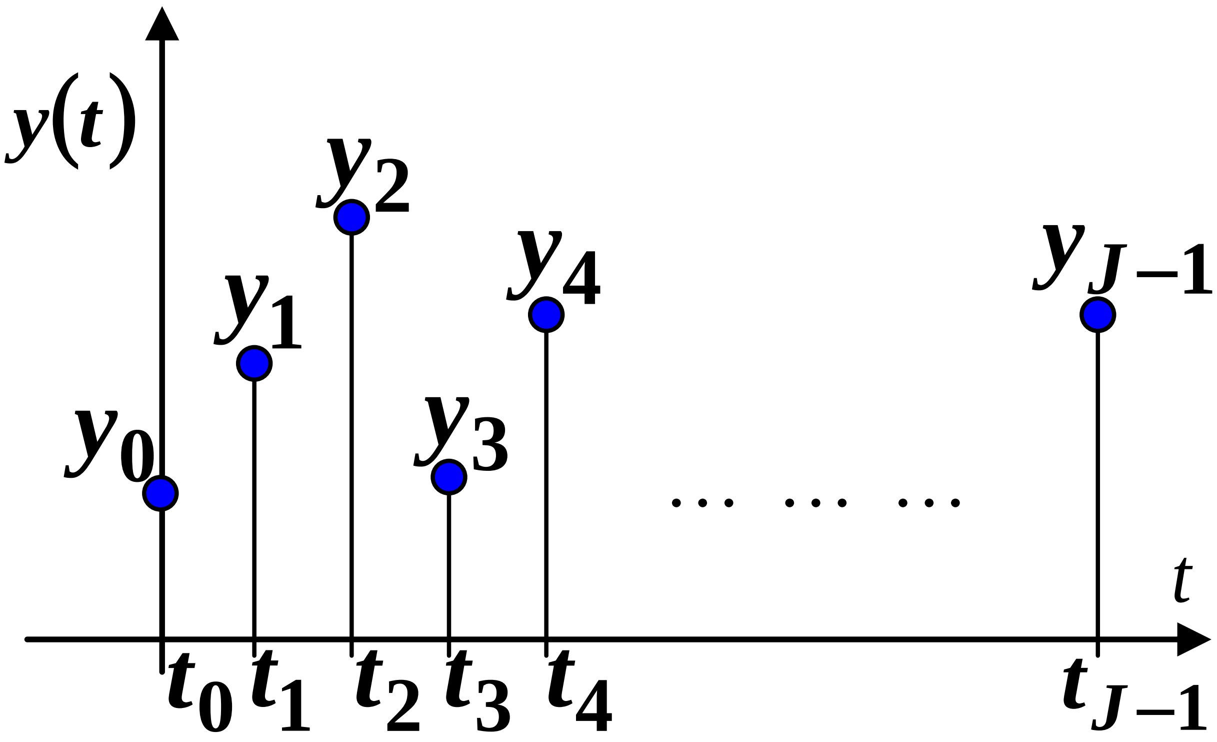
BASIC CONCEPTS OF STOCHASTIC TIME SERIES

EXAMPLES OF STOCHASTIC TIME SERIES





STOCHASTIC TIME SERIES (RANDOM TIME SERIES, STOCHASTIC PROCESS)



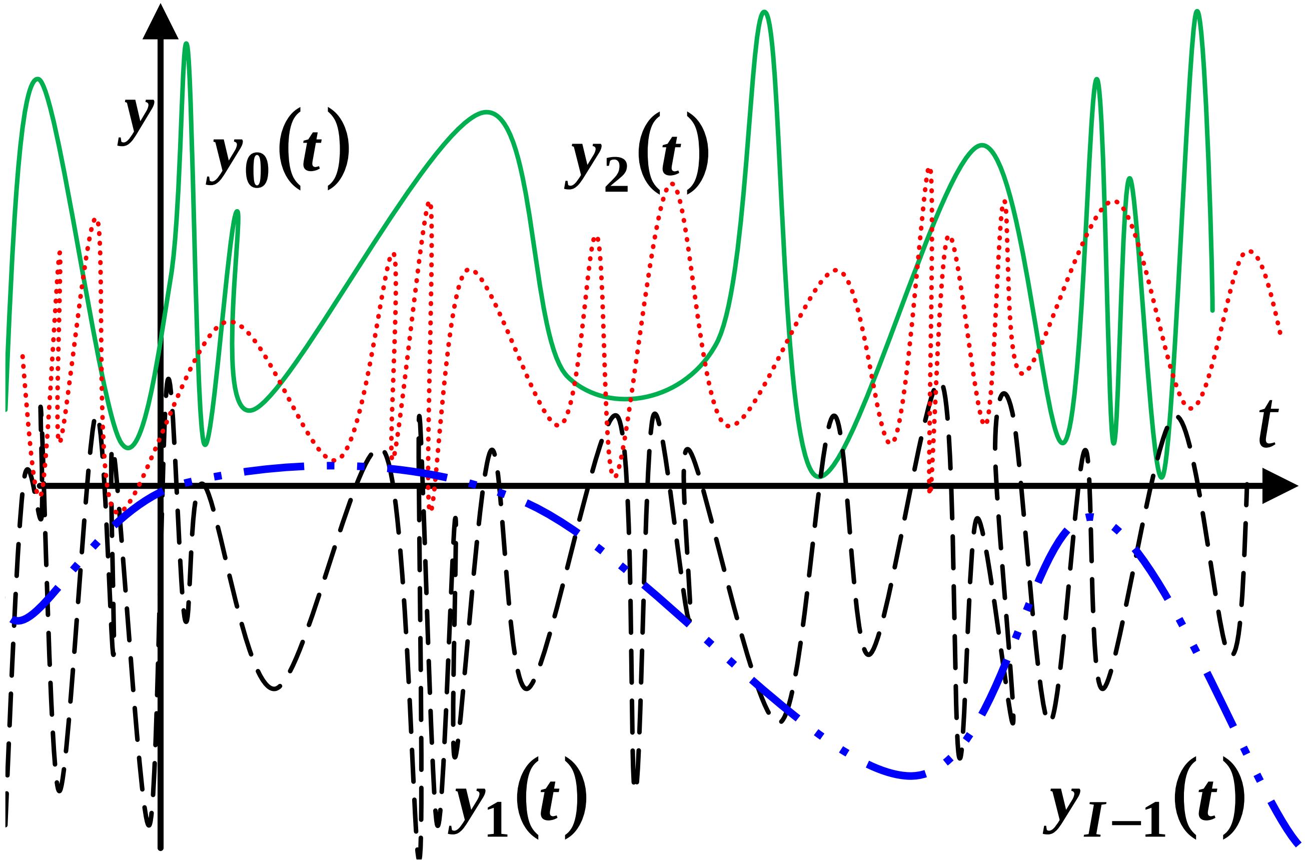
where

$y(t_0) = y_0$, $y(t_1) = y_1$, $y(t_2) = y_2$, … , $y(t_{J-1}) = y_{J-1}$ – **random variables**

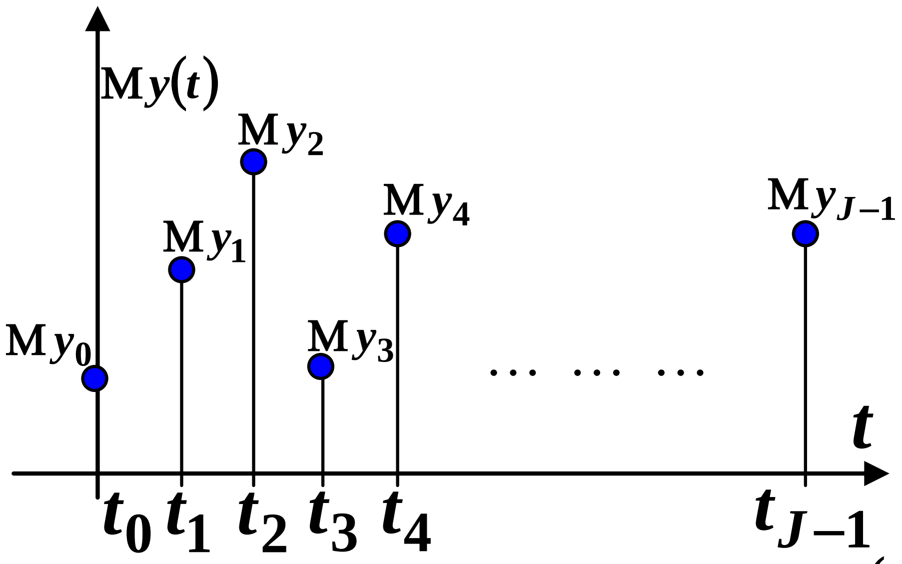
Ensemble of stochastic time series implementations :

$$\left\{ y_i(t):\ i = \overline{0, I-1} \right\}$$

where $y_i(t)$ – deterministic functions (implementations)



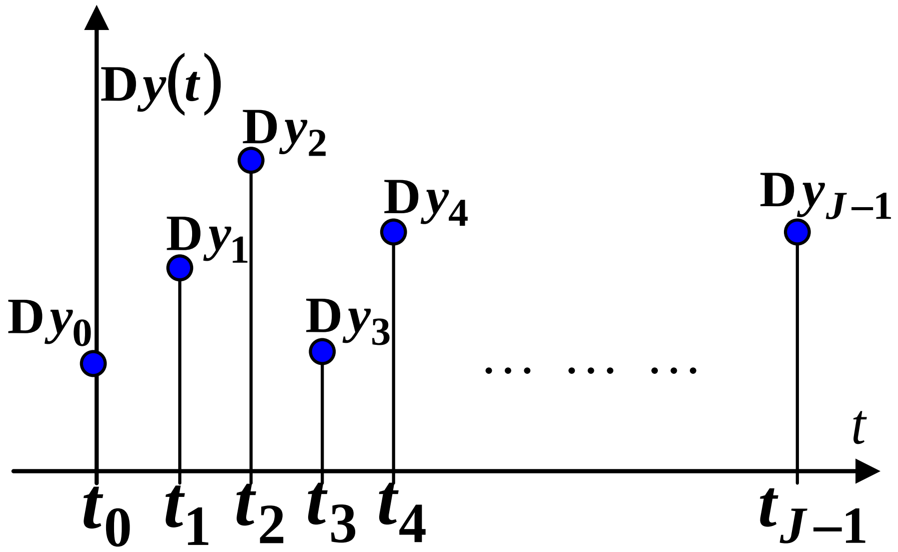
Mathematical expectation $\mathrm{M}\,y(t)$ of stochastic time series $y(t)$ is deterministic function. This deterministic function $\mathrm{M}\,y(t)$ assigns to each moment of time the mathematical expectation of a random variable



Statistical estimation $\hat{m}_{y(t)}$ of mathematical expectation $\mathrm{M}\,y(t)$ (average, mean) of the stochastic time series $y(t)$:

$$\hat{m}_{y(t)} = \frac{1}{I}\sum_{i=0}^{I-1} y_i(t)$$

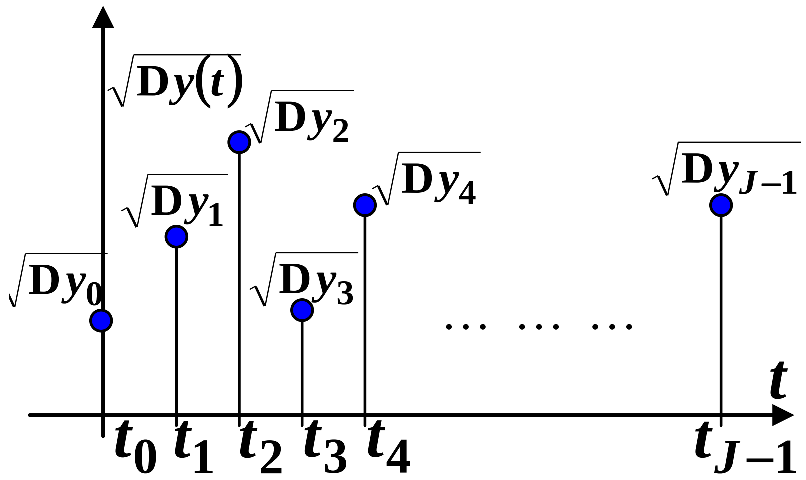
Variance $\mathrm{D}y(t)$ of stochastic time series $y(t)$ is deterministic function. This deterministic function $\mathrm{D}y(t)$ assigns to each moment of time the variance of a random variable



Statistical estimation $\hat{d}_{y(t)}$ of variance $\mathrm{D}y(t)$ of the stochastic time series $y(t)$:

$$\hat{d}_{y(t)} = \frac{1}{I-1}\sum_{i=0}^{I-1}\left(y_i(t) - \hat{m}_{y(t)}\right)^2$$

Mean square deviation (standard deviation, root-mean-square deviation $\sigma = \sqrt{\mathrm{D}y(t)}$ is deterministic function. This deterministic function $\sigma = \sqrt{\mathrm{D}y(t)}$ assigns to each moment of time the mean square deviation of a random variable



Statistical estimation $\hat{\sigma}_{y(t)}$ of mean square deviation (standard deviation, root-mean-square deviation) $\sigma = \sqrt{\mathrm{D}y(t)}$ of the stochastic time series $y(t)$:

$$\hat{\sigma}_{y(t)} = \sqrt{\hat{d}_{y(t)}} = \sqrt{\frac{1}{I-1}\sum_{i=0}^{I-1}\left(y_i(t) - \hat{m}_{y(t)}\right)^2}$$

MIX OF TIME SERIES

Additive mix of time series :

$$y(t) = T(t) + S(t) + N(t)\ ,$$

$T(t)$ – trend time series, $S(t)$ – seasonal time series , $N(t)$ – noise time series

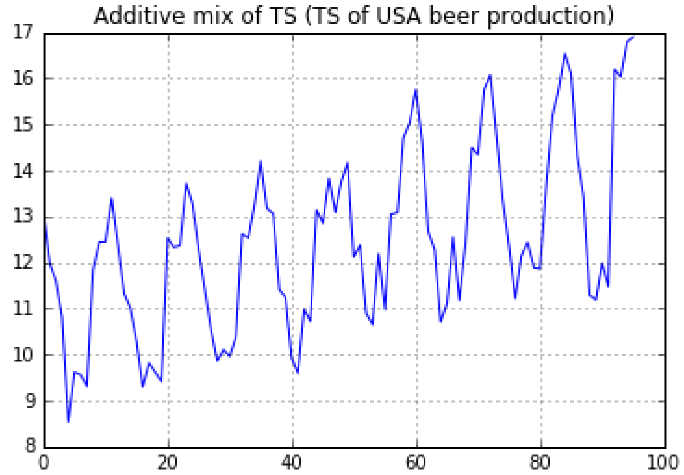
Multiplicative mix of time series :

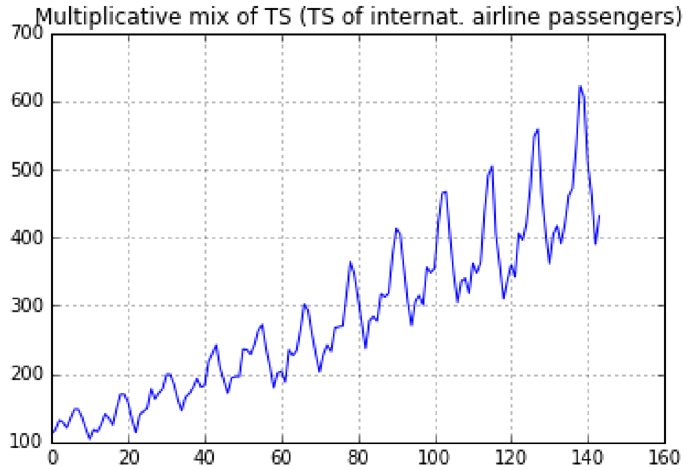
$$y(t) = T(t) \cdot S(t) \cdot N(t)\ ,$$

$T(t)$ – trend time series, $S(t)$ – seasonal time series , $N(t)$ – noise time series

Other mix of time series

***Example*:**





TRANSFORMATIONS OF MULTIPLICATIVE MIX OF TIME SERIES WHICH PENALIZE HIGHER VALUES MORE THAN SMALLER VALUES

$$g_1(t) = \log(y(t)),$$

$$g_2(t) = \sqrt{y(t)},$$

$$g_3(t) = \sqrt[3]{y(t)},$$

•

•

•

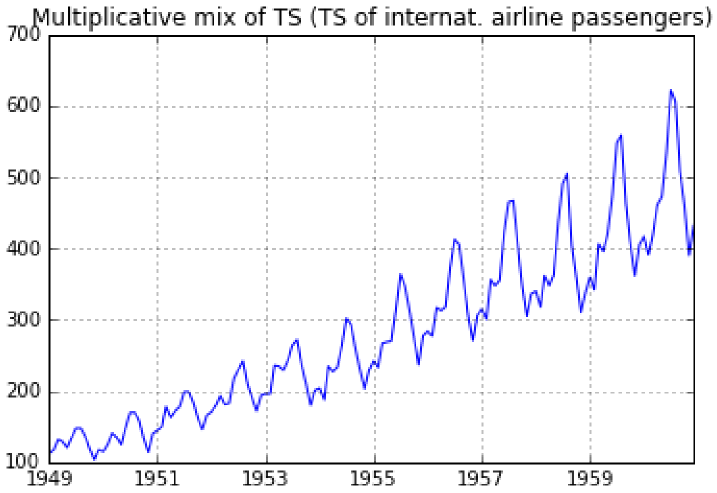
where

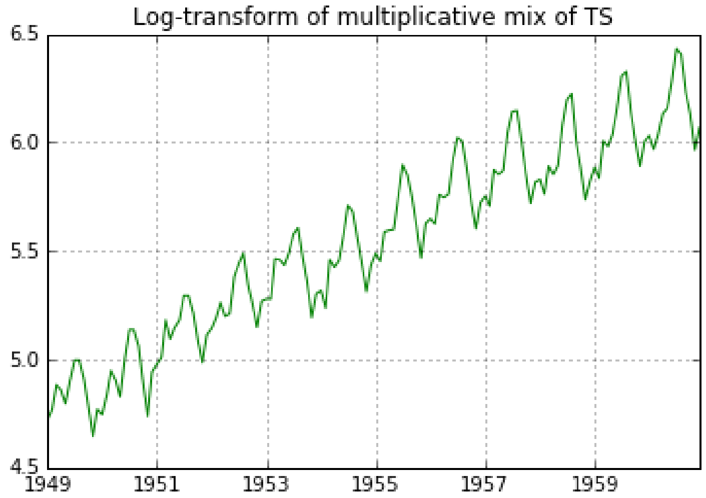
$y(t)$ – multiplicative mix of time series ,

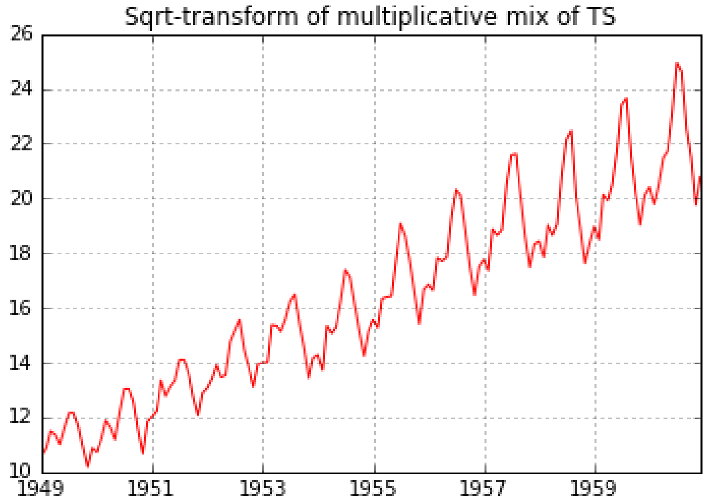
$g_1(t)$, $g_2(t)$, $g_3(t)$, … – transformed multiplicative mix of time series

__Example__:

TRANSFORMATIONS OF MULTIPLICATIVE MIX OF TIME SERIES WHICH PENALIZE HIGHER VALUES MORE THAN SMALLER VALUES



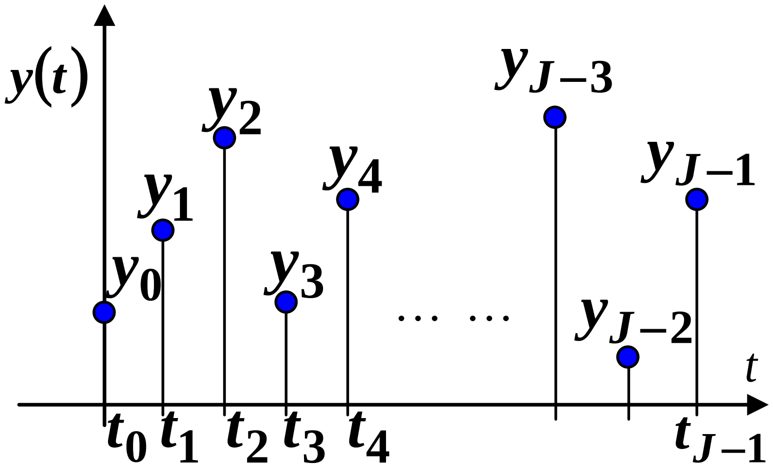




MOVING (ROLLING, RUNNING) AVERAGE
(FILTERING OF NOISE, ESTIMATING TREND) OF STOCHASTIC TIME SERIES

SIMPLE MOVING (ROLLING, RUNNING) AVERAGE (SMA) OF STOCHASTIC TIME SERIES :

Stochastic time series:

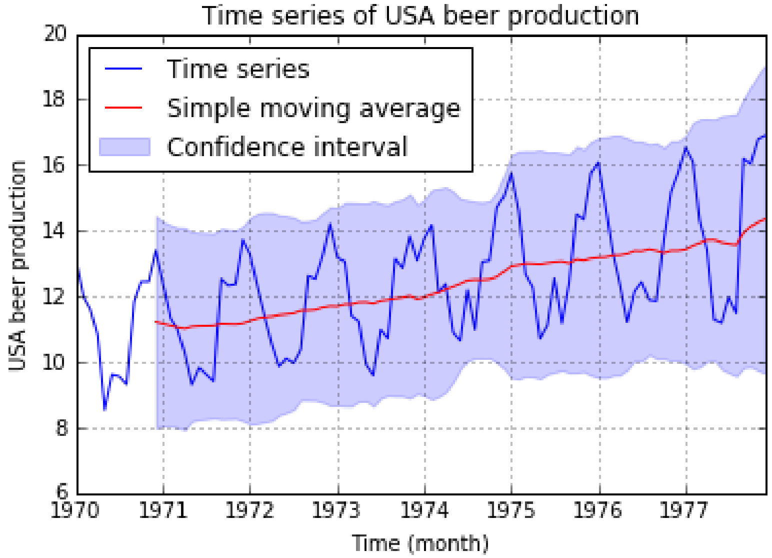


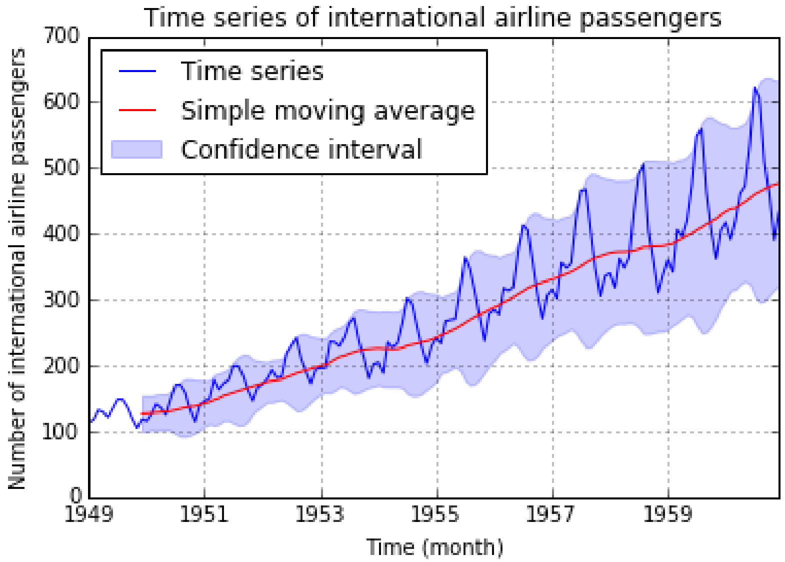
$$SMA_{J-1} = \frac{y_{J-1} + y_{J-2} + \cdots + y_{J-1-(tw-1)}}{tw} = \frac{1}{tw}\sum_{j=0}^{tw-1} y_{J-1-j} \,,$$

where *tw* – time window

__Example__:

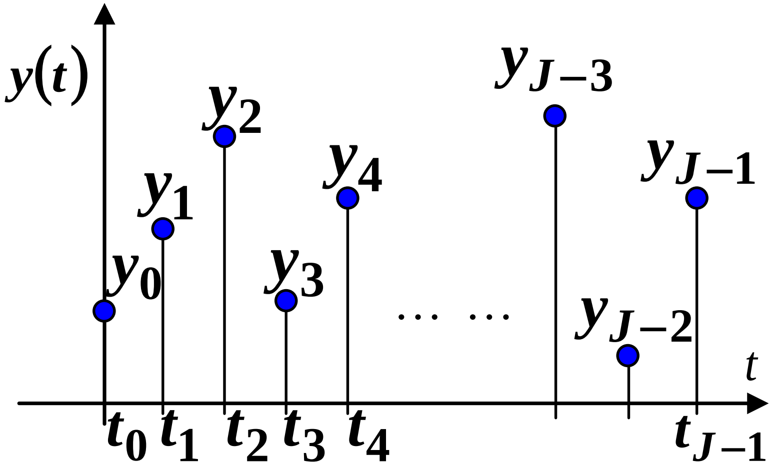
STOCHASTIC TIME SERIES & SIMPLE MOVING AVERAGE OF STOCHASTIC TIME SERIES





EXPONENTIAL (EXPONENTIALLY WEIGHTED) MOVING AVERAGE OF STOCHASTIC TIME SERIES (EMA):

Stochastic time series:



$$EMA_{J-1} = \alpha \cdot y_{J-1} + (1-\alpha) \cdot EMA_{J-2},$$

$$EMA_{J-2} = \alpha \cdot y_{J-2} + (1-\alpha) \cdot EMA_{J-3},$$

$$EMA_{J-3} = \alpha \cdot y_{J-3} + (1-\alpha) \cdot EMA_{J-4},$$

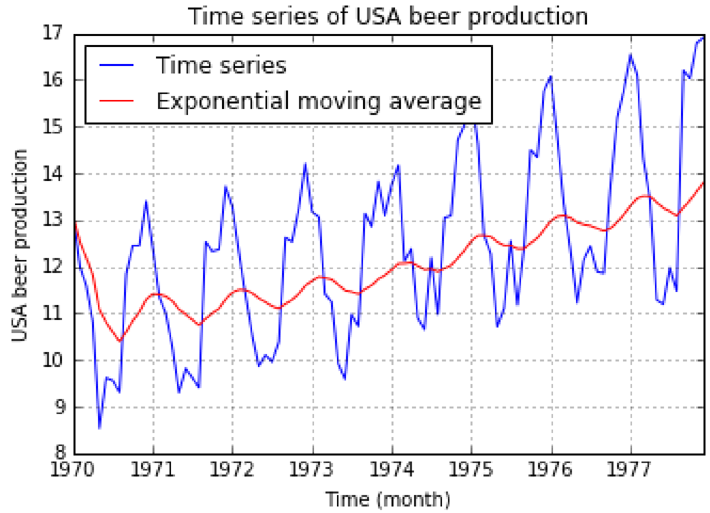
$$\vdots$$

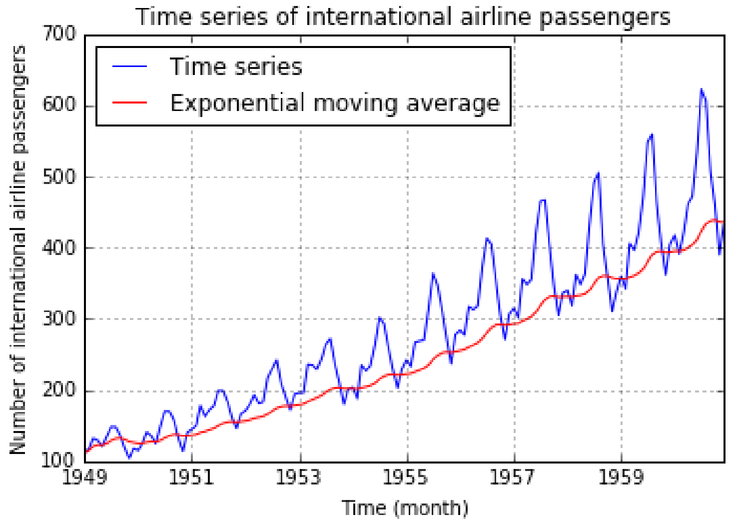
$$EMA_0 = y_0$$

where α **– smoothing constant,** $\alpha \in (0, 1)$, $\alpha \in (0, 1)$)

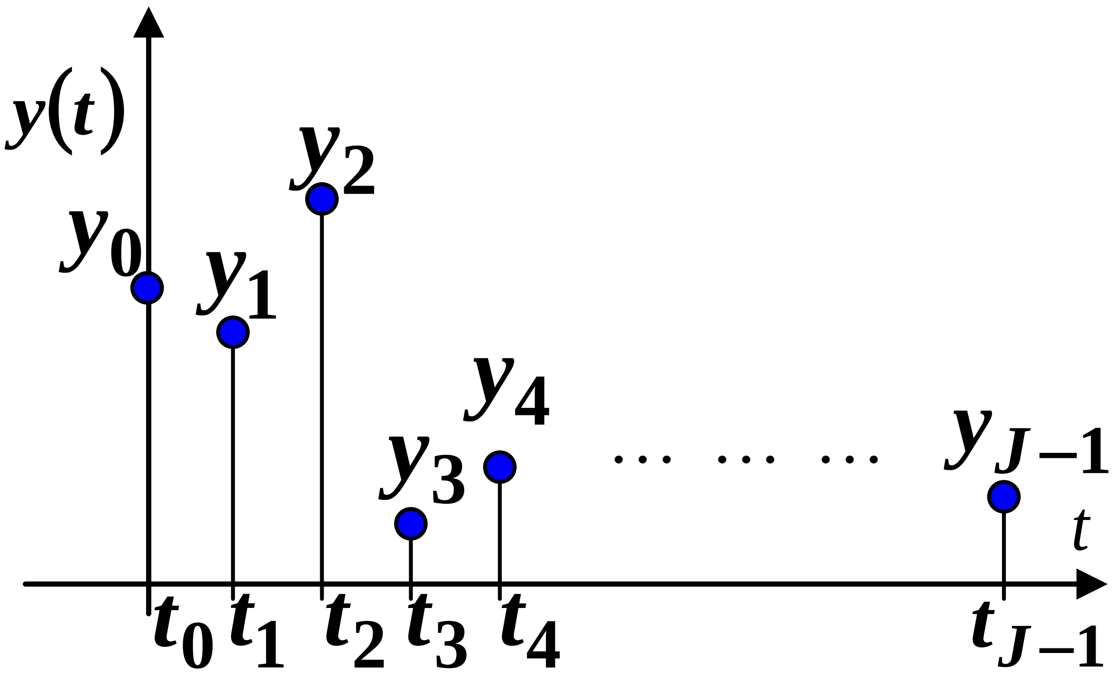
***Example*:**

STOCHASTIC TIME SERIES & EXPONENTIAL MOVING AVERAGE OF STOCHASTIC TIME SERIES





NONSTATIONARY STOCHASTIC TIME SERIES



where $y(t_0) = y_0$, $y(t_1) = y_1$, … , $y(t_{J-1}) = y_{J-1}$ **– different random variables**

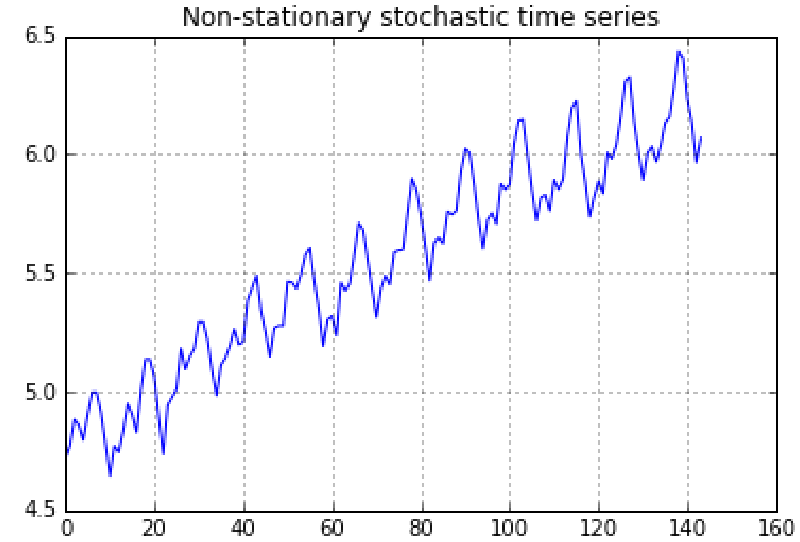
AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA) MODEL OF NONSTATIONARY OR STATIONARY STOCHASTIC TIME SERIES

The autoregressive integrated moving average (ARIMA(p,d,q)) is extension of the ARMA(p,d,q) model for nonstationary time series, which can be made stationary by taking differences of some order from the original time series (so-called integrated or difference-stationary time series). The ARIMA(p,d,q) model indicates that the differences of first order of time series are subject to the ARMA(p,d,q) model:

$$\Delta^d y(t) = const. + \sum_{k_1=1}^{p} a_{k_1} \Delta^d y(t-k_1) + \sum_{k_2=1}^{q} b_{k_2} \varepsilon(t-k_2) + \varepsilon(t),$$

where $const.$, a_{k_1}, b_{k_2}, $k_1 = \overline{1,p}$, $k_2 = \overline{1,q}$ – model parameters, $\varepsilon(t)$ – white noise , Δ^d – the operator of the difference of the time series of order d (successively taking d times the differences of the first order - first from the time series, then from the differences of the first order, then from the second order, etc.)

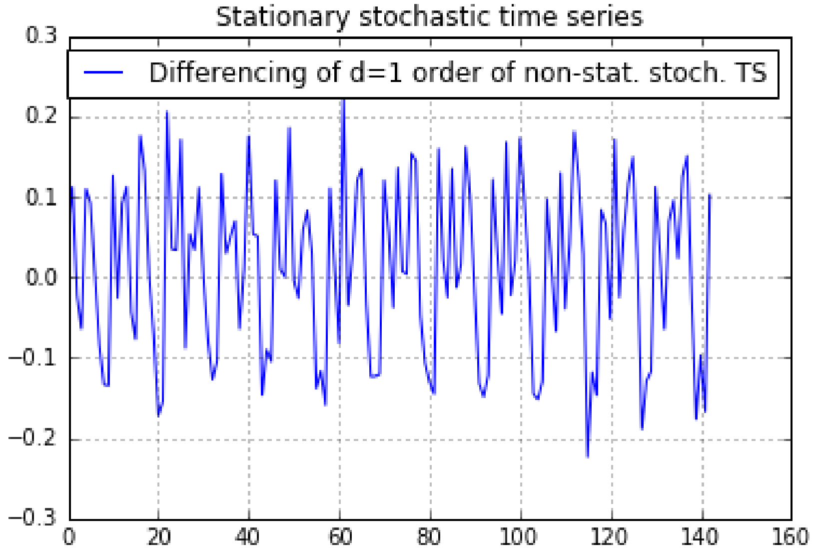
IDENTIFICATION OF ORDER (p,d,q) OF ARIMA(p,d,q)-MODEL



Results of Dickey-Fuller Test:

Test Statistic	**-1.717017**
p-value	**0.422367**
#Lags Used	**13.0**
Number of Observations Used	**130.0**
Critical Value (5%)	**-3.481682**
Critical Value (1%)	**-2.578770**
Critical Value (10%)	**-2.884042**

H_0 hypothesis: Stochastic time series is non-stationary stochastic time series

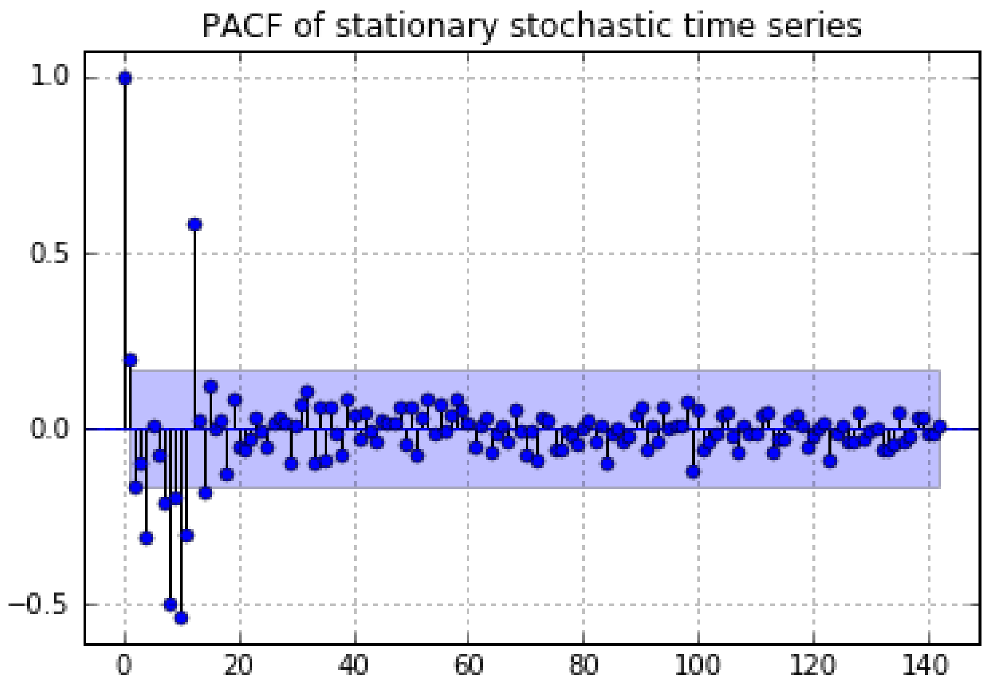


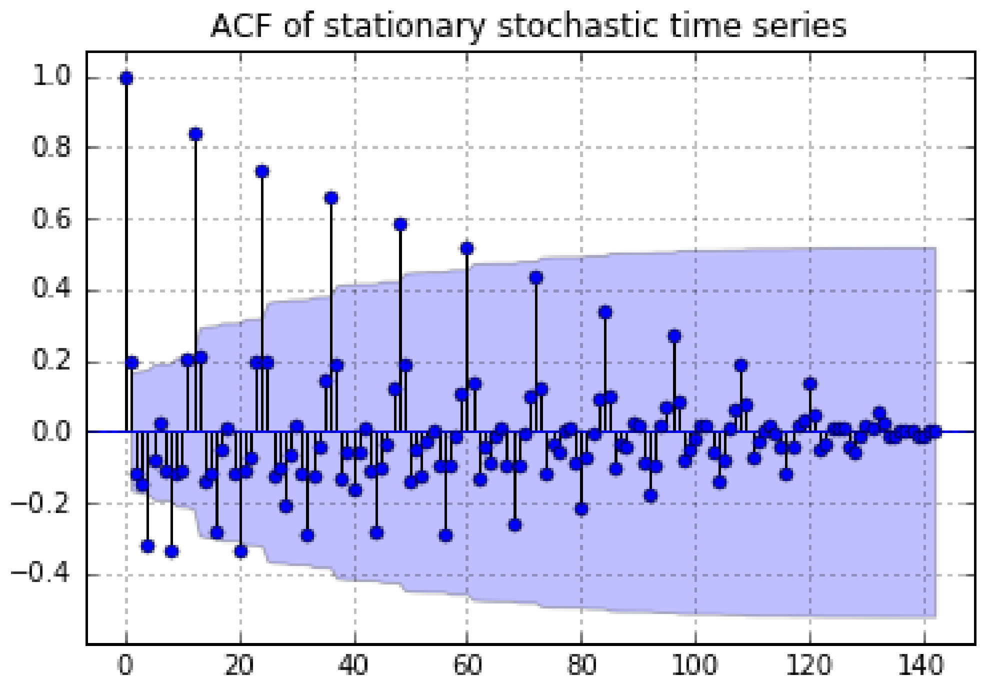
Results of Dickey-Fuller Test:

Test Statistic	**-2.717131**
p-value	**0.071121**
#Lags Used	**14.0**
Number of Observations Used	**128.0**
Critical Value (5%)	**-2.884398**
Critical Value (1%)	**-2.578960**
Critical Value (10%)	**-3.482501**

H_1 hypothesis: Differencing of $d=1$ order of non-stationary stochastic time series is stationary stochastic time series. The Dickey-Fuller test statistic is less than the 10% critical value, thus the TS is stationary with 90% confidence

$$d = 1$$



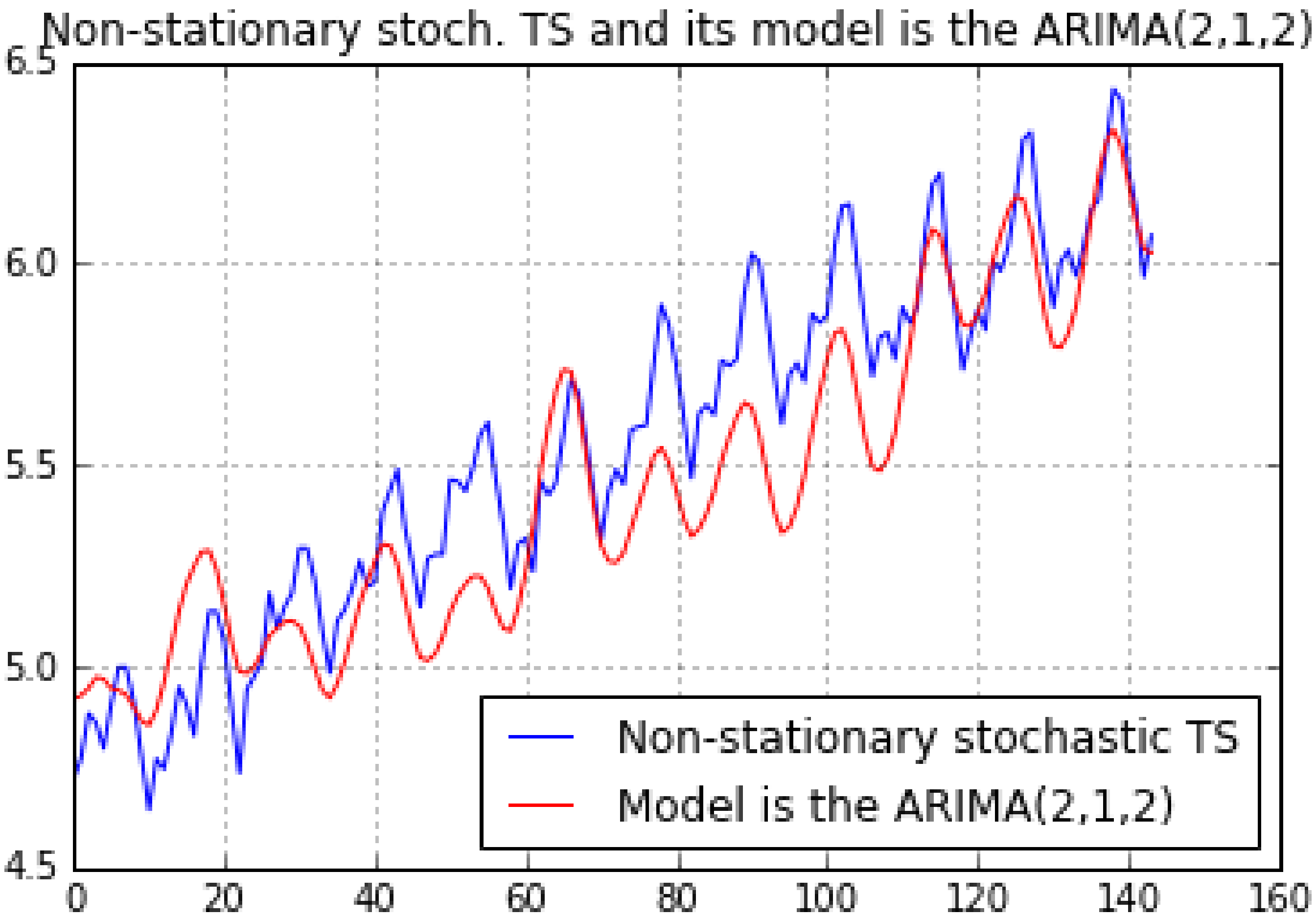


Order p – the lag value where the partial autocorrelation function (PACF) chart crosses the upper confidence interval for the first time

$$p = 2$$

Order q – the lag value where the autocorrelation function (ACF) chart crosses the upper confidence interval for the first time

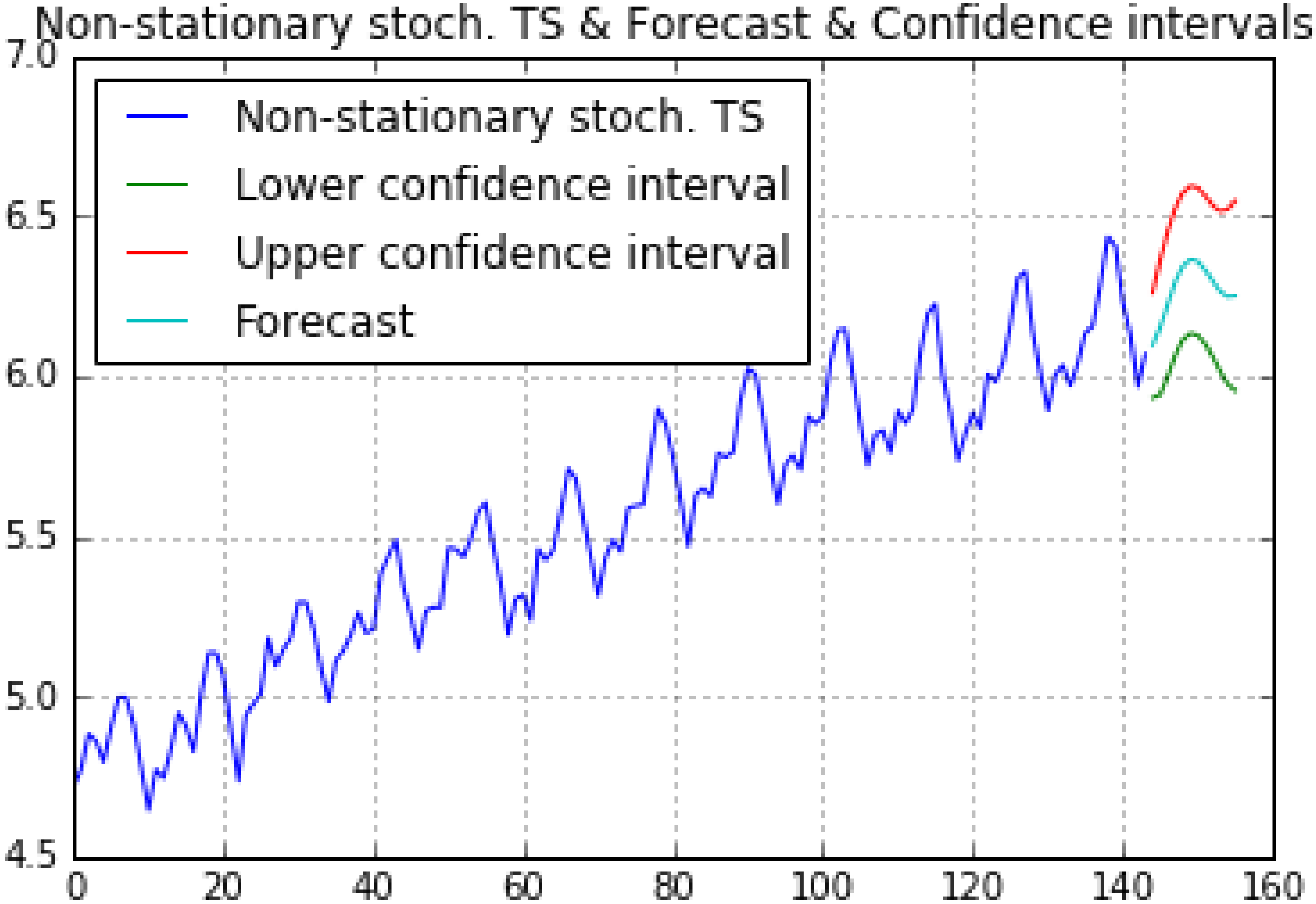
$$q = 2$$



Order ($p = 2$, $d = 1$, $q = 2$)

Sum of squared residuals (SSR) = 16.9

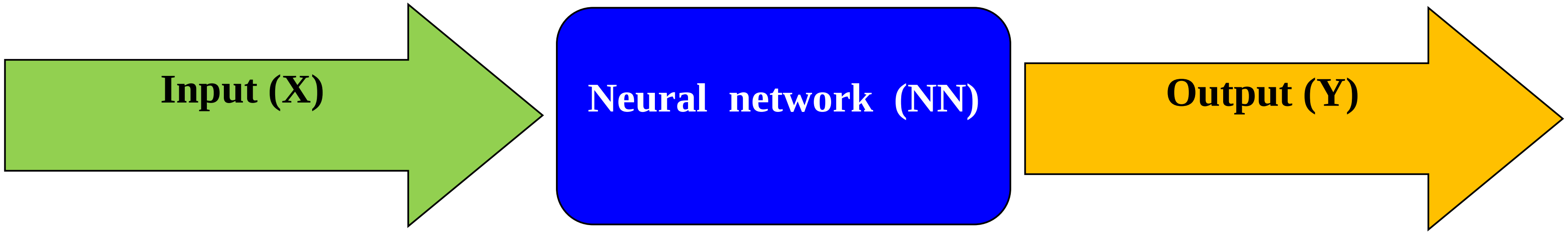
FORECAST



Forecast horizon = 12

TIME SERIES FORECASTING BASED ON NEURAL NETWORKS (NN)

Neural network (NN):



Time series :