Time Series Analysis:
An Introduction

WHAT IS A TIME SERIES?

- Series of observations recorded sequentially over a period of time (i.e. a collection of observations recorded along with the time stamp)
- Represented as

$$\left(t, Y_t\right);\ t = 1, \ldots, n$$

- Y_t may be univariate (single variable) or multivariate (collection of variables of interest)
- Frequency of observations: yearly, quarterly, monthly, weekly, daily or tick (i.e. as and when data changes)

ANALYSIS OF TIME SERIES

- Identify the dominant feature (s) of an observed time series
- Explain the movements of the time series through an "appropriate" stochastic process and study the characteristics of the process
- Identify possible cause-effect or lead-lag relationship between related time series
- Fit (through estimation of parameters) a statistical model and build a predictive model for forecasting
- Build early warning prediction system for possible future adverse events

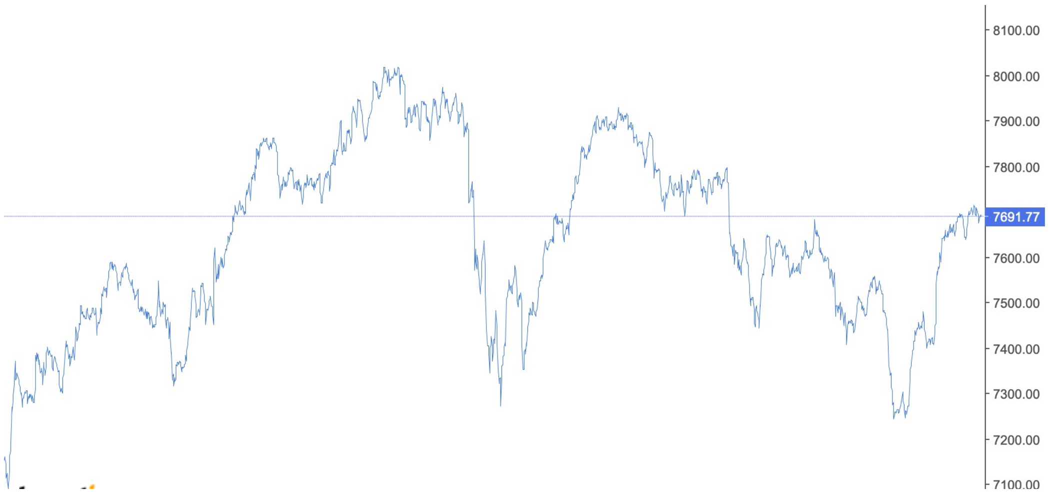
APPLICATION AREAS

- **FINANCIAL:** Exchange rate forecasting, share price movements, credit card fraud detection
- **ECONOMIC:** Modeling of Economic fundamentals data, e.g. GDP, GNP, inflation, employment rate
- **BIO-MEDICAL:** Monitoring/modeling ECG, EEG, Diastolic & systolic pressure
- **SPEECH**: Digitized speech signal data-synthesis and analysis
- **METEOROLOGICAL:** Forecasting and/or relationship between rainfall, temperature, pressure, air pollution
- **Wireless communication, radar signal, nowcasting in aviation, demographic**, ……….. ……..

FINANCIAL TIME SERIES

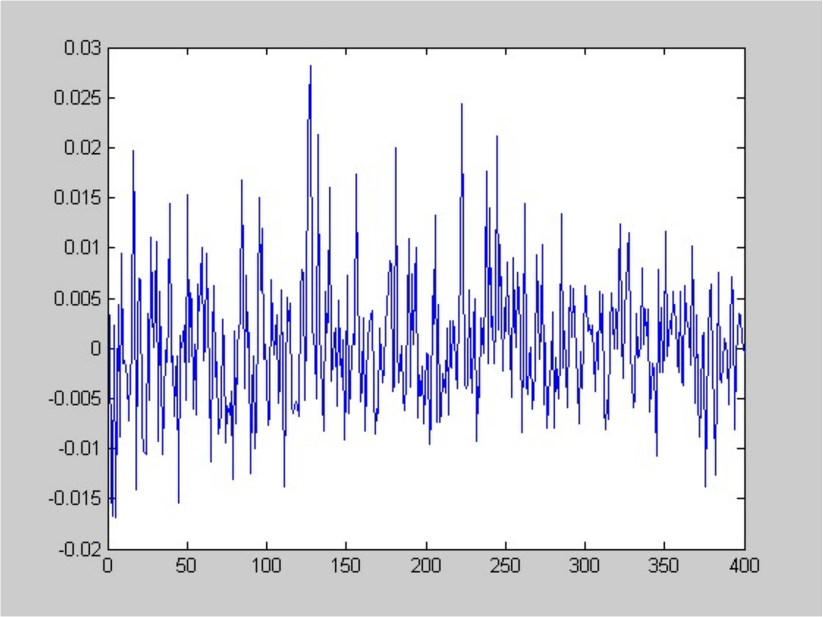
DJIA (Dow Jones): daily Series

FTSE tick series



FINANCIAL TIME SERIES

NSE Nifty: Daily Return Series



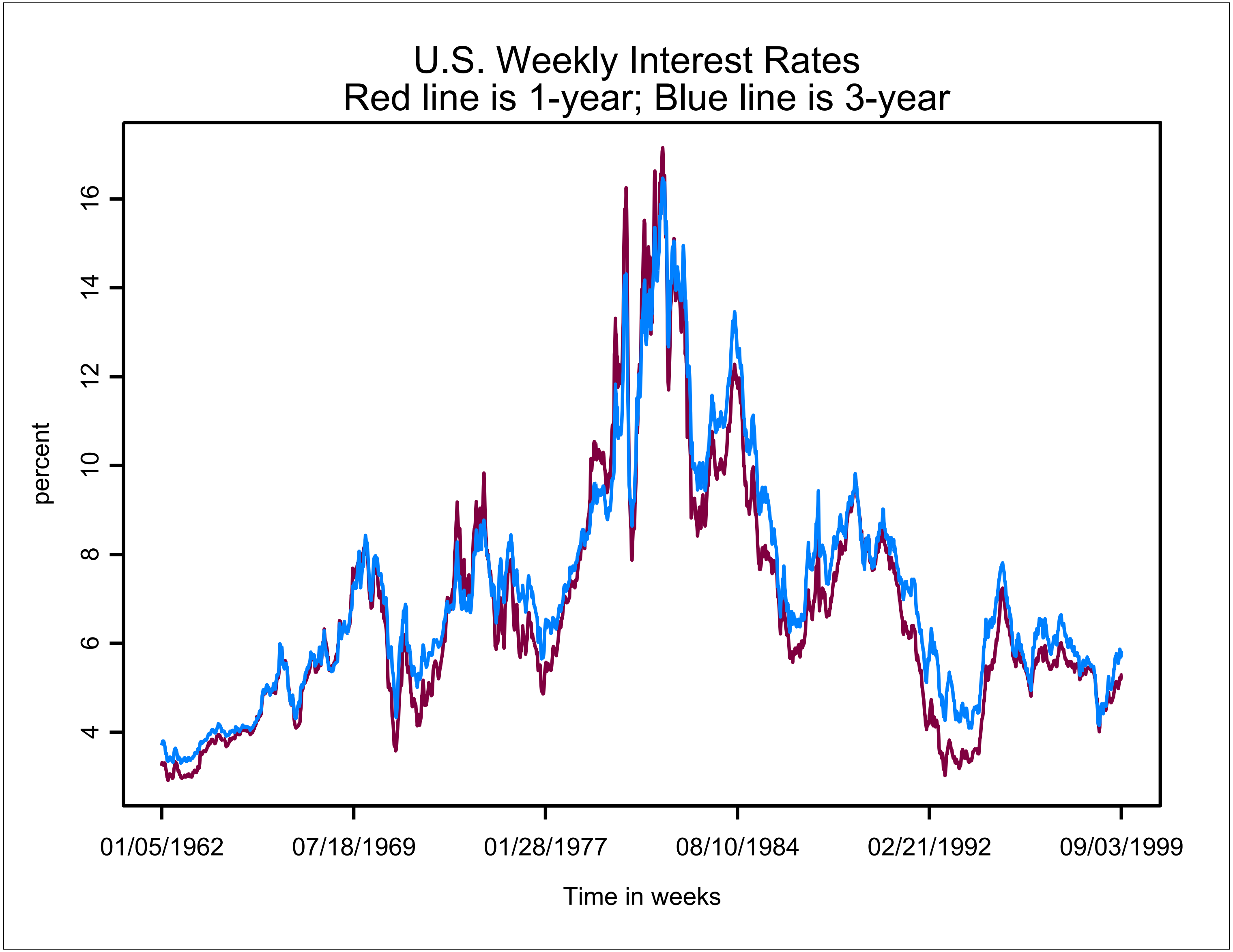
FINANCIAL TIME SERIES

FOREIGN EXCHANGE RATES

INR/EURO Daily

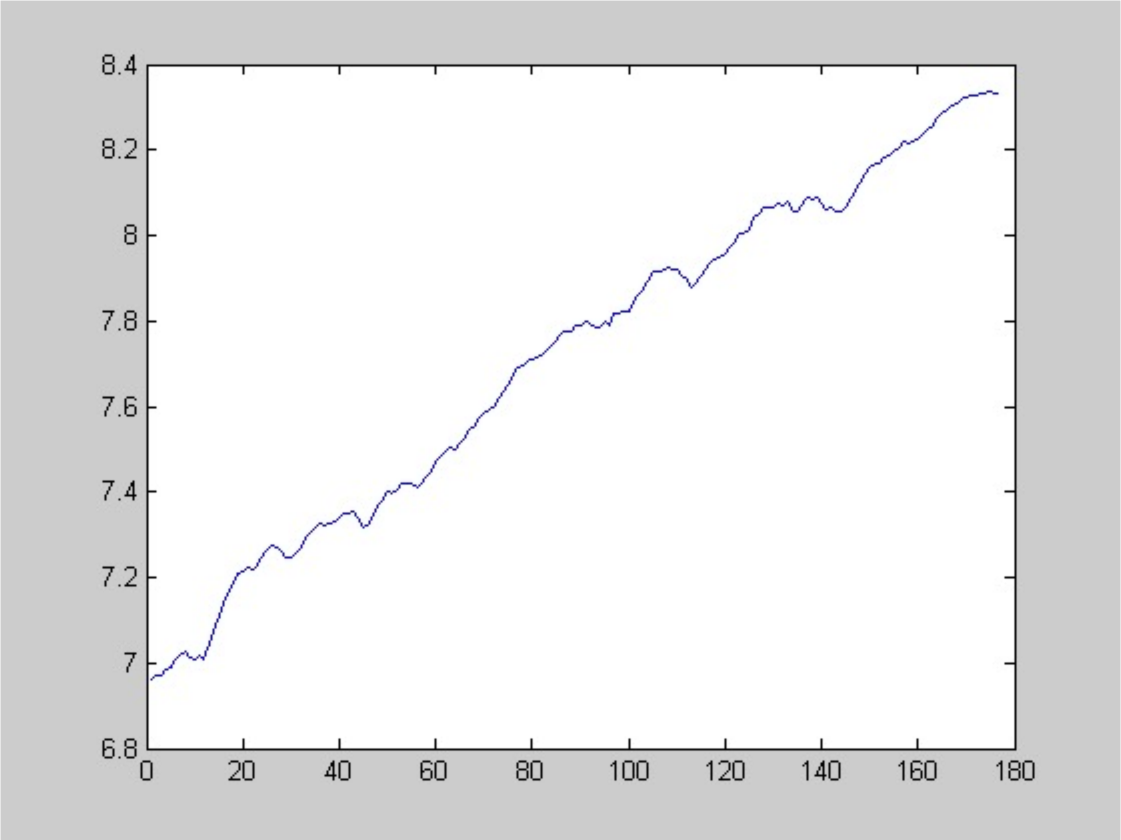
INR/USD Daily

FINANCIAL TIME SERIES

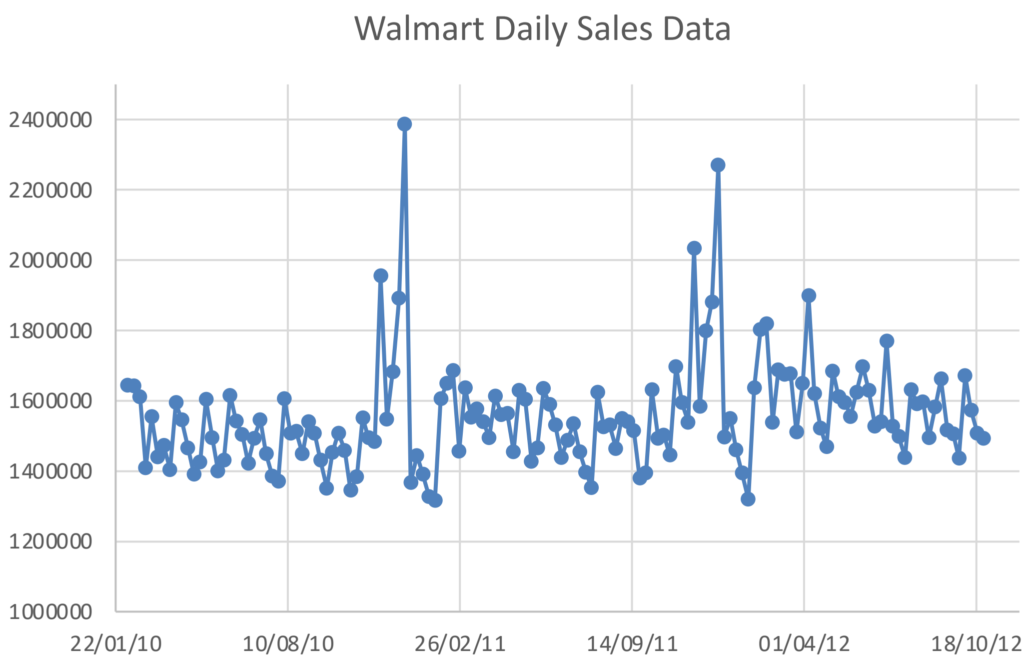


ECONOMIC TIME SERIES

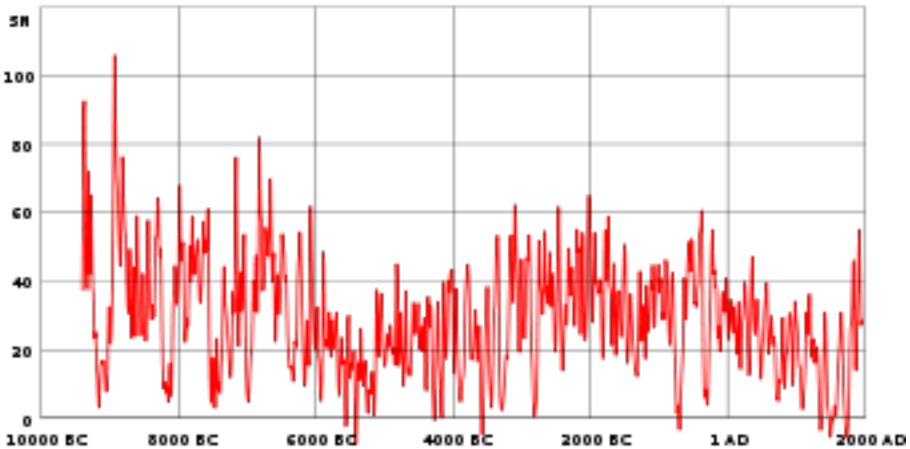
US Gross National Product (GNP) series

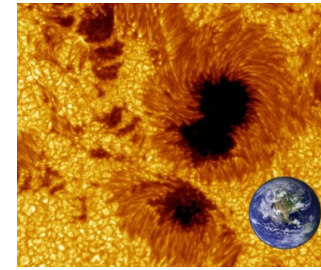


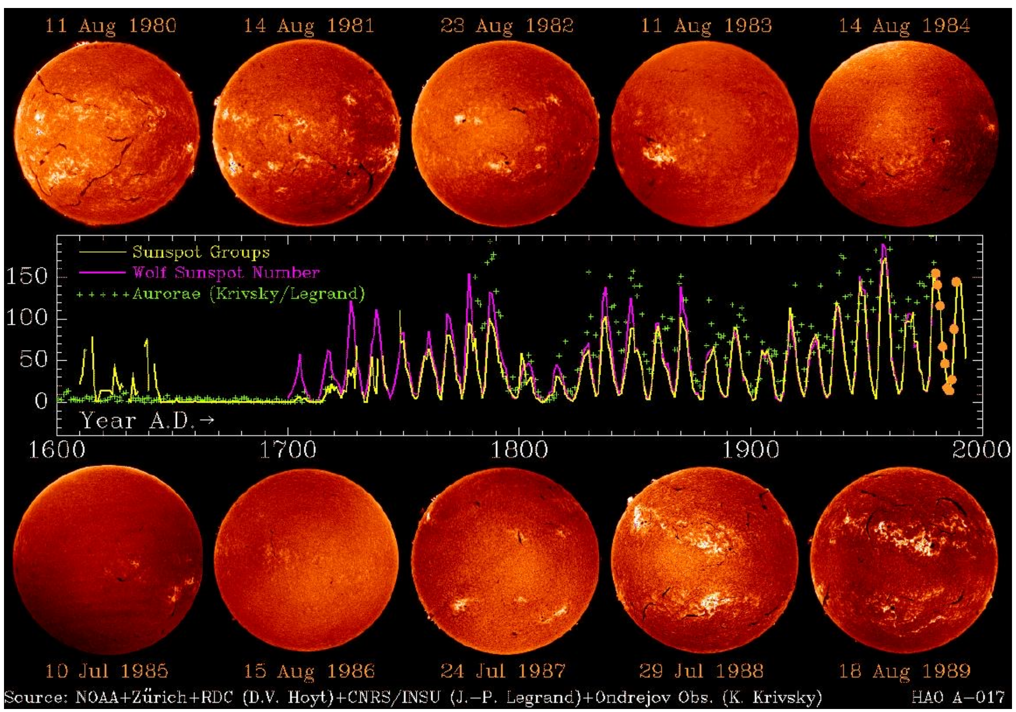
Market Research Time Series



SUNSPOT SERIES



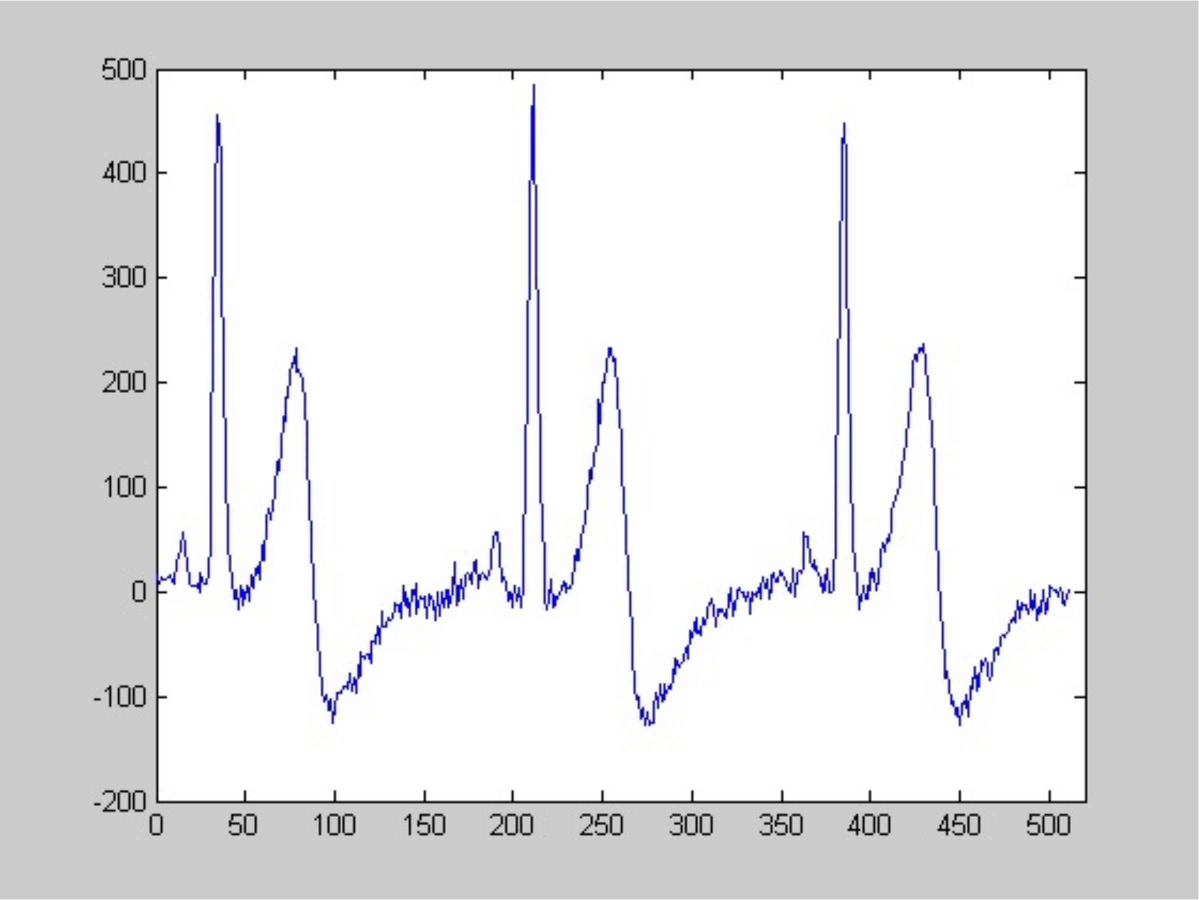




The Wolf number (also known as the International sunspot number, relative sunspot number, or Zürich number) is a quantity that measures the number of sunspots and groups of sunspots present on the surface of the sun. The Sunspot Number is a crucial tool used to study the solar dynamo, space weather and climate change.

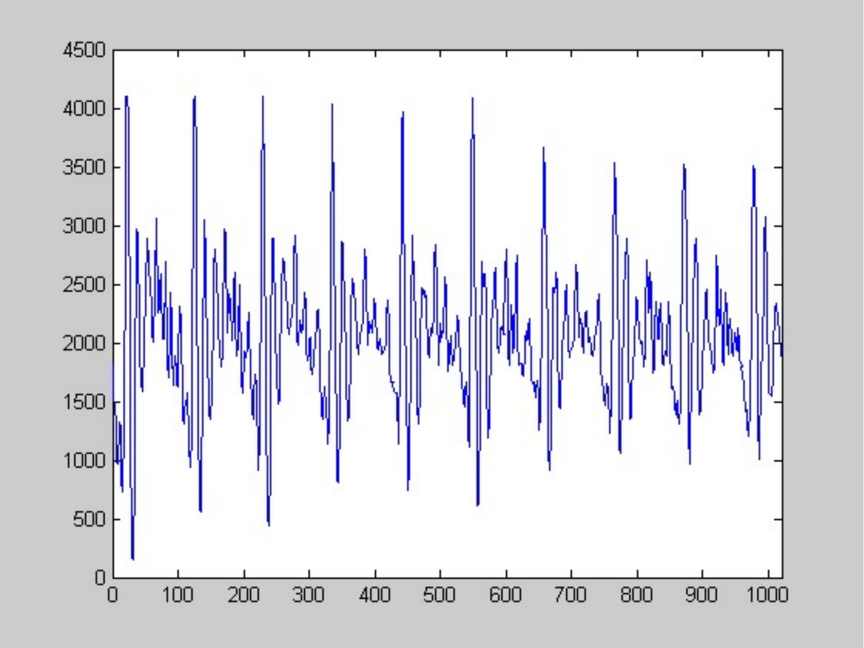
BIO-MEDICAL TIME SERIES

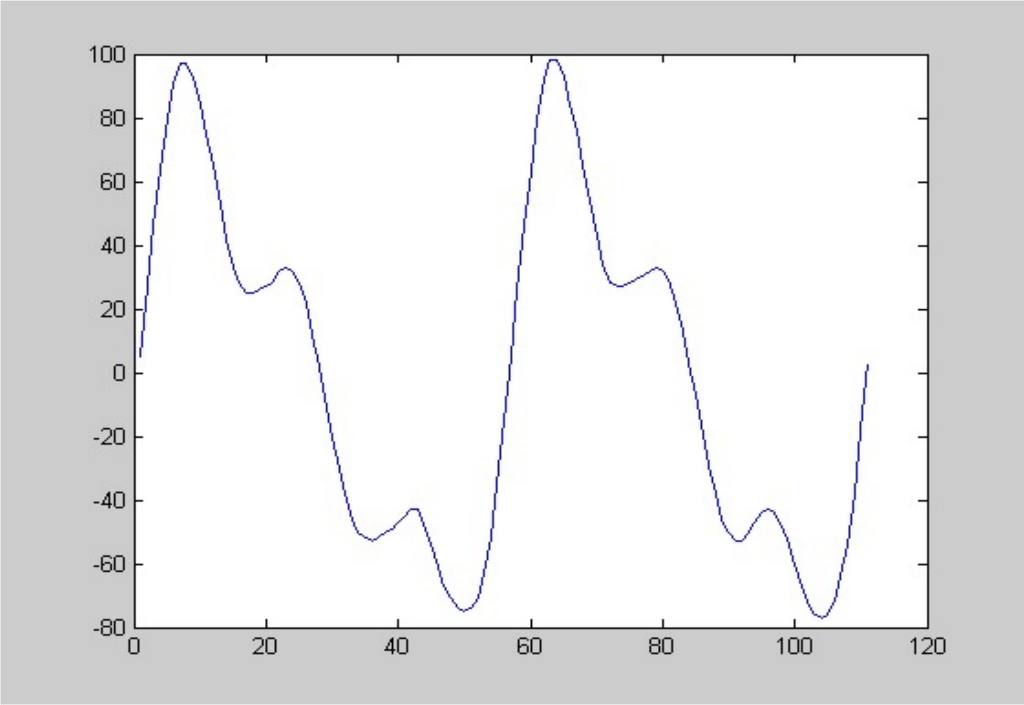
Digitized ECG signal data



SPEECH PROCESSING

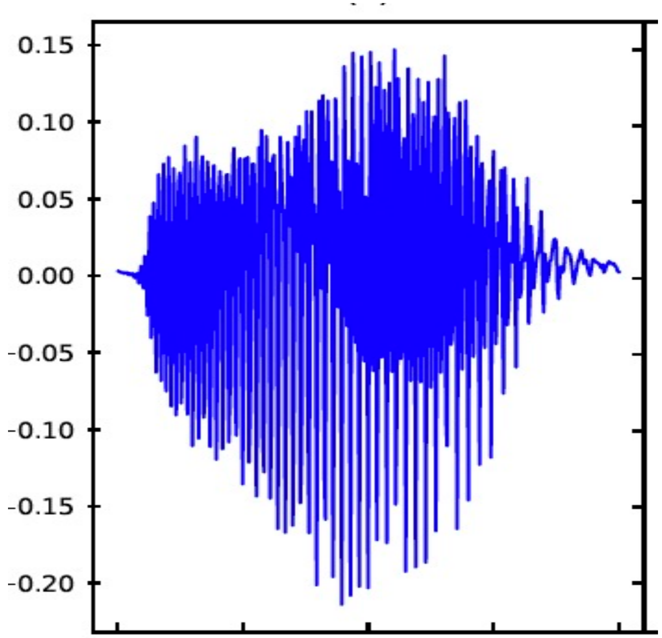
Digitized Speech signal



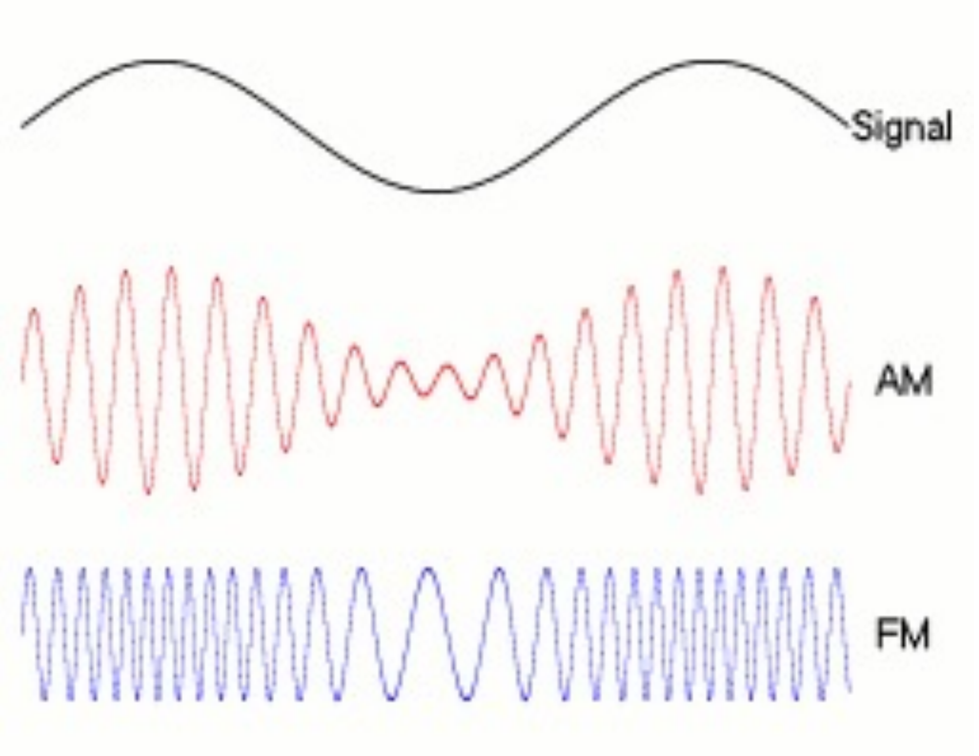


SPEECH PROCESSING

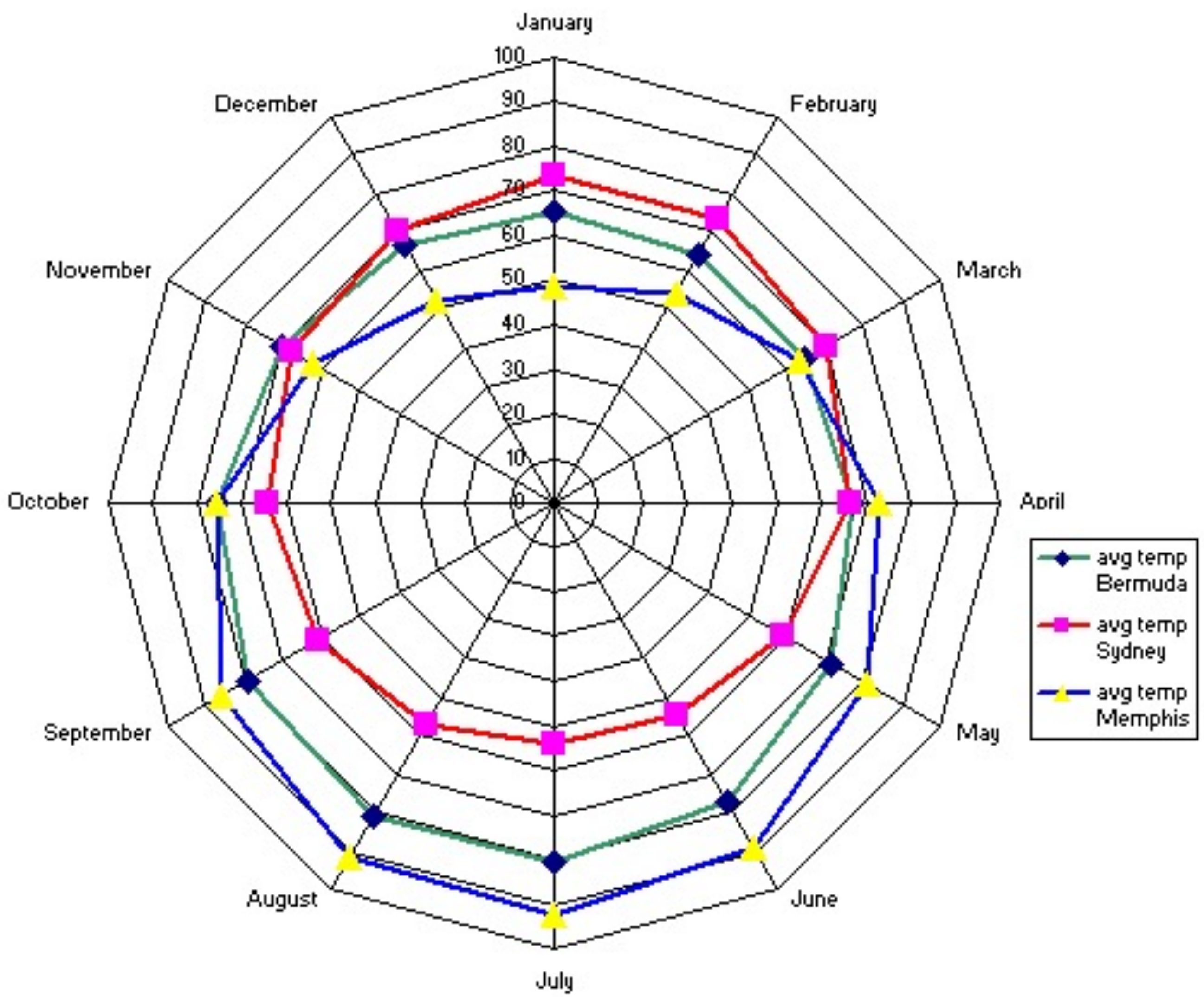
Bat squeak/chirp Digitized Data



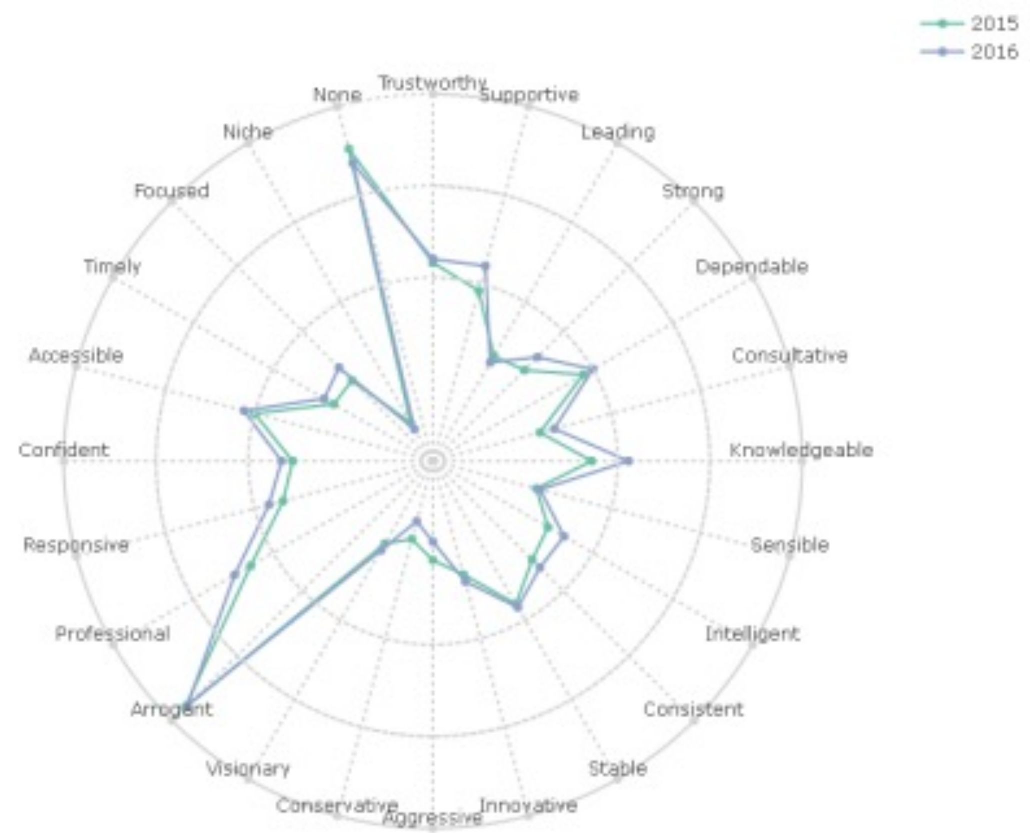
Communication Signals



MULTIVARIATE METEOROLOGICAL TIME SERIES



MULTIVARIATE TIME SERIES



WHAT TYPE OF DOMINANT FEATURES CAN BE PRESENT IN A TIME SERIES?

- TIME TREND COMPONENT
- SEASONAL VARIATION COMPONENT
- LONG TERM CYCLICAL COMPONENT
- IRREGULAR COMPONENT

Components of a Time Series

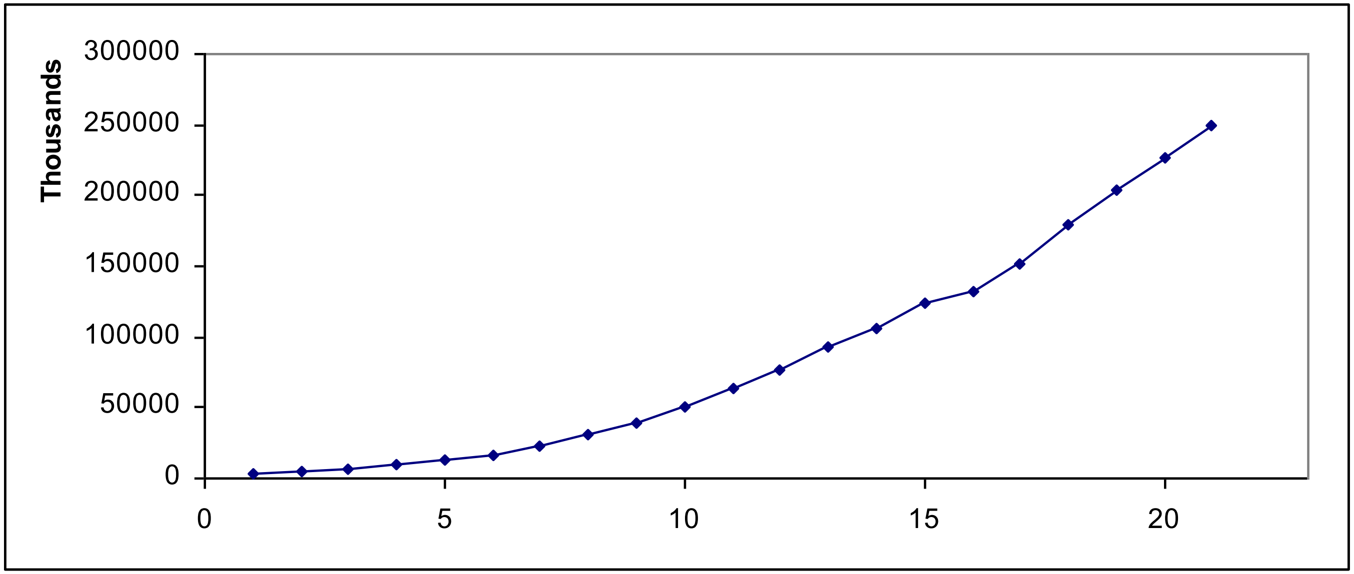
- Trend – smooth long term characteristics of a time series
- Seasonal Variation – Patterns of change in a time series within a year which tends to repeat each year
- Cyclical Variation – the rise and fall of a time series over periods longer than one year
- Irregular Variation – Random stochastic component

Additive Model OR Multiplicative Model of components

- STATISTICAL HYPOTHESIS TESTING APPROACH FOR TESTING EXISTENCE OF THE COMPONENTS
- ESTIMATION OF SIGNIFICANT COMPONENTS

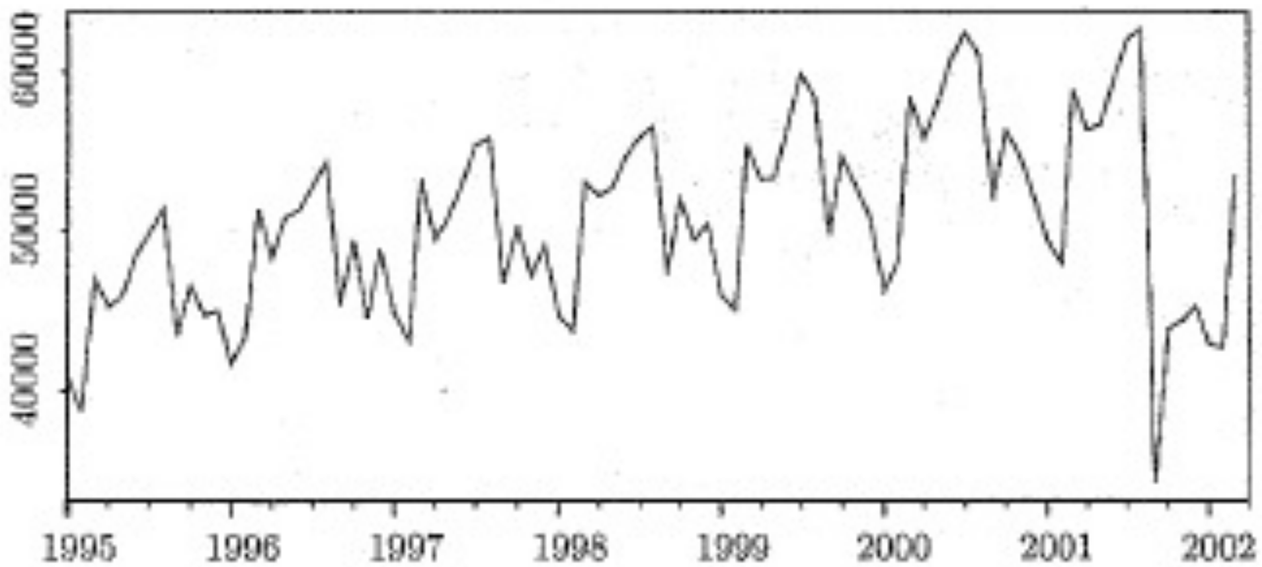
Dominant Trend Component

- US annual population series



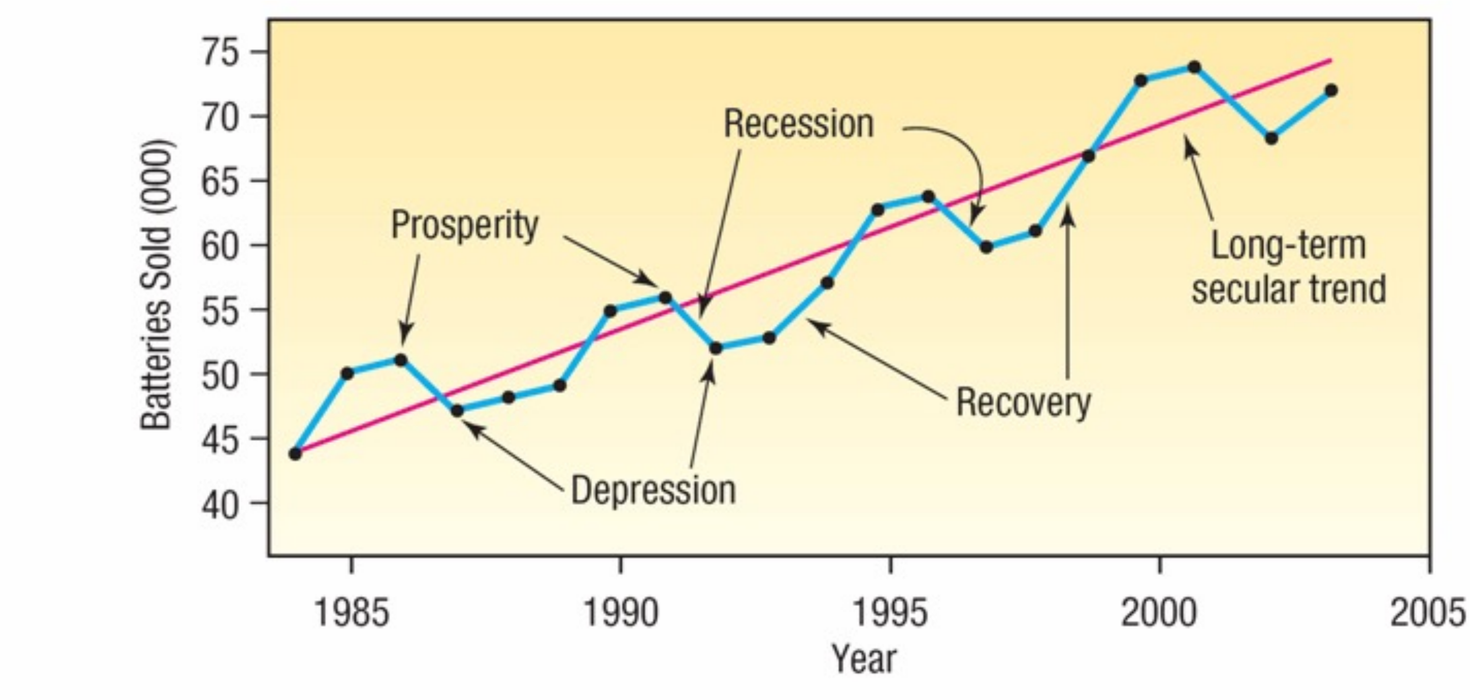
Dominant Seasonal and Trend Component

Monthly Airline Passengers in USA



Observe the “structural break” in the time series

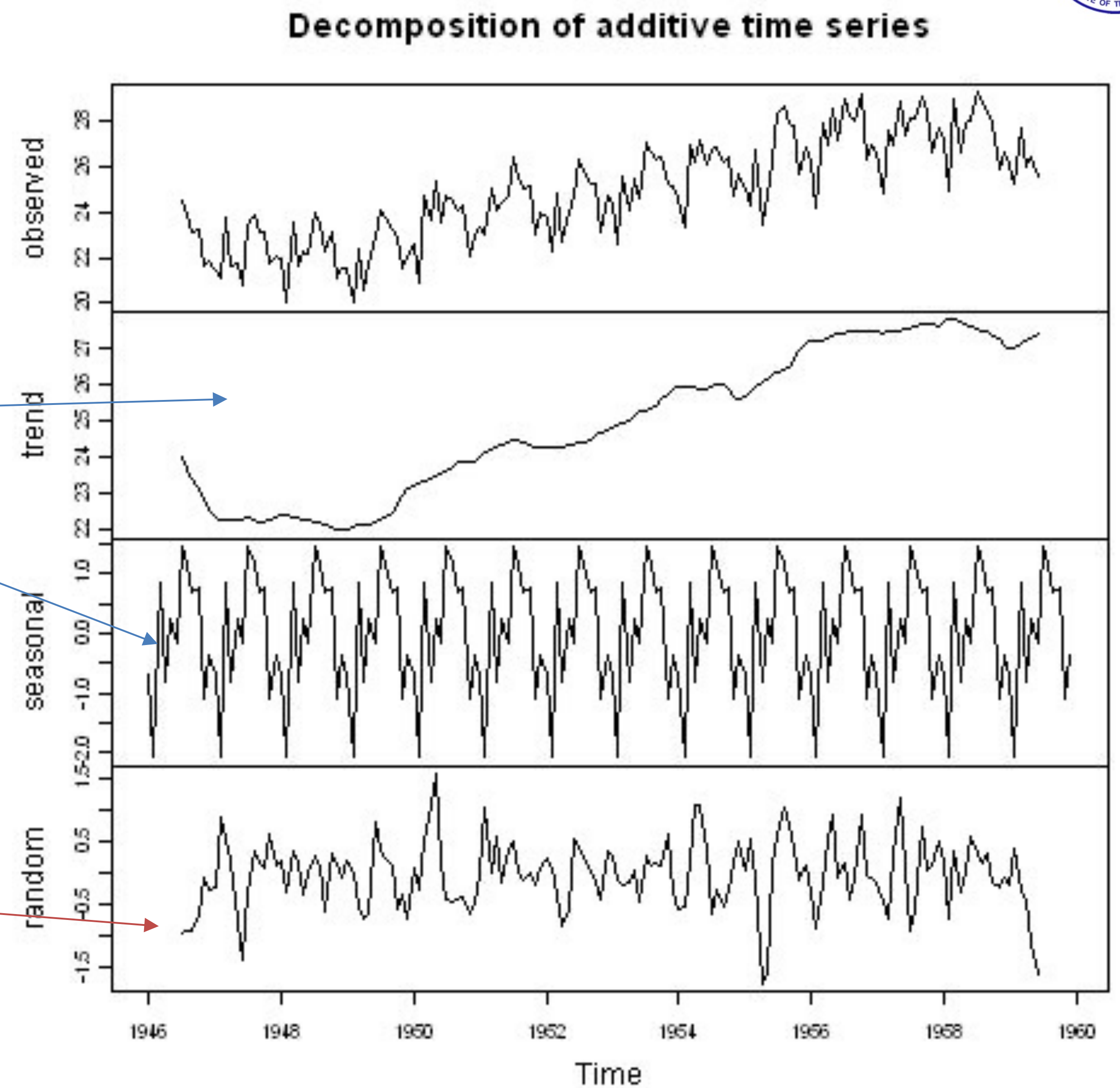
Dominant Cyclical and Trend Components



Positions of cycle:

Prosperity/Peak, recession, depression/ trough, recovery

Aim of decomposition



Deterministic components

Stochastic component:

Trend Estimation

Simple Moving Average Method

- Useful in smoothing time series to find its trend
- Equal importance to all values inside the MA window
- 2 sided or 1-sided
- Padding, if required

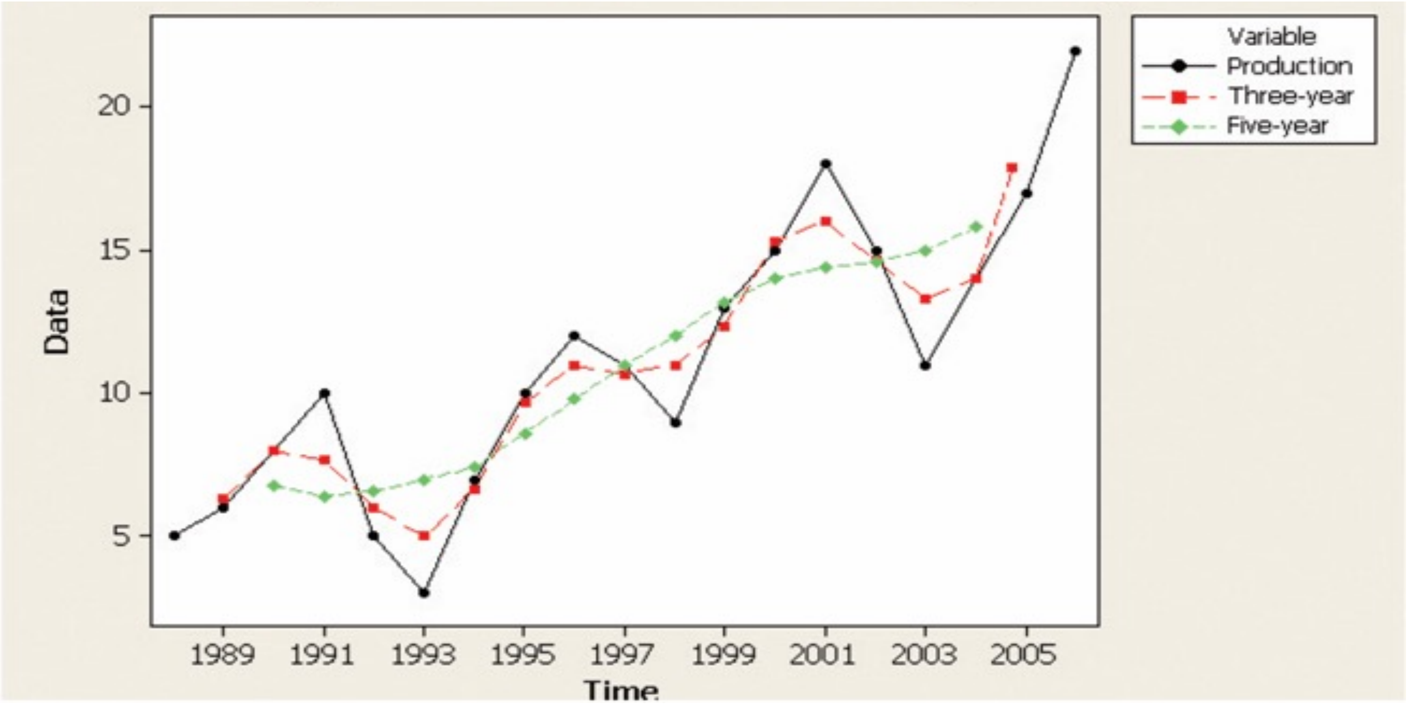
Extensions

- Weighted MA
- Exponential Weighted MA

Simple Moving Average

Year	Values	MA(3)
1995	20	NA
1996	24	(20+24+22)/3 = 22
1997	22	(24+22+26)/3 = 24
1998	26	(22+26+25)/3 = 24
1999	25	NA

Three-year and Five-Year Moving Averages



Weighted Moving Average

- A simple moving average assigns the same weight to each observation inside the MA window
- Weighted moving average assigns different weights to the observations inside the MA window
- Most recent observation receives the highest weight, the weights decrease for older data values
- The sum of the weights = 1

Weighted Moving Average

Year	Attendance (000)	Weighted Moving Average	Found by
1993	5,761		
1994	6,148	6,388	.2(5,761) + .3(6,148) + .5(6,783)
1995	6,783	6,987	.2(6,148) + .3(6,783) + .5(7,445)
1996	7,445	7,293	.2(6,783) + .3(7,445) + .5(7,405)
1997	7,405	9,436	.2(7,445) + .3(7,405) + .5(11,450)
1998	11,450	10,528	.2(7,405) + .3(11,450) + .5(11,224)
1999	11,224	11,509	.2(11,450) + .3(11,224) + .5(11,703)
2000	11,703	11,701	.2(11,224) + .3(11,703) + .5(11,890)
2001	11,890	12,098	.2(11,703) + .3(11,890) + .5(12,380)
2002	12,380	12,183	.2(11,890) + .3(12,380) + .5(12,181)
2003	12,181	12,409	.2(12,380) + .3(12,181) + .5(12,557)
2004	12,557		

Trend estimation using least squares fitting

- A Linear time trend model : long term trend of many time series may often be approximated by a straight line

$$y_t = \alpha + \beta t$$

- Unknown constants estimated using least squares

Beyond Linear Trend

- Quadratic and higher order trends can also be fitted using OLS

- Degree of the polynomial trend line may be arrived at using statistical hypothesis testing approach

- Nonlinear trend lines may also be appropriate and can be fitted using least squares

Seasonal Component Estimation

- Estimation depends on whether trend is present or not.
- In case trend is present, we first estimate trend (slow or fast changing) and then seasonal factors are extracted from the de-trended times series.
- In the absence of trend, seasonal factors from monthly or quarterly data are extracted directly from the observed data.
- Estimation of the seasonal factors would depend on the assumed additive or multiplicative component model

Probabilistic Formulation of Time Series

- Time series is an example of a stochastic process (collection of random variables)

- Stochastic process is an indexed collection of random variables

$$\left\{X_t : t \in T\right\}$$

Probabilistic Formulation of Time Series

$$X(\omega, t)$$

belongs to sample space

In time index set

- For a fixed t, $X(\omega, t)$ *is a random variable*
- *For a given* ω, $X(\omega, t)$ *is a sample function or a realization as a function of* t.

- We are concerned with processes indexed by time, either continuous time or discrete time processes

$$\{X_t : t \in (-\infty, \infty)\} = \{X_t : -\infty < t < \infty\}$$

or

$$\{X_t : t \in \{0,1,2,\ldots\}\} = \{X_0, X_1, X_2, \ldots\}$$

Process Specification

- To describe a stochastic process completely, we must specify all the finite dimensional distributions, i.e. the joint distribution of the random variables for any finite set of time points

$$\left\{X_{t_1}, X_{t_2}, X_{t_3}, \ldots. X_{t_n}\right\}$$

- e.g. A Gaussian process

Process Specification

- A simpler approach is to only specify the moments of the underlying random process
- The mean and variance functions are given by

$$\mu_t = E(X_t) \text{ and } \sigma_t^2 = V(X_t)$$

Auto Covariance Function

- Because the random variables comprising the system are usually not independent (they are observed sequentially over time), we must also specify their covariance to extract the time dependence structure
- ACF $\gamma_{t_1,t_2} = Cov\left(X_{t_1}, X_{t_2}\right)$

Forms of Stationarity

- Strict stationarity: joint distribution function of any collection of random variables does not change with location shift

- Weak stationarity or covariance stationarity: mean and autocovariance functions do not depend on time points of reference

Stochastic nature of Time Series: Stationary Vs. Non-Stationary

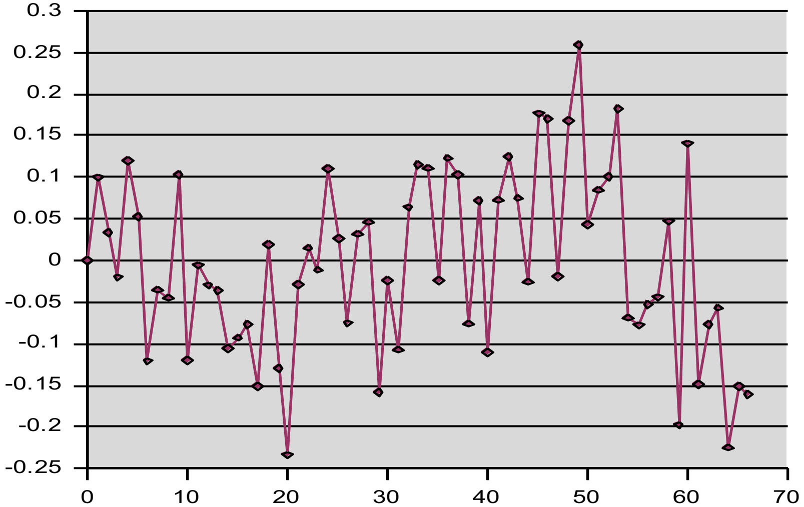
STATIONARITY

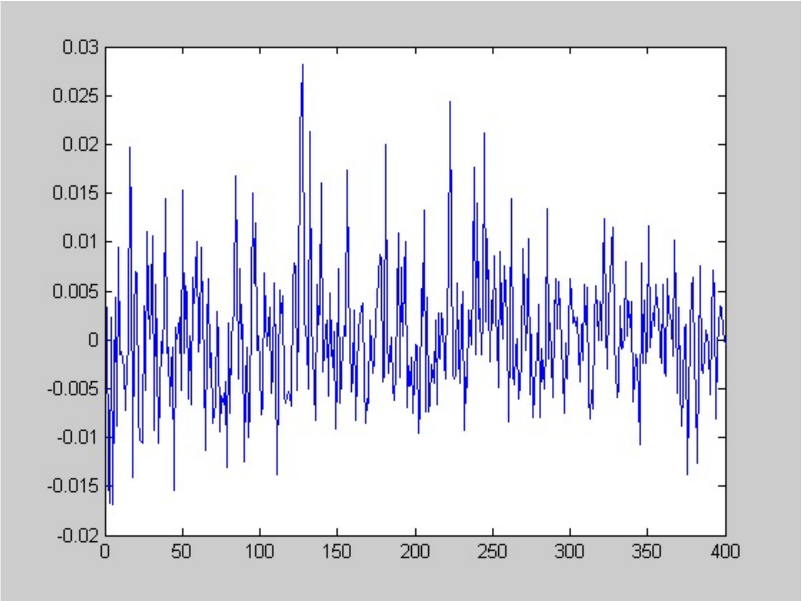
- No change in means, no systematic change in variability
- Process in a state of statistical equilibrium

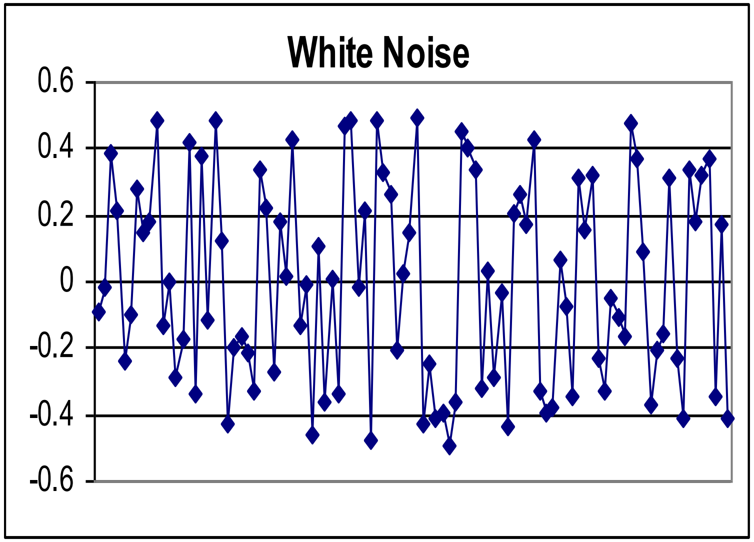
- Analysis and modeling requires the time series to be stationary (through transformation, if required)

- Co-integration analysis uses non-stationarity of a particular nature for meaningful analysis

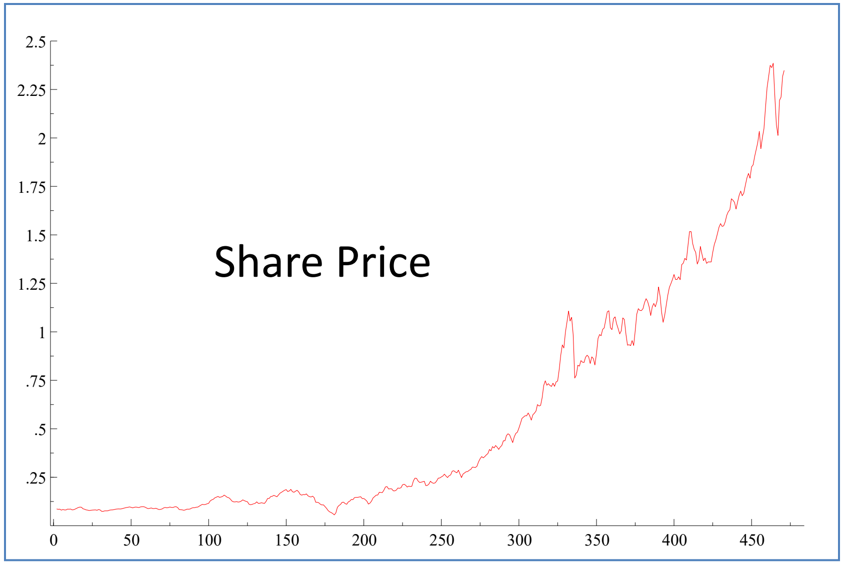
Examples of stationary time series

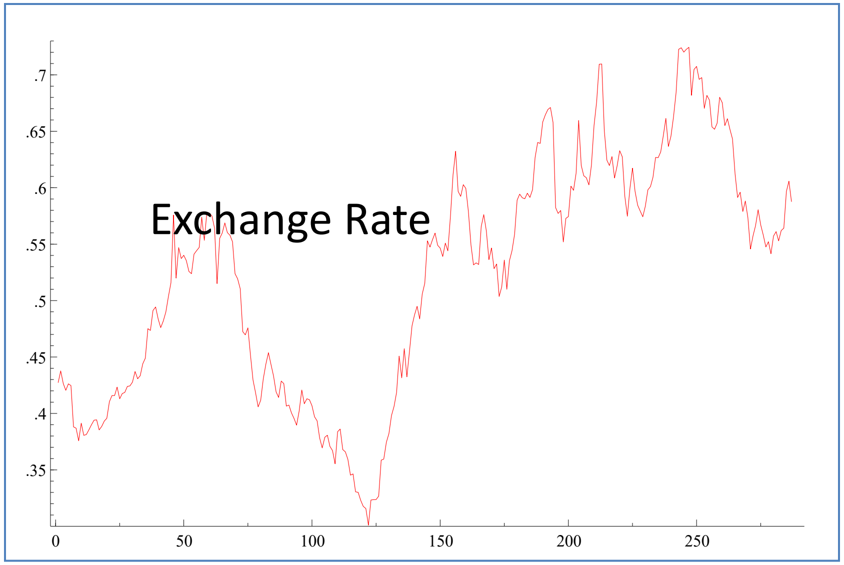


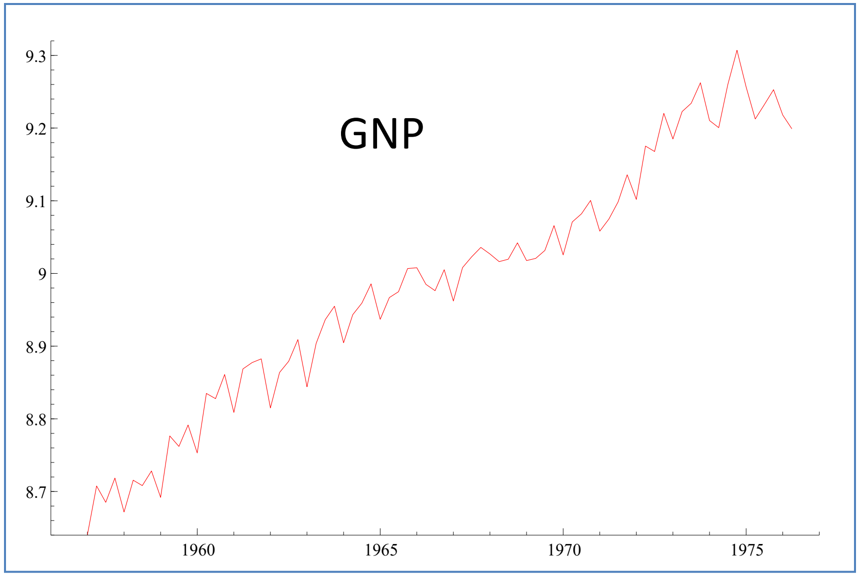




Some non-stationary time series







APPROACHES OF TIME SERIES ANALYSIS

- TIME DOMAIN ANALYSIS
- FREQUENCY DOMAIN ANALYSIS

TIME DOMAIN ANALYSIS

- Study the covariance/correlation structure between observations separated by a fixed unit of time
- Use the above information to build models of the form

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots\ldots + \phi_p Y_{t-p}$$
$$+ \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots\ldots + \theta_q \varepsilon_{t-q}$$

- Auto correlation function (ACF) plays a central role in time domain analysis

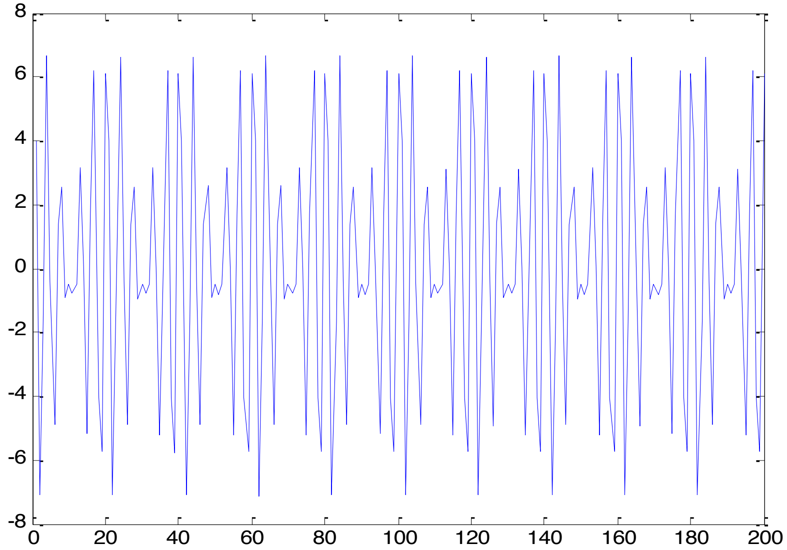
TIME DOMAIN ANALYSIS

- Use the Auto Correlation structure for building forecasting models based on the historical data that can be used for forecasting future path of the observed time series.
- Parameter estimation, model selection, model validation and residual diagnostics are important concepts under this approach.

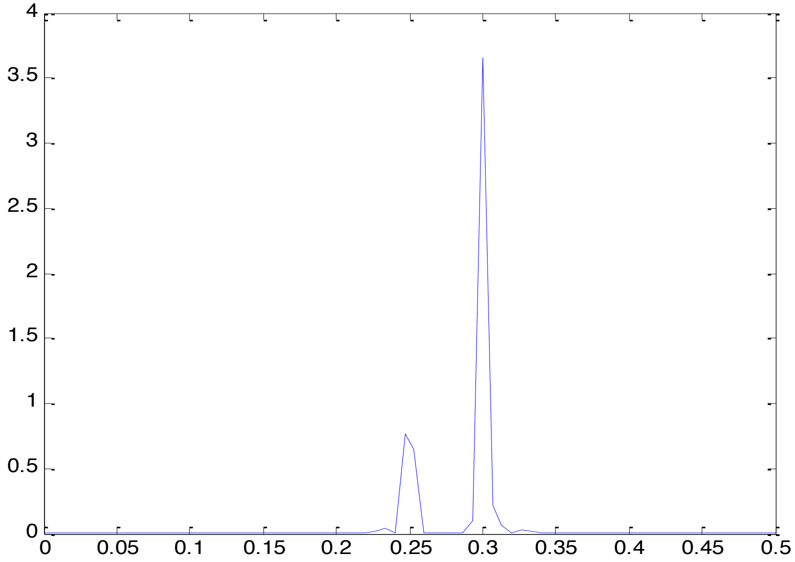
FREQUENCY DOMAIN ANALYSIS

- Extract hidden periodicities in the time series
- Spectral Density Function plays a central role
- Study the structure of spectral density function of standard probability models (WN, AR, MA, ARMA); estimation of spectral density function and periodogram analysis are important concepts.

FREQUENCY DOMAIN ANALYSIS



Time series

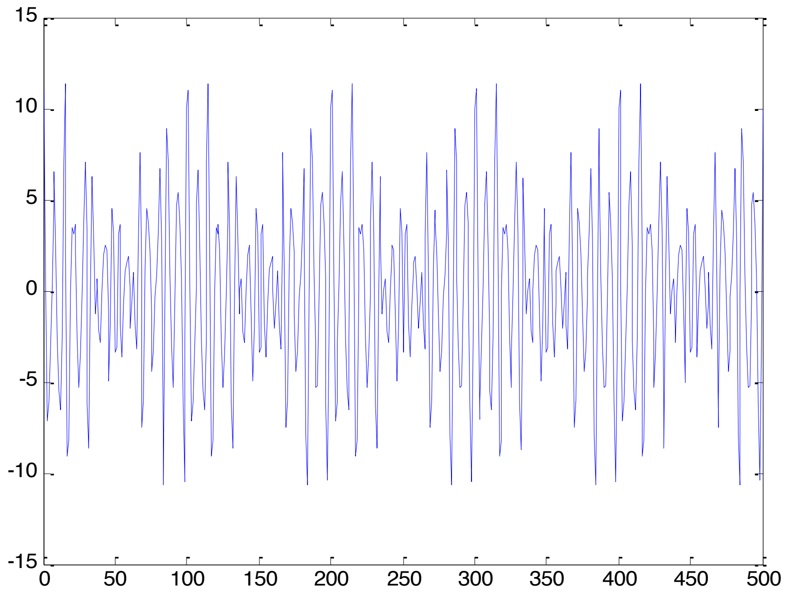


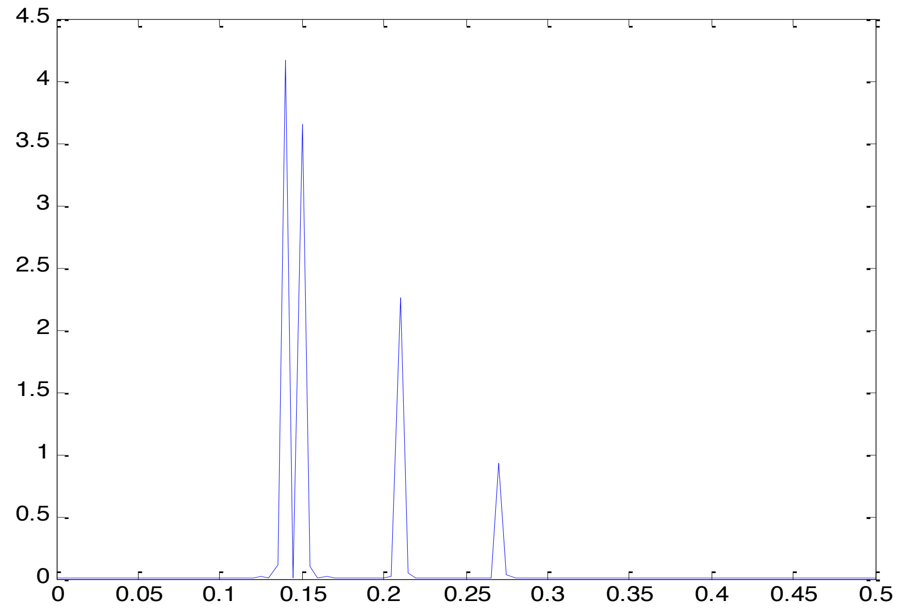
Periodogram

- 2-component sinusoidal model: frequencies at 0.25 & 0.30

$$y(t)=\sum_{k=1}^{2}\left(\alpha_k^0\, Cos\left(2\pi\omega_k^0 t\right)+\beta_k^0\, Sin\left(2\pi\omega_k^0 t\right)\right)+\varepsilon(t); \quad t=1,\ldots,n$$

FREQUENCY DOMAIN ANALYSIS





Time series

Periodogram

- 4-component sinusoidal model: frequencies at 0.14, 0.15, 0.21 & 0.27

FREQUENCY DOMAIN ANALYSIS

- Time domain analysis and frequency domain analysis are not mutually exclusive

- Any covariance stationary process has both time domain as well as frequency domain representation
- Inter relationship between ACF and spectral density function

STATISTICAL MODELS OF TIME SERIES

LINEAR TIME SERIES MODELS

- AUTO REGRESSIVE (AR)
- MOVING AVERAGE (MA)
- AUTO REGRESSIVE MOVING AVERAGE (ARMA)
- AUTO REGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA)
- SEASONAL ARMA (S-ARMA)
- SEASONAL ARIMA (S-ARIMA)
- ARMA EXOGENOUS (ARMAX)

MULTIVARIATE VERSION OF THE ABOVE MODELS

- VAR, VMA………

NONLINEAR MODELS

- THRESHOLD AR (TAR)
- SELF EXCITED TAR (SETAR)
- AUTOREGRESSIVE CONDITIONAL HETEROSKADASTIC (ARCH)
- GENERALIZED ARCH (GARCH)

- SUPERIMPOSED SINUSOIDAL MODEL

- ARTIFICIAL NEURAL NETWORK (ANN) MODELS

Box-Jenkins Approach of Time Series Modelling

- The approach involves the following steps:
  - model identification
  - model fitting
  - diagnostic checking

Forecasting Guidelines

- Simplest (least number of parameters) model to describe the data
  - principle of parsimony

- No pattern in forecast error
  - e_i = (Actual Y_i - Forecast Y_i)
  - Seen in plots of errors over time

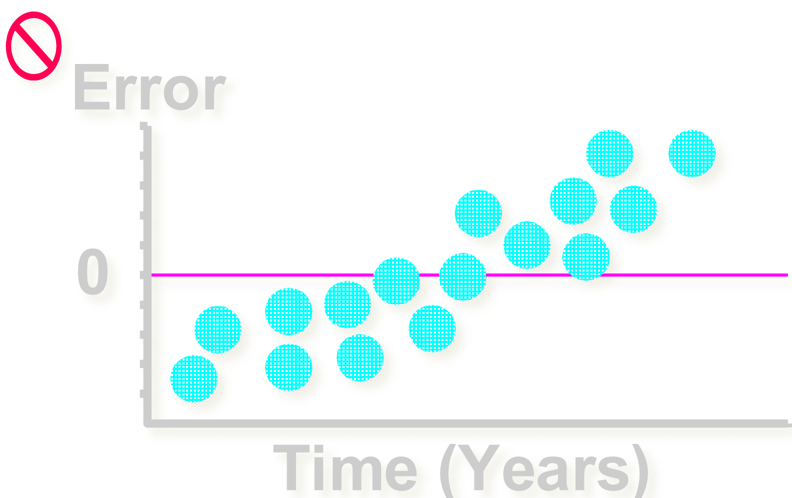
Principal of Parsimony

- Suppose two or more models provide “good fit” for data

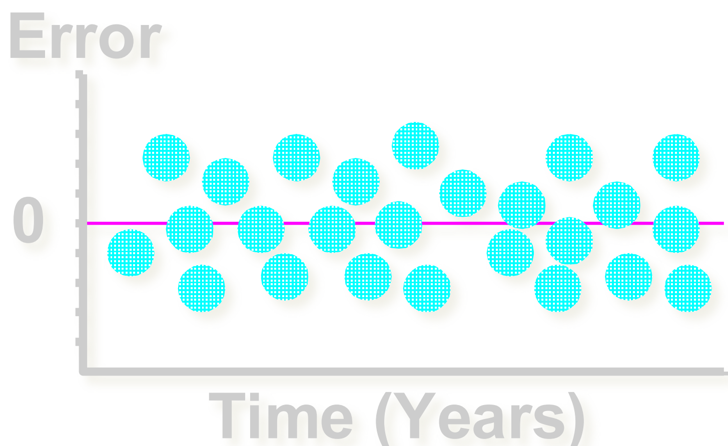
- Select the Simplest Model-less number of model parameters

Pattern of Forecast Error

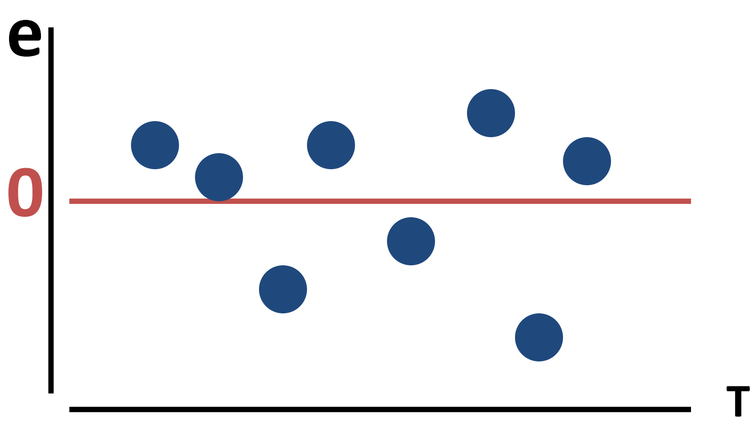
Trend Not Fully Accounted for



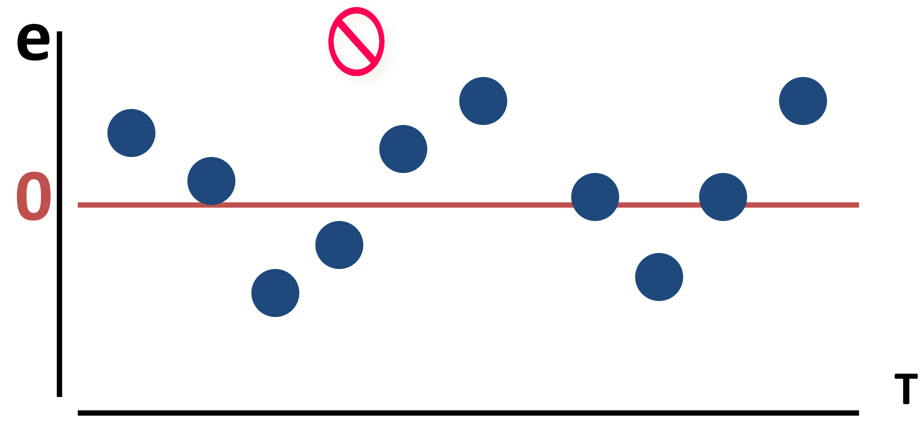
Desired Pattern



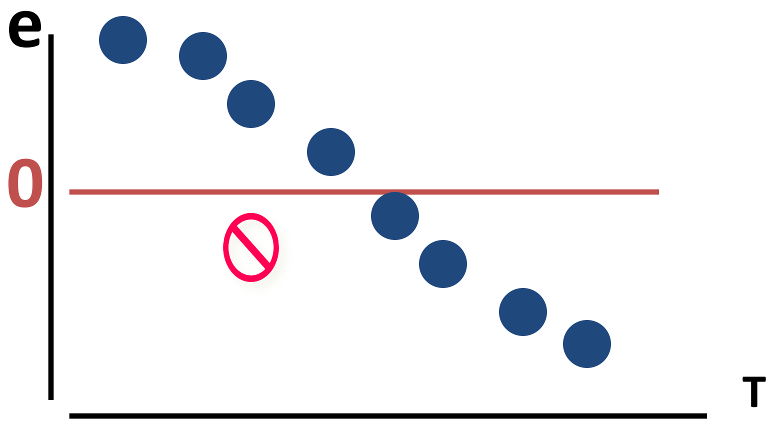
Residual Analysis



Random errors

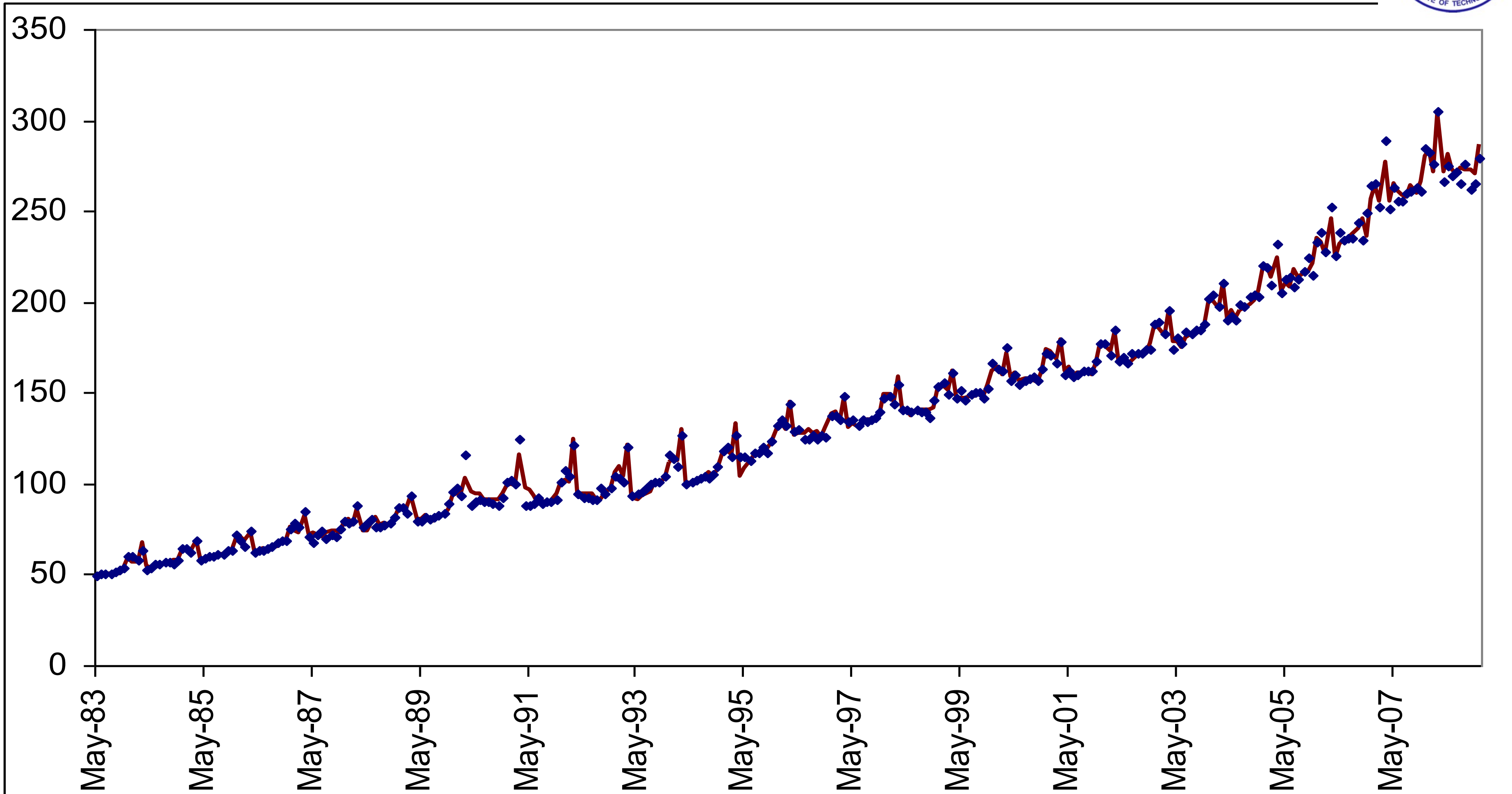


periodic effects not accounted for



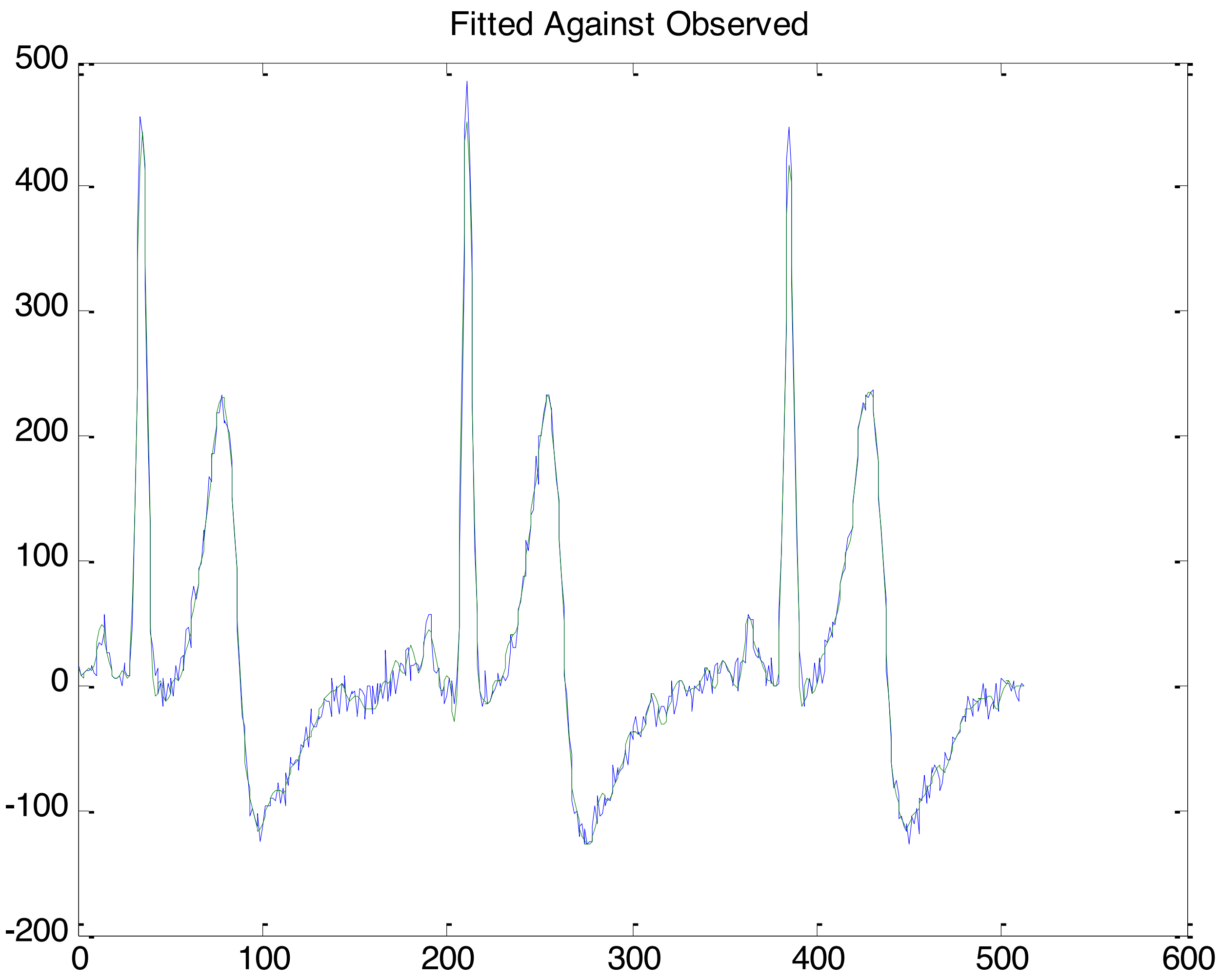
Trend not accounted for

MODELING OF TIME SERIES



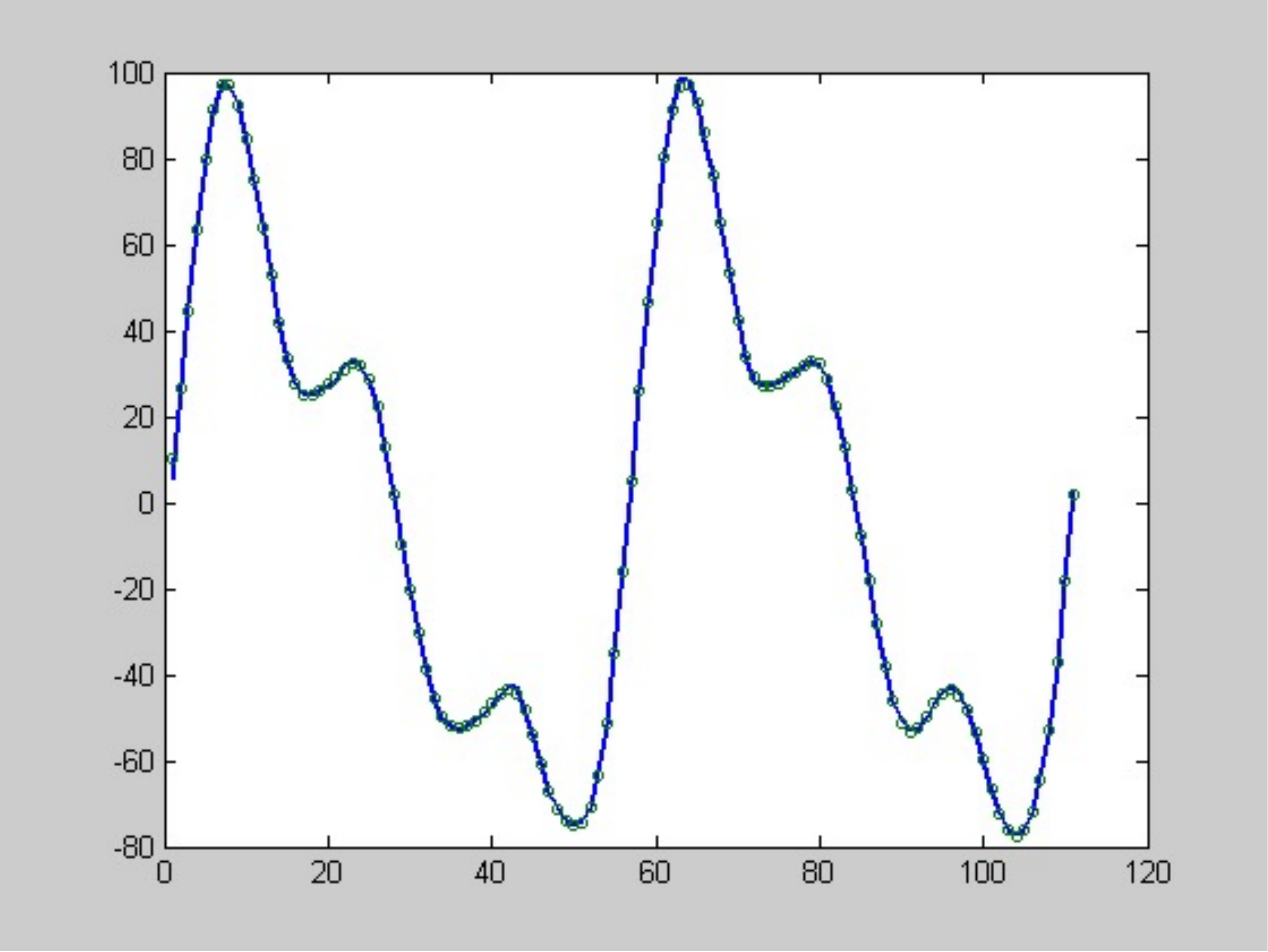
- Actual monthly index for industrial production (IIP) and fitted seasonal-arima model

FITTING OF ECG SIGNAL



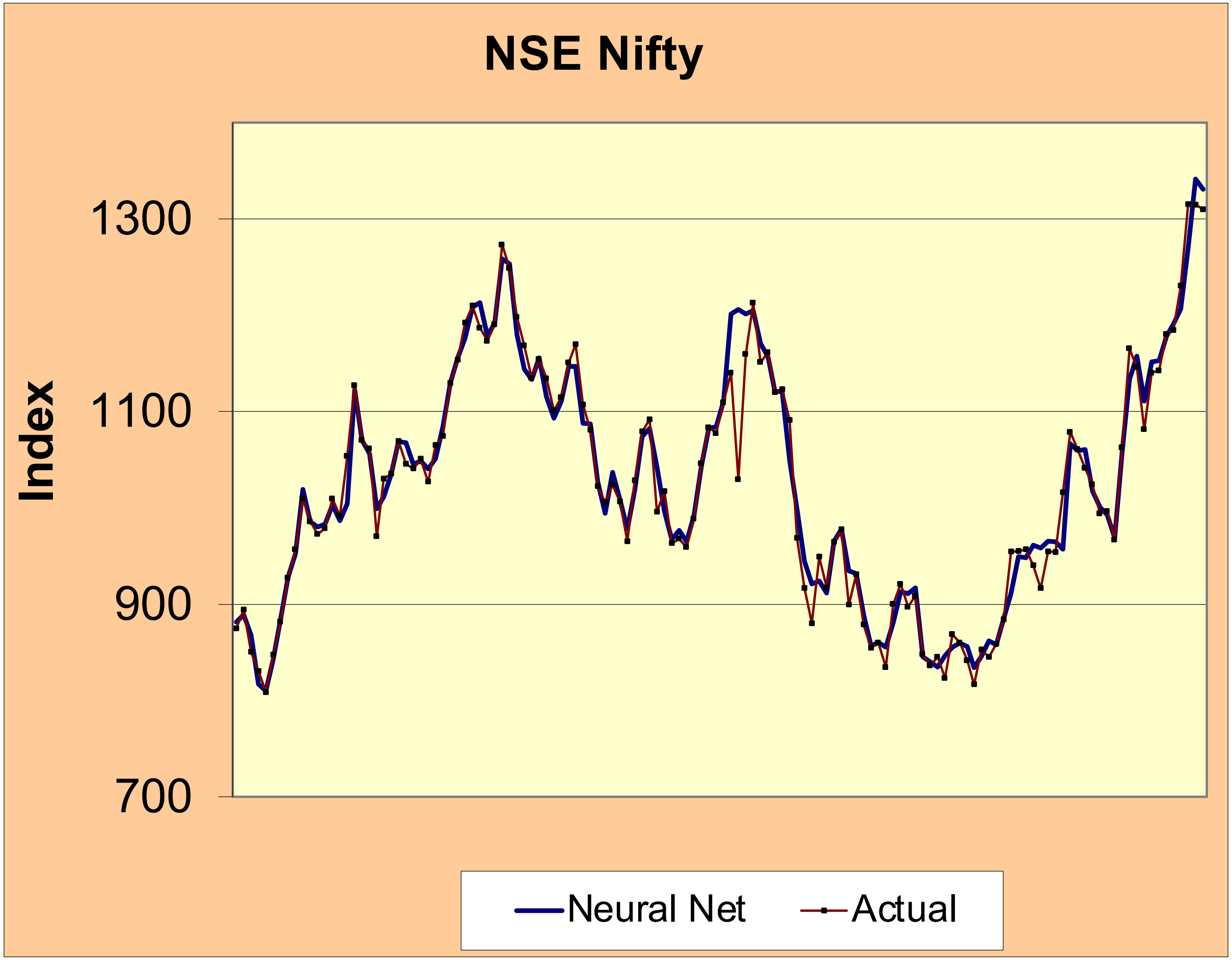
- A MULTIPLE SINUSOIDAL MODEL FIT OF ECG DATA

MODELING OF SPEECH SIGNAL DATA

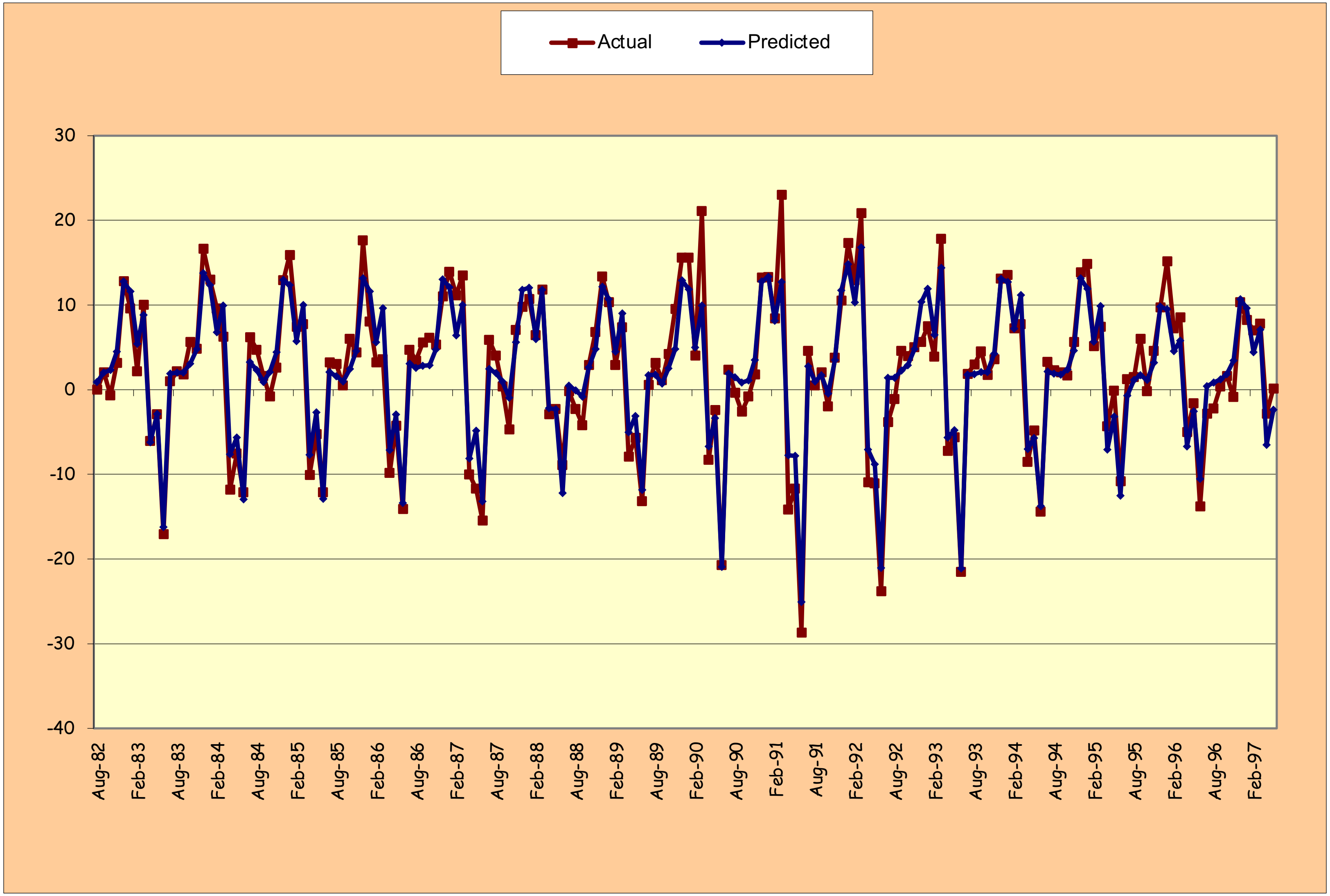


MLE BASED MULTIPLE SINUSOIDAL MODEL FIT

Neuro-Genetic forecasting model of NIFTY



Neuro-Genetic forecasting model of IIP (growth rate)



Some Standard Text Books

- Introduction to Time Series and Forecasting- P.J.Brockwell & R.A.Davies
- Time Series: Theory & Methods- P.J.Brockwell & R.A.Davies
- Time Series Analysis- J.D. Hamilton
- Introduction to Statistical Time Series – Wayne A. Fuller

Thank You & Best Wishes

Amit Mitra