Chapter 1. Characteristics of Time Series

August 25, 2018

Outline

Introduction

- We will be using **Time Series Analysis and Its Applications: With R Examples** by Shumway and Stoffer (SaS) as the textbook. Go to http://www.stat.pitt.edu/stoffer/tsa4/index.html for more details.

- Useful references:

  - *Time Series Analysis with Applications in R* by Cryer and Chan (CaC). Go to http://homepage.divms.uiowa.edu/ kchan/TSA.html for more details.

  - *Time Series: Theory and Methods* by Brockwell and Davis

  - *Time Series Analysis* by William W.S. Wei

R

- **R** is a software for statistical computing and graphics.
- **Appendices of SaS and CaC** give an introduction to R.
- To install R:
  1. Go to http://www.r-project.org.
  2. Click on CRAN on the left panel.
  3. Scroll down the list of CRAN Mirror sites, and click on one of them nearest to you geographically.
  4. Click on the link for Windows (or Linux or Mac OS as appropriate) and click on the link named 'base'.
- Install the `astsa` package for SaS. All R scripts are here: http://www.stat.pitt.edu/stoffer/tsa4/index.html
- Install the `TSA` package that contains all datasets needed to redo all the analyses in CaC.
- All R scripts from `TSA` package are here: http://homepage.stat.uiowa.edu/~kchan/TSA.htm

Introduction

- A **time series** is a sequence of data points, measured typically at uniformly spaced times.

- Some Applications:
  - In business, the daily closing price of a stock or quarterly economic data (say, GDP, CPI, unemployment rate, etc.)
  - In meteorology, annual precipitation and drought indices
  - In the biological sciences, the electrical activity of the heart at millisecond intervals
  - In ecology, the abundance of an animal species

- Big data
  - Large-scale sensor networks (e.g., complex time-stamped events such as web-click logs)
  - Web and social networks (e.g., time-evolving graph)
  - Medical and healthcare records

Monthly Average Temperatures in Dubuque, Iowa (CaC)

- The average monthly temperatures (in degrees Fahrenheit) over a number of years ($n = 144$) recorded in Dubuque, Iowa, are shown in `Exhibit 1.7`. This time series displays a very regular pattern called **seasonality**.

- Seasonality for monthly values occurs when observations twelve months apart are related in some manner or another.

- All Januarys and Februarys are quite cold but they are similar in value and different from the temperatures of the warmer months of June, July, and August, for example. There is still variation among the January values and variation among the June values.

Johnson and Johnson Quarterly Earnings

- The data ($n = 84$) are quarterly earnings per share for the U.S. company Johnson & Johnson.
- There are 84 quarters (21 years) measured from the first quarter of 1960 to the last quarter of 1980.
- Note the gradually increasing underlying trend and the regular variation superimposed on the trend.

Global Warming

- The data ($n = 142$) are the global mean land-ocean temperature index from 1880 to 2009, with the base period 1951-1980. In particular, the data are deviations, measured in degrees centigrade.

- We note an apparent upward trend in the series during the latter part of the twentieth century that has been used as an argument for the global warming hypothesis.

- The question of interest for global warming proponents and opponents is whether the overall trend is natural or whether it is caused by some human-induced interface. It is a more of a question of trend rather than particular periodicities.

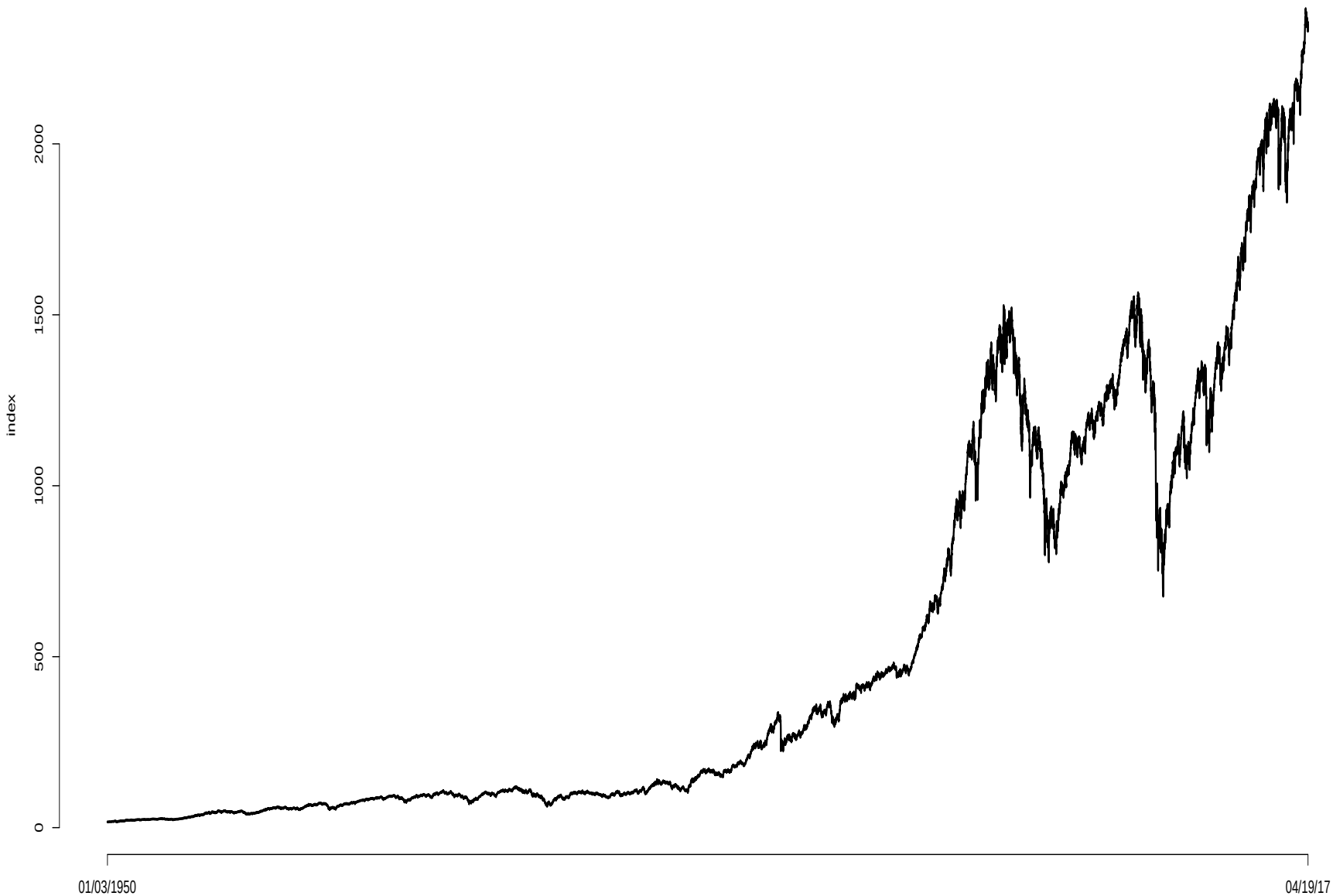
El Ni$\widetilde{n}$o and Fish Population

- Monthly values of an environmental series called the Southern Oscillation Index (SOI) and associated Recruitment (number of new fish).
- Both series are for a period of $n = 453$ months ranging over the years 1950-1987. The SOI measures changes in air pressure, related to sea surface temperatures in the Central Pacific Ocean. The central Pacific warms every three to seven years due to the El N$i\widetilde{n}$o, which has been blamed, in particular, for the 1997 floods in the midwestern portions of the United States.
- Both series tend to exhibit repetitive behavior, with regularly repeating cycles. This periodic behavior is of interest may be regular and the rate or frequency of oscillation characterizing the behavior of the underlying series would help to identify them.

S&P 500 Index

- The S&P 500 Index was launched in March 1957.
- It was the first index to be published daily.
- It contains 500 of the largest stocks in the United States.
- It is a benchmark for gauging the overall health of the large American companies and the U.S. stock market.
- More than $7.8 trillion is benchmarked to the index. (Source: *Investopedia*).
- Data Source: `finance.yahoo.com` and many others.

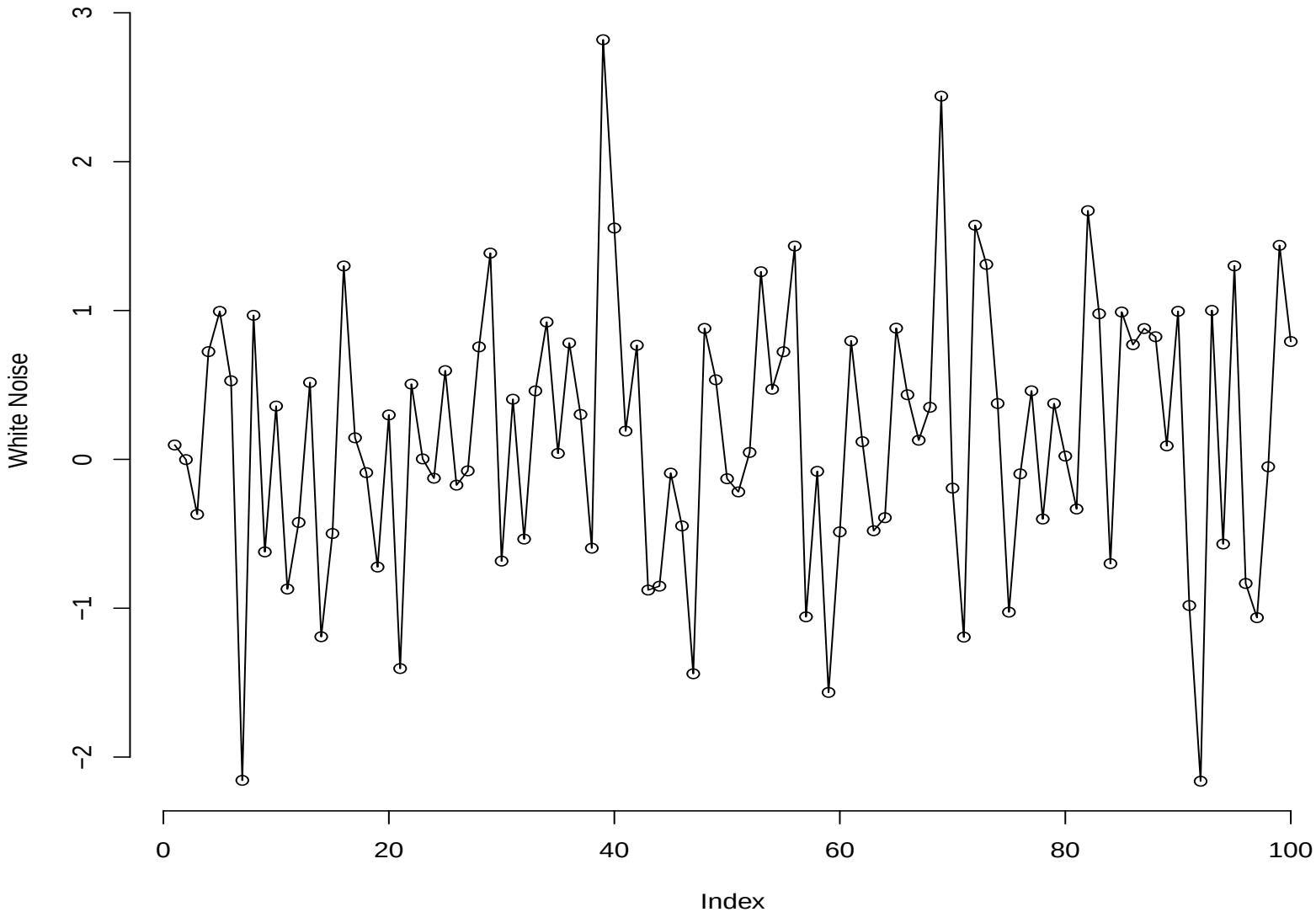
S&P 500's 16,932 daily data: 1950-2017



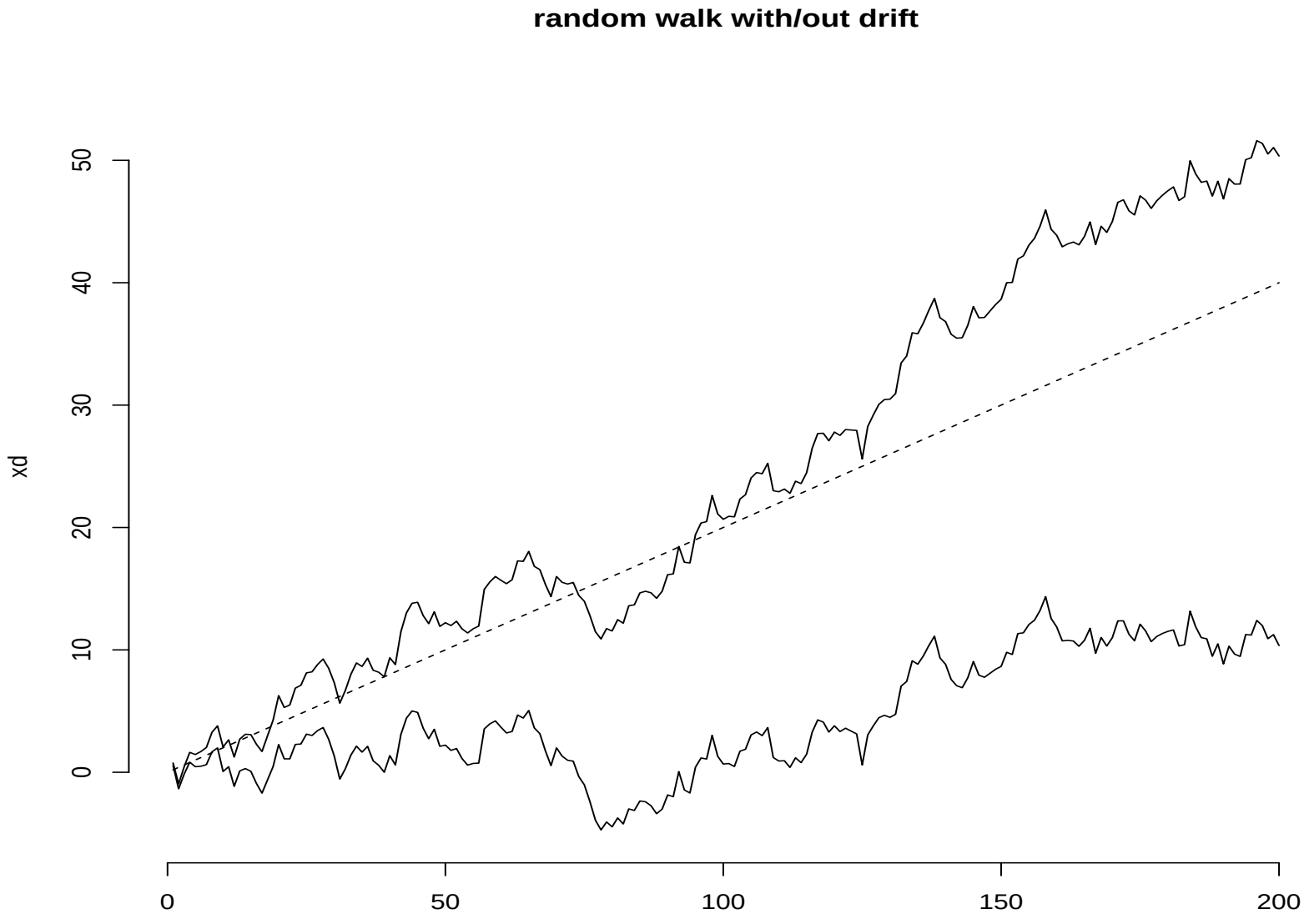
White Noise

- A simple kind of generated series might be a collection of *uncorrelated* random variables, w_t, with constant mean (usually zero) and variance σ_w^2.
- We may also require the noise to be *independent and identically distributed* (**iid**) random variables with constant mean (usually zero) and variance σ_w^2.
- For now, we assume the second kind as **white noise** with mean zero.
- A particularly useful white noise series is Gaussian white noise, wherein w_t's are *independent Gaussian/normal* random variables, mean 0 and variance σ_w^2.
- Often times, we denote $w_t \overset{iid}{=} N(0, \sigma_w^2)$.

White Noise

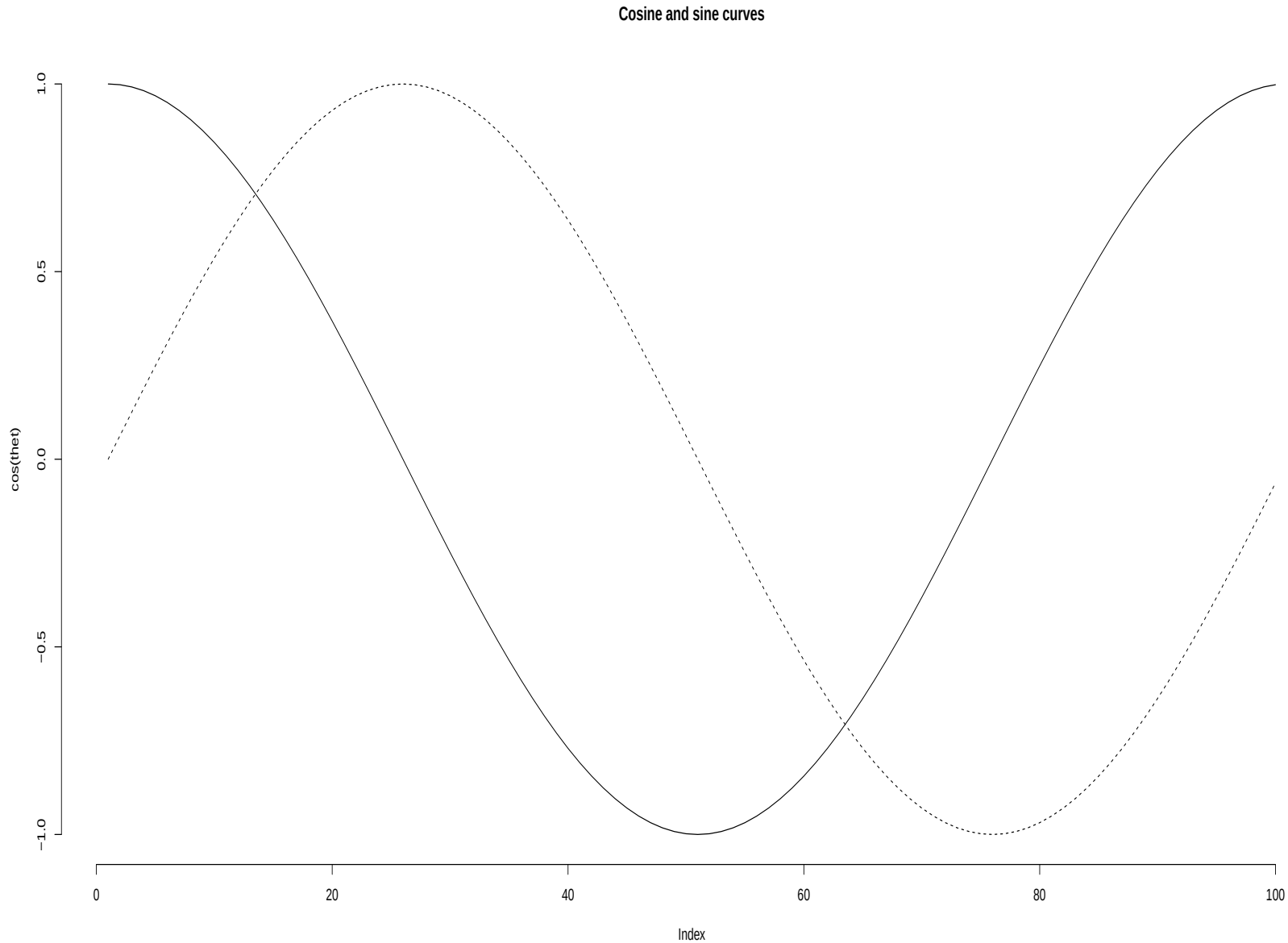


Random Walk



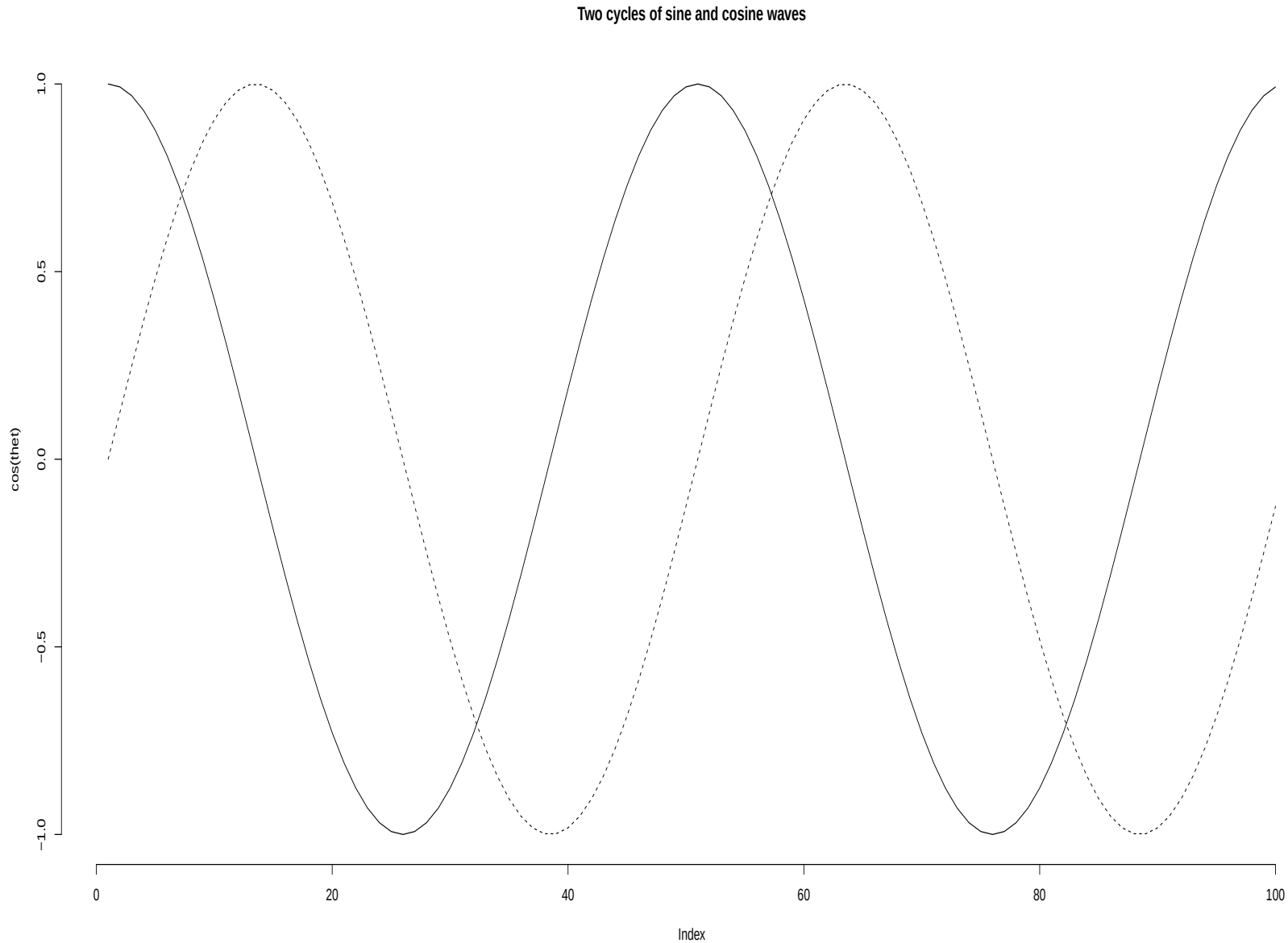
Sine and Cosine Curves

- $\cos(2\pi t/100)$ and $\sin(2\pi t/100)$



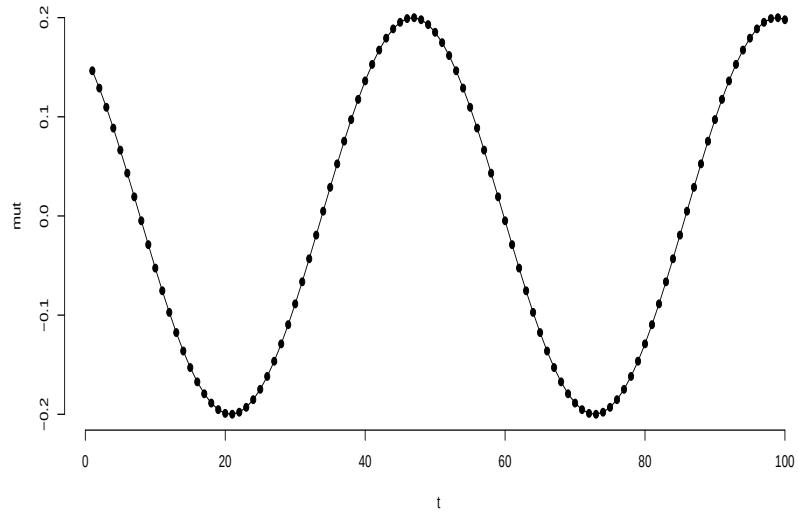
Sine and Cosine Curves

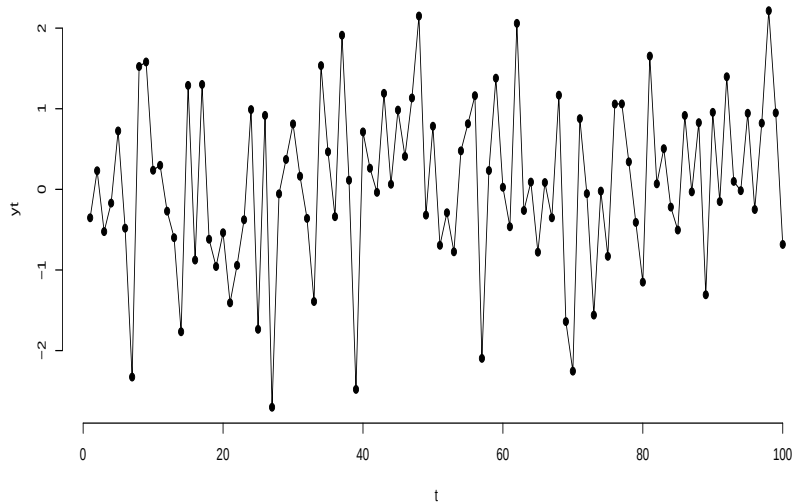
- $\cos(2\pi t/50)$ and $\sin(2\pi t/50)$

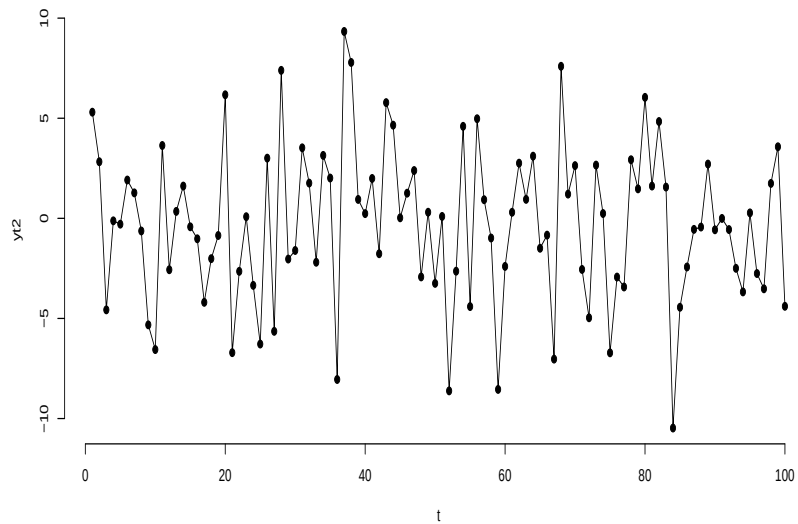


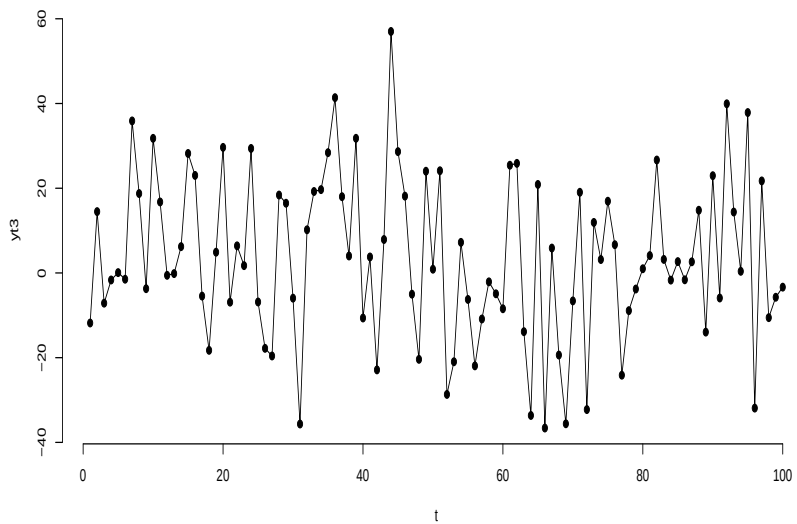
Sine and Cosine Curves

- $0.2\sin(2\pi t/52 + .7\pi) + \mathit{NOISE}$



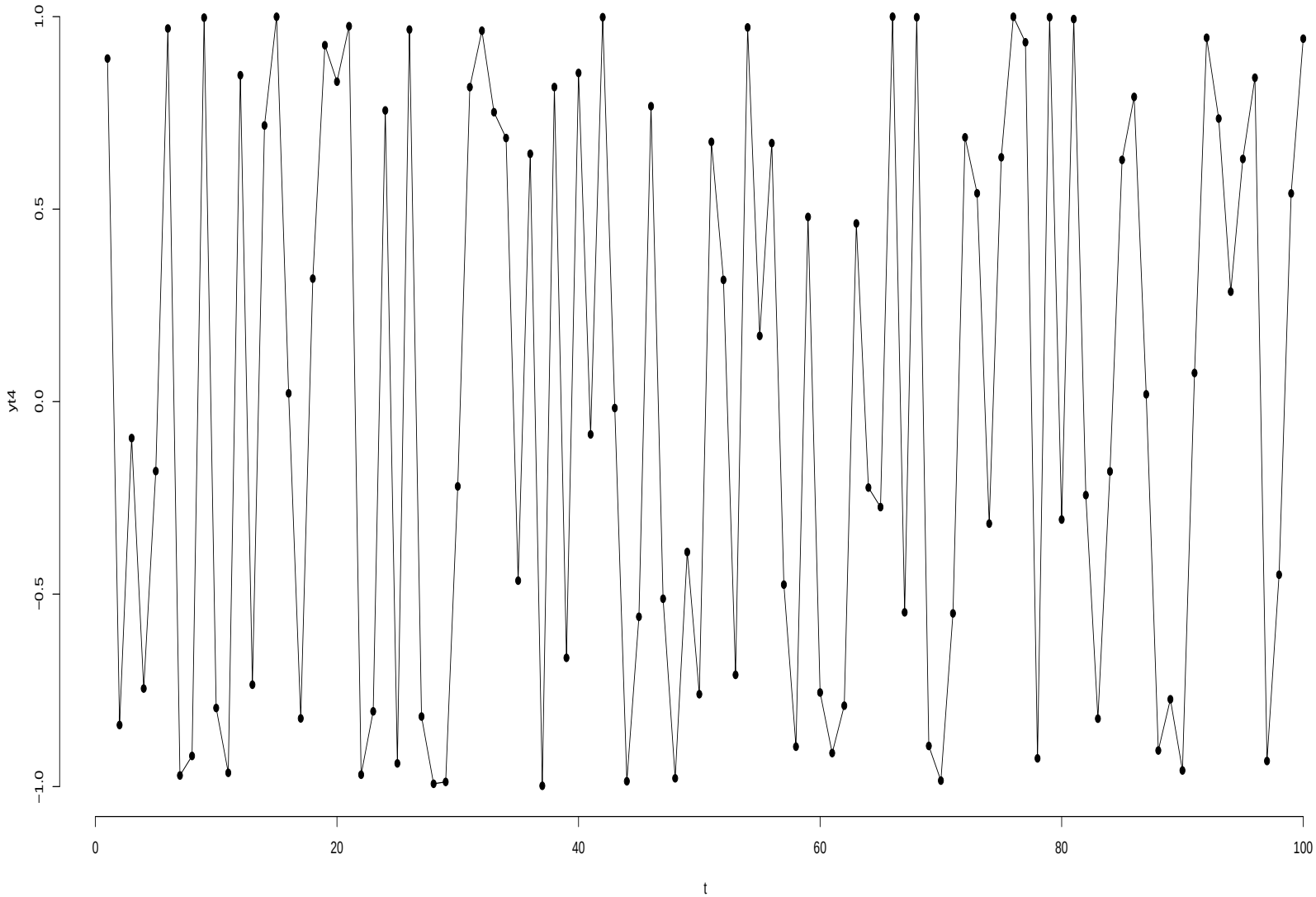






Sine and Cosine Curves

- **Random Cosine Curve**: $\cos(2\pi(t/12 + U(0,1)))$



Objectives and Approaches

- Compact description of data.
- Seasonal adjustment
- Trend detection
- Forecasting of future observations
- Time domain vs Frequency domain

Strategy

- **Model identification:**

  Choose appropriate class of time series models for a given observed series. Explore the data and apply any subject matter knowledge.

- **Model fitting:**

  Model fitting consists of finding the best possible estimates/estimators (e.g., least squares and maximum likelihood) of the unknown parameters of a model.

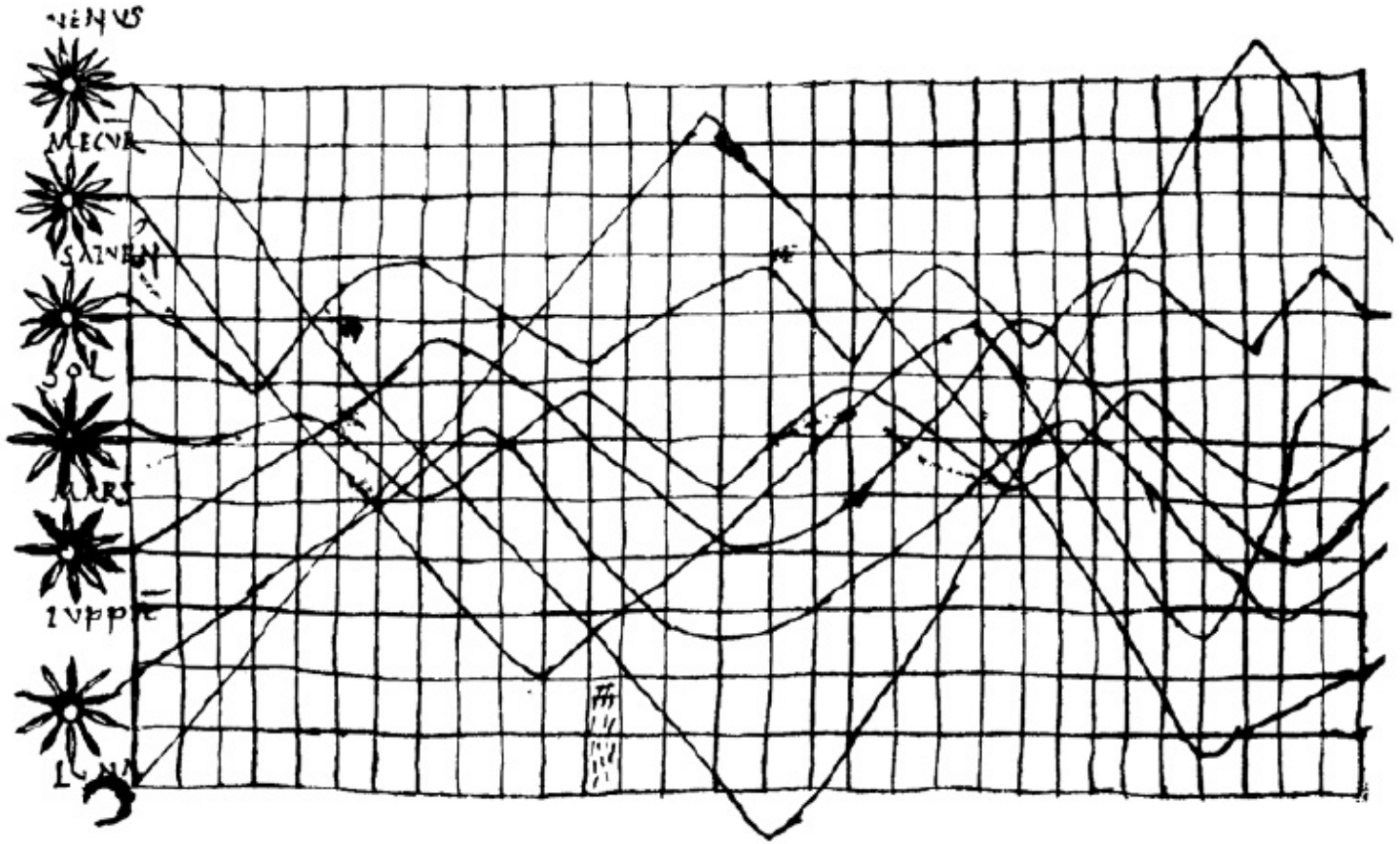
- **Model diagnostics:**

  How well does the model fit the data? Are the assumptions of the model reasonably well met? If no issues are found then the model may be used for its intended purpose (e.g., forecasting).

Oldest Time Series Plot (CaC)

- Exhibit 1.10 reproduces the oldest known example of a time series plot, dating from the tenth (or possibly eleventh) century and showing the inclinations of the planetary orbits.

Exhibit 1.10 A Tenth-Century Time Series Plot



Some Useful Formulas

- The **Taylor Series expansion** of an infinitely differentiable function $f(x)$ at a point a is

$$f(x) = \sum_{j=0}^{\infty} \frac{f^{(j)}(a)}{j!} (x-a)^j.$$

- When $a = 0$, it becomes the Maclaurin series.

- **Geometric Series:**
  For $x \neq 1$,

$$\sum_{j=0}^{n-1} a\, x^j = a\, \frac{1-x^n}{1-x}.$$

- If $|x| < 1$,

$$\sum_{j=0}^{\infty} a\, x^j = a\, \frac{1}{1-x}.$$