MSA 8200 Predictive Analytics
Week1: Time Series Introduction

Yichen Cheng

Spring 2020

Predictive Analytics Course Overview

Objectives:

- Time series analysis.
- Real data problems – missingness, outliers, etc.
- Estimation bias – reasons and remedies.
- Undertsand different data types; use approporiate model for analysis.

Predictive Analytics Course Overview

Topics:

- Time series analysis
- Data Prepresessing: missingness, censoring, outliers, etc.
- Fixed/Random Effect Model, endogoneity.
- Count Data Model, Survival analysis, Bayesian Analysis

Outline:

- What is time series data? Why important?
- Examples of time series data.
- First step in analyzing time series data
- Measure of dependence: ACF

Introduction – What is a time series data?

Examples of time series data include:

- Daily IBM stock prices.
- Monthly rainfall.
- Quarterly sales results for Amazon
- Annual Google profits

Anything that is observed sequentially over time is a time series.

Introduction – Why time series anlaysis?

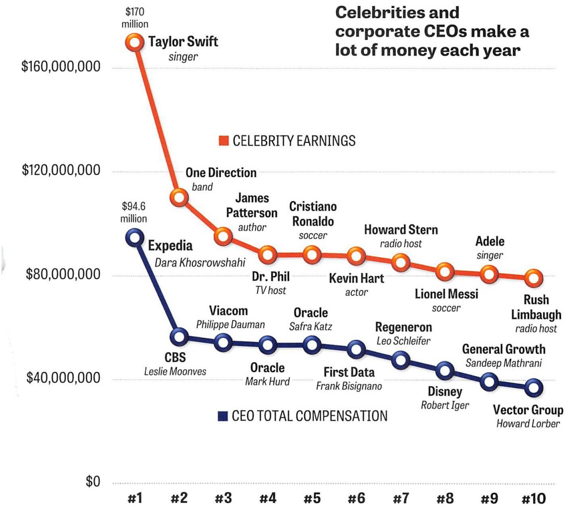
Time series analysis can be applied for various purposes, such as:

- Stock Market Analysis
- Economic Forecasting
- Inventory studies
- Budgetary Analysis
- Yield Projection
- Sales Forecasting

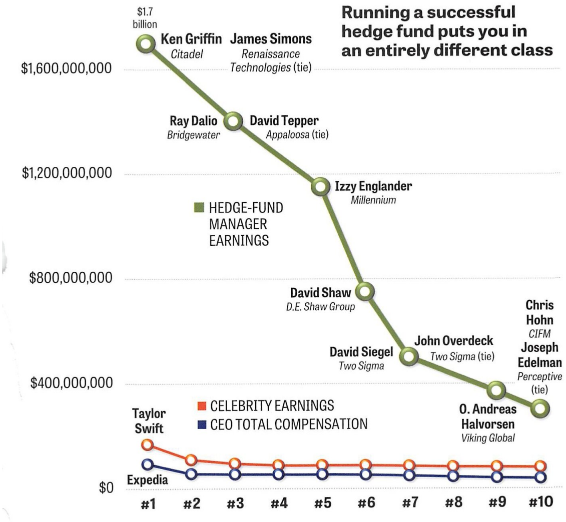
and more.

Introduction – Why time series anlaysis?

https://twitter.com/mebfaber/status/1199091047269187584



https://twitter.com/mebfaber/status/1199091047269187584



Introduction – Why time series anlaysis?

James Simons:

- PhD in Mathematics from UC Berkeley
- Created an investment company called monemetrics: combining "money" and "econometrics".

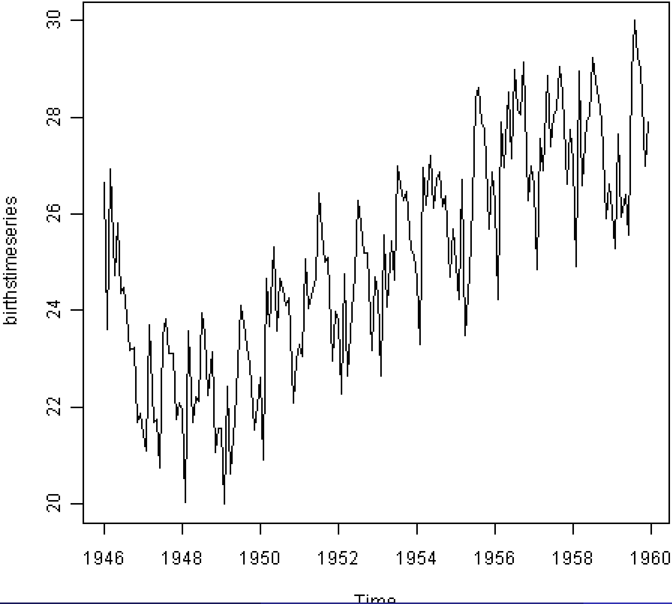
References

- https://www.investopedia.com/financial-edge/0411/the-6-highest-paid-people-on-wall-street.aspx
- https://www.forbes.com/sites/forbesdigitalcovers/2019/11/08/jim-simons-the-man-who-solved-the-market-gregory-zuckerman-book-excerpt/

Time series analysis helps us to capture of the following features of the data:

- Dependence: observations are correlated;
- Trend;
- Seasonality;
- Cyclic;

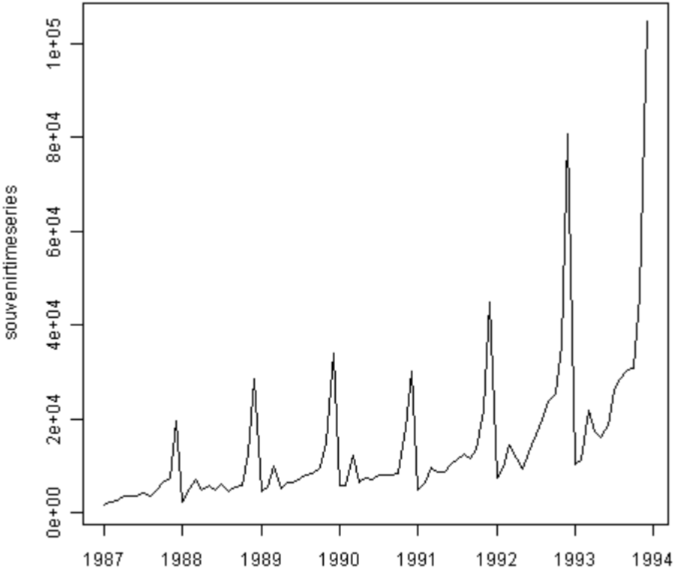
Introduction – example: monthly birth data



Introduction – example: monthly birth data

- Trend
- Seasonality

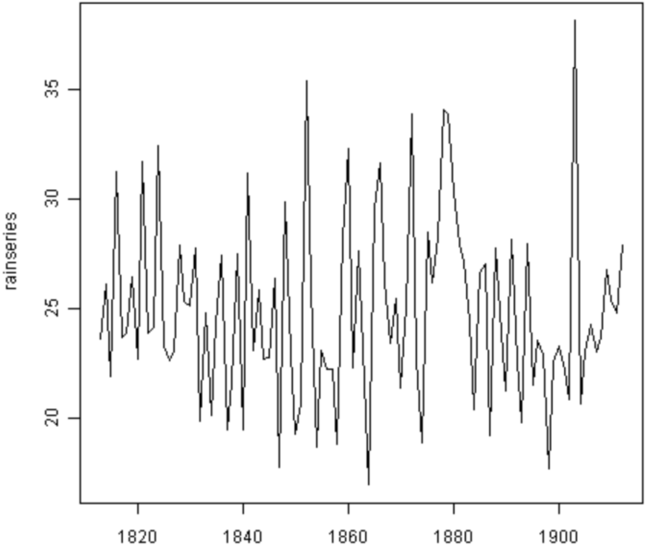
Introduction – example: mothly sales at a souvenir shop



Introduction – example: mothly sales at a souvenir shop

- Trend
- Seasonality
- Increased variation

Introduction – example: annual rainfall in inches for London

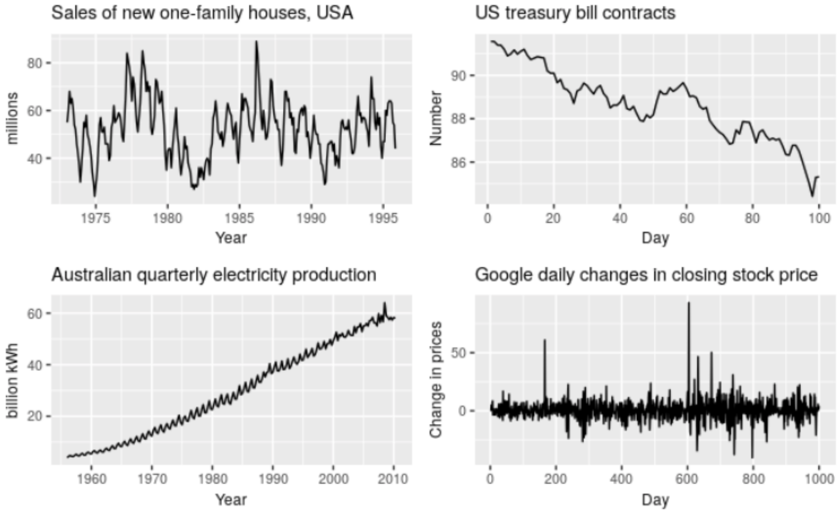


Introduction – example: annual rainfall in inches for London

Key components for a time series model

- Transformation.
- Decomposition:
  - Trend: A long-term increase or decrease. It does not have to be linear.
  - Seasonality:
    Seasonal factors such as the time of the year or the day of the week. Always of a fixed and known frequency
  - Cyclic:
    when the data exhibit rises and falls that are not of a fixed frequency. usually due to economic conditions.

Introduction – More examples



Introduction – Stationarity Time Series and Dependence Measure

Two important concepts for time series data:

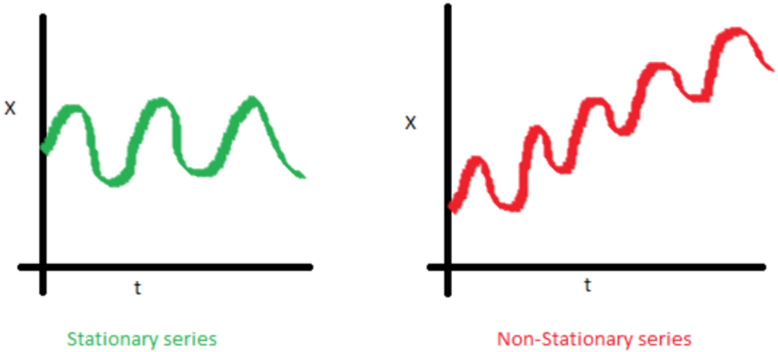
- Stationary time series
- Dependence measure

Introduction – Stationary Time Series and Dependence Measure

A stationary time series:

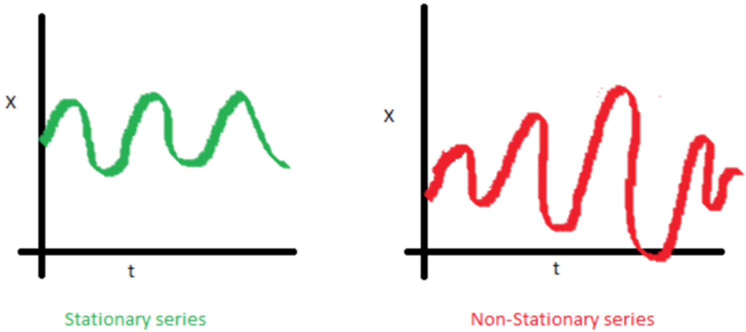
- the mean value function: $E(x_t) = \mu_t$ does not depend on t.
- the autocovariance function: $\gamma(s, t) = cov(x_s, x_t)$ depends on s and t only through $s - t$.

Introduction – Stationarity Time Series and Dependence Measure



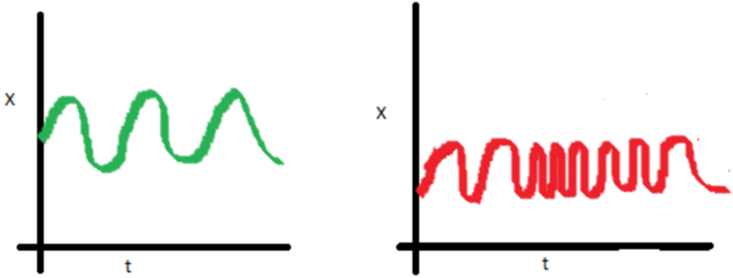
https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/

Introduction – Stationarity Time Series and Dependence Measure



https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/

Introduction – Stationarity Time Series and Dependence Measure



Stationary series

Non-Stationary series

https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/

Introduction – Stationarity Time Series and Dependence Measure

Why stationary?

- Most of the time series models are developed to model a stationary time series.

In case of non-stationary process:

- The first step would be to make a time series stationary.

Introduction – Stationary Time Series and Dependence Measure

Autocovariance function of a stationary time series:

- $\gamma(h) = cov(x_{t+h}, x_t) = E[(x_{t+h} - \mu)(x_t - \mu)]$.

Auto correlation function (ACF) of a stationary time sereis:

$$\rho(h) = \frac{\gamma(t+h,t)}{\sqrt{\gamma(t,t)\gamma(t+h,t+h)}} = \frac{\gamma(h)}{\gamma(0)} \tag{1}$$

Introduction – Stationary Time Series and Dependence Measure

Sample autocovariance function

$$\hat{\gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-h}(x_{t+h} - \bar{x})(x_t - \bar{x})$$

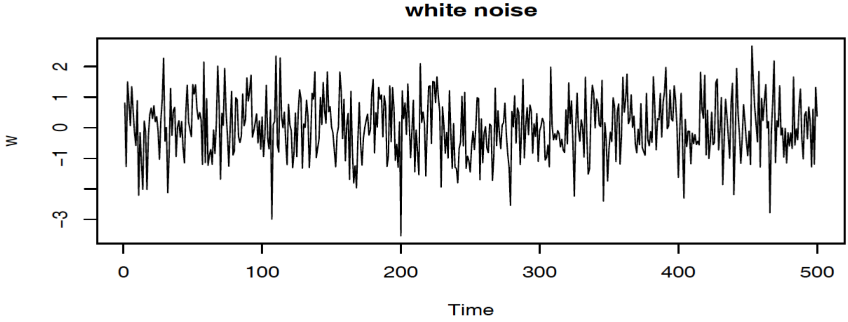
- $\hat{\gamma}(h) = \hat{\gamma}(-h)$ for $h = 0, 1, \ldots, n-1$.
- t goes up to $n - h$ because x_{t+h} not observed for $t + h > n$.
- $\frac{1}{n}$ instead of $\frac{1}{n-h}$ is to ensure the joint autocovariance matrix is non-negative definite.

Sample autocorrelation function (Sample ACF):

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$$

White noise:

- $w_t \overset{iid}{\sim} \mathcal{N}(0, \sigma_w^2)$.
- $E(w_t) = 0,\ Var(w_t) = \sigma_w^2,\ Cov(w_t, w_s) = 0,\ t \neq s.$

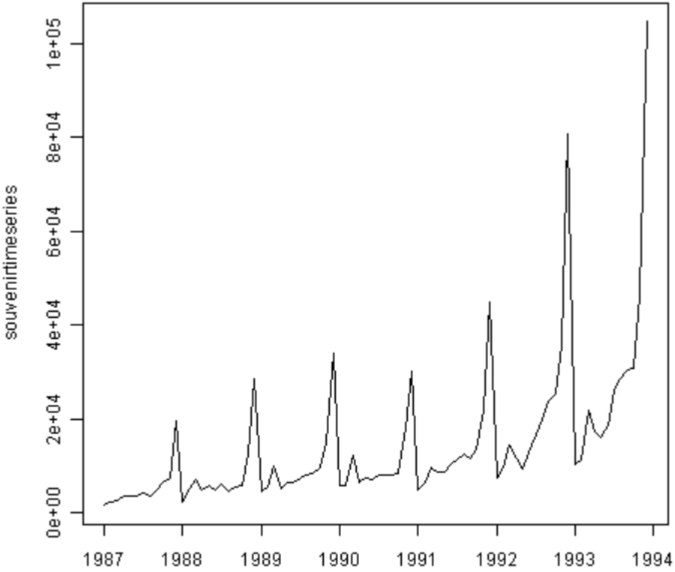


Stationarity of white noise:

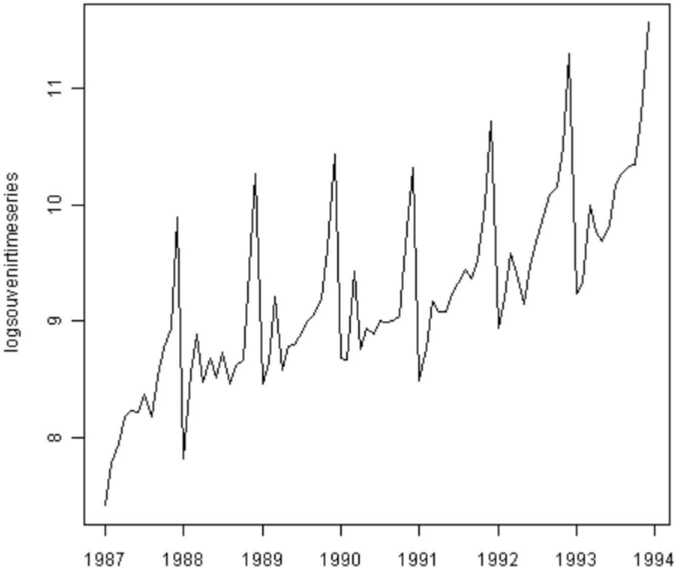
$$x_t = w_t.$$

- $\gamma_w(h) = cov(w_{t+h}, w_t) = \begin{cases} \sigma_w^2 & h = 0 \\ 0 & h \neq 0 \end{cases}$
- $\rho_w(h) = \begin{cases} 1 & h = 0 \\ 0 & h \neq 0 \end{cases}$

Make stationary 1: transformation



Make stationary 1: transformation



Make stationary

After necessary transformation:

- remove trend
- remove seasonality
- remove cyclic patterns

Make stationary 2: – Detrending Analysis

Detrending: two ways.

- Option 1: Regression based.
- Option 2: Differencing.

$$x_t = \mu_t + y_t$$

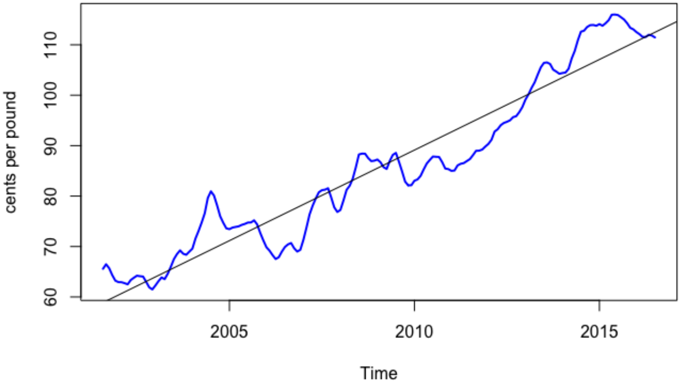
Options 1: $\mu_t = \beta_0 + \beta_1 t$ – regression

Options 2: $\mu_t = \delta + \mu_{t-1} + w_t$ – differencing

Make stationary 2: – Detrending Analysis

Example: Chicken price:

- monthly price (per pound) of a chicken
- mid-2001 to mide-2016 (180 months)



Make stationary 2: – Detrending Analysis

Detrend using regression:

$$x_t = \beta_0 + \beta_1 t + w_t$$

- $\mu_t = \beta_0 + \beta_1 t$ is used to model the trend.
- $\hat{\mu}_t = -7131 + 3.59t$.
- The detrended series is:

$$\hat{y}_t = x_t + 7131 - 3.59t$$

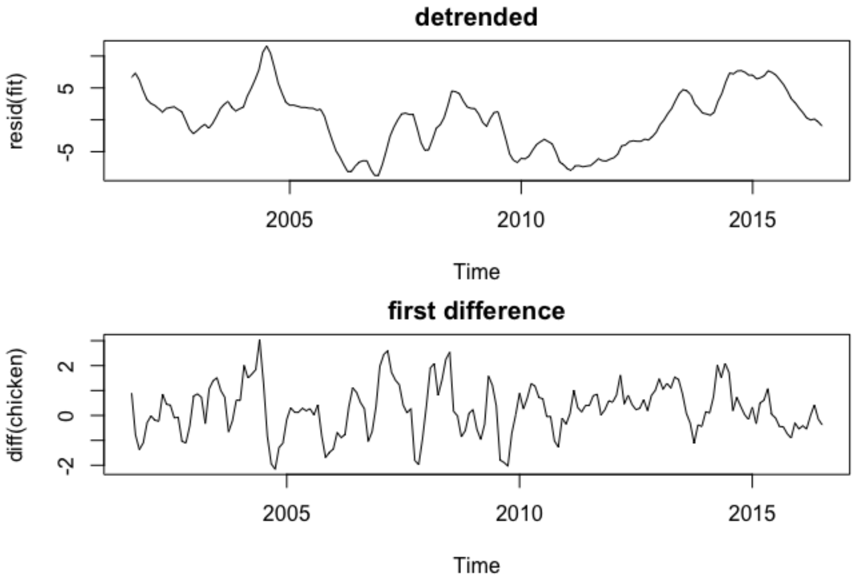
Make stationary 2: – Detrending Analysis

Detrend by differencing:

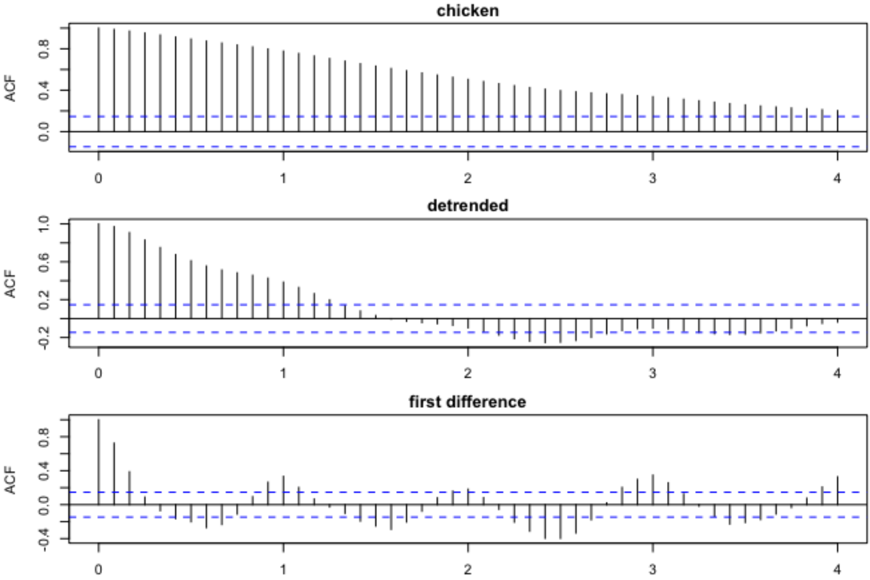
$$x_t = \mu_t + y_t.$$

- $\mu_t = \delta + \mu_{t-1} + w_t$.
- $x_t - x_{t-1} = (\mu_t + y_t) - (\mu_{t-1} + y_{t-1}) = \delta + w_t + y_t - y_{t-1}$.
- Note: no parameters to be estimated.

Make stationary 2: – Detrending Analysis



Make stationary 2: – Detrending Analysis



Make stationary 2: – Detrending Analysis

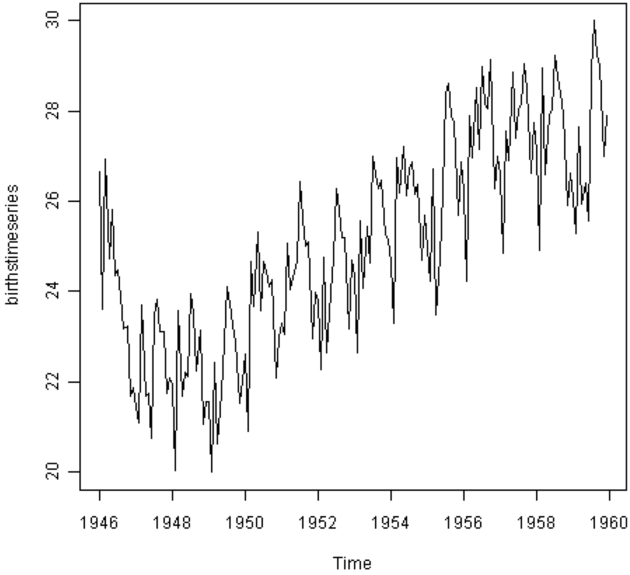
ACF plot:

- ACF stands for autocorrelation function.
- It is a very important tool for us to understand time series data.
- It shows a series of autocorrelation $\rho(0)$, $\rho(1)$, $\rho(2)$, ...

ACF plot can be used to (visually) detect:

- Nonstationary
- Seasonality

Make stationary 2: – Seasonality

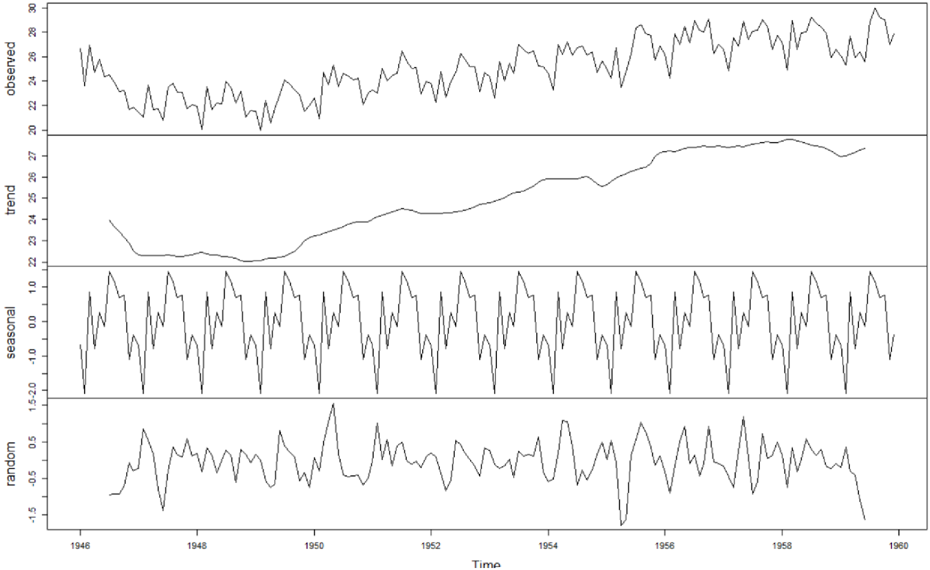


Introduction – Seasonality

How to check seasonality? Eg. yearly seasonality

- Yearly pattern is similary across different years.
- We can calculate a averages using all January data, one for Feb, ...;
    - There should be some difference.

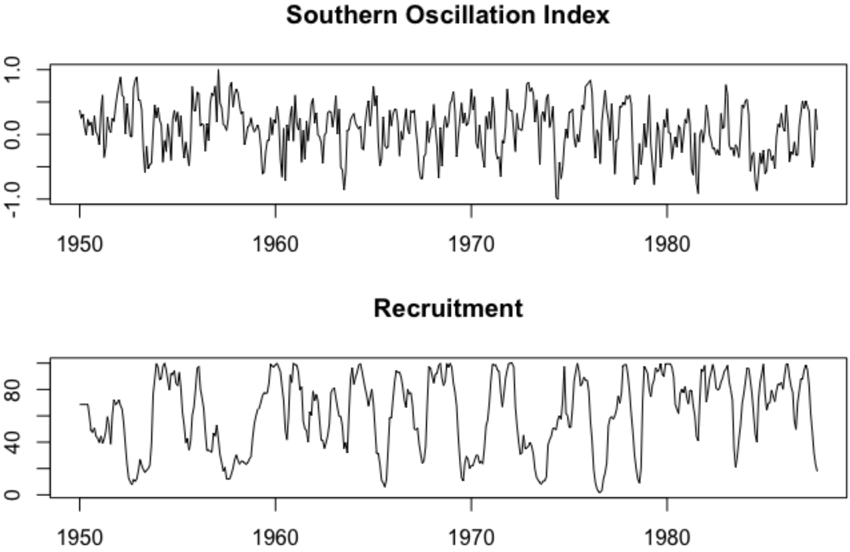
Introduction – Seasonality



Example: EL Nino and Fish Population

- Two monthly time series for a period of 453 months for year 1950-1987.
- SOI: an environmental series called Southern Oscillation Index.
- SOI measures changes in air pressure, related to sea surface temp in the central pacific ocean.
- Central Pacific warms every 3-7 years due to EL Nino effect.
- Recruitment: a series to measure the fish population (number of new fish).

Introduction – Example: EL Nino and Fish Population



Introduction – Example: EL Nino and Fish Population

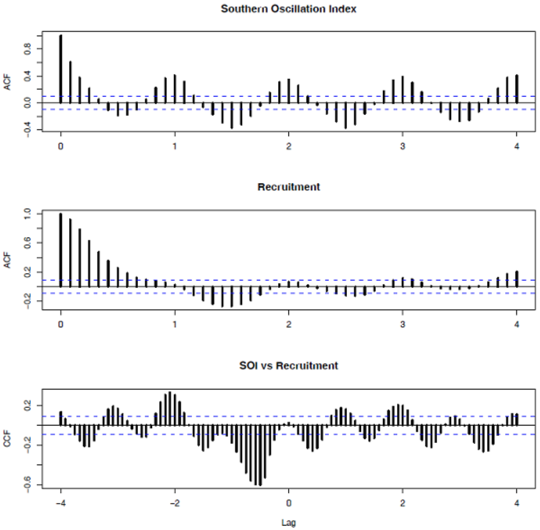
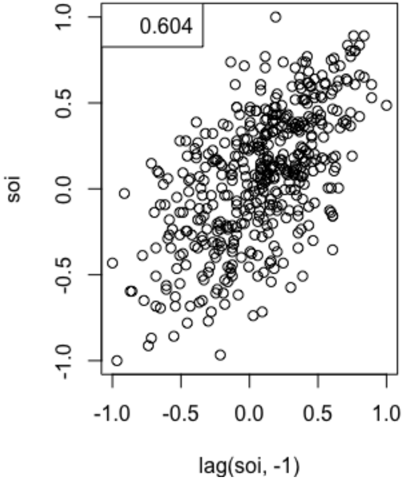


Fig. 1.14. Sample ACFs of the SOI series (top) and of the Recruitment series (middle), and the sample CCF of the two series (bottom); negative lags indicate SOI leads Recruitment. The lag axes are in terms of seasons (12 months).

Introduction – Example: EL Nino and Fish Population



Introduction – Smoothing Analysis

Smoothing helps to discover certain traits such as

- long term trend.
- cyclic pattern.

Introduction – Smoothing Analysis

Moving average smoother:

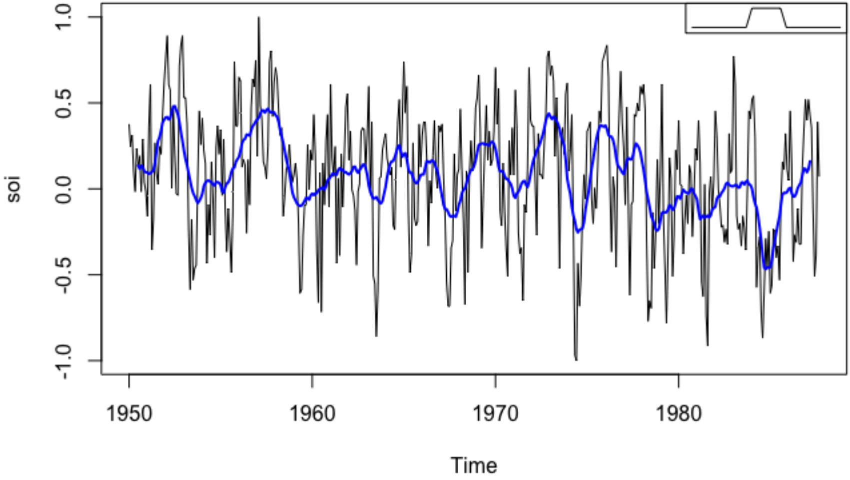
$$m_t = \sum_{j=-1}^{k} a_j x_{t-j}$$

- $a_j = a_{-j} = 0$.
- $\sum_{j=-1}^{k} a_j = 1$.
- This is also called filtering.

Example: SOI series:

- $a_0 = a_{\pm 1} = \ldots = a_{\pm 5} = 1/12$.
- $a_{\pm 6} = 1/24$.
- $k = 6$.
- This method removes the obvious annual temperature cycle and hleps emphasize the EL Nino cycle.

Introduction – Smoothing Analysis



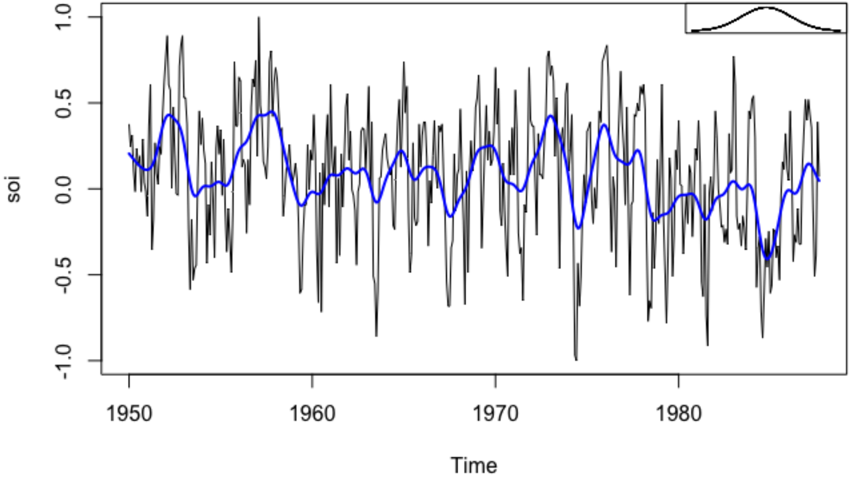
Introduction – Smoothing Analysis

Kernel smoother:

$$m_t = \sum_{i=1}^{n} w_t(t) x_i$$

- Kernel: $K(z) = \frac{1}{\sqrt{2\pi}} \exp(-z^2/2)$.
- $w_i(t) = K(\frac{t-i}{b}) / \sum_{j=1}^{n} K(\frac{t-i}{b})$.

Introduction – Smoothing Analysis



Introduction – References for time series data

References:

- https://otexts.com/fpp2/tspatterns.html
- https://a-little-book-of-r-for-time-series.readthedocs.io/en/latest/src/timeseries.html