Lecture 26

Time Series Analysis

STAT 8020 Statistical Methods II
December 1, 2020

Whitney Huang
Clemson University

Notes

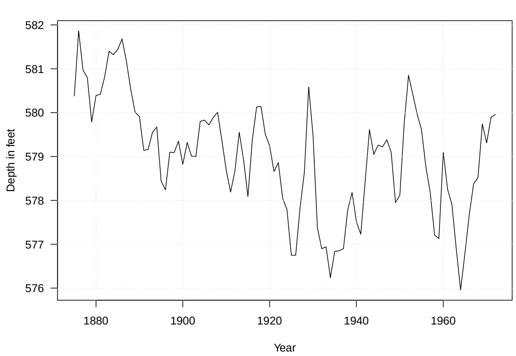
Agenda

1. Time Series Data
2. Objectives of Time Series Analysis
3. Features of Times Series
4. Means & Autocovariances
5. A Case Study

Notes

Level of Lake Huron 1875–1972

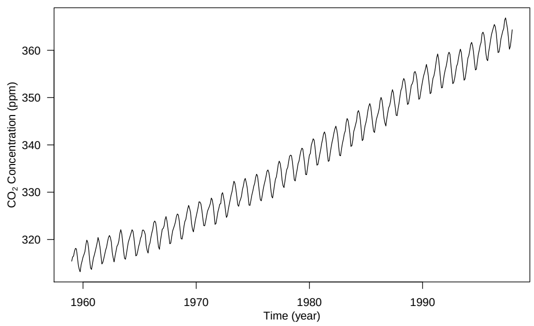
Annual measurements of the level of Lake Huron in feet. [Source: Brockwell & Davis, 1991]



Notes

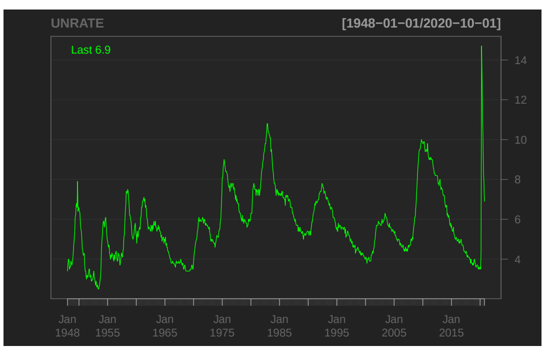
Mauna Loa Atmospheric CO_2 Concentration

Monthly atmospheric concentrations of CO_2 at the Mauna Loa Observatory [Source: Keeling & Whorf, Scripps Institution of Oceanography (SIO)]



Notes

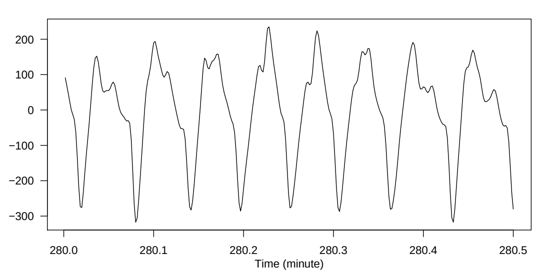
US Unemployment Rate 1948 Jan. – 2020 Oct.



Notes

Airflow Signal

A "normal" patient's 100 Hz sleep airflow signal [Source: Huang et al. 2020+]



Notes

Time Series Data & Models

- A time series is a set of observations made sequentially in "time"
- Time series analysis is the area of statistics which deals with the analysis of dependency between different observations in time series data
- A time series model is a probabilistic model that describes ways that the series data $\{y_t\}$ could have been generated
- More specifically, a time series model is usually a probability model for $\{Y_t : t \in T\}$, a collection of random variables indexed in time

Notes

Some Objectives of Time Series Analysis

- Find a statistical model that adequately explains the dependence observed in a time series
- To conduct statistical inferences, e.g., Is there evidence of a decreasing trend in the Lake Huron depths?
- To forecast future values of the time series based on those we have already observed

Notes

Features of Times Series

- Trends
  - One can think of trend, μ_t as continuous changes, usually in the mean, over longer time scales
  - Usually the form of the trend is unknown and needs to be estimated. When the trend is removed, we obtain a detrended series
- Seasonal or periodic components
  - A seasonal component s_t constantly repeats itself in time, i.e., $s_t = s_{t+kd}$
  - We need to estimate the form and/or the period d of the seasonal component, i.e, to deseasonalize the series
- The "noise" process
  - The noise process, η_t, is the component that is neither trend nor seasonality
  - We will focus on finding plausible (typically stationary) statistical models for this process

Notes

Combining Trend μ_t, Seasonality s_t, and Noise η_t Together

There are two commonly used approaches

- Additive model:

$$y_t = \mu_t + s_t + \eta_t$$

- Multiplicative model:

$$y_t = \mu_t s_t \eta_t$$

If all $\{y_t\}$ are positive then we obtain the additive model by taking logarithms:

$$\log y_t = \log \mu_t + \log s_t + \log \eta_t$$

Notes

Means, Autocovariances, and Stationary Processes

- The mean function of $\{Y_t\}$ is

$$\mu_t = \mathrm{E}[Y_t], \quad t \in T$$

- The autocovariance function of $\{Y_t\}$ is

$$\gamma(t, t') = \mathrm{Cov}(Y_t, Y_{t'}) = \mathrm{E}[(Y_t - \mu_t)(Y_{t'} - \mu_{t'})], \quad t, t' \in T$$

When $t = t'$ we obtain $\gamma(t, t') = \mathrm{Cov}(Y_t, Y_t) = \mathrm{Var}(Y_t) = \sigma_t^2$, the variance function of Y_t

Notes

Autocorrelation Function

The autocorrelation function (ACF) of $\{Y_t\}$ is

$$\rho(t, t') = \mathrm{Corr}(Y_t, Y_{t'}) = \frac{\gamma(t, t')}{\sqrt{\gamma(t, t)\gamma(t', t')}}$$

It measures the strength of linear association between Y_t and $Y_{t'}$

Properties:

1. $-1 \leq \rho(t, t') \leq 1, \quad t, t' \in T$
2. $\rho(t, t') = \rho(t', t), \quad \forall t, t' \in T; \rho(t, t) = 1, \quad \forall t \in T$
3. $\rho(t, t')$ is a non-negative definite function

Notes

Stationary Processes

We will still try to keep our models for $\{\eta_t\}$ as simple as possible by assuming stationarity, meaning that some characteristic of $\{\eta_t\}$ does not depend on the time points, only on the "time lag" between time points:

- $\mathrm{E}[\eta_t] = 0, \quad \forall t \in T$
- $\mathrm{Cov}(\eta_t, \eta_{t'}) = \gamma(t' - t) = \mathrm{Cov}(\eta_{t+s}, \eta_{t'+s})$

$\Rightarrow$ autocorrelation function (ACF):

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

Notes

Autoregressive Moving Average (ARMA) Models

Time Series Analysis
CLEMSON UNIVERSITY
Time Series Data
Objectives of Time Series Analysis
Features of Times Series
Means & Autocovariances
A Case Study

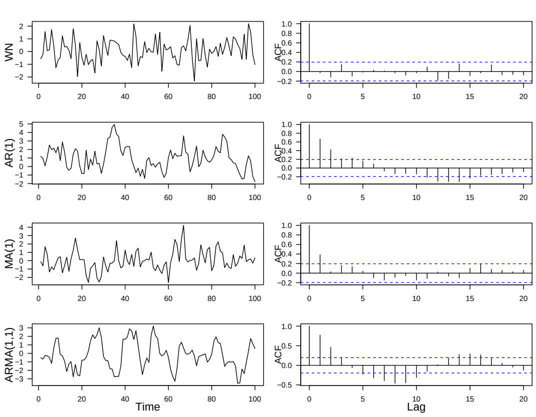
Let $\{Z_t\}$ be independent and identical random variables that follow $\mathrm{N}(0, \sigma^2)$

- Moving Average Processes (MA(q)): $\eta_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} \cdots + \theta_q Z_{t-q}$
- Autoregressive Processes (AR(p)): $\eta_t = \phi_1 \eta_{t-1} + \phi_2 \eta_{t-2} + \cdots + \phi_p \eta_{t-p} + Z_t$
- Autoregressive Moving Average Processes ARMA(p,q): $\eta_t = \phi_1 \eta_{t-1} + \phi_2 \eta_{t-2} + \cdots + \phi_p \eta_{t-p} + Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \cdots + \theta_q Z_{t-q}$

Notes

Autocorrelation Plot

Time Series Analysis
CLEMSON UNIVERSITY
Time Series Data
Objectives of Time Series Analysis
Features of Times Series
Means & Autocovariances
A Case Study



Notes

Lake Huron Case Study

- Detrending
- Model selection and fitting
- Forecasting

See R lab 22 for a demo

Notes

Notes

Notes