# Time Series (MATH5845)

Dr. Atefeh Zamani
Based on the notes by Prof. William T.M. Dunsmuir

T2 2025

# Chapter 2

# Introduction to Time Series

# Contents


**2.1 Some Motivating Examples . . . . . . . . . . . . . . . . . 42**

**2.2 Discrete time series: Definition and Basics . . . . . . . 50**

2.2.1 Weakly Stationary Time Series . . . . . . . . . . . . . . . 55

2.2.2 Strictly Stationary Time Series . . . . . . . . . . . . . . . 72

**2.3 Estimation of Means and Covariances. . . . . . . . . . . 76**

2.3.1 The sample mean for a stationary time series. . . . . . . 77

2.3.2 Estimating autocovariances and autocorrelations. . . . . 79

2.3.3 Approximate Distribution of the Sample ACF . . . . . 84

**2.4 Simple ways to obtain stationarity . . . . . . . . . . . . . 88**

2.4.1 Differencing and Seasonal Differencing . . . . . . . . . . 90

## 2.1 Some Motivating Examples

**Example 2.1 (European countries retail trade, Lima et al. [2023])** *The dataset* `TOVT_2015.csv` *covers total turnover indexes for European countries from 2000 to 2023, focusing on retail trade. These indexes measure market development for goods and services, including all charges passed on to customers but excluding VAT and other similar taxes. The base year for comparison is 2015, with an index value of 100.*

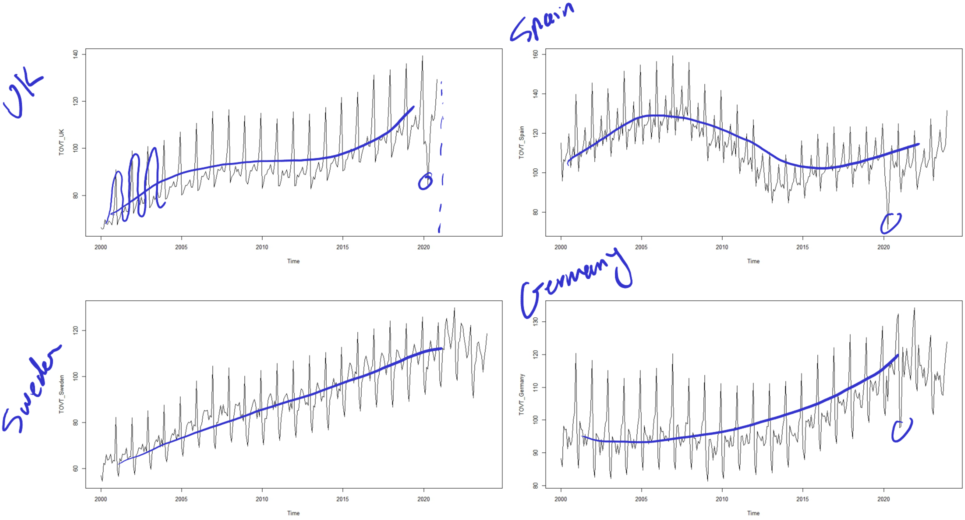



Figure 2.1: Index of total turnover for some European countries.

**Example 2.2 (Sunspot Data, Brockwell and Davis [2002, 1991] )** *The Wolf number quantifies the presence of sunspots and sunspot groups on the Sun's surface. The series is referred to as the international sunspot number series, which exhibits an approximate 11-year periodicity. The series extends back to 1700 with annual values, while daily values exist only since 1818. A revised and updated series has been available since July 1, 2015.*
*The file named* `Sunspot.csv` *presents the yearly mean total sunspot number from 1700 to 2023, along with some other data. Figure 2.2 presents this dataset.*

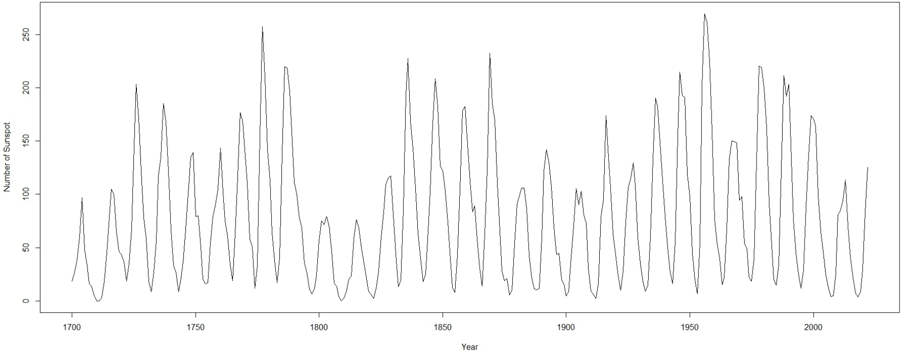



Figure 2.2: The sunspot numbers, 1700-2023.

**Example 2.3 (Speech data, Shumway et al. [2000])** *The* `speech` *dataset, which is part of the* `astsa` *package, is a small 0.1 second (1000 point) sample of recorded speech for the phrase aaa ... hhh, and we note the repetitive nature of the signal and the rather regular periodicities. One current problem of great interest is computer recognition of speech, which would require converting this particular signal into the recorded phrase aaa ... hhh. From Figure 2.3, one can immediately notice the rather regular repetition of small wavelets. The separation between the packets is known as the pitch period and represents the response of the vocal tract filter to a periodic sequence of pulses.*

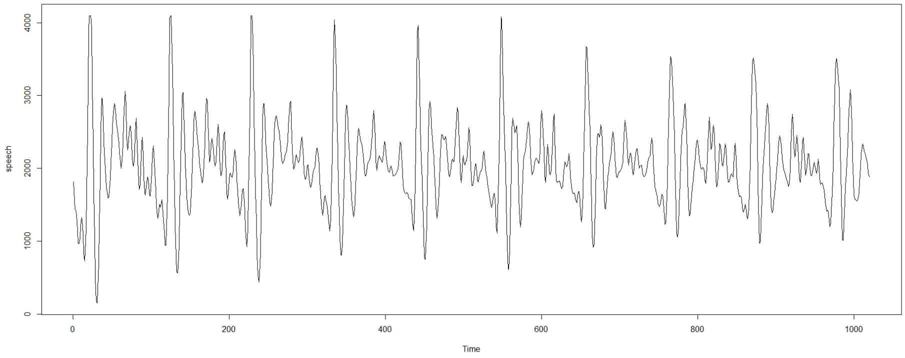



Figure 2.3: Speech recording of the syllable aaa ... hhh sampled at 10,000 points per second with n = 1020 points

**Example 2.4 (Global Warming, Shumway et al. [2000])** *Figure 2.4 shows the global mean land-ocean temperature index from 1850 to 2023. The datasets, named* `gtemp_land` *and* `gtemp_ocean`*, are part of the* `astsa` *package; they indicate a noticeable increase in temperature towards the end of the 20th century, supporting the global warming hypothesis. There was a stabilization around 1935 followed by a significant rise starting around 1970. The ongoing debate among those concerned with global warming is whether this trend is a natural occurrence or driven by human activities.*

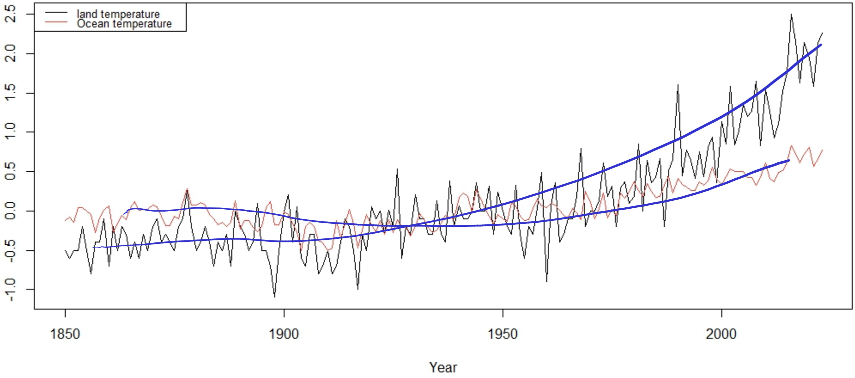



Figure 2.4: Yearly average land-ocean temperature deviations (1850–2023) in degrees centigrade

## 2.2 Discrete time series: Definition and Basics

- The primary objective of time series analysis is to develop mathematical models that provide plausible descriptions for data.

- To allow for the possibly unpredictable nature of future observations, suppose that the time series $\{x_t,\ t \in T\}$ is a realization (sample path) of the stochastic process $\{X_t,\ t \in T\}$.

**Definition 2.1** *A time series is a collection of random variables $\{X_t\}$ indexed by $t \in T$, where the index set $T$ is a set of time points. Generally, $T$ is defined to be $\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$ or $\mathbb{R} = (-\infty, \infty)$.*

- $t \in \mathbb{Z}$ : *discrete time, time series or "***discrete time series***" for short (focus of this course).*

- $t \in \mathbb{R}$: *continuous time, time series or "continuous time series" for short.*

not covered in this course

- The probabilistic properties of these stochastic processes should be known for inference and forecasting.

- A complete probabilistic specification requires the specification of the joint distributions of $\mathbf{X} = (X_{t_1}, \ldots, X_{t_n})'$ with $t_1 < t_2 < \ldots < t_n$.

- These multidimensional distribution functions cannot usually be written easily unless the random variables are **jointly Gaussian**(normal), or the time series is **stationary**.

Let $\{X_t\}$ be a time series with $E(X_t^2) < \infty$ for all $t \in \mathbb{Z}$ .

**Definition 2.2** *The mean function is defined as*

$$\mu_X(t) = E(X_t) = \int_{-\infty}^{\infty} x f_{X_t}(x)dx \tag{2.1}$$

*provided it exists.*

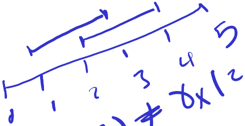
The lack of independence between two adjacent values $X_s$ and $X_t$ can be assessed using the notions of covariance and correlation.

**Definition 2.3 (The Covariance Function)** *If $\{X_t,\ t \in T\}$ is a process such that $E(X_t^2) < \infty$ for each $t \in T$, then the covariance function $\gamma_X(\cdot, \cdot)$ of $\{X_t,\ t \in T\}$ is defined by*

time-stamps

$$\gamma_X(r, s) = cov(X_r, X_s) = E[(X_r - \mu_X(r))(X_s - \mu_X(s))], \quad r, s \in T. \tag{2.2}$$

*Note that $\gamma_X(r, s) = \gamma_X(s, r)$ for all time points $r$ and $s$.*




$\gamma_X(1,3) \neq \gamma_X(2,4)$

The covariance measures the linear dependence between two points in the same series at different times.

- Very smooth series $\Rightarrow$ covariance functions is large even when $s$ and $r$ are far apart.
- Choppy series $\Rightarrow$ covariance functions are nearly zero for large separations.

**Note 2.1**

- *If $\gamma_X(r,s) = 0$, $X_r$ and $X_s$ are not linearly related, but there still may be some dependence structure between them.*
- *If $X_r$ and $X_s$ are bivariate normal, $\gamma_X(r,s) = 0$ ensures their independence.*
- *For $r = s$, $\gamma_X(r,r) = E[(X_r - \mu_X(r))^2] = var(X_r)$.*
- *Sometimes in time series, the focus is on the properties of the sequence $\{X_t\}$ that depend only on the first- and second-order moments of the joint distributions, i.e., $E(X_t)$ and $E(X_rX_s)$. Such properties are referred to as second-order properties.*

**A Gaussian time series** is one for which any finite collection of the series has a multivariate normal distribution. That is, for $t_1 < t_2 < \ldots < t_n$, the joint distribution of $\mathbf{X} = (X_{t_1}, \ldots, X_{t_n})^T$ is

$$\mathbf{X} \sim N(\mu, \Gamma)$$

where

$$\mu = \begin{pmatrix} \mu_{t_1} \\ \vdots \\ \mu_{t_n} \end{pmatrix}, \quad \Gamma_{i,j} = cov(X_{t_i}, X_{t_j}) = \gamma_X(t_i, t_j)$$

Note that in a Gaussian time series all distributions are **completely specified** by the mean and covariance functions (second-order properties of $\mathbf{X}$).

### 2.2.1 Weakly Stationary Time Series

**Definition 2.4 (Weakly stationary)** *A weakly stationary time series, $X_t$, has mean and covariance functions which are* **invariant in time***:*

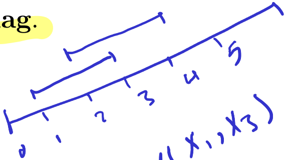
*(i)* $E(|X_t|^2) < \infty$, *for all* $t$,

*(ii)* $\mu_X$ *is constant and does not depend on time* $t$, *i.e.*,

$$\mu_X(t) = \mu_X, \quad t \in \mathbb{Z},$$

*(iii)* *For* $t \in \mathbb{Z}$, $\gamma_X(t+h, t) = \gamma_X(h)$, *i.e., the covariance function depends* **only on the time separation** $h$ **and not the actual time** $t$.

Sometimes, we refer to $h$ as the time shift or **lag**.




$\mathrm{cov}(X_1, X_3) = \mathrm{cov}(X_2, X_4)$

**Definition 2.5 (Autocovariance and autocorrelation functions)** *For a weakly stationary time series,*

- *autocovariance function (ACVF):* $\{\gamma_X(h) : \; h = 0, \pm 1, \ldots\}$
- *autocorrelation function (ACF):*

$$\rho_X(h) = \frac{\gamma_X(t+h, t)}{\sqrt{\gamma_X(t+h, t+h)\gamma_X(t, t)}} = \frac{\gamma_X(h)}{\gamma_X(0)}, \quad h = 0, \pm 1, \ldots.$$

The autocovariance or, equivalently, the autocorrelation function describes the **second order dependence properties** in a stationary time series.

**Theorem 2.1** *For stationary time series,*

1. *Nonnegative variance:* $\gamma(0) \geq 0$.
2. *Autocovariances are dominated by the variance.* $|\gamma(h)| \leq \gamma(0)$, $\forall h$.
3. *Even function:* $\gamma(-h) = \gamma(h)$, $\forall h$.
4. *Non-negative definite function: For any integer* $n$ *and vector of constants* $a = (a_1, \ldots, a_n)^T$ *we have* $\sum_{j=1}^{n} \sum_{k=1}^{n} a_j a_k \gamma(j-k) \geq 0$.

$$\gamma_X(h) = cov(X_{t+h}, X_t)$$

① $\gamma_X(0) = cov(X_t, X_t) = Var(X_t) \geq 0$

② $|\rho_X(h)| \leq 1 \longrightarrow |\gamma_X(h)| \leq \gamma(0)$

cauchy-schwartz inequality

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma(0)}$$

③

$$\operatorname{cov}(\overbrace{X_{t+h}, X_t}^{t+h-t=h \text{ lag}}) = \operatorname{cov}(\underbrace{X_t, X_{t+h}}_{t-(t+h)=-h \text{ lag}})$$

$$\gamma_X(h) = \gamma_X(-h)$$

④

$$Y = \sum_{j=1}^{n} a_j (X_j - \mu)$$

$$0 \leq Var(Y) = Var\left(\sum_{j=1}^{n} a_j (X_j - \mu)\right)$$

$$= \operatorname{cov}\left(\sum_{j=1}^{n} a_j (X_j - \mu), \sum a_k (X_k - \mu)\right)$$

$$= \sum_{j=1}^{n} \sum_{k=1}^{n} a_j a_k \underbrace{\operatorname{cov}(X_j - \mu, X_k - \mu)}_{\gamma(j-k)}$$

**Example 2.5 (White Noise)** • *A simple kind of generated series is a collection of* **uncorrelated** *random variables,* $W_t$*, with* $\mu_W(t) = 0$ *and* $var(W_t) = \sigma^2_W$*. This series is called* **white noise** *(*$W_t \sim WN(0, \sigma^2_W)$*).*

- *If the noise process is a sequence of iid random variables with* $\mu_W(t) = 0$ *and* $var(W_t) = \sigma^2_W$*, it is called* **white independent noise** *or* **iid noise** *(*$W_t \sim iid(0, \sigma^2_W)$*).*

- **Gaussian white noise** *is a sequence of* **independent normal** *random variables,* $W_t \sim N(0, \sigma^2_W)$*.*

- *For white noise processes,* $E(W_t) = 0$ *and*

variance of the process

$$\gamma_W(s,t) = cov(W_s, W_t) = \begin{cases} \sigma^2_W & s = t \\ 0 & s \neq t \end{cases} \tag{2.3}$$

non-zero lag

**White noise processes are weakly stationary**.

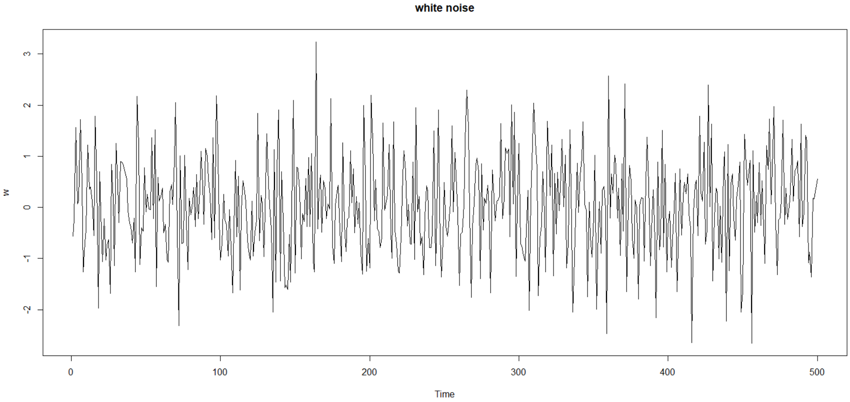



Figure 2.5: 500 observations from a Gaussian white noise series

**Example 2.6 (Random walk with drift)** *The random walk with drift model is given by*

position in time t-1

position in time t

$$X_t = \delta + X_{t-1} + W_t \tag{2.4}$$

noise

constant (drift)

*successively for* $t = 1, 2, \ldots,$ *with*

- $X_0 = 0$,
- $W_t \sim WN(0, \sigma_W^2)$,
- $\delta$ *is constant and it is called the drift*
  - *when* $\delta = 0$, (2.4) *is called a random walk.*

*The term random walk comes from the fact that, when* $\delta = 0$, *the value of the time series at time* $t$ *is the value of the series at time* $t - 1$ *plus a completely random movement determined by* $W_t$. *We may rewrite* (2.4) *as a cumulative sum of white*

$E(X_t) = E(t\delta + \sum_{i=1}^{t} W_i) = t\delta + \sum \underbrace{E(W_i)}_{0}$

$= t\delta$

*noise variates:*

$$X_t = t\delta + \sum_{i=1}^{t} W_i, \qquad t = 1, 2, \ldots. \tag{2.5}$$

↳ function of time ⇒ not weakly stationary

*Since* $E(W_t) = 0$, *for all* $t$, *and* $\delta$ *is constant, we have*

$$\mu_X(t) = E(X_t) = \delta t,$$

*which is a function of* $t$. *Besides,*

$$cov(X_r, X_s) = \min\{r, s\}\sigma_W^2, \quad ✓$$

*and*

$$var(X_t) = t\sigma_W^2. \quad ✓$$

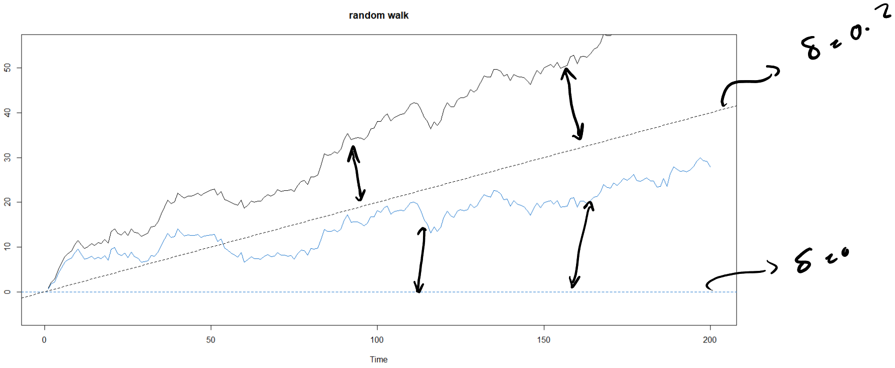



Figure 2.6: Random walk, $\sigma_W = 1$, with drift $\delta = 0.2$ (upper jagged line), without drift, $\delta = 0$ (lower jagged line), and straight (dashed) lines with slope $\delta$.

- *Random walk with drift is not weakly stationary.*

- *In this case, variance increases without bound as time $t$ increases. The effect of this variance increase can be seen in Fig. 2.6 where the processes start to move away from their mean functions $\delta t$.*

## Last Session

Weak Stationary
- Constant mean
- Constant variance
- Autocovariance function is a function of lag (h)

White Noise
- Uncorrelated random variables with mean 0 and constant variance $\sigma^2$

iid Noise
- Independent and identically distributed random variables with mean 0 and constant variance $\sigma^2$

Gaussian White Noise

Random Walk with drift

**Example 2.7 (Signal in Noise)** *Many realistic models for generating time series assume an underlying signal with some consistent periodic variation (first term in Equation* (2.6)*), contaminated by adding a random noise:*

$$X_t = 2\cos\left(\frac{2\pi(t+15)}{50}\right) + W_t, \qquad t = 1, 2, \ldots. \tag{2.6}$$

*We note that a sinusoidal waveform can be written as*

phase

$$A\cos(2\pi\omega t + \phi), \tag{2.7}$$

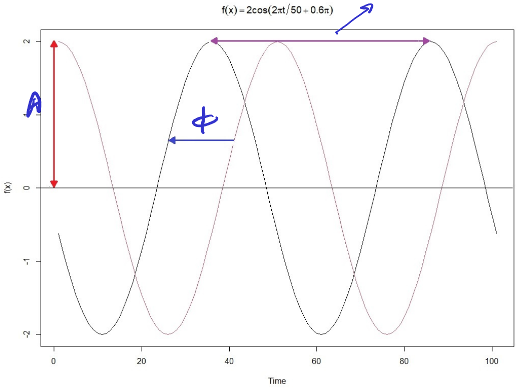
amplitude frequency

*where*

- $A$ *is the amplitude (in* (2.6), $A = 2$*),*
- $\omega$ *is the frequency of oscillation (in* (2.6), $\omega = 1/50$ *(one cycle per 50 time points)),*
- $\phi$ *is a phase shift (in* (2.6), $\phi = 2\pi 15/50 = 0.6\pi$*.)*

1 period = $\frac{1}{\omega}$

$A\cos(2\pi\omega t + \phi)$




**Question**: *Can you label the amplitude, ω or period, and φ in this figure?*

$X_t$ **is not stationary**, *since*

$$\mu_X(t) = E(X_t) = 2\cos\left(\frac{2\pi(t+15)}{50}\right),$$

to check weakly stationarity

- Is $\mu_x$ constant?! → Not stationary
- Is the autocovariance function just a function of lag?!

$$E(X_t) = E\left(2\cos\left(\frac{2\pi(t+15)}{50}\right) + w_t\right)$$

$$= 2\cos(\;--\;) + \cancel{E(w_t)}^{0}$$

$\hookrightarrow$ $w_t \sim WN(0, \sigma_w^2)$

$$= 2\cos\left(\frac{2\pi(t+15)}{50}\right)$$

$\hookrightarrow$ is a function of time

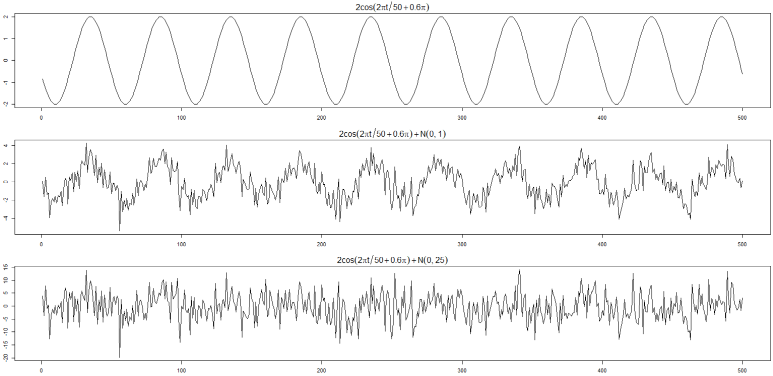



Figure 2.7: Cosine wave (signal) with period 50 points (top panel) contaminated with additive white Gaussian noise, $\sigma_W = 1$ (middle panel) and $\sigma_W = 5$ (bottom panel)

**Example 2.8 (Trend Stationarity)** *If $X_t = \alpha + \beta t + Y_t$, where $Y_t$ is stationary, then*

$E(Y_t) = \mu$
$cov(Y_t, Y_s) = \gamma(t-s)$

- $\mu_X(t) = \alpha + \beta t + \mu_Y(t) = \alpha + \beta t + \mu_Y$, *which is not independent of time.*

*Therefore, $X_t$ is not stationary.*
*However, $cov(X_r, X_s)$ is independent of time, because*

$$\begin{aligned} cov(X_r, X_s) &= cov(\cancel{\alpha + \beta r} + Y_r, \cancel{\alpha + \beta s} + Y_s) \\ &= cov(Y_r, Y_s) \\ &= \gamma_Y(|r-s|). \checkmark \end{aligned}$$

- *The model has stationary behavior around a linear trend.*
- *This behavior is sometimes called* **trend stationarity**.

**Question**: *Can you suggest a transformation to make $X_t$ stationary?*

Differencing

$$Z_t = X_t - X_{t-1}$$

$$= \alpha + \beta(t) + Y_t - (\alpha + \beta(t-1) + Y_{t-1})$$

$$= \beta + Y_t - Y_{t-1}$$

$$E(Z_t) = \beta + E(Y_t) - E(Y_{t-1})$$

$$= \beta + \mu - \mu = \beta \checkmark$$

$$\operatorname{cov}(Z_t, Z_s) = \operatorname{cov}(\beta + Y_t - Y_{t-1}, \beta + Y_s - Y_{s-1})$$

$$= \operatorname{cov}(Y_t - Y_{t-1}, Y_s - Y_{s-1})$$

$Y_t$ is stationary

$$= \underset{t-s}{\operatorname{cov}(Y_t, Y_s)} - \underset{t-1-s}{\operatorname{cov}(Y_{t-1}, Y_s)}$$

$$- \underset{t-s+1}{\operatorname{cov}(Y_t, Y_{s-1})} + \underset{t-s}{\operatorname{cov}(Y_{t-1}, Y_{s-1})}$$

### 2.2.2 Strictly Stationary Time Series

**Definition 2.6 (strictly stationary time series)** *A strictly stationary time series has joint distributions which are invariant to time translation so that*

$$(X_{t_1}, \ldots, X_{t_n}) \stackrel{d}{=} (X_{t_1+h}, \ldots, X_{t_n+h}), \quad h \in \mathbb{Z}.$$

↳ distribution is the same

*That is,*

$$P(X_{t_1} \leq c_1, X_{t_2} \leq c_2, \ldots, X_{t_n} \leq c_n) = P(X_{t_1+h} \leq c_1, X_{t_2+h} \leq c_2, \ldots, X_{t_n+h} \leq c_n\}) \tag{2.8}$$

*for the time points* $t_1, t_2, \ldots, t_n$, *all numbers* $c_1, c_2, \ldots, c_n$, *and all time shifts* $h = 0, \pm 1, \pm 2, \ldots$.

**Note 2.2**

$$\textit{strictly stationary} \not\Rightarrow \textit{weakly stationary}.$$

*However,*

$$\textit{strictly stationary with } E(X_t^2) < \infty \Rightarrow \textit{weakly stationary},$$

*but, the converse is not true:*

$$\textit{strictly stationary} \not\Leftarrow \textit{weakly stationary} \ ,$$

*(Refer to Example 2.9).*
*For a* **Gaussian time series**,

$$\textit{strictly stationary} \Leftrightarrow \textit{weakly stationary}.$$

*For a weakly stationary time series, the mean vector and covariance matrices of* $\mathbf{X} = (X_{t_1}, \ldots, X_{t_n})'$ *simplify to*

$$\mu = \begin{pmatrix} \mu \\ \vdots \\ \mu \end{pmatrix}, \qquad \Gamma_{i,j} = \gamma_X(t_i - t_j)$$

*and when* $(t_1, \ldots, t_n) = (1, \ldots, n)$ *the covariance matrix takes on the banded (Toeplitz) form:*

$$\Gamma = \begin{pmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \ddots & \vdots \\ & & \ddots & \gamma(1) \\ \gamma(n-1) & \cdots & \gamma(1) & \gamma(0) \end{pmatrix}$$

Variance

**Example 2.9** *Let $X_t$ be a sequence of independent random variables such that it is exponentially distributed with mean one when t is odd and normally distributed with mean one and variance one when t is even, then $X_t$ is stationary with $\gamma_X(0) = 1$ and $\gamma_X(h) = 0$ for $h \neq 0$. However since $X_1$ and $X_2$ have different distributions, $X_t$ cannot be strictly stationary.*

Not strongly stationary
$F_{X_1}(x) \neq F_{X_2}(x)$

$E(1)$ $\quad$ $N(1,1)$

$$X_t \sim \begin{cases} E(1) & t = 2k-1 \\ N(1,1) & t = 2k \end{cases}$$

$E(\lambda) \to f_X(x) = \lambda e^{-\lambda x} \quad x > 0 \qquad E(X) = \frac{1}{\lambda} \qquad Var(X) = \frac{1}{\lambda^2}$

weakly stationary

$$\begin{cases} E(X_t) = 1 \checkmark \\ Var(X_t) = 1 \checkmark \\ Cov(X_t, X_s) = \begin{cases} 0 \\ 1 \end{cases} \end{cases}$$

$t \neq s$ independent seq

$t = s \implies$ variance

## 2.3 Estimation of Means and Covariances.

In this section, we consider

- estimation of means and autocovariance for a *weakly stationary time series*,
- described their large sample ($n \to \infty$) distributions
- present a test statistic to test if there is significant lag serial correlation in a time series.

### 2.3.1 The sample mean for a stationary time series.

Let $\{X_t\}$ be a stationary time series with mean $\mu$ and ACVF $\gamma(h)$.

we know that $E(X_t) = \mu$

- The sample mean

$$\bar{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$$

  is an unbiased estimator of $\mu$. $E(\bar{X}_n) = \mu$

- $var(\bar{X}_n)$ is given by

weight

Tutorial.

$$var(\bar{X}_n) = \frac{1}{n} \sum_{h=-n+1}^{n-1} (1 - \frac{|h|}{n})\gamma(h). \tag{2.9}$$

- If $\gamma(h) \to 0$ as $h \to \infty$ the right-hand side of (3.11) converges to zero, so that $\bar{X}_n$ converges in mean square to $\mu$.

- If $\sum |\gamma(h)| < \infty$, then (3.11) gives

$$\lim_{n\to\infty} n\, var(\bar{X}_n) = \sum_{-\infty}^{\infty} \gamma(h).$$

- This formula for the approximate variance of the sample mean can be simplified in some cases.

### 2.3.2 Estimating autocovariances and autocorrelations.

The usual estimator of the ACVF is

$$\hat{\gamma}(-h) = \hat{\gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-h}(X_t - \bar{X}_n)(X_{t+h} - \bar{X}_n), \quad 0 \leq h < n,$$

$\hat{Var}(X_t)$ $\ast$

and the estimator of the ACF is

$$\hat{\rho}(h) = \hat{\gamma}(h)/\hat{\gamma}(0), \qquad |h| < n.$$

$$\mathrm{cov}(X_t, X_s) = E(X_t - \mu)(X_s - \mu)$$

$\frac{1}{n}\sum$ $\bar{X}$ $\bar{X}$

**Properties**

1. $\hat{\gamma}(h)$ and $\hat{\rho}(h)$ are not unbiased estimators for $\gamma(h)$ and $\rho(h)$ respectively.

2. The bias can be reduced substantially by replacing $n$ by $n-h$ as divisor in the definition of $\hat{\gamma}(h)$.

3. In either case, for large $n$, the bias is small.

4. The matrix
$$\hat{\Gamma}_n = [\hat{\gamma}(i-j)]_{i,j=1}^{n},$$
is positive definite with probability equal to 1.
**Proof**: Let $Y_i = X_i - \bar{X}_n$ and

$$T = \begin{pmatrix} 0 & \cdots & 0 & Y_1 & \cdots & Y_{n-1} & Y_n \\ 0 & \cdots & Y_1 & Y_2\cancel{0} & \cdots & Y_n & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ Y_1 & \cdots & Y_{n-1} & Y_n & \cdots & 0 & 0 \end{pmatrix}.$$

be an $n \times (2n-1)$ dimensional matrix. It is easy to show

$$\hat{\Gamma}_n = \frac{1}{n} T T',$$

which is positive definite ($a'TT'a > 0$, for all $a$) unless all the data are identical in value.

week 2,
week 3

- This is important for obtaining Yule-Walker estimates for autoregressive processes and ensuring that the autoregressive operator is causal.
- This is not true if $n$ is replaced by $n-h$ as divisor in the definition of $\hat{\gamma}(h)$.

5. In particular, since $\hat{\Gamma}_n$ is positive definite the sample ACF values are bounded as
$$|\hat{\rho}(h)| < 1,$$
so they are proper estimates of correlations.

6. If $h$ is too large relative to the sample size $n$ then the estimates $\hat{\gamma}(h)$ and $\hat{\rho}(h)$ may be unreliable. A general "rule of thumb" is to restrict calculation of $\hat{\gamma}(h)$ and $\hat{\rho}(h)$ to cases where $n \geq 50$ and $h \leq n/4$.

### 2.3.3 Approximate Distribution of the Sample ACF

A **linear process**, $X_t$, is defined to be a linear combination of white noise variates $W_t$, and is given by

weight

$E(X_t) = \mu$

$$X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j W_{t-j}, \qquad \sum_{j=-\infty}^{\infty} |\psi_j| < \infty, \tag{2.10}$$

For the linear process, we may show that

$$\gamma_X(h) = \sigma_W^2 \sum_{j=-\infty}^{\infty} \psi_{j+h}\psi_j \tag{2.11}$$

online tutorial

for $h \geq 0$; recall that $\gamma_X(-h) = \gamma_x(h)$.

weakly stationary

- The linear process (2.10) depends on the future ($j < 0$), the present ( $j = 0$), and the past ( $j > 0$).
- For the purpose of forecasting, a future dependent model will be useless.
- We will focus on processes that do not depend on the future. Such models are called **causal**.
- For linear processes, the central limit theorem holds for the sample ACF's:

$$\begin{pmatrix} \hat{\rho}(1) \\ \vdots \\ \hat{\rho}(k) \end{pmatrix} \approx N_k\left(\begin{pmatrix} \rho(1) \\ \vdots \\ \rho(k) \end{pmatrix}, \frac{1}{n}W\right)$$

where

$$W_{ij} = \sum_{k=1}^{\infty}[\rho(k+i) + \rho(k-i) - 2\rho(i)\rho(k)][\rho(k+j) + \rho(k-j) - 2\rho(j)\rho(k)].$$

**Example 2.10** *under the null hypothesis that* $\{X_t\} \sim WN(0, \sigma^2)$, *we have* $\rho(h) = 0$ *for* $|h| > 0$ *so that*

$$W_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases},$$

*and* $\hat{\rho}(j) \sim \ AN(0, \frac{1}{n})$ *(AN: Approximately normally distributed).*

- *If* $\hat{\rho}(h)$ *falls outside of the range* $\pm 1.96/\sqrt{n}$ *then we would declare* $\rho(h)$ *to be significantly different from zero at the 5% level.*

- *For a white noise sequence, approximately 95% of the sample ACFs should be within these limits.*

- *Many statistical modeling procedures depend on reducing a time series to a white noise series using various kinds of transformations. After such a procedure is applied, the plotted ACFs of the residuals should then lie roughly within the limits given above.*

**Portmanteau test statistic**
A "portmanteau" test statistic used to test if all autocorrelations are zero up to a maximum specified lag $h$ is

$$Q = n \sum_{j=1}^{h} \hat{\rho}(j)^2 \sim \chi^2(h).$$

- A large value of $Q$ suggests that the sample autocorrelations of the data are too large for the data to be a sample from a white noise sequence.
- We reject the white noise hypothesis at level $\alpha$ if $Q > \chi^2_{1-\alpha}(h)$.

A modification, which is better approximated by the $\chi^2$ distribution, is the Ljung-Box statistic

$$Q_{LB} = n(n+2) \sum_{j=1}^{h} \frac{\hat{\rho}(j)^2}{n-j} \sim \chi^2(h).$$

## 2.4 Simple ways to obtain stationarity

Nonstationarity arises in many ways.

- The mean function $\mu_X(t) \neq \mu_X$ for all $t$.
  - a deterministic trend in time, such as linear (Example 2.8), exponential, polynomial etc.,
  - a function of other processes which are not stationary.
- The variance is not constant through time or autocovariance function does not only depend on time separation in the process
- The distributions themselves might not be time invariant.
- Processes like random walks and their variants, where the current value of the process is a random increment from the previous value of the process.

- The simplest way to **detect non-stationarity** is to **examine a time plot of the data**.
- A simple method to **remove many common forms of non-stationarity** is to **difference the observations at sensible time lags**

### 2.4.1 Differencing and Seasonal Differencing

Often time series exhibit forms of non-stationary behaviour that can be removed by taking successive differences at certain lags, typically one time lag or at a seasonal lag.

The **backshift operator,** $B$, takes as input a time series and produces as output the series shifted backwards in time by one time unit:

- $BX_t = X_{t-1}$
- $B^j X_t = X_{t-j}$ (shift the series backwards in time by $j$ time units)

The **difference operator**, $\nabla$, is defined as

$$\nabla = (1 - B)$$

and its action on a time series $\{Y_t\}$ is

$$X_t = \nabla Y_t = (1 - B)Y_t = Y_t - BY_t = Y_t - Y_{t-1}$$

which are referred to as **lag 1 differences**.

**Example 2.11 (Differencing a Time Series)** *Figure 2.8 shows a time series plot of the number of employees in the Fabricated Metals industries in Wisconsin measured each month over five years as well as the lag 1 differences. Note that lag 1 differencing appears to remove the smooth underlying trend obvious in the original series.*

*Figure 2.9 shows (left panel) estimates of the autocorrelation function as well as that for the lag 1 differences (right panel). Note that lag 1 differencing appears to remove the smooth underlying trend.*


Figure 2.8: Monthly Employment in Metals Industries (top) and Monthly differences (bottom)

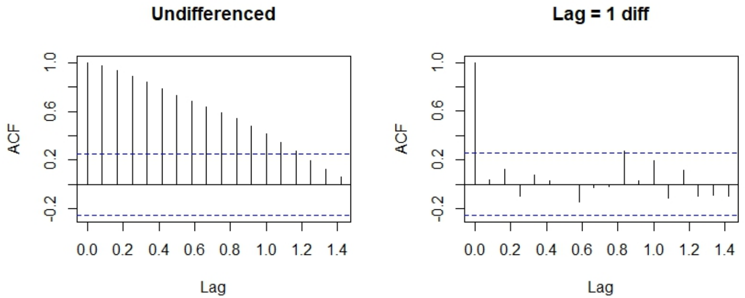



Figure 2.9: Autocorrelation function for Monthly Employment in Metals Industries (left) and Monthly differences (right)

For seasonal time series with period, $S$, the **seasonal difference operator**, $\nabla_S$, can be useful in obtaining a series that is free of seasonal patterns and is stationary. This is defined as

$$\nabla_S = (1 - B^S)$$

and its action on a time series $\{Y_t\}$ is

$$X_t = \nabla_S Y_t = (1 - B^S)Y_t = Y_t - B^S Y_t = Y_t - Y_{t-S}$$

which are referred to as **seasonal differences**.

**Example 2.12 (Seasonally differenced series)** *Figure 2.10 is a time series plot of employment in Trades in Wisconsin measured each month over five years (top panel), the seasonal (lag=12) differenced data (middle panel) and the seasonal differenced followed by lag one differenced data (bottom panel).*

*Figure 2.11 shows (left panel) estimates of the autocorrelation function for the time series of employment in Trades in Wisconsin, that of the seasonal lag ($= 12$) differenced data (middle panel), and that of the seasonally differenced then lag 1 differenced data (right panel). Note that lag 12 differencing appears to remove the seasonal pattern obvious in the original series but that there is evidence of non-stationarity of the mean for these differences which is removed when additional lag 1 differencing is applied.*

***Note 2.3*** *The differencing operators commute: $\nabla_S \nabla = \nabla \nabla_S$ so it does not matter in which order they are applied.*

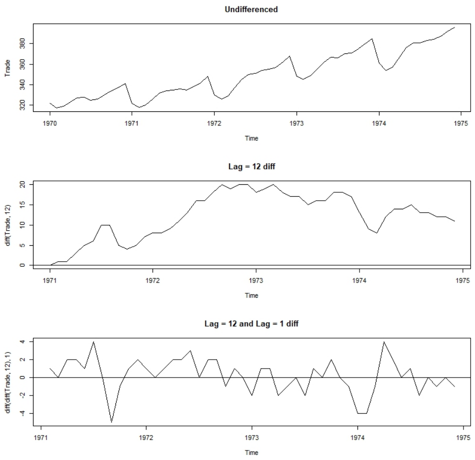



Figure 2.10: Monthly Employment in Trades in Wisconsin (top), Seasonal differences (middle) and combined seasonal and lag 1 differences (bottom)

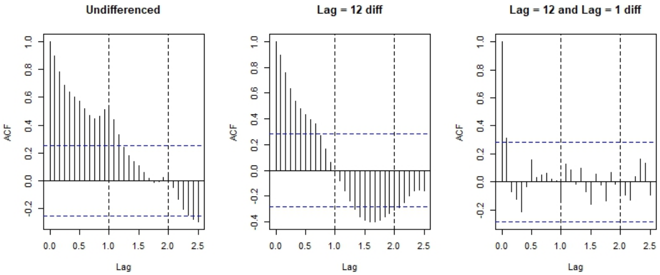



Figure 2.11: Autocorrelation function for Monthly Employment in Trades (left) and seasonal (lag =12) differences (middle) and combined seasonal and monthly differences (right)