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

In this section, we will explore several motivating examples in time series analysis to demonstrate its practical applications and key concepts.

Example 2.1 (European countries retail trade, Lima et al. [2023]) *The trade sector is crucial to Europe's economy. In 2015, it employed 33 million people and accounted for 9.9% of the EU's total gross value added. Retail trade is vital, complex, and varies with cultural and social influences. It plays a significant role in understanding a country's economic health. In 2015, retail trade comprised nearly one-third of the trade sector's turnover, employed 8.7% of the EU workforce, or roughly 18.8 million jobs.*

The dataset `TOVT_2015.csv` *covers the total turnover indices of 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 value-added tax (VAT) and other similar taxes. The base year for comparison is 2015, with an index value of 100. Figure 2.1 shows the time series for some European countries.*

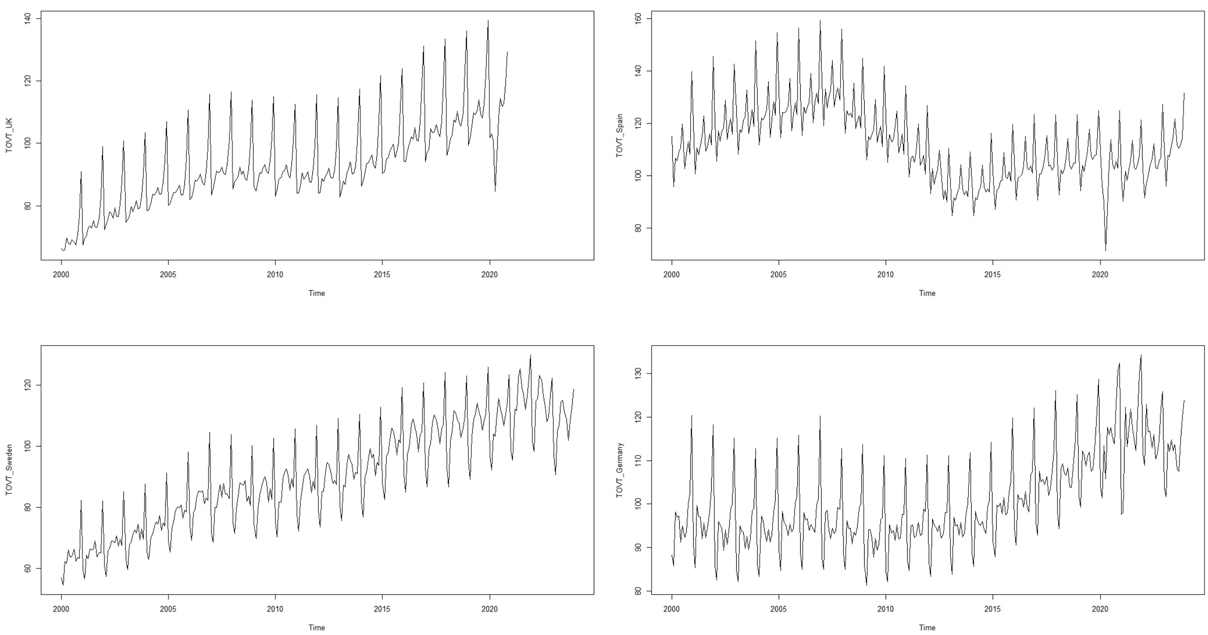

Figure 2.1: Index of total turnover in the context of retail trade for UK, Spain, Sweden and Germany.

Example 2.2 (Sunspot Data, Brockwell and Davis [2002, 1991]) *The Wolf number, also known as the relative sunspot number, quantifies the presence of sunspots and sunspot groups on the Sun's surface. Astronomers have been observing sunspots since the invention of the telescope in 1609. However, the idea of compiling sunspot data from various observers originated with Rudolf Wolf in 1848. The series is now commonly referred to as the international sunspot number series, which exhibits an approximate 11-year periodicity, corresponding to the solar cycle. This cycle was first discovered by Heinrich Schwabe in 1843 and is sometimes called the Schwabe cycle. 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, with an overall*

increase by a factor of 1.6 to the entire series. Modern counts now closely reflect raw values, without the traditional scaling applied after 1893.
The file named `Sunspot.csv` *presents the yearly mean total sunspot number from 1700 to 2023 and its contents are structured as follows:*

- *Column 1: contains the Gregorian calendar year, marking the mid-year date.*
- *Column 2: records the yearly mean of total number of sunspots.*
- *Column 3: details the yearly mean standard deviation of the sunspot numbers collected from individual stations.*
- *Column 4: lists the number of observations used to compute the yearly mean of total sunspot number.*
- *Column 5: serves as a definitive/provisional marker, where '1' indicates a definitive value and '0' denotes a provisional value.*

Figure 2.2 presents this dataset.

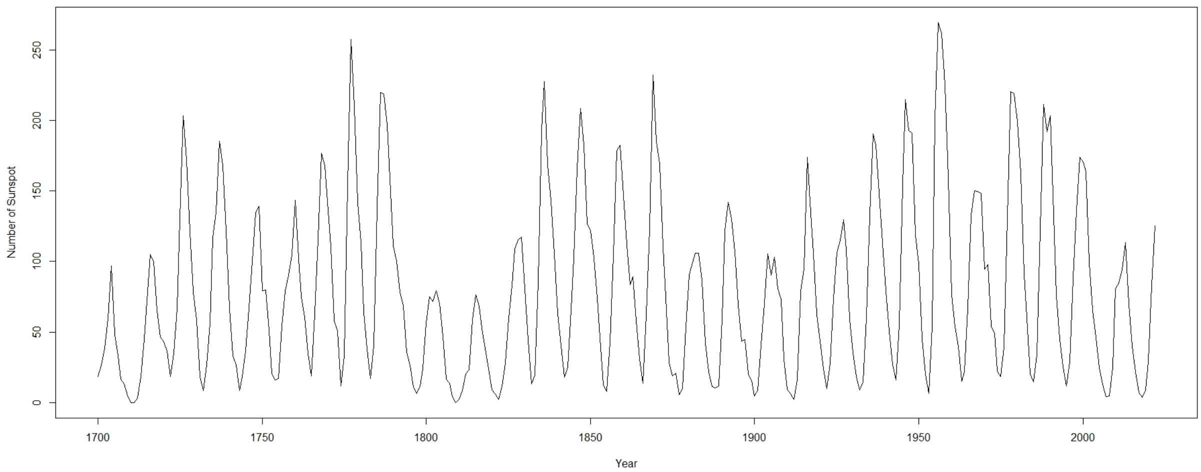

Figure 2.2: The sunspot numbers, 1700-2023.

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. We note the repetitive nature of the signal and the rather regular periodicities, along with a decrease in the power as the time increases. One current problem of great interest is the recognition of speech by computers, which would require converting this particular signal into the recorded phrase aaa ... hhh. Spectral analysis can be used in this context to produce a signature of this phrase that can be compared with signatures of various library syllables to look for a match. 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 stimulated by the opening and closing of the glottis.*

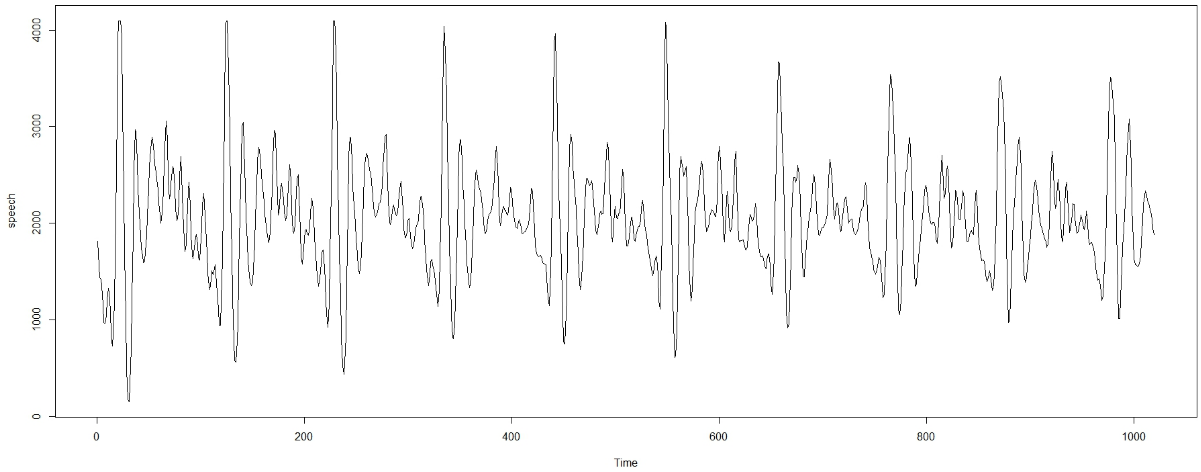

Figure 2.3: Speech recording of the syllable aaa ... hhh sampled at 10,000 points per second with n = 1020 points

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 toward the end of the 20th century, providing evidence in support of 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.*

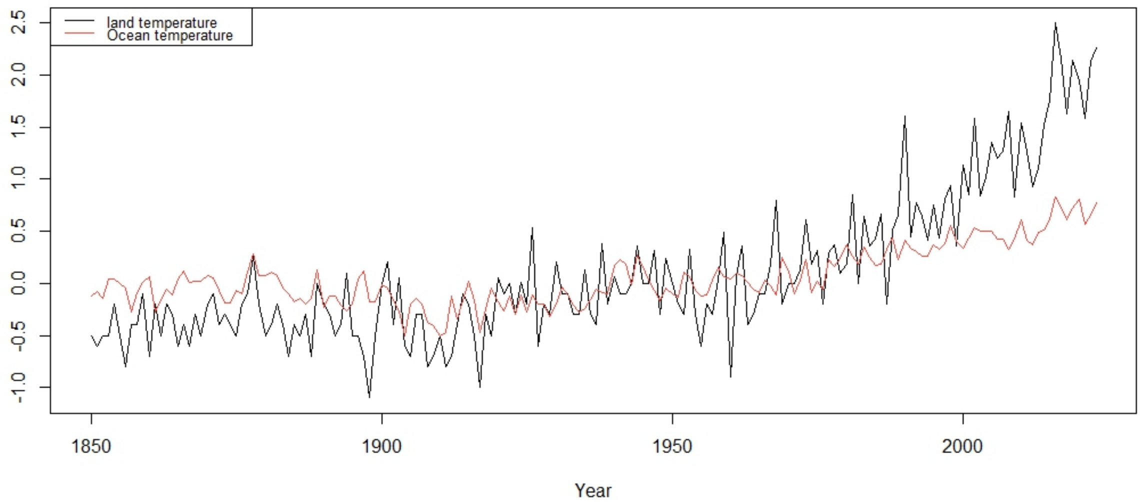

Figure 2.4: Yearly average land-ocean temperature deviations (1850–2023) in degrees centigrade

2.2 Discrete time series: Definition and Basics

The primary objective of time series analysis is to develop mathematical models that provide plausible descriptions of the data. To account for the possibly unpredictable nature of future observations, it is natural to 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 $\mathbb{R} = (-\infty, \infty)$.*

- *A discrete time, time series ("**discrete time series**" for short) is a collection of random variables $\{X_t\}$ indexed by $t \in \mathbb{Z}$. In this course we will exclusively focus on discrete time series.*

- *A continuous time, time series ("continuous time series" for short) is a collection of random variables $\{X_t\}$ indexed by $t \in \mathbb{R}$.*

The probabilistic properties of these types of stochastic processes need to be prescribed for inference and forecasting. A complete probabilistic specification requires the specification of the joint distributions of any collection of random variables $\mathbf{X} = (X_{t_1}, \ldots, X_{t_n})'$ where $t_1 < t_2 < \ldots < t_n$ is any collection of ordered integers. Unfortunately, these multidimensional distribution functions cannot usually be written easily unless the random variables are jointly Gaussian (normal), in which case the joint density has the well-known form. Along with being Gaussian, such a specification can be considerably simplified by the notion of stationarity, which is defined in this chapter. Later in the course, we will consider models for which these properties are relaxed.

Let $\{X_t\}$ be a time series with $E(X_t^2) < \infty$ for all $t \in \mathbb{Z}$. We define the mean and covariance functions as follows.

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 numerically, as in classical statistics, using the notions of covariance and correlation. Assuming that the variance of X_t is finite, we have the following definition.

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*

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

When there is no possible confusion about which time series we are referring to, we will drop the subscript and write $\gamma_X(r, s)$ as $\gamma(r, s)$. Note that $\gamma_X(r, s) = \gamma_X(s, r)$ for all time points r and s.

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

- Very smooth series exhibit covariance functions that remain large even when the s and r are far apart.
- Choppy series tend to have covariance functions that are nearly zero for large separations.

Note 2.1 *Recall from classical statistics that if $\gamma_X(r,s) = 0$, X_r and X_s are not linearly related, but there may still be some dependence structure between them. However, if X_r and X_s are bivariate normal, $\gamma_X(r,s) = 0$ ensures their independence.*

Note 2.2 *For $r = s$, the covariance reduces to the (assumed finite) variance, because $\gamma_X(r,r) = E[(X_r - \mu_X(r))^2] = var(X_r)$.*

Note 2.3 *Sometimes in the analysis of 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., the expected values $E(X_t)$ and the expected products $E(X_r X_s)$. Such properties of $\{X_t\}$ 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, let $t_1 < t_2 < \ldots < t_n$ be a collection of ordered integers. 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. This means that X_t is a process such that*

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

(ii) the mean value function, μ_X, is constant and does not depend on time t, i.e.,

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

(iii) For $t \in \mathbb{Z}$, $\gamma_X(t+h, t) = \gamma_X(h)$, meaning that 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**.

Definition 2.5 (Autocovariance and autocorrelation functions) *For a weakly stationary time series, we define the autocovariance function (ACVF) as* $\{\gamma_X(h):\ h = 0, \pm 1, \ldots\}$ *and the autocorrelation function (ACF) as*

$$\rho_X(h) = \frac{\gamma_X(t+h,t)}{\sqrt{\gamma_X(t+h,t+h)\gamma_X(t,t)}} = \frac{\gamma_X(h)}{\gamma_X(0)}, \quad h = 0, \pm 1, \ldots. \tag{2.3}$$

Some general properties of autocovariance functions: The autocovariance or, equivalently, the autocorrelation function describes the second order dependence properties in a stationary time series.

Theorem 2.1 *The autocovariance function for stationary time series satisfies the following properties:*

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* $\sum_{j=1}^{n}\sum_{k=1}^{n} a_j a_k \gamma(j-k) \geq 0$.

Example 2.5 (White Noise) *A simple kind of generated series might be a collection of uncorrelated random variables,* W_t*, with mean 0 and finite variance* σ_W^2*. The time series generated from uncorrelated variables is used as a model for noise in engineering applications, where it is called* **white noise***; denoted as* $W_t \sim WN(0, \sigma_W^2)$.
We will sometimes require the noise to be independent and identically distributed (iid) random variables with mean 0 and variance σ_W^2*. We distinguish this by writing* $W_t \sim iid(0, \sigma_W^2)$ *or by saying* **white independent noise** *or* **iid noise***. A particularly useful white noise series is* **Gaussian white noise***, wherein the* W_t *are* **independent normal** *random variables, with mean 0 and variance* σ_W^2*; or more succinctly,* $W_t \sim N(0, \sigma_W^2)$*. Figure 2.5 demonstrates 500 observations from a Gaussian white noise.*

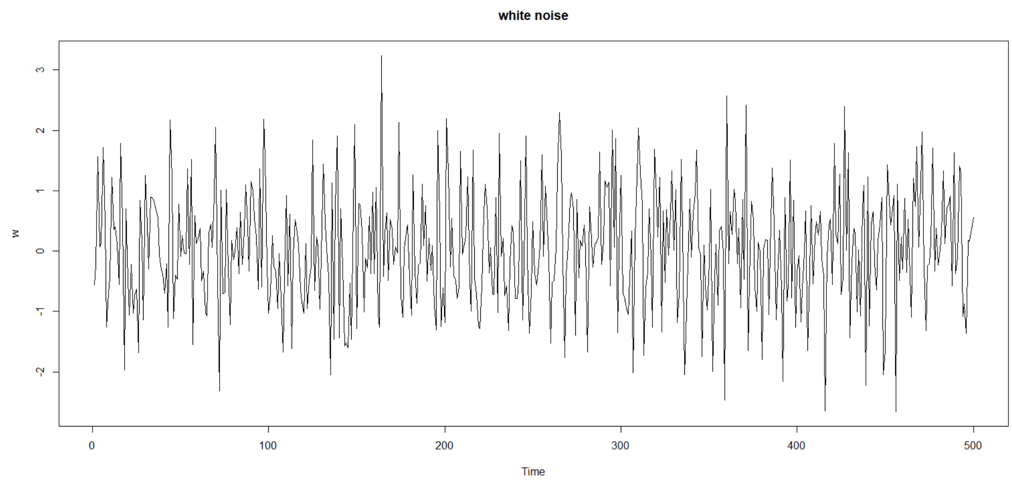

Figure 2.5: Gaussian white noise series

Based on the definition of white noise processes, $E(W_t) = 0$ and by uncorrelatedness of the observations, we have

$$\gamma_W(s,t) = cov(W_s, W_t) = \begin{cases} \sigma_W^2 & s = t \\ 0 & s \neq t \end{cases} \tag{2.4}$$

Consequently, white noise processes are weakly stationary.

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

$$X_t = \delta + X_{t-1} + W_t \tag{2.5}$$

successively for $t = 1, 2, \ldots$ with initial condition $X_0 = 0$, and where W_t is white noise. The constant δ is called the drift, and when $\delta = 0$, (2.5) is called simply 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. Note that we may rewrite (2.5) as a cumulative sum of white noise variates:

$$X_t = t\delta + \sum_{i=1}^{t} W_i \tag{2.6}$$

for $t = 1, 2, \ldots$. Since $E(W_t) = 0$, for all t, and δ is constant, we have

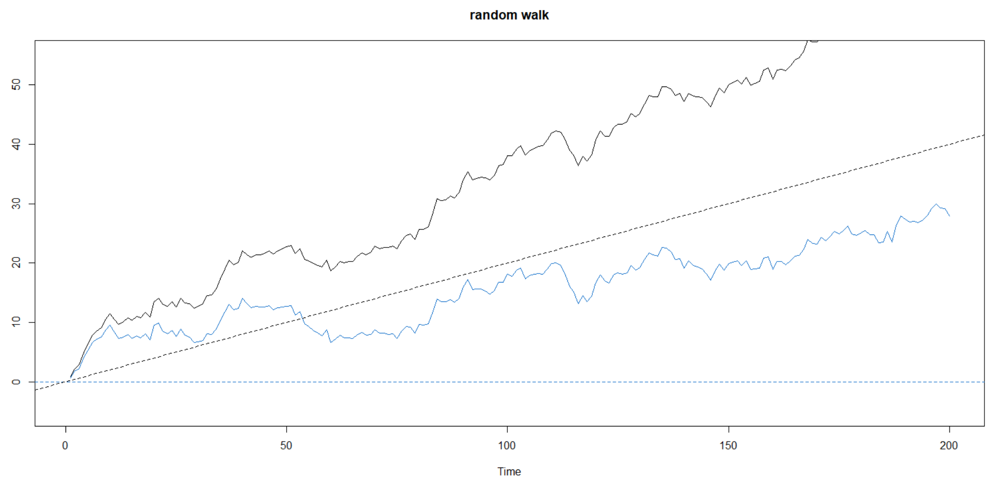

Figure 2.6: Random walk, $\sigma_W = 1$, with drift $\delta = 0.2$ (upper jagged line), without drift, $\delta = 0$ (lower jagged line), and straight (dashed) lines with slope δ.

$$\mu_X(t) = E(X_t) = \delta t + \sum_{i=1}^{t} E(W_i) = \delta t,$$

which is a function of t. Besides, the autocovariance function can be calculated as follows:

$$\begin{aligned} cov(X_r, X_s) &= cov(\delta s + \sum_{i=1}^{r} W_i, \delta t + \sum_{j=1}^{s} W_j) \\ &= cov(\sum_{i=1}^{r} W_i, \sum_{j=1}^{s} W_j) \\ &= \sum_{i=1}^{\min\{r,s\}} var(W_i) \\ &= \min\{r, s\}\sigma_W^2 \end{aligned}$$

because W_t are uncorrelated random variables. Note that, as opposed to the previous example, the autocovariance function of a random walk depends on the particular time values r and s, and not on the time separation or lag. Also, notice that the variance of the random walk, $var(X_t) = t\sigma_W^2$, increases without bound as time t increases. The effect of this increase in variance can be seen in Fig. 2.6 where the processes start to move away from their mean functions δt.

Example 2.7 (Signal in Noise) *Many realistic models for generating time series assume an underlying signal with some consistent periodic variation, contaminated by adding a random noise (As an example, refer to Example 1.6 fMRI Imaging, in Shumway et al. [2000]). Consider the model*

$$X_t = 2\cos\left(\frac{2\pi(t+15)}{50}\right) + W_t \tag{2.7}$$

for $t = 1, 2, \dots$, where the first term is regarded as the signal, shown in the upper panel of Fig. 2.7. We note that a sinusoidal waveform can be written as

$$A\cos(2\pi\omega t + \phi), \tag{2.8}$$

where A is the amplitude, ω is the frequency of oscillation, and ϕ is a phase shift. In (2.7), $A = 2$, $\omega = 1/50$ (one cycle every 50 time points), and $\phi = 2\pi 15/50 = 0.6\pi$. It is clear, because the signal in (2.7) is a fixed function of time, we will have

$$\begin{aligned} \mu_X(t) = E(X_t) &= 2\cos\left(\frac{2\pi(t+15)}{50}\right) + E(W_t) \\ &= 2\cos\left(\frac{2\pi(t+15)}{50}\right). \end{aligned}$$

Therefore, the mean function is just the cosine wave, which depends on t. So, X_t is not stationary.

Example 2.8 (Trend Stationarity) *If $X_t = \alpha + \beta t + Y_t$, where Y_t is stationary, then the mean function is $\mu_X(t) = \alpha + \beta t + \mu_Y(t)$, which is not independent of time. Therefore, the process is not stationary. However, $cov(X_r, X_s)$ is independent of time, because*

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

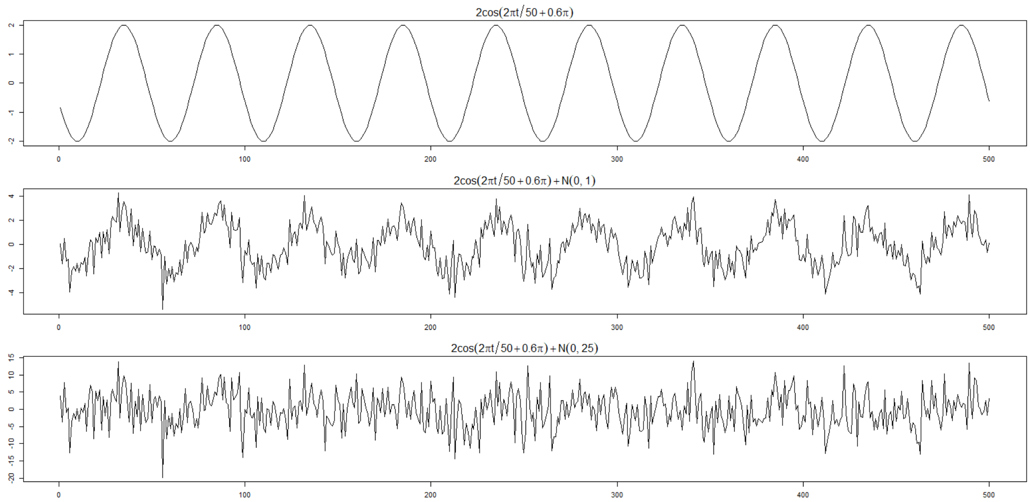

Figure 2.7: Cosine wave with period 50 points (top panel) compared with the cosine wave contaminated with additive white Gaussian noise, $\sigma_W = 1$ (middle panel) and $\sigma_W = 5$ (bottom panel)

Thus, the model may be considered as having stationary behavior around a linear trend; this behavior is sometimes called **trend stationarity**. *Can you suggest a transformation to make X_t stationary?*

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*

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

That is,

$$P(X_{t_1} \le c_1, X_{t_2} \le c_2, \ldots, X_{t_n} \le c_n) = P(X_{t_1+h} \le c_1, X_{t_2+h} \le c_2, \ldots, X_{t_n+h} \le c_n\}) \qquad (2.9)$$

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 2.4

1. *A strictly stationary time series with finite second moment $E(X_t^2) < \infty$ is also weakly stationary.*
2. *There exist strictly stationary processes which are not weakly stationary.*
3. *The converse is not true: weak stationarity does not imply strict stationarity in general (Example 2.9).*
4. *However, for a Gaussian time series, weak and strict stationarity are equivalent.*
5. *For a weakly stationary time series, the mean vector and covariance matrices of $\mathbf{X} = (X_{t_1}, \ldots, X_{t_n})'$ simplify to*

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

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

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

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.*

2.3 Estimation of Means and Covariances.

In this section,, we consider estimation of means and autocovariance for a *weakly stationary time series.* We also describe their large sample ($n \to \infty$) distributions and a test statistic, which is commonly used 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 μ and ACVF $\gamma(h)$. The sample mean

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

is an unbiased estimator of μ, i.e., $E(\bar{X}_n) = \mu$, and it has variance given by

$$var(\bar{X}_n) = \frac{1}{n} \sum_{h=-n+1}^{n-1} (1 - \frac{|h|}{n})\gamma(h). \tag{2.10}$$

Now, if $\gamma(h) \to 0$ as $h \to \infty$, the right-hand side of (2.10) converges to zero, so that $\bar{X}_n$ converges in mean square to μ. If $\sum |\gamma(h)| < \infty$, then (2.10) 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

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

and the estimator of the ACF is

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

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. This is easy to show. Let $Y_i = X_i - \bar{X}_n$ and
$$T = \begin{pmatrix} 0 & \cdots & 0 & Y_1 & \cdots & Y_{n-1} & Y_n \\ 0 & \cdots & Y_1 & 0 & \cdots & Y_n & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ Y_1 & \cdots & Y_{n-1} & Y_n & \cdots & 0 & 0 \end{pmatrix}.$$
be an $n \times (2n-1)$ dimensional matrix. It is easy to show
$$\hat{\Gamma}_n = \frac{1}{n} TT',$$
which is positive definite ($a'TT'a > 0, \quad$ for all a) unless all the data are identical in value. 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

$$X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j W_{t-j}, \qquad \sum_{j=-\infty}^{\infty} |\psi_j| < \infty, \tag{2.11}$$

For the linear process, we may show that

$$\gamma_X(h) = \sigma_W^2 \sum_{j=-\infty}^{\infty} \psi_{j+h}\psi_j \tag{2.12}$$

for $h \geq 0$; recall that $\gamma_X(-h) = \gamma_x(h)$.
Notice that the linear process (2.11) is dependent 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. Consequently, we will focus on processes that do not depend on the future. Such models are called **causal** (Will be explained more accurately in the next chapters).

For linear processes, it can be shown that 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*

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

Therefore, for large n, $\hat{\rho}(1), \cdots, \hat{\rho}(k)$ are approximately independent and identically distributed with $\hat{\rho}(j) \sim AN(0, \frac{1}{n})$, $j = 1, \cdots, k$. Note that "AN" stands for approximately normally distributed.
This property helps us to obtain a rough method of assessing whether peaks in $\hat{\rho}(h)$ are significant by determining whether the observed peak is outside the interval $\pm 1.96/\sqrt{n}$ (Why?). For a white noise sequence, approximately 95% of the sample ACFs should be within these limits. The applications of this property develop because 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.

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

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

and this has a χ^2 distribution with h degrees of freedom. 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. Therefore, we reject the white noise hypothesis at level α if $Q > \chi^2_{1-\alpha}(h)$, where $\chi^2_{1-\alpha}(h)$ is the $1 - \alpha$ quantile of the chi-squared distribution with h degrees of freedom.

A modification of this test statistic, which is better approximated by the χ^2 distribution, is the Ljung-Box statistic:

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

2.4 Simple ways to obtain stationarity

Nonstationarity arises in many ways. The first obvious way in which a process is non-stationary is when the mean function $\mu_X(t) \neq \mu_X$ for all t. For example it could be a deterministic trend in time, such as linear (Example 2.8), exponential, polynomial etc., or a function of other processes which are not stationary. Another obvious way in which a process is non-stationary is when the variance is not constant through time or autocovariance function does not only depend on time separation in the process. Also the distributions themselves might not be time invariant. Besides, as mentioned before, a random walk and its variants are non-stationary, in which the current value of the process is a random increment from the previous value of the process. We will encounter these types of non-stationarity throughout the course and learn methods to deal with the most common situations.

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. Examples follow.

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. Examples will be used to illustrate this. First some notation.

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

$$BX_t = X_{t-1}$$

and iterating this we get

$$B^j X_t = X_{t-j}$$

which shifts the series backwards in time by j time units.

The difference operator, ∇, 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 for the time series of employment in the metal industries in Wisconsin measured each month over five years as well as that for the lag 1 differences (right panel). Note that lag 1 differencing appears to remove the smooth underlying trend obvious in the original series.

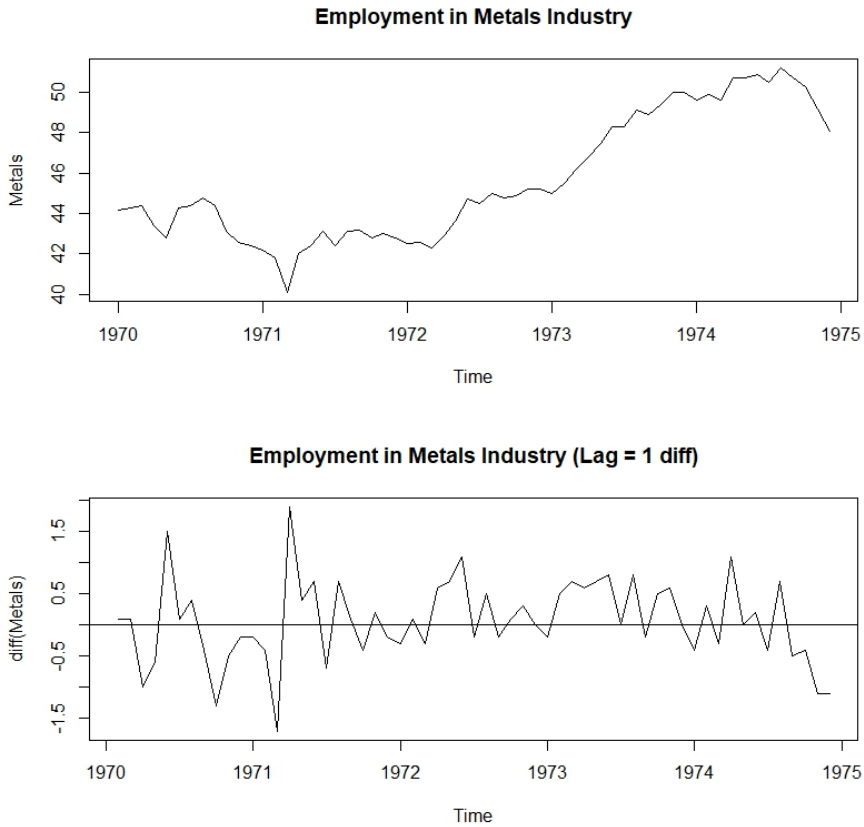

Figure 2.8: Monthly Employment in Metals Industries (top) and Monthly differences (bottom)

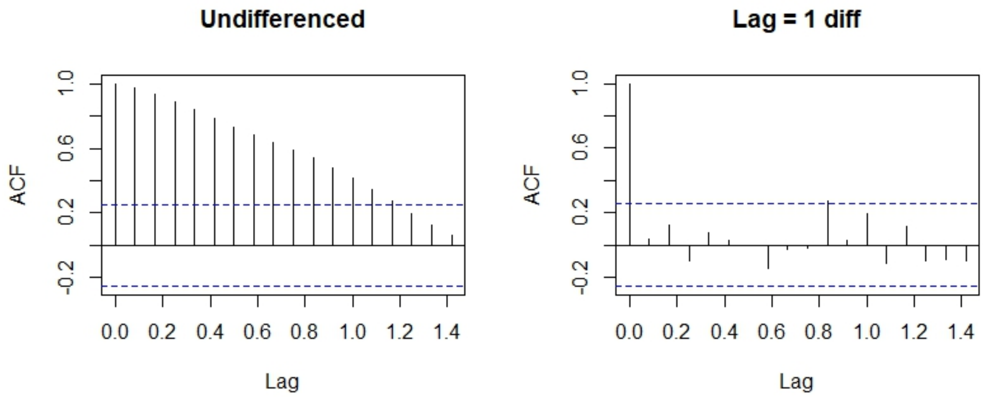

Figure 2.9: Autocorrelation function for Monthly Employment in Metals Industries (left) and Monthly differences (right)

For seasonal time series with period, S, the **seasonal difference operator**, ∇_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).*

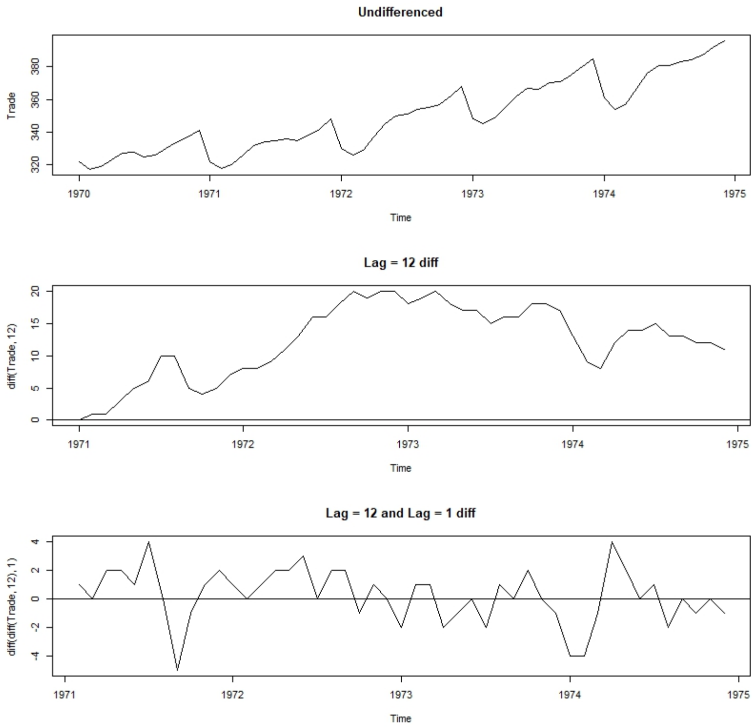

Figure 2.10: Monthly Employment in Trades in Wisconsin (top), Seasonal differences (middle) and combined seasonal and lag 1 differences (bottom)

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

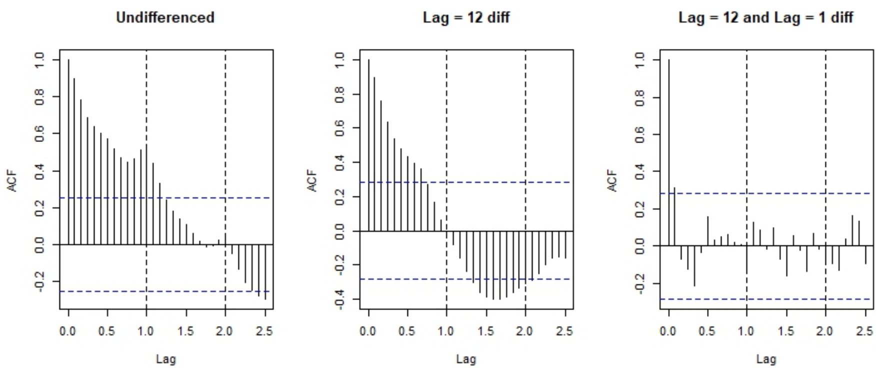

Figure 2.11: Autocorrelation function for Monthly Employment in Trades (left) and seasonal (lag =12) differences (middle) and combined seasonal and monthly differences (right)

2.5 Exercises

Exercise 2.1 *Prove Theorem 2.1.*

Exercise 2.2 *Based on Theorem 2.1, what are the properties of autocorrelation function $\rho_X(\cdot)$?*

Exercise 2.3 *In Example 2.7, find $cov(X_r, X_s)$?*

Exercise 2.4 *Derive formula (2.10) for $var(\bar{X}_n)$.*

Exercise 2.5 *Let $Z_t \sim$i.i.d. $N(0,1)$ and define the process*

$$X_t = Z_t Z_{t-1} \text{ for } t = \ldots, -1, 0, 1, \ldots$$

i. *Is $\{X_t\}$ a weakly stationary time series? To establish this you will need to calculate the mean and covariance functions for $\{X_t\}$.*

ii. *Is $\{X_t\}$ strictly stationary? A sequence of identically distributed random variables?*

iii. *Is $\{X_t\}$ a white noise process?*

iv. *Let $W_t = X_t^2$ be the squared process. Show that $E(W_t) = 1$, $var(W_t) = 8$, $cov(W_t, W_{t+l}) = 0, |l| > 1$, $cov(W_t, W_{t+l}) = 2, |l| = 1$. [Hint: $E(Z_t^4) = 3$.]*

v. *Calculate the autocorrelations for the squared process $\{X_t^2\}$.*

vi. *Is $\{X_t\}$ an independent sequence? Give reasons.*

vii. *Is $\{X_t\}$ a Gaussian process? Give reasons.*

2.6 Tutorial: Week 1

2.6.1 Use of R for Basic Time Series Analysis

The use of R for time series analysis is growing very rapidly and the number of available libraries, packages and functions is getting very large. This course will introduce some of these.

This section aims to introduce you to the use of R for basic time series analysis - plotting, smoothing, least squares regression and estimation of autocorrelations using two examples.

First, we look at the basics.

Getting Started

As a starting point we will access the `R` material that is provided by David Stoffer for his book, Shumway and Stoffer (2017) Time Series Analysis and its Applications: with R Examples, 4rd Edition, Springer-Verlag. This is one of the recommended texts for this course.

First steps:

1. Install `R`. It runs on Linux, Windows and MacOS. You can find the link to download `R` and some more resources here.

2. I strongly recommend you use RStudio (unless you have other preferences), which can be installed from Rstudio website.

3. Familiarize yourselves with the various material available through links at Time Series Analysis and Its Applications homepage.

4. Invoke `R` or `RStudio` and start using it. Decide on the directory in which you want to do your work and change to that directory.

5. `R` Script files: It is strongly recommended that you get in the habit of using the script file feature of `R`. You can start a new script file (a plain text file that has an extension '.r' or '.R') easily with the pull down menus in R or R Studio. Create your R commands in that file (save it periodically - the best place may be the same directory as the current workspace) and run all commands or a selection as you need to. In this way you will be creating and saving `R` program scripts which can be recalled, edited or expanded in a subsequent session. By the end of the course you should have quite a collection of such scripts.

6. R Markdown: I will be using this for many demonstrations and you should also consider this - it is easy to use and produces nice enough reports. If you are a latex user you can also write maths in your reports using `R` Markdown. To start with, you can find some Tutorial and cheat sheet for R markdown here. Another, shorter, tutorial is available to help you get started.

Useful R Functions for This Week

- **Reading the data from a file**:
  - *read.csv*

- **Define the data as time series**
  - *ts*
- **Plot the time series**
  - *plot*
  - *ts.plot*
  - *plot.ts*
- **Differencing**
  - *diff*
- **(Partial) Auto-Correlation Function**
  - *acf*
  - *pacf* refers to Partial Auto-Correlation Function. We will learn more about this function in the coming weeks.
  - *acf1* from *astsa* package
  - *acf2* from *astsa* package

2.6.2 Practical Exercises

Exercise 2.6 (Analysis of Iowa Wine Sales using `R` Markdown) *In the United States the alcohol industry has traditionally been a heavily regulated industry. Wine and spirits were sold only in a limited number of state-owned and operated stores for as many as fifty years after the end of Prohibition. Over the past thirty to forty years several states have abandoned their retail monopoly structure, with the extent and timing of the deregulation varying between state to state. Consequently, abrupt and dramatic increases in alcohol availability ensued. In Iowa for instance, retail outlets rapidly increased from 200 state stores to 800 private wine outlets and 400 spirits outlets on the 1st of July 1985 (time of intervention). Nearly all grocery and convenience stores entered the market, Sunday sales were legalised, hours of sales were extended, advertising was allowed and alcohol could be purchased on credit.*

Data on monthly consumption of wine in Iowa was obtained from January 1968 to December 1997 ($n = 360$). Source of wine data is the Wine Institute, based on reports from the Bureau of Alcohol, Tobacco and Firearms, U.S. Department of the Treasury. You need to download the data set `IOWA.csv` (a comma seperated file) from Moodle.

The file `IowaWineAnalysis.Rmd` is an `R Markdown` script for producing some graphs for this dataset.

(i) Initially execute each code chunk in sequence, examine the output (all graphs in this case) and write some brief summary notes on what you have done and what you observe.

(ii) Create a new `R markdown` script which reads in the data and presents a time series plot and an autocorrelation function for the combined lag 1 and lag 12 differences of log of wine sales per head for the post-intervention period (from July 1985).

(iii) *Compare the new graphs with the corresponding graphs of pre-intervention data produced by the commands in part (i) and give a brief summary of your conclusions. You need to compare the time series graphs and their autocorrelation functions noting any differences and similarities between the time series plots and autocorrelation functions between the pre- and post- intervention data.*

Exercise 2.7 (Trends in Global Temperatures) *The data set* `gtemp_both` *presents "annual temperature anomalies (in degress centigrade) averaged over the Earth's land and ocean area from 1850 to 2023. Anomalies are with respect to the 1991-2020 average". One primary question is whether there is a significant (statistically and physically) upward trend in the data.*

(i) *Plot the data and summarize the basic features of this graph.*

(ii) *Consider fitting a straight line regression to these data. Review the output appearing in the* `R` *Console, which summarizes the regression fit. In particular consider the significance of the trend on time.*

(iii) *List the assumptions that are required for valid use of the t-statistic in assessing the significance of regression coefficients. In particular do you think that the assumption of independent residuals is valid? Assess this using the time series plot and autocorrelation function of the residuals.*

In particular, answer the following questions:

(a) *Do you think that the residuals from fitting a straight line are independent?*

(b) *Describe any dependence in terms of the behaviour of the autocorrelation function.*