Chapter 1

Time Series Modelling

1.1 Introduction

An observed *time series* is a set of observations $(x_t)_{t\in\mathcal{T}}$ where $\mathcal{T}$ denotes an indexing set of finite size; each observation x_t is recorded at a specific time, related to the index t. x_t may be vector-valued if we are considering a multivariate time series, where the components may be influencing each other.

This course deals with Time Series models and their applications. We consider four areas of application:

1. The *forecasting* of future values of a time series from current and past values.

2. Computing a *transfer function* of a system, which shows the effect on the output of a system on any given series of inputs.

3. The use of *indicator* input variables in transfer function models to represent and assess the effects of unusual *intervention* events on the behaviour of a time series.

4. Examining relationships between several related time series of interest and establishing multivariate dynamic models to represent these joint relationships over time.

For now, we consider *univariate* time series; each $x_t \in \mathbb{R}$. We consider vector valued time series later.

1.2 Time Series Models

Definition 1.1 (Time Series Model)**.** *A* time series model *for the observed data* $\{x_t : t \in \mathcal{T}\}$ *is the hypothesis that the observed data is an observation of a sequence of random variables* $\{X_t : t \in \mathcal{T}\}$ *and the specification of its joint probability distribution, or possibly only its expectations and covariances.*

A time series can only be observed at a finite number of times, $(x_t)_{t=1}^n$ and the n observations are a realisation of an n dimensional random vector $X = (X_1, X_2, \ldots, X_n)$. These random variables may be considered to come from an infinite sequence $\{X_t,\, t \in \mathbb{Z}_+ \quad \text{or} \quad \mathbb{Z}\}$, a *stochastic process*.

Definition 1.2 (Stochastic Process)**.** *A* stochastic process *is a family of random variables* $\{X_t : t \in \mathcal{T}\}$*, indexed by a set* $\mathcal{T}$*, which is defined on a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$*.*

Example 1.1 (The binary process)**.**

A simple example of a stochastic process $\{X_t,\, t \in \mathbb{Z}_+\}$ is a process where the variables are i.i.d. (independent identically distributed) satisfying

$$\mathbb{P}(X_t = 1) = \mathbb{P}(X_t = -1) = \frac{1}{2}.$$

For this process, the finite dimensional marginals are well defined; for any $i_1 < \ldots < i_n$,

$$\mathbb{P}(X_{i_1} = j_1,\, X_{i_2} = j_2,\, \ldots, X_{i_n} = j_n) = 2^{-n}$$

for any $\{j_1, \ldots, j_n\} \in \{-1, 1\}^n$. □

Definition 1.3 (IID noise)**.** *A process* $\{X_t,\, t \in \mathbb{Z}\}$ *is said to be an IID noise with mean* μ *and variance* σ^2*, written*

$$\{X_t\} \sim IID(\mu, \sigma^2),$$

if the random variables X_t *are independent and identically distributed with* $\mathbb{E}[X_t] = \mu$ *and* $Var(X_t) = \sigma^2$*.*

Usually, we are interested in $\text{IID}(0, \sigma^2)$ noise.

Notation Througout, Var(.) will be used to denote variance.
The binary process is clearly an example of an $\text{IID}(0, 1)$ noise, since the variables are independent, $\mathbb{E}[X_t] = -1 \times \frac{1}{2} + 1 \times \frac{1}{2} = 0$ and $\text{Var}(X_t) = \mathbb{E}[X_t^2] - \mathbb{E}[X_t^2] = \mathbb{E}[X_t^2] = 1$.

In many situations, the complete specification of the underlying stochastic process is not required; the methods will generally rely only on its means and covariances. Sometimes even less general assumptions are needed, but these will not be treated here.

Definition 1.4 (Mean function, Covariance function)**.** *Let* $\{X_t,\, t \in \mathcal{T}\}$ *be a stochastic process with* $Var(X_t) < \infty$ *for each* $t \in \mathcal{T}$*. The* mean function *of* $\{X_t\}$ *is denoted by* μ_X*, or simply* μ *when there is no danger of ambiguity:*

$$\mu_X(t) := \mathbb{E}[X_t], \quad t \in \mathcal{T} \tag{1.1}$$

The covariance function *of* $\{X_t\}$ *is denoted by* C_X *or* C *when there is no danger of ambiguity and is defined as:*

$$C_X(r, s) := Cov(X_r, X_s), \quad r, s \in \mathcal{T}. \tag{1.2}$$

The symbol Cov will be used to denote covariance.

1.3 Stationarity and Strict Stationarity

A stochastic process is said to be *stationary*, if its statistical properties do not change with time. Formally, stationarity is defined in the following way.

Definition 1.5 (Stationary, Strictly Stationary, Wide sense stationary). *A time series* $\{X_t,\, t \in \mathbb{Z}\}$ *is said to be* weakly stationary*, or* wide sense stationary*, or simply* stationary *if*

1. $Var(X_t) < \infty \quad$ *for all* $t \in \mathbb{Z}$,
2. $\mu_X(t) = \mu \quad$ *for all* $t \in \mathbb{Z}$,
3. $C_X(r, r+h) = C_X(0, h) \quad$ *for all* $r, h \in \mathbb{Z}$.

A process is said to be strictly stationary *if any finite collection* $(X_{n_1}, \ldots, X_{n_k})$ *has the same distribution as* $(X_{n_1+t}, \ldots, X_{n_k+t})$ *for any* $k \geq 1$ *and any* $(n_1, \ldots, n_k, t) \in \mathbb{Z}$.

In many practical situations, only weak stationarity is considered; usually only expectation and covariance, at most, can reasonably be assessed from data. In some situations, though (for example ARCH and GARCH processes, which arise in the analysis of financial time series) it is worthwhile placing additional modelling assumptions on the data generation mechanism.

The third point in the definition of weak stationarity implies that $C_X(r, s)$ depends on r and s only through $r - s$. It is therefore convenient to define

$$\gamma_X(h) := C_X(h, 0).$$

When only one time argument appears in γ, then it denotes the autocovariance (ACVF) function of a stationary process. The value h is referred to as the *lag*.

Definition 1.6. *Let* $\{X_t,\, t \in \mathbb{Z}\}$ *be a stationary time series. The* autocovariance function (ACVF) *of* $\{X_t\}$ *is defined as*

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

The autocorrelation function (ACF) *is defined as:*

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

A simple example of a stationary process is the so-called *white noise.*

Definition 1.7 (White noise). *A process* $\{X_t,\, t \in \mathbb{Z}\}$ *is said to be a white noise with mean* μ *and variance* σ^2*, written*

$$\{X_t\} \sim WN(\mu, \sigma^2),$$

if $\mathbb{E}[X_t] = \mu$ *for all* $t \in \mathbb{Z}$ *and*

$$\gamma(h) = \begin{cases} \sigma^2 & \textit{if } h = 0, \\ 0 & \textit{if } h \neq 0. \end{cases}$$

Note that IID noise is an example of white noise, but not necessarily vice versa; the underlying distribution can be different even if the mean and covariance structures are the same; strictly stationary time series $\{X_t,\, t \in \mathbb{Z}\}$ with $\text{Var}(X_t) < \infty$ is stationary, but a stationary time series $\{X_t,\, t \in \mathbb{Z}\}$ does not need to be strictly stationary

From now on, the term 'stationary' will be used to denote 'weakly' or 'wide sense stationary'; the term *strictly stationary* will be used for the stronger assumption.

Example 1.2 (AR(1) process)**.**

Autoregressive (AR) processes will be considered in more detail later. A process $\{X_t, t \in \mathbb{Z}\}$ is said to be AR(1) if it stationary and satisfies:

$$X_t = \phi X_{t-1} + Z_t \qquad \{Z_t\} \sim WN(0, \sigma^2).$$

For this process, the autocovariance may be computed as follows: by squaring up both sides and using $\gamma_X(0) = \text{Var}(X_t)$,

$$\gamma_X(0) = \phi^2 \gamma_X(0) + \sigma^2 \Rightarrow \gamma_X(0) = \frac{\sigma^2}{1-\phi^2}.$$

for $h \geq 1$,

$$\gamma_X(h) = \text{Cov}(X_{t+h}, X_t) = \phi \text{Cov}(X_{t+h-1}, X_t) + \text{Cov}(Z_{t+h}, X_t) = \phi \gamma_X(h-1)$$

so that, since $\gamma_X(-h) = \gamma_X(h)$,

$$\gamma_X(h) = \frac{\sigma^2}{(1-\phi^2)} \phi^{|h|}.$$

Its autocorrelation function (ACF) is

$$\rho_X(h) = \phi^{|h|}.$$

Note that the AR(1) process is not well defined if $|\phi| \geq 1$. □

1.4 Linear filters

A linear process may be regarded as a *linear filter*. Let $\{X_t\}$ be a time series. A *filter* is an operation on a time series in order to obtain a new time series $\{Y_t\}$. $\{X_t\}$ is called the *input* and $\{Y_t\}$ the *output*. A linear filter $\mathcal{C}$ is the following operation:

$$\mathcal{C}(X)_t := Y_t = \sum_{k=-\infty}^{\infty} c_{t,k} X_k. \tag{1.3}$$

We only consider the situation were $\mathbb{E}\left[X_t^2\right] < \infty$ and $\mathbb{E}\left[Y_t^2\right] < \infty$.

A linear filter is said to be *time-invariant* if $c_{t,k} = c_{t-k}$, in which case it may be written as:

$$Y_t = \sum_{j=-\infty}^{\infty} c_j X_{t-j}.$$

A time-invariant linear filter (TLF) is said to by *causal* if

$$c_j = 0 \quad \text{for } j < 0,$$

When the input $\{X_t\}$ of a time invariant linear filter is stationary, then the output $\{Y_t\}$ is also stationary provided $\sum_k |c_k| < +\infty$.

Definition 1.8 (Stable Linear Filter)**.** *A TLF of the form* (1.3) *is* stable *if* $\sum_{k=-\infty}^{\infty} |c_k| < \infty$.

Definition 1.9 (Transfer function, Power function)**.** *Consider a stable linear filter and set*

$$c(z) = \sum_{j=-\infty}^{\infty} c_j z^j.$$

The function $c(e^{-i\lambda}) := \sum_{j=-\infty}^{\infty} c_j e^{-i\lambda j}$ *is known as the* transfer function*, while the function* $|c(e^{-i\lambda})|^2$ *is known as the* power transfer function.

A filter may be written as $c(B)$, in the sense that

$$Y_t = c(B) X_t$$

where B as usual denotes the backward shift operator.

A linear process is a linear filter where the input is $\text{WN}(0, \sigma^2)$.

Impulse Response Function In general, for a stationary process $\{X_t : t \in \mathbb{Z}\}$, where the variables $\{X_t\}$ are functions of impulses $\{\epsilon_t : t \in \mathbb{Z}\}$, the *impulse response function* $g(s)$ is defined as:

$$g(t;s) = \frac{\partial X_{t+s}}{\partial \epsilon_t} \tag{1.4}$$

In the case of a causal linear filter $X_t = \sum_j c_j \epsilon_{t-j}$, $g(t;s) = g(s) = c_s$.

The impulse response function may be extended to vectors; if $\{\underline{X}_t : t \in \mathbb{Z}\}$ is an m-vector valued process which is a function of vector impulses $\{\underline{\epsilon}_t : t \in \mathbb{Z}\}$, then

$$g_{ij}(t,s) = \frac{\partial X_{t+s,i}}{\partial \epsilon_{t,j}}. \tag{1.5}$$

If $\{\underline{X}_t\}$ is a linear causal vector valued process satisfying $X_{t.j} = \sum_{s\geq 0}\sum_k c_{jk;s}\epsilon_{t-s,k}$ then

$$g_{ij}(t,s) = g_{ij}(s) = c_{ij;s}.$$

1.5 Trends and Seasonal Components

The classical decomposition model is:

$$X_t = \mu_t + s_t + \epsilon_t,$$

where

- μ_t is a slowly changing function (the *trend*);
- s_t is a function with known period d (the 'seasonal component');
- ϵ_t is a stationary time series.

The aim is to extract the deterministic components μ_t and s_t and estimate them and then check whether or not the residual component ϵ_t is a stationary time series.

1.5.1 No Seasonal Component

Assume that

$$X_t = \mu_t + \epsilon_t, \quad t = 1,\dots,n$$

where, without loss of generality, $\mathbb{E}[\epsilon_t] = 0$ (since the mean of the stationary process is systematic and is therefore considered to be part of the trend).

There are several methods for estimating μ. Three are considered here; *least squares*, *moving average* and *differencing*.

Method 1 : Least Squares estimation of μ_t The function μ_t is modelled by a function with as few parameters as necessary for accurate modelling and the parameters are estimated by the least squares technique. For example, suppose that μ_t can be modelled by a quadratic function, $\mu_t = a_0 + a_1 t + a_2 t^2$. The parameters $(a_k)_{k=0}^2$ are estimated by $(\widehat{a}_k)_{k=0}^2$, chosen to minimise

$$\sum_{t=1}^{n}(x_t - a_0 - a_1 t - a_2 t^2)^2.$$

Method 2 : Smoothing by means of a moving average Let q be a non-negative integer and consider a smoothed version of X defined by

$$W_t := \frac{1}{2q+1}\sum_{j=-q}^{q} X_{t+j}, \quad q+1 \leq t \leq n-q.$$

If it turns out that μ is approximately linear over the time interval $[t-q, t+q]$ and also that q is sufficiently large so that $\frac{1}{2q+1}\sum_{j=-q}^{q} Y_{t+j} \simeq 0$, then

$$W_t = \frac{1}{2q+1}\sum_{j=-q}^{q} \mu_{t+j} + \frac{1}{2q+1}\sum_{j=-q}^{q} Y_{t+j} \simeq \mu_t.$$

For $t \leq q$ and $t > n-q$, W has to be defined in a different way. For example,

$$W_t = \begin{cases} \frac{1}{2t+1}\sum_{j=-t}^{t} X_{t+j} & t = 1, \dots, q \\ \frac{1}{2(n-t)+1}\sum_{j=-(n-t)}^{n-t} X_{t-j} & t = n-q+1, \dots, n. \end{cases}$$

Unless μ_t is a straight line and the stationary time series component Y is very small, it will not be possible to find a q satisfying both the conditions that μ is approximately linear over the interval $[t-q, t+q]$ (requiring small q) and such that $\frac{1}{2q+1}\sum_{t-q}^{t+q} Y_s \simeq 0$ (requiring large q).

Definition 1.10 (Linear Filter)**.** *A* linear filter *is defined as a linear combination:*

$$\widehat{\mu}_t = \sum_j a_j X_{t+j},$$

where $\sum a_j = 1$ *and* $a_j = a_{-j}$*.*

A linear filter will allow a *linear trend* $\mu_t = \alpha_0 + \alpha_1 t$ to pass without distortion since

$$\sum_j a_j(\alpha_0 + \alpha_1(t+j)) = (\alpha_0 + \alpha_1 t)\sum_j a_j + \alpha_1 \sum_j a_j j = \alpha_0 + \alpha_1 t.$$

It is possible to choose the weights $\{a_j\}$ so that a larger class of trend functions pass without distortion. For example, the *Spencer 15-point moving average*, defined as

$$\begin{cases} [a_0, a_{\pm 1}, \dots, a_{\pm 7}] = \frac{1}{320}[74, 67, 46, 21, 3, -5, -6, -3] \\ a_j = 0 \text{ for } |j| > 7 \end{cases}$$

allows a cubic trend to pass without distortion. That is, applied to $\mu_t = at^3 + bt^2 + ct + d$,

$$\widehat{\mu}_t = \sum a_j X_{t+j} = \sum a_j \mu_{t+j} + \sum a_j Y_{t+j} \simeq \sum a_j \mu_{t+j} = \mu_t.$$

Conditions required for a filter to pass a trend which is polynomial of degree k without distortion are found in the exercises. Moving average methods will be considered in greater detail later.

Method 3: Differencing to generate stationarity Let B denote the *backward shift* operator; $(BX)_t = X_{t-1}$, with powers given by $(B^j X)_t = X_{t-j}$. In other words, applying B^j to X pushes it back j time units. Strict stationarity means that $B^h X$ has the same distribution for all $h \in \mathbb{Z}_+$.

The *difference operator* ∇ is defined by

$$\nabla X_t = X_t - X_{t-1} = (1-B)X_t,$$

where B is the *backward shift operator.* That is, $(BX)_t = X_{t-1}$. For positive integer k, ∇^k is defined by: by:

$$\nabla^k X_t = \nabla(\nabla^{k-1}X)_t.$$

For example,

$$\nabla^2 X_t = \nabla X_t - \nabla X_{t-1} = (X_t - X_{t-1}) - (X_{t-1} - X_{t-2}) = X_t - 2X_{t-1} + X_{t-2}.$$

Using the backward shift operator, this may be expressed as:

$$\nabla^2 X_t = (1-B)^2 X_t = (1 - 2B + B^2)X_t = X_t - 2X_{t-1} + X_{t-2}.$$

For a linear trend $\mu_t = a + bt$,

$$\nabla X_t = \nabla\mu_t + \nabla Y_t = a + bt - a - b(t-1) + \nabla Y_t = b + \nabla Y_t.$$

For the covariance,

$$\begin{aligned}
\mathrm{Cov}(\nabla Y_t, \nabla Y_s) &= \mathrm{Cov}(Y_t, Y_s) - \mathrm{Cov}(Y_{t-1}, Y_s) - \mathrm{Cov}(Y_t, Y_{s-1}) + \mathrm{Cov}Y_{t-1}, Y_{s-1}) \\
&= \gamma_Y(t-s) - \gamma_Y(t-s-1) - \gamma_Y(t-s+1) + \gamma_Y(t-s) \\
&= 2\gamma_Y(t-s) - \gamma_Y(t-s+1) - \gamma_Y(t-s-1).
\end{aligned}$$

It follows that ∇X_t is stationary with

$$\mu_{\nabla X} = b \qquad \gamma_{\nabla X}(h) = 2\gamma_Y(h) - \gamma_Y(h+1) - \gamma_Y(h-1).$$

In general, if $\mu_t = \sum_{j=0}^k c_j t^j$, then

$$\nabla^k X_t = k! c_k + \nabla^k Y_t,$$

which is stationary.

1.5.2 Trend and Seasonality

Now consider the model with a seasonal component:

$$X_t = \mu_t + s_t + Y_t,$$

where $\mathbb{E}[Y_t] = 0$, $s_{t+d} = s_t$ and $\sum_{k=1}^d s_k = 0$. For simplicity in the representation, assume that n/d is an integer; in any reasonable modelling situation, n and d will be chosen so that n/d is an integer.

In models with a seasonal component, the data is often indexed by period and time-unit;

$$x_{j,k} = x_{k+d(j-1)}, \quad k = 1, \ldots, d, \; j = 1, \ldots, \frac{n}{d}.$$

In this notation, $x_{j,k}$ is the observation at the k:th time-unit of the j:th period.

Three methods for dealing with seasonal components will be considered; the *small trend* method, the *moving average estimation* method and the *differencing at lag d* method.

Method S1: Small trends If the trend is considered to be constant during each period, the model may be written as:

$$X_{j,k} = \mu_j + s_k + Y_{j,k}.$$

A natural way to estimate the trend is:

$$\widehat{\mu}_j = \frac{1}{d}\sum_{k=1}^{d} x_{j,k}$$

and a natural method for the seasonal component is:

$$\widehat{s}_k = \frac{d}{n}\sum_{j=1}^{n/d}(x_{j,k} - \widehat{\mu}_j).$$

Method S2: Moving average estimation For a known period d, the trend is estimated by applying a moving average to eliminate the seasonal component and to reduce the noise. For d even set $q = d/2$. The trend is estimated by:

$$\widehat{\mu}_t = \frac{0.5x_{t-q} + x_{t-q+1} + \cdots + x_{t+q-1} + 0.5x_{t+q}}{d}.$$

For for d odd, set $q = (d-1)/2$. The trend is estimated by:

$$\widehat{\mu}_t = \frac{x_{t-q} + x_{t-q+1} + \cdots + x_{t+q-1} + x_{t+q}}{d},$$

for $q + 1 \leq t \leq n - q$.

The seasonal component s_k is then estimated in the following way. Set

$$w_k = \frac{1}{\text{number of summands}} \sum_{\frac{q-k}{d} < j \leq \frac{n-q-k}{d}} (x_{k+jd} - \widehat{\mu}_{k+jd}).$$

The seasonal component satisfies $\sum_{k=1}^{d} \widehat{s}_k = 0$ and therefore the estimates are:

$$\widehat{s}_k = w_k - \frac{1}{d}\sum_{i=1}^{d} w_i, \quad k = 1, \ldots, d.$$

Method S3: Differencing at lag d Define the lag-d difference operator ∇_d by

$$\nabla_d X_t = X_t - X_{t-d} = (1 - B^d)X_t.$$

Then

$$\nabla_d X_t = \nabla_d \mu_t + \nabla_d Y_t.$$

This has no seasonal component and the methods for dealing with time series without a seasonal component may be applied.

1.6 The LOESS Algorithm

LOESS, or Locally Estimated Scatterplot Smoothing (the acronym is also sometimes understood as LOcal regrESSion) is a non-parametric regression method used for smoothing data and fitting curves to scatterplots. It is particularly useful for time series data where the relationship between variables may be non-linear and complex. In LOESS, a regression model is computed for *each point* in the dataset. The method fits a simple model to *localised subsets* of the data, using a *weighted* regression approach. The weights are determined by a kernel function, such as the tricubic function, which assigns higher weights to points closer to the point of interest and lower weights to points further away.

1. **Weighting**: Each data point is assigned a weight based on its distance from the point of interest. We can select a value q (which defines the bandwidth) and then set

$$w(s) = \begin{cases} (1 - \left|\frac{s}{q+1}\right|)^3 & -q \leq s \leq q \\ 0 & s > q \end{cases}$$

   This is known as the *tricubic weighting function* and is commonly used. Note that the weight is 0 for points outside a distance q from the point of interest; the bandwidth is $2q + 1$.

2. **Local Regression**: A weighted regression is performed on the subset of data points within the bandwidth. This can be a linear regression, or quadratic regression, but higher-degree polynomials can also be used. For quadratic, the regression then takes the form of: minimise

$$\sum_{t=t_0-q}^{t_0+q} w(|t - t_0|)(y_t - \beta_0^{(t_0)} - \sum_{j=1}^{d} \beta_j^{(t_0)}(t - t_0)^j)^2.$$

   The default is $d = 1$.

   The LOESS smoothed estimate of y_{t_0} is then

$$\widehat{y}_{t_0} = \widehat{\beta}_0^{(t_0)}.$$

3. **Smoothing**: The fitted values from the local regression are used as the initial estimates. These estimates are then refined by iteratively adjusting the weights of outliers and re-fitting the model. This is done by introducing so-called *robustness* weights ρ_t. If $|y_t - \widehat{y}_t|$ is small, then t is assigned a large robustness weight; if it is large, then t is assigned a small robustness weight. Let $R_t = y_t - \widehat{y}_t$ (the residual), then

$$\rho_t = B\left(\frac{|R_t|}{h}\right)$$

   where $h = 6 * \text{median}(|R_t|)$ (the median of the absolute value is considered to be more robust than a standard deviation) and B is the bi-square weight function. This is:

$$B(u) = (1 - u^2)^2 \mathbf{1}_{[-1,1]}(u).$$

1.7 STL LOESS

We now apply LOESS to the problem of decomposing a time series into trend, seasonal and a stationary process. STL stands for Seasonal and Trend using LOESS. In Time Series, the goal is to separate a time series $X_t : t \in \{1, ..., N\}$ into $\{X_t = T_t + S_t + R_t\}$; trend, seasonal and remainder components. This is done through two loops. In the outer loop, the robustness weights are assigned to each data point depending on the size of the remainder. This allows us to reduce (or even eliminate) the effects of outliers. The inner loop iteratively updates the trend and seasonal components. This is done by subtracting the current estimate of the trend from the raw series. The time series is then partition into cycle-subseries (e.g. if it is monthly data with a yearly season, then there will be 12 cycle subseries: all Januarys will be one TS, all February a second, etc.). The cycle-subseries are LOESS smoothed and then passed through a low-pass filter (a filter that allows linear, quadratic or cubic trends to pass without distortion, but reduces higher frequencies). The seasonal components are the smoothed cycle-subseries minus the result from the low-pass filter. The seasonal components are subtracted from the raw data. The result is LOESS smoothed, which becomes the trend. What is left is the remainder. The algorithm is now outlined; see below for the parameters that are introduced.

1. Initialize trend as $T_t^{(0)} = 0$ and $R_t^{(0)} = 0$
2. **Outer loop** Run $n_{(o)}$ times. Initially, $\rho_t^{(0)} = 1$ for all t.
    - At iteration j, using $R_t^{(j)}$ calculate robustness weights $\rho^{(j+1)}(t)$.
    - Calculate $R_t^{(j+1)}$ (from inner loop)
    - Calculate robustness weights $\rho^{(j+1)}(t)$
    - Repeat until convergence.
3. **Inner loop**
    - Iteratively calculate trend and seasonal terms. Run $n_{(i)}$ times
    - **Detrend**: Let $D_t^{(k)} := Y_t - T_t^{(k)}$ where k is the loop number and $T^{(k)}$ is the current trend estimate. Then $\{D_t^{(k)}\}$ is the current de-trended time series. If the observed value Y_t is missing, then the detrended term is also missing
    - **Cycle-subseries smoothing**: The detrended time series is broken into cycle-subseries. For example, monthly data with a periodicity of twelve months would yield twelve cycle-subseries, one of which would be all of the months of January. Each cycle-subseries is then LOESS smoothed with $q = n_{(s)}$ (recall that $2q + 1$ is the bandwidth) and $d = 1$ (the degree of the polynomial in t). The smoothed values are then put together to give a seasonal time series which we call $C^{(k+1)}$.
    - **Low-pass filter**: We now apply a low pass filter on C^{k+1} to yield L^{k+1}.
        - Let $2m + 1$ be the smallest odd integer greater than or equal to the period and apply a moving average $\widetilde{C}_t^{(k+1)} = \frac{1}{2m+1} \sum_{j=-m}^{m} C_{t+j}^{(k+1)}$. If the period is even, use:

$$\widetilde{C}_t^{(k+1)} = \frac{1}{2m+1}\left(\frac{1}{2}(C_{t+m}^{(k+1)} + C_{t-m}^{(k+1)}) + \sum_{j=-m}^{m} C_{t+j}^{(k+1)}\right)$$

(e.g. for monthly data, we use $m = 6$ and the $\frac{1}{2}$ correction) to get L^{k+1}.

– The output of the low-pass filter is L^{k+1}.

The parameter $n_{(l)} = 2m + 1$. Note that L^{k+1} has had the seasonal component removed.

- **Detrending of smoothed cycle-subseries**: $S^{k+1} = C^{k+1} - L^{k+1}$. This is the $k+1$-th estimate of seasonal component. Importantly, the low-pass filter causes this seasonal time series average to be nearly zero.
- **Deseasonalising**: To estimate the trend, we de-seasonalise Y by: $D^{(k+1)} = Y - S^{k+1}$
- **Trend smoothing**: LOESS smooth $D^{(k+1)}$, the deseasonalised time series with $q = n_{(t)}$.

This results in $\{T^{k+1}\}$, the $k+1$-th estimate of the trend component.

Model parameters

There are six major parameters in the model.

- $n_{(p)}$ (n.p) This is the periodicity of the seasonality.
- $n_{(i)}$ (inner) This is the number of cycles through the inner loop. The number of cycles should be large enough to reach convergence, which is typically only two or three. When there are multiple outer cycles, the number of inner cycles can be smaller as they do not necessarily help get overall convergence. The default value in stlplus is 2.
- $n_{(o)}$ (outer) This is the number of cycles through the outer loop. More cycles here reduce the affect of outliers. For most situations this can be quite small (even 0 if there are no significant outliers). The default value in `stlplus` is 1.
- $n_{(l)}$ (l.window) This is the span in lags for the low-pass filter. Almost always taken as the least odd integer greater than or equal to $n_{(p)}$.
- $n_{(s)}$ (s.window) This is the smoothing parameter for the seasonal filter. As $n_{(s)}$ increases, each cycle subseries becomes smoother. This is one of the parameters with the greeatest freedom of choice from the modeller. It looks as if it can become a question of what the modeller believes to be changes in seasonal behaviour versus aberrant behaviour. In `stlplus`, s.window can accept the keyword 'periodic' instead. The package notes say this makes smoothing 'effectively replaced by taking the mean.'
- $n_{(t)}$ (t.window) This is the smoothing parameter of the trend behaviour. As this increases, the trend is increasingly smoothed. The authors recommend 'consider the trend to be a component whose estimation is needed to form an estimate of the seasonal.' If more careful trend modelling is needed, they recommend first extracting the seasonal, then model the sum of $T_t + R_t$.

In addition to these six primary parameters, the degree of the LOESS smoothing can be changed, though this is hardly ever needed. The default is typically $d = 1$ (i.e. a linear function is used to compute $\widehat{y}(t_0)$).

1.8 Autocovariance and Spectral Density of a stationary time series

Recall Definition 1.5 of a weakly stationary time series. It follows directly from the definition that:

$$\begin{cases} \gamma(0) \geq 0, \\ |\gamma(h)| \leq \gamma(0) & \text{for all } h \in \mathbb{Z}, \\ \gamma(h) = \gamma(-h) & \text{for all } h \in \mathbb{Z}. \end{cases} \tag{1.6}$$

An autocovariance function is clearly non-negative definite, since $\sum_{j=1}^{n}\sum_{k=1}^{n} a_j a_k \gamma(t_j - t_k)$ is the *variance* of $\sum_{j=1}^{n} a_j X_{t_j}$.

Definition 1.11. *A function* $\kappa : \mathbb{Z} \to \mathbb{R}$ *is said to be* non-negative definite*, or* positive semi-definite*, if*

$$\sum_{i,j=1}^{n} a_i a_j \kappa(t_i - t_j) \geq 0$$

for all n *and all vectors* $\underline{a} \in \mathbb{R}^n$ *and* $\underline{t} \in \mathbb{Z}^n$.

Theorem 1.12. *The autocorrelation function of a stationary time series is a real valued, even non negative definite function defined on* $\mathbb{Z}$*. For any real valued even non negative definite function* κ *defined on* $\mathbb{Z}$ *and for any* $N \geq 1$*, there exists a sequence of random variables* $(X_{-N}, \ldots, X_N)$ *such that Cov*$(X_i, X_j) = \kappa(i - j)$.

Proof Let $\gamma(\cdot)$ be the autocovariance function of a stationary time series X_t. Then for any $(t_1, \ldots, t_n)$ and any $(a_1, \ldots, a_n)$,

$$0 \leq \text{Var}\left(\sum_{j=1}^{n} a_j X_{t_j}\right) = \sum_{j,k} a_j a_k \gamma(t_j - t_k).$$

For the other way, Let $\underline{Z} = (Z_{-N}, \ldots, Z_N)$ be a vector of i.i.d. $N(0,1)$ variables. Let K denote the $2N + 1 \times 2N + 1$ matrix with entries $K_{i,j} = \kappa(i - j)$. Then K is a non negative definite matrix. It follows that K has a decomposition $P\Lambda P^t$ where P is an orthonormal matrix and Λ is a diagonal matrix whose entries are the eigenvalues. Let $K^{1/2} = P\Lambda^{1/2}P^t$, then $(K^{1/2})^2 = K$. Let $\underline{X} = K^{1/2}\underline{Z}$, then $\underline{X}$ is a random vector with covariance K as required. □

1.9 Holt Winters Filtering

No trend, no seasonal component Given observations $X_1, X_2, \ldots, X_n$ from the model:

$$X_t = \mu + Z_t \qquad \{Z_t\} \sim \text{WN}(0, \sigma^2)$$

where μ is considered to be approximately constant. The method of *exponential smoothing* is to compute a *smoothed* series:

$$\widetilde{X}_t = \lambda X_t + (1-\lambda)\widetilde{X}_{t-1} \qquad \lambda \in (0,1) \tag{1.7}$$

where λ is the *smoothing parameter*. The forecast for time $t+h$ given the series up to time t is

$$\widehat{X}_{t+h|t} = \widetilde{X}_t.$$

The quantity $\widetilde{X}_t$ is the estimate of μ at time t; the assumption is that the underlying value of μ will not change between t and $t+h$.

Linear trend, no seasonal component Holt and Winters independently extended this idea (Holt (1959) and Winters (1960)) to deal with the model

$$X_t = \mu_t + Z_t \qquad \{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$$

under the assumption that the trend is approximately linear. Let $m_t = \mu_t - \mu_{t-1}$. Then the equations suggested by Holt and Winters are:

$$\begin{cases} \widetilde{X}_t = \lambda_1 X_t + (1-\lambda_1)\left(\widetilde{X}_{t-1} + \widetilde{m}_{t-1}\right) \\ \widetilde{m}_t = \lambda_2\left(\widetilde{X}_t - \widetilde{X}_{t-1}\right) + (1-\lambda_2)\widetilde{m}_{t-1} \end{cases} \tag{1.8}$$

where $\widetilde{m}_t$ is the estimate of m_t at time t. The h-step ahead forecasts are then given by:

$$\widehat{X}_{t+h|t} = \widetilde{X}_t + h\widetilde{m}_t.$$

Holt Winters with linear trend and additive seasonal component Now suppose that $\{X_t\}$ is a time series with both trend and seasonal component where the seasonal component $\{s_t\}$ has period d:

$$X_t = \mu_t + s_t + Z_t \qquad \{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$$

The Holt-Winters algorithm accommodates the seasonal component in the following way: let $Y_t = X_t - s_t$, then $\widetilde{Y}_t$ is an approximation of μ_t and

$$\begin{cases} \widetilde{Y}_t = \lambda_1(X_t - \widetilde{s}_{t-d}) + (1-\lambda_1)(\widetilde{Y}_{t-1} + \widetilde{m}_{t-1}) \\ \widetilde{m}_t = \lambda_2(\widetilde{Y}_t - \widetilde{Y}_{t-1}) + (1-\lambda_2)\widetilde{m}_{t-1} \\ \widetilde{s}_t = \lambda_3(X_t - \widetilde{Y}_t) + (1-\lambda_3)\widetilde{s}_{t-d} \end{cases}$$

The initial conditions are:

$$\begin{cases} \widetilde{Y}_{d+1} = X_{d+1} \\ \widetilde{m}_{d+1} = \frac{1}{d}(X_{d+1} - X_1) \\ \widetilde{s}_i = X_i - (X_1 + \widetilde{m}_{d+1}(i-1)) \quad i = 1, \ldots, d+1 \end{cases}$$

The predictors are:

$$\widehat{X}_{t+h|t} = \widetilde{Y}_t + h\widetilde{m}_t + \widetilde{s}_{t+h} \qquad h = 1, 2, \ldots$$

The parameters $\lambda_1, \lambda_2, \lambda_3 \in (0,1)$ may be chosen by minimising the sum of squares of the one-step prediction error on data that has already been observed:

$$\sum_{i=d+2}^{n} \left(X_i - \widehat{X}_{i|i-1}\right)^2$$

Holt Winters Seasonal Multiplicative The structural equation is:

$$\widehat{X}_{t+h|t} = (\widetilde{Y}_t + h\widetilde{m}_t)s_{t+h-m(k+1)}$$

$\widetilde{Y}_t$ the estimate of μ_t, $\widetilde{m}_t$ the trend, s the seasonal component, m the seasonal period and $k = \lfloor \frac{h-1}{m} \rfloor$. The *seasonal* values s_t are *relative percentages*, that is they represent percentage deviations from the level of the series. For example, $s_t = 0.95$ means the series is at 95% of the level it would have been at if there were no seasonal effect and the s_t's typically average to 1 over the length of a cycle.

$$\widetilde{Y}_t = \lambda_1 \frac{X_t}{s_{t-m}} + (1-\lambda_1)(\widetilde{Y}_{t-1} + \widetilde{m}_{t-1})$$

$$\widetilde{m}_t = \lambda_2(\widetilde{Y}_t - \widetilde{Y}_{t-1}) + (1-\lambda_2)\widetilde{m}_{t-1}$$

$$s_t = \lambda_3 \frac{X_t}{\widetilde{Y}_{t-1} + \widetilde{m}_{t-1}} + (1-\lambda_3)s_{t-m}$$

1.10 Time Series in R

1.10.1 Extracting the Trend, Seasonal Component and Noise in R

The `stl` command may be used to decompose a time series into trend, seasonal component and noise. The computation of 'trend' is based on moving average. For illustration, consider the carbon dioxide data from Mauna Loa in the file `atmospheric-carbon-dioxide-recor.csv` .

```
> www =
"https://www.mimuw.edu.pl/~noble/courses/TimeSeries/data/atmospheric-
carbon-dioxide-recor.csv"
> carbon = read.csv(www)
```

Delete observation 611 which is 'na':

```
> carbon = carbon[-611,]
```

(this deletes the last row, which is 'na').

```
> y = carbon$MaunaLoaCO2
> MaunLoaCo2 = ts(data = y, frequency = 12)
```

(this gets it into an appropriate format - each row represents a year)

```
> output.stl = stl(MaunLoaCo2, s.window = "periodic")
> plot(output.stl)
```

This gives a plot of the original data, the seasonal component, the trend and the 'remainder'.

```
> a <- output.stl$time.series
> acf(a)
```

The time.series part of the stl output gives a decomposition into trend, the seasonal and the noise. The acf gives the autocorrelation for each of these; the trend, seasonal and noise, while the off-diagonals show the cross autocorrelations.

The dotted blue lines indicate 'error' bars. The plot of interest is the residual (or 'remainder'). The acf indicates clear correlations between the residuals; they are not $\mathrm{WN}(0, \sigma^2)$. The plot is in Figure 1.1

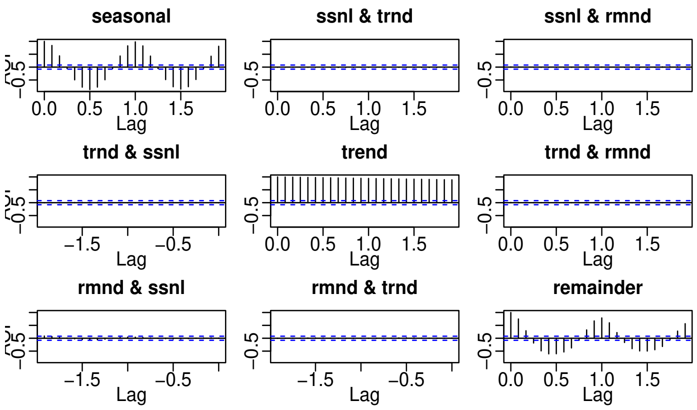

Figure 1.1: Mauna Loa: estimated acf for decomposition

To get the sample standard deviation of each column in the time series, try:

```
> apply(a,2,sd)
  seasonal       trend  remainder
 2.0402413 21.0085895  0.2735003
```

This indicates that the remainder is small compared with the trend and seasonal components.

1.10.2 Holt Winters Filtering: Implementation in R

The 'Air Passengers' data set is included in the data sets that come with R. Implementation of Holt-Winters can be carried out as follows: Type

```
> data(AirPassengers)
> AP <- AirPassengers
> str(AP)
 Time-Series [1:144] from 1949 to 1961: 112 118 132 129 121 135 148 148 136 119
...
```

The data set is now loaded. Type

```
> ?HoltWinters
```

to obtain the syntax for the command. Note that values for λ_1, λ_2 and λ_3 may be given; if the user does not give the values, then they are computed by minimising the sum of squares of the one-step prediction errors as outlined above. To make a *multiplicative seasonal* Holt Winters, try:

```
> AP.hw <- HoltWinters(AP,seasonal="mult")
> plot(AP.hw)
> legend("topleft",c("observed","fitted"),lty=1,col=1:2)
```

This gives a plot showing both the original data and the one-step predictors. The plot is in Figure 1.2.

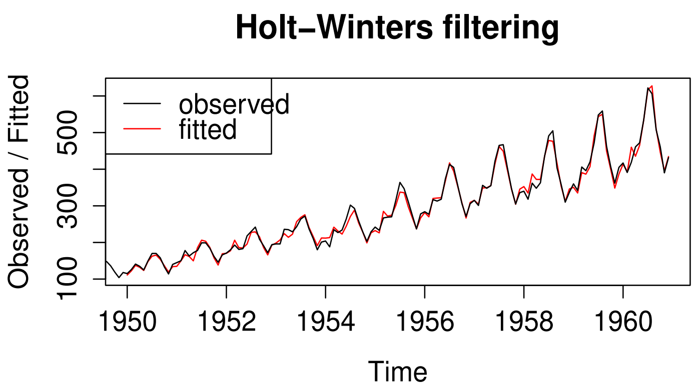

Figure 1.2: Air Passenger data with Holt Winters filtering

Prediction is made quite simply using the 'predict' command, which makes the arithmetical computations from the Holt Winters object. The following shows the predictions for the next four years.

```
> AP.predict <-predict(AP.hw,n.ahead=4*12)
> ts.plot(AP,AP.predict,lty=1:2)
```

The plot is found in Figure 1.3.

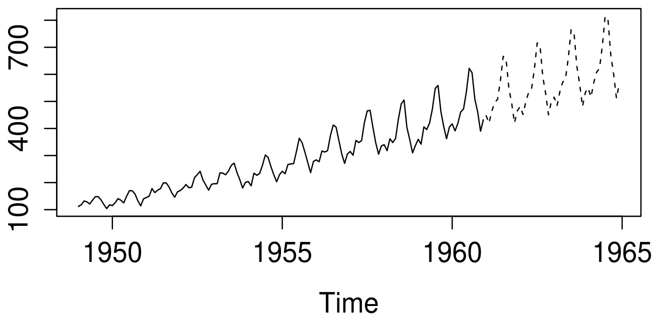

Figure 1.3: Air Passenger data: Holt Winters prediction