

MATH 681 Time Series Analysis

Dr David A. Stephens

Department of Mathematics & Statistics
Room 1225, Burnside Hall

`d.stephens@math.mcgill.ca`
`www.math.mcgill.ca/~dstephens/681/`

Time Series: Basic Prerequisites

- Basic distribution theory
- Expectations and other moments
- Linear algebra (matrix manipulations and decomposition)
- Computing an advantage (knowledge of [R](#)).

Time Series: Objectives

The basic objective is to construct models for observed data

$$x_1, x_2, \dots, x_t, \dots$$

with t indexing time, with the goals of

- **explaining** the underlying data generating mechanism
- **forecasting** future values of the underlying process.

We will mainly focus on **discrete** time modelling, and denote the underlying stochastic process $\{X_t\}$.

Time Series: Objectives

The stochastic process can be (partly) characterized via its

- ▶ finite dimensional distributions, that is the joint distributions of

$$(X_{t_1}, \dots, X_{t_n})$$

for each index set $\{t_1, \dots, t_n\}$

- ▶ moment properties (expectation, variance, covariance)

Typically $\{X_t\}$ will not be independent, and inference about the parameters of the joint distribution will be of central interest.

Time Series: Forecasting

Given observed values $X_1 = x_1, X_2 = x_2, \dots, X_T = x_T$, we aim to predict or *forecast* future values of the process

$$X_{T+1}, \dots, X_{T+K}$$

say.

For example, could use a simple linear predictor

$$\hat{X}_{T+1} = \sum_{t=1}^T a_t x_t$$

that is, a linear combination of past values of the process. The objective then is to choose the optimal values for the a_t .

1. Simple Stochastic Models

1.1 Simple Stochastic Models

A *time series* is a sequence of random variables

$$\{X_t\}_{t=1,\dots,T} = (X_1, \dots, X_T) = X_{1:T}$$

or a realization

$$\{x_t\}_{t=1,\dots,T} = (x_1, \dots, x_T) = x_{1:T}$$

collected over time index t representing some unit time period (hours, days, months etc.). A *time series model* is a probability model specifying

- the joint distribution
- the second-order moment structure (mean, variance, covariance)

of $\{X_t\}$. The joint distribution specifies the moment structure.

1.1 Simple Stochastic Models

Special Case: Independence

$$F_{X_{1:T}}(x_{1:T}) = \prod_{t=1}^T F_X(x_t) = \prod_{t=1}^T \mathbb{P}[X_t \leq x_t]$$

or equivalently

$$\mathbb{P}[X_t \leq x_t | X_{1:(t-1)} = x_{1:(t-1)}] = \mathbb{P}[X_t \leq x_t]$$

where the past values of X contain no information about future X .

1.1 Simple Stochastic Models

For the specification via moments, interest focusses on

$$\text{Expectation} \quad : \quad \mu_t = \mathbb{E}[X_t]$$

$$\text{Variance} \quad : \quad \sigma_t^2 = \mathbb{V}ar[X_t]$$

$$\text{Covariance} \quad : \quad \gamma(t, s) = \mathbb{E}[(X_t - \mu_t)(X_s - \mu_s)]$$

for any indices t, s . As the X_t usually represent repeated measurements of the same phenomenon over time, the term *autocovariance* is used.

1.1 Simple Stochastic Models

Example (IID Process)

Let $0 < p < 1$, and suppose

$$\mathbb{P}[X_t = 1] = p \quad \mathbb{P}[X_t = -1] = 1 - p$$

with the X_t mutually independent. Then $\{X_t\}$ is an *IID* (independent, identically distributed) process.

1.1 Simple Stochastic Models

Example (Random Walk)

Suppose $\{X_t\}$ is an IID process, and let

$$S_t = \sum_{i=1}^t X_i = S_{t-1} + X_t$$

Then $\{S_t\}$ is a *random walk*. Also, setting $S_0 = 0$, we have

$$X_t = S_t - S_{t-1}$$

that is, $\{X_t\}$ can be obtained from $\{S_t\}$ by *differencing*.

1.1 Simple Stochastic Models

Example (Random walk)

Suppose $\{X_t\}$ is an IID process with $\mathbb{P}[X_t = 1] = \mathbb{P}[X_t = -1] = 1/2$ and zero otherwise, and let

$$S_t = \sum_{i=1}^t X_i.$$

Then both $\{X_t\}$ and $\{S_t\}$ are zero-mean processes.

If $p \neq 1/2$, then $\mathbb{E}[X_t], \mathbb{E}[S_t]$ are non-zero; if $p < 1/2$, then $\mathbb{E}[S_t] < 0$ and we have a downward drift.

1.1 Simple Stochastic Models

Example (Random walk)

Note that for any (time-homogeneous) random walk

$$\mathbb{V}ar[S_t] = \mathbb{V}ar\left[\sum_{i=1}^t X_i\right] = \sum_{i=1}^t \mathbb{V}ar[X_i] = t\mathbb{V}ar[X_1]$$

by the independence of the X_t . Therefore the variance grows linearly with t .

Thus even if $\mathbb{E}[S_t] = 0$, the probability distribution of S_t has increasing variability around zero, and will (almost surely) diverge to an infinite value.

1.2 Trends and Seasonality

The models from the previous section are purely *stochastic*. It is also possible to incorporate *deterministic* components.

Trends: Suppose that for each t ,

$$X_t = m_t + E_t$$

where

- m_t is a deterministic function of t ,
- E_t is a purely stochastic, zero mean time series.

For example, might have

$$m_t = a_0 + a_1 t$$

that is, a *linear trend* in time. Constants a_0, a_1 are in general unknown and must be estimated.

1.2 Trends and Seasonality

Ordinary Least-Squares (OLS): estimate a_0, a_1 by minimizing the sum of squared errors

$$\hat{a} = \arg \min \left\{ \sum_{t=1}^T (x_t - a_0 - a_1 t)^2 \right\}.$$

Provided $\mathbb{E}[E_t] = 0$, OLS estimators are consistent estimators of the true parameters (if the trend truly is linear)

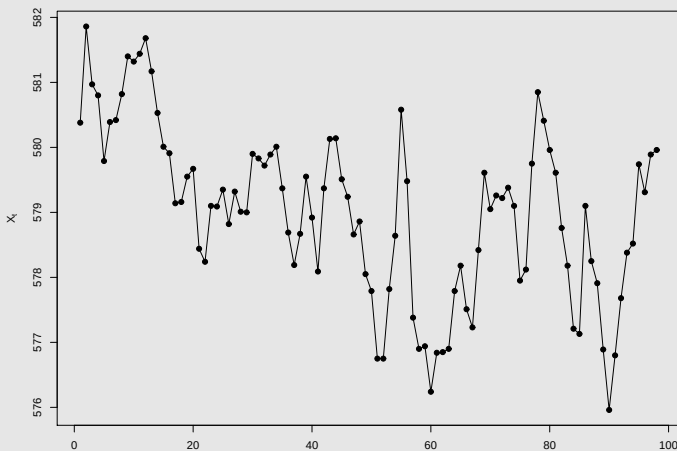
$$\hat{a} \xrightarrow{p} a$$

as $n \rightarrow \infty$. OLS is easily applied for many types of deterministic trend (e.g. polynomial trend).

1.2 Trends and Seasonality

Example (Lake Huron data)

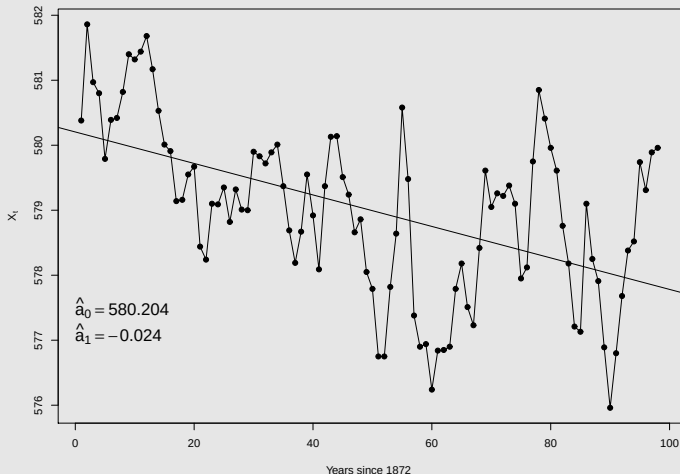
X_t is the annual level (in feet) of Lake Huron 1875-1972 ($T = 98$).



1.2 Trends and Seasonality

Example (Lake Huron data)

Fitted linear trend: $\hat{a}_0 = 580.204$, $\hat{a}_1 = -0.024$.



1.2 Trends and Seasonality

Example (Lake Huron data)

After fitting the linear trend model

$$X_t = a_0 + a_1 t + E_t$$

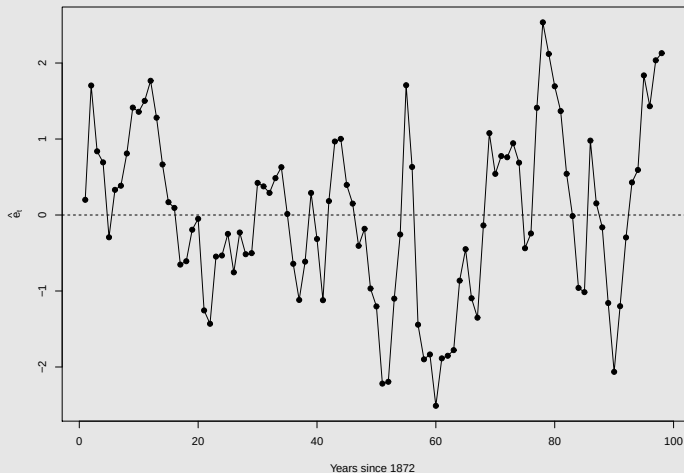
by OLS to obtain $\hat{a}_0, \hat{a}_1$, we may examine the *residual series*

$$\hat{e}_t = x_t - \hat{a}_0 - \hat{a}_1 t$$

1.2 Trends and Seasonality

Example (Lake Huron data)

Residual series: $\hat{e}_t = x_t - \hat{a}_0 - \hat{a}_1 t$.



1.2 Trends and Seasonality

Example (Lake Huron data)

The residual series is approximately zero mean. However, there is some structure:

- the variance is *not constant* over time ?
- successive residuals $(\hat{e}_{t-1}, \hat{e}_t)$ are *positively correlated*. The sample correlation is

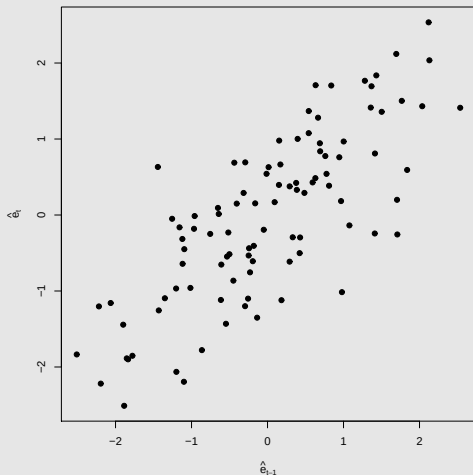
$$\text{corr}[\hat{e}_{t-1}, \hat{e}_t] \approx 0.775$$

That is, the residual series **does not** appear to be a realization of an IID process.

1.2 Trends and Seasonality

Example (Lake Huron data)

Residual series: $\hat{e}_t$ vs $\hat{e}_{t-1}$.



1.2 Trends and Seasonality

Example (Lake Huron data)

The positive dependence is potentially useful for **forecasting**; if

$$\hat{e}_{93}, \hat{e}_{94}, \dots, \hat{e}_{98}$$

are positive, it is likely that

$$\hat{e}_{99}$$

will also be positive. That is, conditional on the available data, the prediction for the next data point X_{99} will be different if the residuals apparently exhibit correlation.

R Code

```
1 Huron<-scan("huron.txt")
2 Huron<-Huron+570
3 n.H<-length(Huron); t.H<-c(1:n.H); x<-c(1875:1972)
4
5 plot(t.H,Huron,type="l",
6       xlab="Years since 1872",ylab=expression(X[t]))
7 points(t.H,Huron,pch=19)
8 abline(lm(Huron~t.H))
9 a0<-lm(Huron~t.H)$coeff[1]; a1<-lm(Huron~t.H)$coeff[2]
10
11 mt.H<-a0+a1*t.H; et.H<-Huron-mt.H
12 plot(t.H,et.H,type="l",xlab="Years since 1872",
13       ylab=expression(hat(e)[t]))
14 points(t.H,et.H,pch=19); abline(h=0,lty=2)
15
16 y1<-et.H[2:98];y2<-et.H[1:97];cor(y1,y2)
17 par(pty="s")
18 plot(y1,y2,xlab=expression(hat(e)[t-1]),
19       ylab=expression(hat(e)[t]),pch=19)
```

1.2 Trends and Seasonality

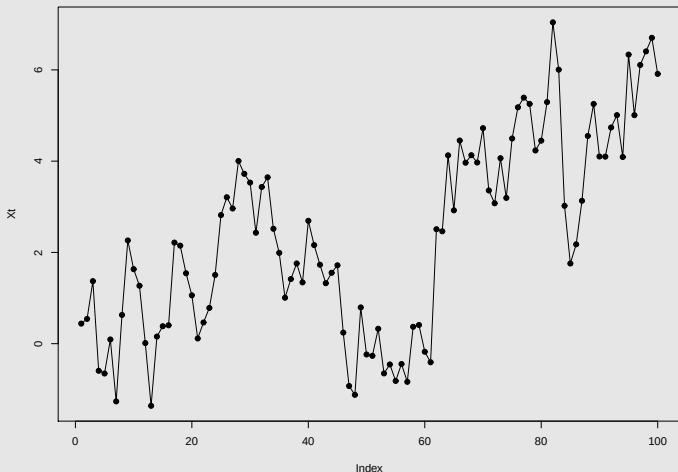
Notes:

1. Other models for trend m_t may be fitted using OLS.
 - polynomial
 - spline
 - local regression
2. An apparent trend is not necessarily a trend, so trend removal must be carried out with care.

1.2 Trends and Seasonality

Example (Trend ?)

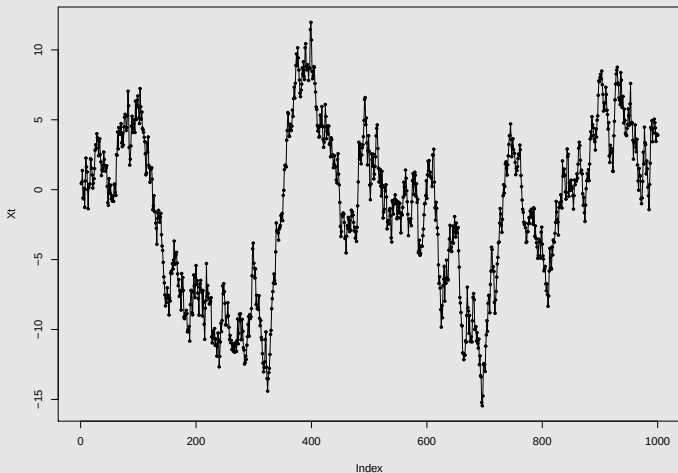
100 observations



1.2 Trends and Seasonality

Example (Trend ?)

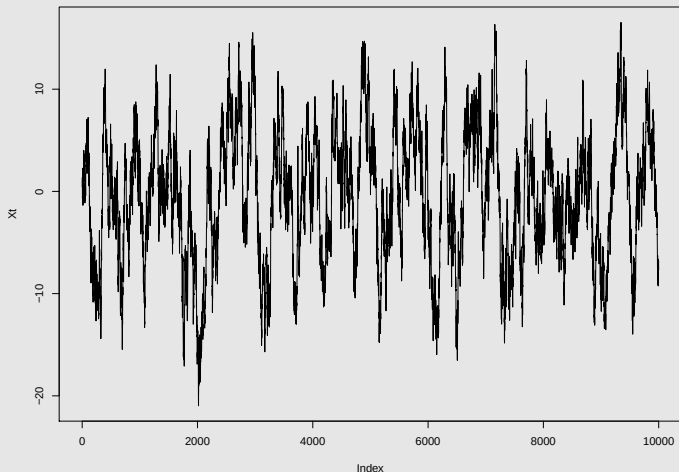
1000 observations



1.2 Trends and Seasonality

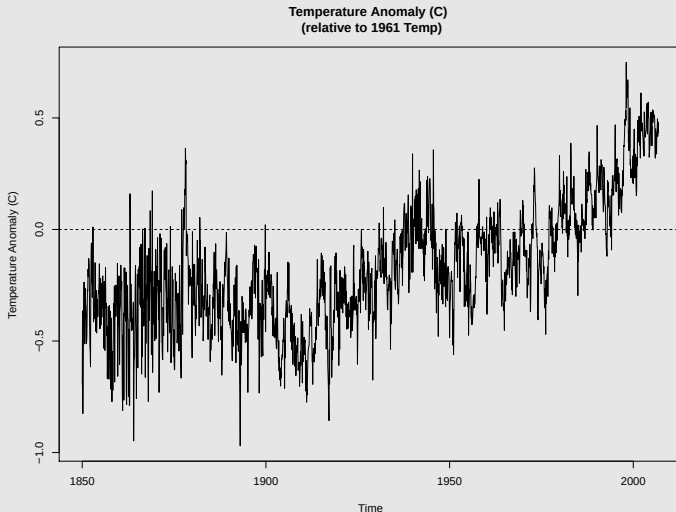
Example (Trend ?)

10000 observations



1.2 Trends and Seasonality

Example (Temperature Anomaly: 1850-2006)



1.2 Trends and Seasonality

Seasonality: Many time series are influenced by seasonally-varying factors

- calendar
- climate
- economic cycles

A deterministic seasonal model can be constructed. Suppose that for each t ,

$$X_t = s_t + E_t$$

where s_t is a periodic function with period d say, so that

$$s_{t-d} = s_t$$

for all t .

1.2 Trends and Seasonality

For example, could use a *harmonic regression*

$$s_t = a_0 + \sum_{j=1}^k [a_j \cos(\lambda_j t) + b_j \sin(\lambda_j t)]$$

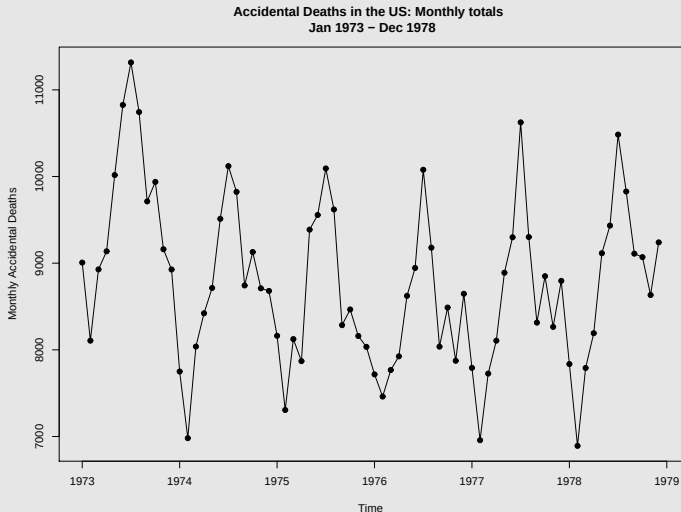
where $a_0, a_1, \dots, a_k$ and $b_1, \dots, b_k$ are unknown coefficients to be estimated, and

$$\lambda_1, \dots, \lambda_k$$

are known constants that define the periodic nature of the function.

1.2 Trends and Seasonality

Example (Accidental Deaths in the US)



1.2 Trends and Seasonality

Example (Accidental Deaths in the US)

For this data set $T = 72$, for monthly data. To fit a model with $k = 2$

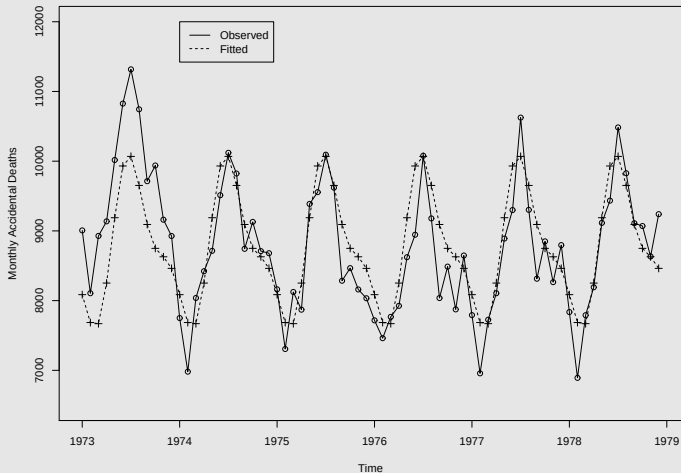
- 12 month cycle
- 6 month cycle

set $\lambda_1 = 2\pi/12$, $\lambda_2 = 2\pi/6$. Then fit using OLS.

1.2 Trends and Seasonality

Example (Accidental Deaths in the US)

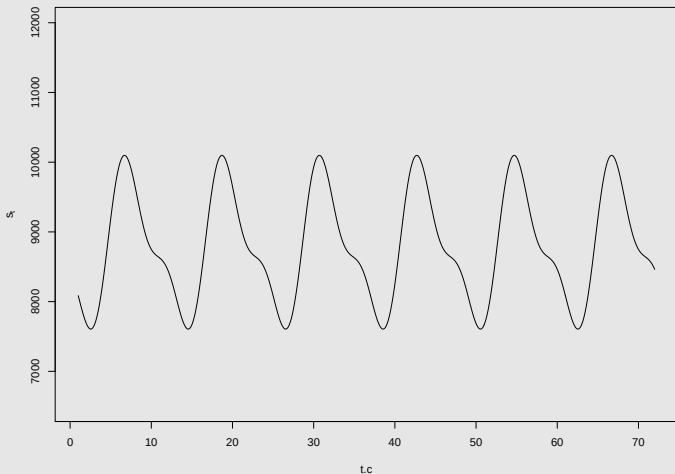
Fitted components.



1.2 Trends and Seasonality

Example (Accidental Deaths in the US)

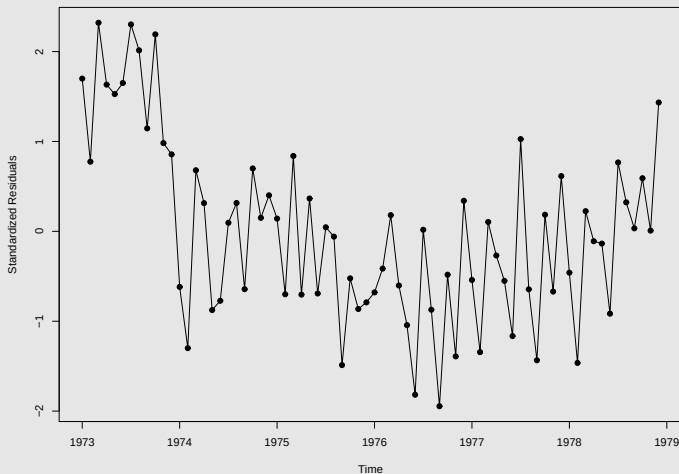
Harmonic components.



1.2 Trends and Seasonality

Example (Accidental Deaths in the US)

Residuals.



R Code

```
1 library(datasets)
2 t.v<-c(1:72)
3 C1<-cos(2*pi*t.v/12); S1<-sin(2*pi*t.v/12)
4 C2<-cos(2*pi*t.v/6); S2<-sin(2*pi*t.v/6)
5 USA.ACC<-data.frame(as.numeric(USAccDeaths), S1, C1, S2, C2)
6 USA.ACC.fit<-lm(USAccDeaths~S1+C1+S2+C2,
7                 data=USA.ACC)$fitted
8 USA.ACC.coeff<-lm(USAccDeaths~S1+C1+S2+C2,
9                  data=USA.ACC)$coeff
10 t.c<-seq(from=1,to=72,by=0.01)
11 C1.c<-cos(2*pi*t.c/12); S1.c<-sin(2*pi*t.c/12)
12 C2.c<-cos(2*pi*t.c/6); S2.c<-sin(2*pi*t.c/6)
13
14 y.c<-USA.ACC.coeff[1]+USA.ACC.coeff[2]*S1.c+
15     USA.ACC.coeff[3]*C1.c +USA.ACC.coeff[4]*S2.c+
16     USA.ACC.coeff[5]*C2.c
```

1.3 Stationarity

Suppose $\{X_t\}$ is a time series process, and suppose that for all t

$$\mathbb{E}[X_t^2] < \infty.$$

Define

$$\text{Mean function : } \mu_X(t) = \mathbb{E}[X_t]$$

$$\text{Covariance function : } \gamma_X(t, s) = \mathbb{E}[(X_t - \mu_X(t))(X_s - \mu_X(s))]$$

for integers t, s . Then $\{X_t\}$ is *weakly stationary* if

- (i) $\mu_X(t)$ does not depend on t
- (ii) $\gamma_X(t + h, t)$ does not depend on t for each integer h .

1.3 Stationarity

$\{X_t\}$ is *strongly stationary* if

$$X_{t:(t+n)} = (X_{t+1}, \dots, X_{t+n})$$

and

$$X_{(t+h+1):(t+h+n)} = (X_{t+h+1}, \dots, X_{t+h+n})$$

have the same joint distribution for all t, h, n .

Note: In practice, the (multivariate) distributions required to model such strongly stationary processes are limited in number.

1.3 Stationarity

If $\{X_t\}$ is weakly stationary process define the *autocovariance function* (ACVF), $\gamma_X(h)$, by

$$\gamma_X(h) \equiv \gamma_X(h, 0) = \gamma_X(t + h, t)$$

and the the *autocorrelation function* (ACF), $\rho_X(h)$, by

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

That is

$$\gamma_X(h) = \text{Cov}[X_{t+h}, X_t]$$

$$\rho_X(h) = \text{Corr}[X_{t+h}, X_t]$$

1.3 Stationarity

Example (IID Process)

Denoted $\{X_t\} \sim \text{IID}(0, \sigma^2)$, an IID (noise) process is characterized by

$$\mathbb{E}[X_t] = 0$$

$$\text{Var}[X_t] = \mathbb{E}[X_t^2] = \sigma^2 < \infty.$$

with the X_t mutually **independent**. This process is stationary as

$$\gamma_X(h) = \text{Cov}[X_{t+h}, X_t] = \begin{cases} \sigma^2 & h = 0 \\ 0 & h > 0 \end{cases}$$

$$\rho_X(h) = \text{Corr}[X_{t+h}, X_t] = \begin{cases} 1 & h = 0 \\ 0 & h > 0 \end{cases}$$

1.3 Stationarity

Example (White Noise Process)

Denoted $\{X_t\} \sim \text{WN}(0, \sigma^2)$, a white noise process is characterized by

$$\mathbb{E}[X_t] = 0$$

$$\text{Var}[X_t] = \mathbb{E}[X_t^2] = \sigma^2 < \infty.$$

with the X_t **uncorrelated**. This process is stationary as

$$\gamma_X(h) = \text{Cov}[X_{t+h}, X_t] = \begin{cases} \sigma^2 & h = 0 \\ 0 & h > 0 \end{cases}$$

$$\rho_X(h) = \text{Corr}[X_{t+h}, X_t] = \begin{cases} 1 & h = 0 \\ 0 & h > 0 \end{cases}$$

1.3 Stationarity

Example (Random Walk)

If $\{X_t\} \sim \text{WN}(0, \sigma^2)$, then the process $\{S_t\}$ defined by

$$S_t = \sum_{i=1}^t X_i$$

is **non-stationary** as

$$\mathbb{E}[X_t] = 0$$

$$\text{Var}[X_t] = t\sigma^2$$

which does depend on t .

1.3 Stationarity

Example (Random Walk)

$$\begin{aligned}\gamma_S(t+h, t) &= \text{Cov}[S_{t+h}, S_t] \\ &= \text{Cov}[S_t + X_{t+1} + \cdots + X_{t+h}, S_t] \\ &= \text{Cov}[S_t, S_t] + \sum_{i=1}^h \text{Cov}[X_{t+i}, S_t] \\ &= \text{Cov}[S_t, S_t] = \text{Var}[S_t] \\ &= t\sigma^2.\end{aligned}$$

1.3 Stationarity

Example (Moving Average)

Let $\{Z_t\} \sim \text{WN}(0, \sigma^2)$, and consider the process $\{X_t\}$ defined by

$$X_t = Z_t + \theta_1 Z_{t-1} \quad t \in \mathbb{Z}$$

where θ_1 is a real-valued parameter. As $\mathbb{E}[Z_t] = 0$, we also have $\mathbb{E}[X_t] = 0$, and by uncorrelatedness

$$\text{Var}[X_t] = \text{Var}[Z_t] + \theta_1^2 \text{Var}[Z_{t-1}] = (1 + \theta_1^2) \sigma^2$$

which does not depend on t .

1.3 Stationarity

Example (Moving Average)

$$\begin{aligned}\gamma_X(t+h, t) &= \text{Cov}[X_{t+h}, X_t] = \mathbb{E}[X_t X_{t+h}] \\ &= \mathbb{E}[(Z_t + \theta_1 Z_{t-1})(Z_{t+h} + \theta_1 Z_{t+h-1})] \\ &= \mathbb{E}[Z_{t+h} Z_t] + \theta_1 (\mathbb{E}[Z_{t+h} Z_{t-1}] + \mathbb{E}[Z_{t+h-1} Z_t]) \\ &\quad + \theta_1^2 \mathbb{E}[Z_{t-1} Z_{t+h-1}]\end{aligned}$$

Now,

$$\mathbb{E}[Z_j Z_k] = \begin{cases} \sigma^2 & j = k \\ 0 & j \neq k \end{cases}.$$

1.3 Stationarity

Example (Moving Average)

Hence

$$\gamma_X(t+h, t) = \begin{cases} \sigma^2(1 + \theta_1^2) & h = 0 \\ \sigma^2\theta_1 & h = \pm 1 \\ 0 & \text{otherwise} \end{cases}.$$

$$\rho_X(t+h, t) = \begin{cases} 1 & h = 0 \\ \theta_1/(1 + \theta_1^2) & h = \pm 1 \\ 0 & \text{otherwise} \end{cases}$$

These functions do not depend on t , so $\{X_t\}$ is a stationary process

$$\{X_t\} \sim \text{MA}(1)$$

1.3 Stationarity

Example (Autoregression)

Let $\{Z_t\} \sim \text{WN}(0, \sigma^2)$, and consider the process $\{X_t\}$ defined by

$$X_t = \phi_1 X_{t-1} + Z_t \quad t \in \mathbb{Z}$$

where ϕ_1 is a real-valued parameter with $|\phi_1| < 1$. **Suppose** that $\{X_t\}$ is stationary, and that

$$\text{Cov}[Z_t, X_s] = 0 \quad t > s.$$

Then

$$\mathbb{E}[X_t] = \phi_1 \mathbb{E}[X_{t-1}] + \mathbb{E}[Z_t] = \phi_1 \mathbb{E}[X_{t-1}]$$

so therefore by the stationarity assumption $\mathbb{E}[X_t] = 0$.

1.3 Stationarity

Example (Autoregression)

$$\begin{aligned}\gamma_X(h) &= \mathbb{Cov}[X_{t+h}, X_t] = \mathbb{Cov}[X_{t-h}, X_t] \\ &= \mathbb{Cov}[\phi_1 X_{t-1}, X_{t-h}] + \mathbb{Cov}[Z_t, X_{t-h}] \\ &= \phi_1 \mathbb{Cov}[X_{t-1}, X_{t-h}] + 0 \\ &= \phi_1 \gamma_X(h-1) \\ &= \phi_1^h \gamma_X(0)\end{aligned}$$

by recursion. Note that $\gamma_X(-h) = \gamma_X(h)$, so

$$\rho_X(h) = \phi_1^{|h|} \quad h = 0, \pm 1, \pm 2$$

that is, a geometrically decaying ACF.

1.3 Stationarity

Example (Autoregression)

Note also that

$$\begin{aligned}\gamma_X(0) &= \text{Cov}[X_t, X_t] \\ &= \text{Cov}[(\phi_1 X_{t-1} + Z_t)(\phi_1 X_{t-1} + Z_t)] \\ &= \phi_1^2 \text{Cov}[X_{t-1}, X_{t-1}] + \text{Cov}[Z_t, Z_t] \\ &= \phi_1^2 \gamma_X(0) + \sigma^2\end{aligned}$$

so therefore

$$\gamma_X(0) = \frac{\sigma^2}{1 - \phi_1^2}$$

$\{X_t\}$ is known as the *autoregressive* process of order 1

$$\{X_t\} \sim \text{AR}(1).$$

1.3 Stationarity

Estimation: For a stationary process, moment estimators can be used to estimate the mean/covariance structure. The estimates are

$$\hat{\mu} = \bar{x} = \frac{1}{T} \sum_{t=1}^T x_t$$

$$\hat{\gamma}(h) = \frac{1}{T} \sum_{t=1}^T (x_t - \hat{\mu})(x_{t+|h|} - \hat{\mu}) \quad -T < h < T$$

These are consistent estimators of the two functions. The $T \times T$ matrix estimate

$$\hat{\Gamma}_T = [\hat{\gamma}(i-j)]_{ij}$$

is non-negative definite.

1.3 Stationarity

Example (Lake Huron data)

Recall that the model

$$X_t = m_t + E_t$$

was fitted using a linear trend $m_t = a_0 + a_1 t$ leaving residuals that were positively correlated (see plot on page 20). We may model

$$E_t \sim \text{AR}(1),$$

that is, assume

$$E_t = \phi_1 E_{t-1} + Z_t$$

and estimate ϕ_1 using OLS from the fitted residuals e_t . This assumes that $\{Z_t\} \sim \text{WN}(0, \sigma^2)$.

1.3 Stationarity

Example (Lake Huron data)

This procedure yields $\hat{\phi}_1 = 0.791$. By inspecting the fitted values, we may construct the residual quantities

$$\hat{z}_t = e_t - \hat{\phi}_1 e_{t-1}$$

we may estimate σ^2 via the sample variance of the $\hat{z}_t$. This yields $\hat{\sigma}^2 = 0.502$.

If the AR(1) model is correct, then the residual quantities $\hat{z}_t$ should resemble a realization of a white noise process; however, they do not - there is **still** positive correlation between successive $\hat{z}_t$ s.

1.4 A Decomposition Strategy

Recap: We seek to decompose an observed series $\{Y_t\}$ as

$$Y_t = m_t + s_t + X_t$$

where

- m_t is a deterministic trend
- s_t is a seasonal component with period d
- X_t is a zero mean stationary process.

It may be necessary to replace the *global* model for trend (linear, polynomial etc.) by a *local* model

1.4 A Decomposition Strategy

Smoothing by averaging: Suppose

$$\hat{m}_t = \frac{1}{2q+1} \sum_{i=-q}^q x_{t+i}$$

with $x_s = 1$ for $s < 1$ and $s > T$. More generally,

$$\hat{m}_t = \sum_{i=-q}^q a_i x_{t+i}$$

for constants $a_{-q}, \dots, a_q$.

This is a form of *low-pass filter*.

1.4 A Decomposition Strategy

Exponential Smoothing: Suppose $\hat{m}_t = x_1$, and for $t \geq 2$

$$\hat{m}_t = \alpha x_t + (1 - \alpha)\hat{m}_{t-1}$$

for $0 < \alpha < 1$. It follows that

$$\hat{m}_t = (1 - \alpha)^{t-1}x_1 + \sum_{j=1}^{t-2} \alpha(1 - \alpha)x_{t-j} + \alpha x_t$$

If α is near one, there is strong dependence on recent x values.

1.4 A Decomposition Strategy

General Strategy: For series $\{Y_t\}$

$$Y_t = m_t + s_t + X_t$$

where s_t has period d .

Step 1. If d is even, set $q = d/2$, else set $q = (d - 1)/2$.

Step 2. Use moving average for detrending respecting the seasonality

$$\hat{m}_t = \begin{cases} \frac{y_{t-q} + y_{t+q}}{2d} + \frac{1}{d} \sum_{j=-q+1}^{q-1} y_{t+j} & d \text{ even} \\ \frac{1}{d} \sum_{j=-q}^q y_{t+j} & d \text{ odd} \end{cases}$$

1.4 A Decomposition Strategy

Step 3. Remove seasonality: for $k = 1, \dots, d$, compute

$$w_k = \frac{1}{n_d} \sum_j (y_{k+jd} - \hat{m}_{k+jd})$$

where n_d is the number of terms in the sum (essentially the number of cycles in the series). For $k = 1, \dots, d$, set

$$\hat{s}_k = w_k - \frac{1}{d} \sum_{j=1}^d w_j$$

and set $\hat{s}_k = \hat{s}_{k-d}$ for $k = d+1, d+2, \dots, T$. The deseasonalized data are $y_t^\star = y_t - \hat{s}_t$.

1.4 A Decomposition Strategy

Step 4. Recompute the deseasonalized trend $\hat{m}_t^\star$ from the deseasonalized data $y_t^\star$ using the procedure in Step 2.

Step 5. Compute the residual series

$$\hat{x}_t = y_t - \hat{m}_t^\star - \hat{s}_t.$$

The resulting process $\hat{x}_t$ should be a zero mean, but potentially autocorrelated process.

1.5 Differencing

Differencing: Another method for removing non-stationarity is *differencing*. The *lag-1 difference operator*, ∇ , acts on X_t as follows

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

where B is the backshift operator

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

Note that

$$B^2X_t = B(BX_t) = BX_{t-1} = X_{t-2}$$

etc.

1.5 Differencing

Similarly

$$\begin{aligned}\nabla^2 X_t &= (1 - B)^2 X_t \\ &= (1 - B)((1 - B)X_t) \\ &= (1 - B)(X_t - X_{t-1}) \\ &= (X_t - X_{t-1}) - (X_{t-1} - X_{t-2}) \\ &= X_t - 2X_{t-1} + X_{t-2}\end{aligned}$$

Note that k times differencing removes a polynomial trend of order k .

1.5 Differencing

Seasonal Differencing: The *lag- d difference operator*, ∇_d , acts on X_t as follows

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

This operator removes a *seasonality* with period d .

∇ and ∇_d can be applied consecutively; they commute

$$\nabla \nabla_d X_t = \nabla_d \nabla X_t$$