Introduction to Time Series and Forecasting
(Chapter 1)

1.1 Examples of time series

Ex 1.1.1 (Australian red wine sales; WINE.TSM)

x_t = monthly sales of red wine (in kilolitres)

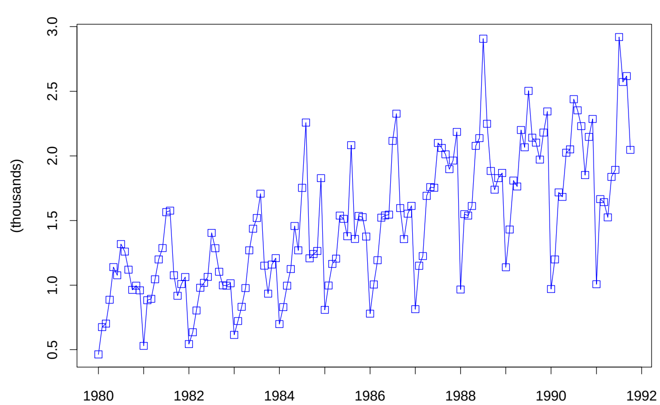
t = (Jan, 1980), (Feb, 1980), . . . , (Oct, 1991)

or

t=1, 2, . . . , 142.

1

Figure 1.1: Australian red wine sales



Features: upward trend

seasonal pattern (peak in July, trough in Jan)

increase in variability

2

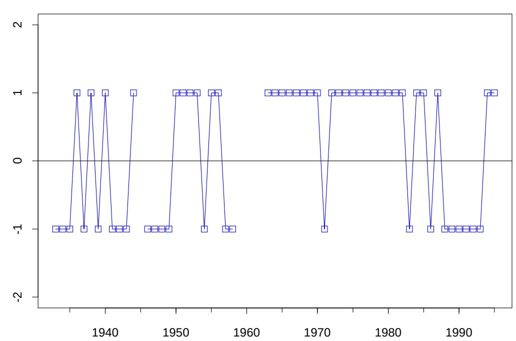
Time Series Plots: Examine plot for:

- trend over time (does the series increase or decrease with time)
- regular seasonal (or cyclical) components
- constant variability over time
- other systematic features of the data

3

Ex 1.1.2 (All-star baseball games, 1933-1995)

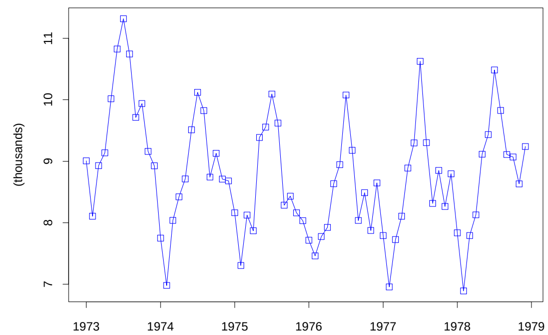
$$x_t = \begin{cases} 1 & \text{if the National League won in year } t \\ -1 & \text{if the American League won in year } t \end{cases}$$



4

Ex 1.1.3 (Accidental deaths, USA; DEATHS.TSM)

Figure 1.3 : Monthly accidental deaths



Features: slight trend

seasonal component (peak in July)

5

Ex. 1.1.4 (Signal Detection; SIGNAL.TSM)

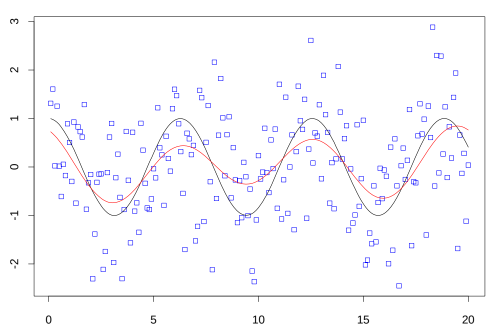
Model:

$X_t = \cos(t/10) + N_t, \quad t = 1, 2, \ldots, 200$

where $\{N_t\}$ is an IID sequence of N(0,.25) rv's.

Figure 1.4: red = estimated signal

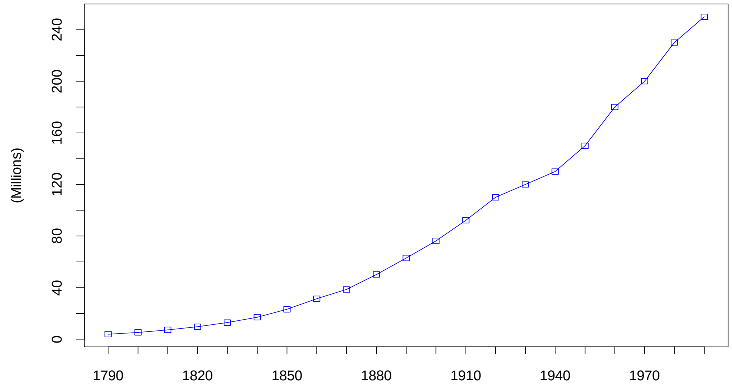
black= true signal



6

Ex. 1.1.5 (Population of USA.; USPOP.TSM)

Figure 1.5.



1.2 Objectives of Time Series Analysis

Modelling paradigm :

- set up family of probability models to represent data
- estimate parameters of model
- check model for goodness of fit

Applications of models:

- provides a compact description of the data
- interpretation
- prediction
- hypothesis testing

1.3 Some Simple Time Series Models

DEFINITION 1.3.1. A **time series model** for the observed data $\{x_t\}$ is a specification of the joint distributions of a sequence of random variables $\{X_t\}$ of which $\{x_t\}$ is postulated to be a realization.

2nd Order Properties.

means: $E(X_t)$

2nd -order moments: $E(X_{t+h}X_t)$

1.3.1 Zero-mean Models

Ex 1.3.1 (IID NOISE).

$$\{X_t\} \sim \text{IID}(0, \sigma^2)$$

if $\{X_t\}$ is an IID sequence with mean 0 and variance σ^2.

Ex 1.3.2 (Binary Process).

$$\{X_t\} \sim \text{IID}$$

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

where p=.5. (Model for All Star baseball games??)

1.3.2 Models with trend and seasonality

Model with no seasonal component.

$$X_t = m_t + Y_t \quad ,$$

where m_t is a slowly varying function called the **trend function.**

Estimation via least squares.

e.g. $\quad m_t = a_0 + a_1 t + a_2 t^2$

where coefficients are estimated by minimizing

$$\Sigma(x_t - m_t)^2$$

11

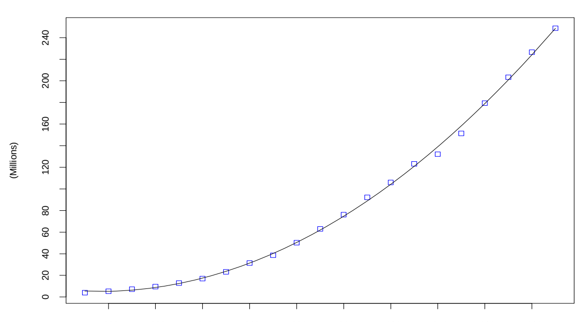
Ex. 1.3.4 (Population of USA.; USPOP.TSM)

Model: $X_t = a_0 + a_1 t + a_2 t^2 + Y_t$

$$\boxed{\hat{a}_0 = 6.96\text{x}10^5 , \quad \hat{a}_1 = -2.16 \text{ x } 10^6, \hat{a}_2 = 6.51\text{x } 10^5}$$

Forecast for year 2000: $\hat{m}_{2000} = 274.35 \text{ x } 10^6$

Figure 1.8.



12

Ex 1.3.5 (Lake Huron Levels (1875-1972); LAKE.TSM)

Model: $X_t = a_0 + a_1 t + Y_t$

Figure 1.9.

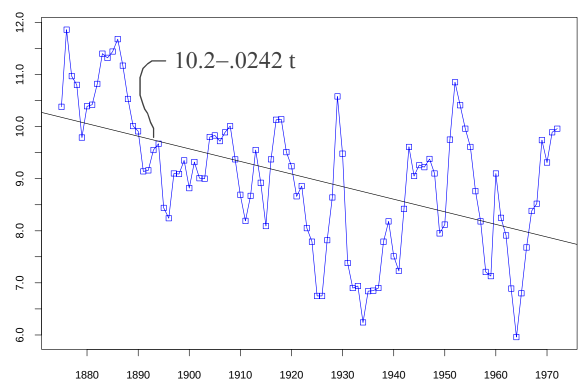
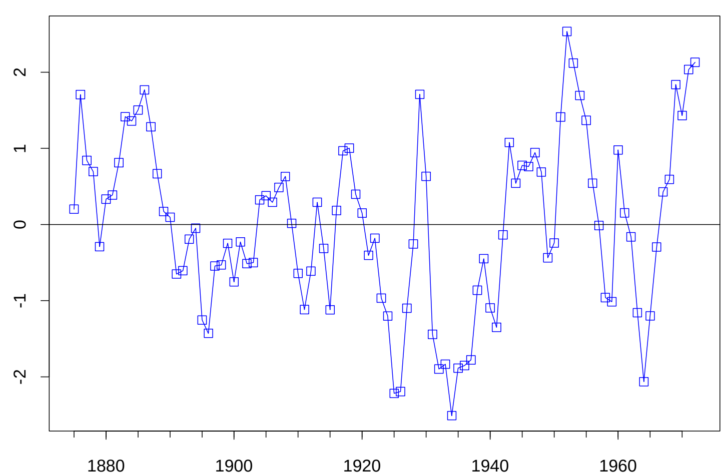


Figure 1.10. Residuals from the LS fit in previous figure.

Note: residuals do not look IID



Harmonic Regression

Useful for data exhibiting a clear periodic component.

Model: $X_t = s_t + Y_t$,

$s_t = s_{t-d}$ (periodic component)

Convenient choice:

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

where a_j, b_j are unknown parameters

λ_j are fixed frequencies, multiple of $2\pi/d$ (usually a **Fourier** frequency $2\pi k/n$ for some $k=1,\ldots,[n/2]$.) For daily data, $\lambda = 2\pi/365$.

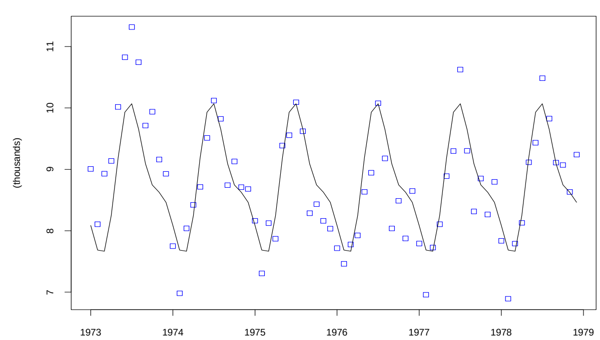
15

Ex 1.1.6 (Accidental deaths, USA; DEATHS.TSM)
Model: $X_t = s_t + Y_t$,

$$s_t = a_0 + \sum_{j=1}^{2} (a_j \cos(\lambda_j t) + b_j \sin(\lambda_j t))$$

$\lambda_1 = 2\pi/12$ (period 12), $\lambda_2 = 2\pi/6$ (period 6)

Figure 1.11 : Monthly accidental deaths



16

1.3.3 General Approach to TS Modelling

- Plot the series. Check for
  - (a) a trend
  - (b) a seasonal component
  - (c) any apparent sharp changes in behavior
  - (d) any outlying observations
- Remove trend and seasonal components to get **stationary** residuals
- Choose model to fit residuals

17

1.4 Stationary Models and the ACF

Let $\{X_t\}$ be a time series with $\mathrm{E}\, X_t^2 < \infty$.

Mean function: $\mu(t) = \mathrm{E}\, X_t$

Covariance function: $\gamma(r,s) = \mathrm{Cov}(X_r, X_s)$

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

(i) $\mu(t)$ is independent of t

(ii) $\gamma(t+h,t)$ is independent of t for each h.

18

Let $\{X_t\}$ be a stationary time series.

Autocovariance function (ACVF):

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

Autocorrelation function (ACF):

$$\rho(h) = \mathrm{Cor}(X_{t+h}, X_t) = \frac{\gamma(h)}{\gamma(0)}$$

19

Ex 1.4.1 (IID noise). $\{X_t\} \sim \mathrm{IID}\ (0,\sigma^2)$

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

Ex 1.4.2 (White noise). $\{X_t\} \sim \mathrm{WN}\ (0,\sigma^2)$

In this case the rv's are only **uncorrelated** and not necessarily IID.

$$\mathrm{IID} \Rightarrow \mathrm{WN}$$

$$\mathrm{IID} \not\Leftarrow \mathrm{WN}$$

20

Problem 1.8. Let $\{Z_t\} \sim$ IID N(0,1) and define

$$X_t = \begin{cases} Z_t\,, & \text{if } t \text{ even} \\ (Z_{t-1}^2 - 1)/\sqrt{2}\,, & \text{if } t \text{ odd} \end{cases}$$

(a) Show $\{X_t\}$ is WN but not IID.

$\mathrm{E}\, X_t = 0\;, \;\gamma(0) = 0$

$\gamma(1) = \mathrm{Cov}(X_{t+1}, X_t)$

$= \mathrm{Cov}((Z_t^2 - 1)/\sqrt{2}\,, Z_t)$ if t even

$= 0$

(b) $\mathrm{E}(X_{n+1} | X_1, \ldots, X_n) = (X_n^2 - 1)/\sqrt{2}$, if n even

$= 0,$ if n odd.

Ex 1.4.3 (Random Walk).

$$S_t = Z_1 + Z_2 + \cdots + Z_t\,,$$

where $\{Z_t\} \sim \mathrm{WN}(0,\sigma^2)$. Then

- $\mathrm{E}(S_t) = 0$
- $\mathrm{Var}(S_t) = t\sigma^2$
- $\mathrm{Cov}(S_{t+h}, S_t) = t\sigma^2$

Conclude that $\{S_t\}$ is not stationary.

Ex 1.4.4 (Moving Average; MA(1)).

$$X_t = Z_t + \theta Z_{t-1}, \quad \{Z_t\} \sim \mathrm{WN}(0,\sigma^2)$$

$$\rho(h) = \begin{cases} 1, & \text{if } h=0, \\ \theta/(1+\theta^2), & \text{if } h=\pm 1, \\ 0, & \text{otherwise.} \end{cases}$$

23

Ex 1.4.5 (Autoregression; AR(1)).

Assume $\{X_t\}$ is a stationary series satisfying

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

and Z_t is uncorrelated with X_s for each $s < t$.
Taking expectations of each side of the above equation, we find that the mean is 0.
Multiplying through by X_{t-h} and taking expectations, we obtain

$$\begin{aligned} \gamma(h) &= \mathrm{Cov}(X_t, X_{t-h}) \\ &= \mathrm{Cov}(\phi X_{t-1}, X_{t-h}) + \mathrm{Cov}(Z_t, X_{t-h}) \\ &= \phi\gamma(h-1) + 0 \\ &= \phi^h \gamma(0) \end{aligned}$$

$$\rho(h) = \gamma(h)/\gamma(0) = \phi^h$$

24

1.4.1 The Sample Autocorrelation Function

Observed data: $x_1, \ldots, x_n$

Sample mean : $\overline{x} = n^{-1} \sum_{t=1}^{n} x_t$

Sample autocovariance function :

$$\hat{\gamma}(h) = n^{-1} \sum_{t=1}^{n-h} (x_{t+h} - \overline{x})(x_t - \overline{x})$$

Sample autocorrelation function :

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

25

Remarks:

(1) $\hat{\gamma}(h)$ is approximately the sample covariance function of $(x_1, x_{1+h}), \ldots, (x_{n-h}, x_n)$

(2) The covariance matrix $\hat{\Gamma}_n = [\hat{\gamma}(i-j)]$, i,j=1, . . . ,n

is non-negative definite (positive definite).

(3) If data are observations from IID noise, then

$\hat{\rho}(h)$ is approx $N(0, n^{-1})$

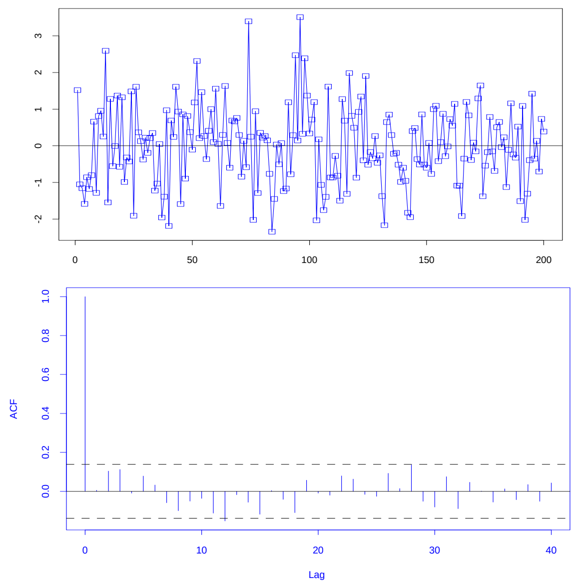
and are independent for all $h \geq 1$.

For IID noise,

$|\hat{\rho}(h)| < 1.96\, n^{-.5}$ with probability .95

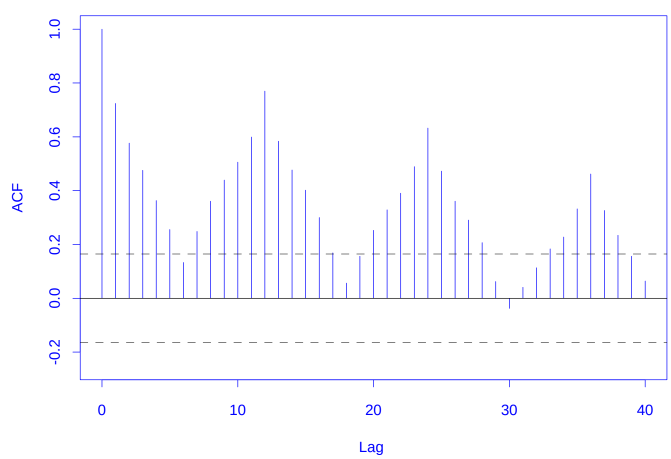
26

Ex 1.4.6 (200 observations from IID N(0,1)).



27

Figure 1.14. Sample ACF of WINE.TSM



28

1.4.2 A Model for the Lake Huron Data

Let residuals from the LS fit be denoted by

$$y_t = x_t - a_0 - a_1 t , \quad t=1,...,98$$

Figure 1.15. ACF of residuals $\{y_t\}$

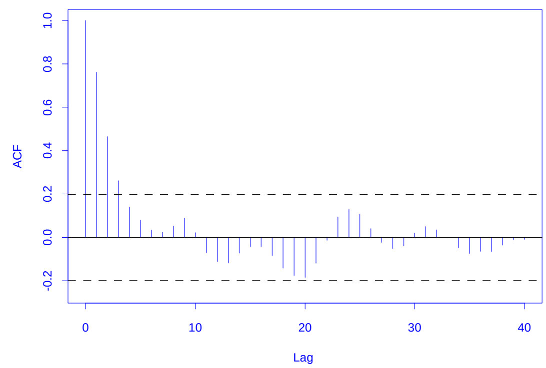
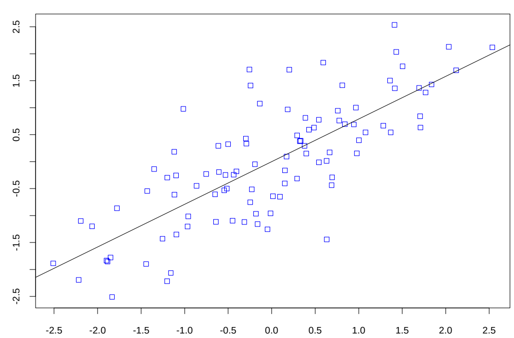


Figure 1.16. Scatter plot of residuals (y_{t-1}, y_t) showing regression line y = .791 x.



Model: $Y_t = .791\, Y_{t-1} + Z_t, \quad \{Z_t\} \sim \mathrm{WN}(0,\sigma^2)$
(Autoregressive or AR(1) model)

1.5 Trend and Seasonal Components

Classical Decomposition:

$$X_t = m_t + s_t + Y_t$$

m_t trend component (slowly changing function of t)

s_t seasonal component (periodic with period d)

Y_t random noise component

31

1.5.1 Estimating trend w/o Seasonal Components

Nonseasonal Model with Trend:

$$X_t = m_t + Y_t$$

$\mathrm{E}\ Y_t = 0$

Smoothing to estimate m_t

(a) finite moving average filter

$$W_t = (2q+1)^{-1} \sum_{|j| \leq q} X_{t-j}$$

$$W_t = (2q+1)^{-1} \sum_{|j| < q} m_{t-j} + (2q+1)^{-1} \sum_{|j| \leq q} Y_{t-j}$$

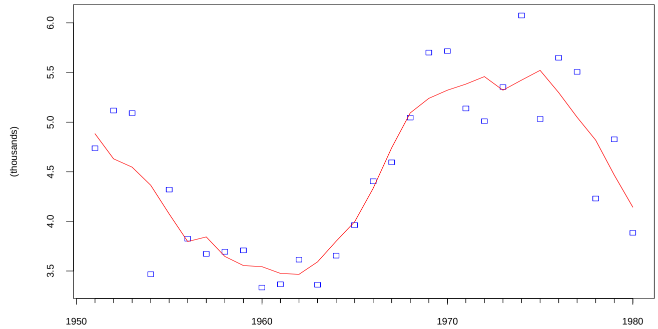
$$\sim m_t + 0$$

32

Application of a linear filter:

$\{x_t\}$ → linear filter → $\{\hat{m}_t = \Sigma\ a_j x_{t-j}\}$

Figure 1.19. Simple 5-term moving average of strikes data STRIKES.TSM



(b) exponential smoothing

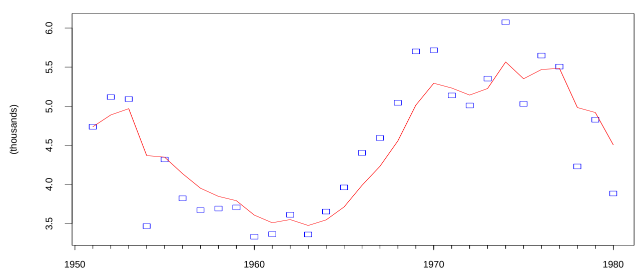
$$\hat{m}_t = a\ X_t + (1- a)\ \hat{m}_{t-1},\ \ t = 2, \ldots, n$$

$$\hat{m}_1 = X_1$$

0 a 1

max smoothing ⟷ no smoothing

Fig 1.21. Strike data a = .4



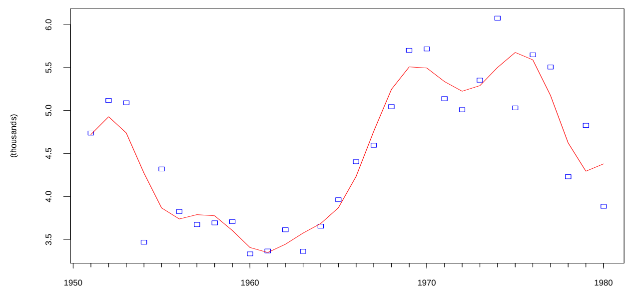
(c) smoothing by elimination of high frequency components

Retain a fraction f of the frequency components in the Fourier expansion of $\{X_t\}$, eliminating the top frequencies.

0 f 1

max smoothing $\longleftrightarrow$ no smoothing

Fig 1.22. Strike data $f = .40$



1.5.2 Estimation of Trend and Seasonality

Classical Decomposition Model:

$$X_t = m_t + s_t + Y_t$$

where $EY_t = 0$, $s_{t+d} = s_t$, and $\sum_{t=1}^{d} s_t = 0.$

Step 1: Estimate the trend using a simple moving average of length $q = d/2$ (or $(d\text{-}1)/2$).

Step 2: Estimate s_k, $k=1,\ldots,d$ using the average deviations from trend for each season.

Step 3: Deseasonalize the data by forming

$$d_t = x_t - \hat{s}_t, \quad t=1,\ldots,n$$

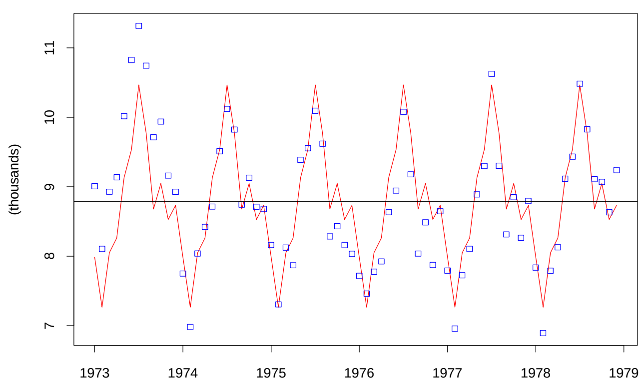
Step 4: Fit a parametric function m_t to the deseasonalized data $\{d_t\}$.

Step 5: Calculate the estimated noise

$$\hat{Y}_t = x_t - \hat{m}_t - \hat{s}_t$$

37

Figure 1.24. The deaths data.
data = blue boxes
s_t = red lines



38

Differencing to eliminate trend and seasonal components.

Backward Shift Operator B :

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

$$B^s X_t = X_{t-s}, \quad s=0, \pm 1, \ldots.$$

Difference Operator $\nabla = 1 - B$:

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

$$\nabla^2 X_t = (1-B)^2 X_t = (1-2B+B^2) X_t$$

$$= X_t - 2 X_{t-1} + X_{t-2}$$

∇ applied to a trend function.

$$m_t = a_0 + a_1 t$$

$$\nabla m_t = m_t - m_{t-1} = a_0 + a_1 t - (a_0 + a_1 (t-1))$$

$$= a_1$$

In particular, ∇^k applied to a polynomial of degree k gives a constant.

Seasonal Differencing $\nabla_d = (1 - B^d)$:

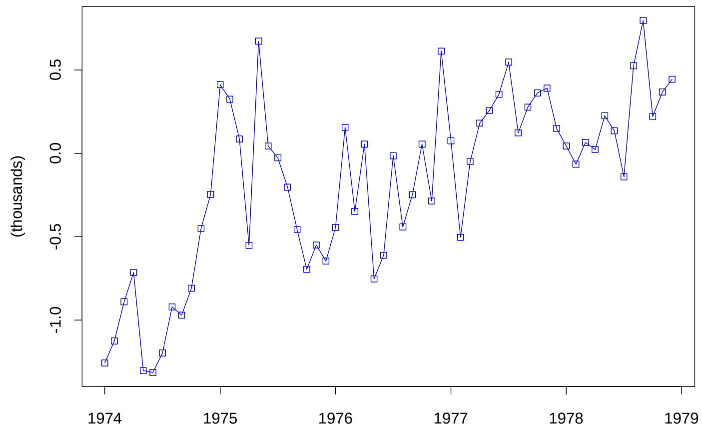
$$\nabla_d X_t = (1- B^d) X_t = X_t - X_{t-d}$$

If s_t is a seasonal component with period d, then

$$\nabla_d s_t = s_t - s_{t-d} = 0.$$

Figure 1.26. Differenced monthly accidental deaths.

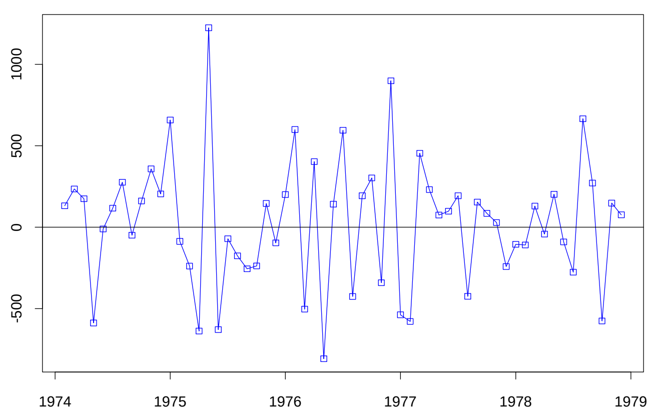
$$\nabla_{12} x_t = x_t - x_{t-12}, \quad t = 13, \ldots, 72.$$



Remark: Deseasonalized series still exhibits trend which we attempt to remove by differencing

Figure 1.27. Detrended and deseasonalized accidental deaths.

$$\nabla \nabla_{12} x_t = x_t - x_{t-1} - x_{t-12} + x_{t-13}, \quad t = 14, \ldots, 72.$$



1.6. Testing the Estimated Noise Sequence

Suppose $y_1, \ldots, y_n$ are observations from random variables $Y_1, \ldots, Y_n$ and you wish to test the hypothesis that the variables are IID.

(a) Sample ACF

Check to see if

$$|\hat{\rho}(h)| < 1.96/\sqrt{n} \ .$$

Reject if more than 5% fall outside bounds.

(b) Portmanteau Test

Under the IID hypothesis, $\sqrt{n}\,\hat{\rho}(h)$ is approx N(0,1)

so that

$$Q := n \sum_{j=1}^{k} \hat{\rho}^2(j) \text{ approx } \chi^2 \text{ with k d.f.}$$

Reject IID hypothesis if

$$Q > \chi^2_{1-\alpha}(k)$$

Two refinements:

(1) $Q_{LB} := n(n+2) \sum_{j=1}^{k} \hat{\rho}^2(j)/(n-j)$ (Ljung and Box)

(2) $Q := n(n+2)\sum_{j=1}^{k} \hat{\rho}^2_{WW}(j)/(n-j)$ (McLeod & Li)

where $\hat{\rho}_{WW}(j)$ is the sample ACF of the squared data $Y_1^2, \ldots, Y_n^2$. This test is designed for testing

data are IID $N(0,\sigma^2)$.

(c) Turning Point Test

Data has a **turning** point at time i if

$\{y_{i-1} < y_i$ and $y_{i+1} > y_i\}$ or $\{y_{i-1} > y_i$ and $y_{i+1} < y_i\}$

T = number of turning pts

is approx $N(\mu_T, \sigma^2_T)$

$\mu_T = 2(n-2)/3$, $\sigma^2_T = (16n-29)/90$

45

(d) Difference sign test

$S = \#\{i: y_i > y_{i-1}\} = \#\{i: \nabla y_i > 0\}$

is approx $N(\mu_S, \sigma^2_S)$

$\mu_S = (n-1)/2$, $\sigma^2_S = (n+1)/12$

A large positive (negative) value of $S - \mu_S$ implies presence of increasing (decreasing) trend.

46

(e) Rank Test

$$P = \#\{(i,j):\ y_j > y_i,\ i<j\}$$

is approx $N(\mu_P, \sigma^2_P)$

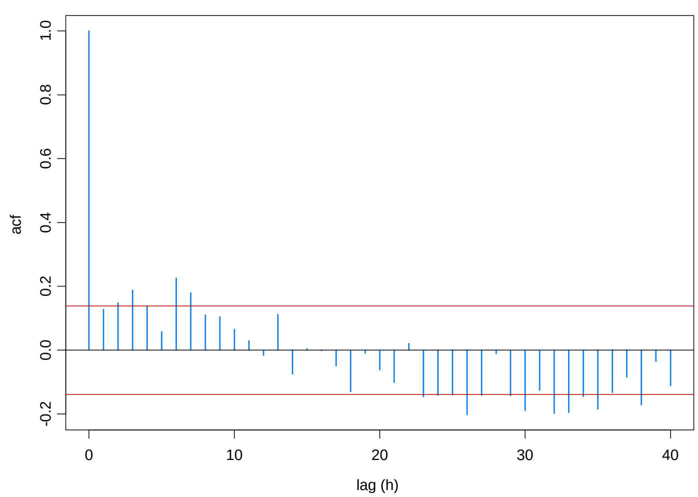
$\mu_P = n(n-1)/4$, $\sigma^2_P = n(n-1)(2n + 5)/72$

A large positive (negative) value of $P - \mu_P$ implies presence of increasing (decreasing) trend. Rank test is useful for detecting linear trends.

(f) Fitting an autoregressive model. Fit AR models and choose order that minimizes AICC (p=0 implies WN).

47

Example: Signal.TSM (Ex 1.1.4).



48

Example (cont): Signal.TSM (Ex 1.1.4).

Select Statistics > Residual analysis > Tests of Randomness

==

ITSM::(Tests of randomness on residuals)

==

Ljung - Box statistic = 51.841 Chi-Square (20), p-value = .00012
McLeod - Li statistic = 15.987 Chi-Square (20), p-value = .71746
\# Turning points = .13800E+03~AN(.13200E+03,sd = 5.9358),
p-value = .31210
\# Diff sign points = .10100E+03~AN(99.500,sd = 4.0927),
p-value = .71399
Rank test statistic = .10310E+05~AN(.99500E+04,sd = .47315E+03),
p-value = .44675
Jarque-Bera test statistic (for normality) = .86432 Chi-Square (2),
p-value = .64911

Order of Min AICC YW Model for Residuals = 7

49

Example: Lake Huron Data

Apply the foregoing tests to the residuals from the LS fit to the Lake Huron data,

$$y_t = x_t - a_0 - a_1 t, \quad t=1,\ldots,98.$$

Select Statistics > Residual analysis > Tests of Randomness

==

ITSM:: (Tests of randomness on residuals)

==

Ljung - Box statistic = .10783E+03 Chi-Square (20), p-value = .00000

McLeod - Li statistic = 68.714 Chi-Square (20), p-value = .00000

\# of Turning points = 40.000 ~ AN(64.000, sd = 4.1352), p-value = .00000

\# of Diff sign points = 50.000 ~ AN(48.500, sd = 2.8723), p-value = .60151

Rank test statistic = .23440E+04~AN(.23765E+04,sd = .16290E+03),
p-value = .84187

Order of Min AICC YW Model for Residuals = 2

50