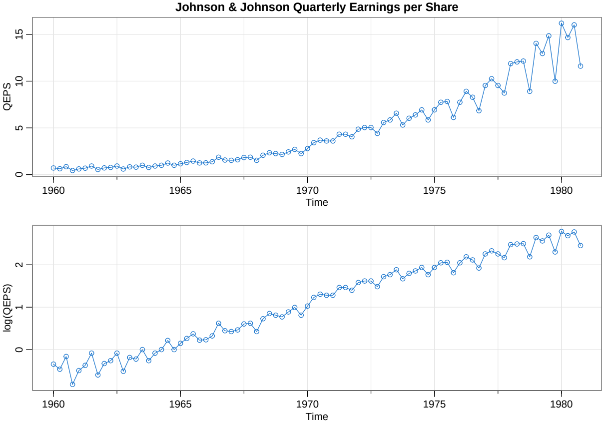
Lecture 18. Time Series Data

R and Data Visualization

BIG2006, Hanyang University, Fall 2022

Example 1. Johnson & Johnson Quarterly Earnings

- 84 quarters (21 years) from 1960 to 1980

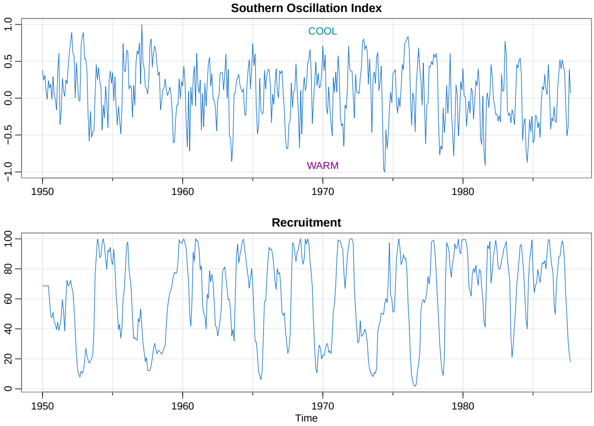


- Gradually increasing underlying trend
- Regular variation superimposed on the trend that seems ro repeat over quarters.

```r
# library(astsa)
par(mfrow=2:1)
tsplot(jj, ylab="QEPS", type="o", col=4,
       main="Johnson & Johnson Quarterly Earnings per Share")
tsplot(log(jj), ylab="log(QEPS)", type="o", col=4)
```

Example 2. El Niño and Fish Population

- Monthly values of the Southern Oscillation Index (SOI) and associated Recruitment (number of new fish) over 1950-1987



- Exhibit repetitive behavior, with regularly repeating cycle
- Two series tend to be somewhat related; the fish population is dependent on the SOI.

```
par(mfrow = c(2,1))
tsplot(soi, ylab="", xlab="", main="Southern Oscillation Index", col=4)
text(1970, .91, "COOL", col="cyan4")
text(1970,-.91, "WARM", col="darkmagenta")
tsplot(rec, ylab="", main="Recruitment", col=4)
```

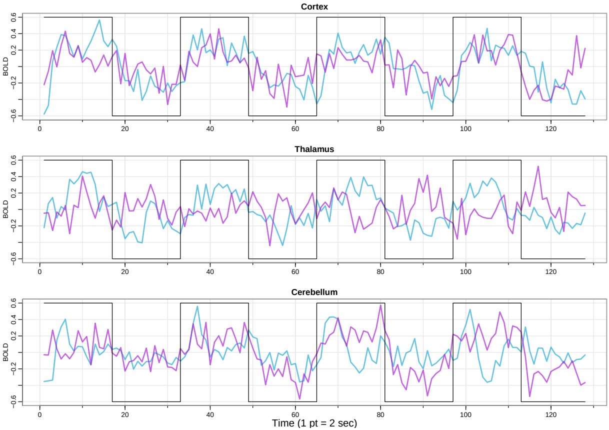
Note: The SOI measures changes in air pressure, related to sea surface temperatures in the central Pacific Ocean.

The central Pacific warms every three to seven years due to the El Niño effect, which has been blamed, in particular, for the 1997 floods in the midwestern portions of the United States.

Example 3. fMRI Imaging

- Data collected from various locations in the brain via functional magnetic resonance imaging (fMRI)
- In this example, two subjects were given periodic brushing on the hand. The stimulus was applied for 32 seconds and then stopped for 32 seconds; thus, the signal period is 64 seconds.
- The sampling rate was one observation every 2 seconds for 256 seconds ($n = 128$).
- Blood oxygenation level dependent (BOLD) signal intensity: measures areas of activation in the brain

- The periodicities appear strongly in the motor cortex series and less strongly in the thalamus and cerebellum.
- One has series from different areas of the brain suggests testing whether the areas are responding differently to the brush stimulus.



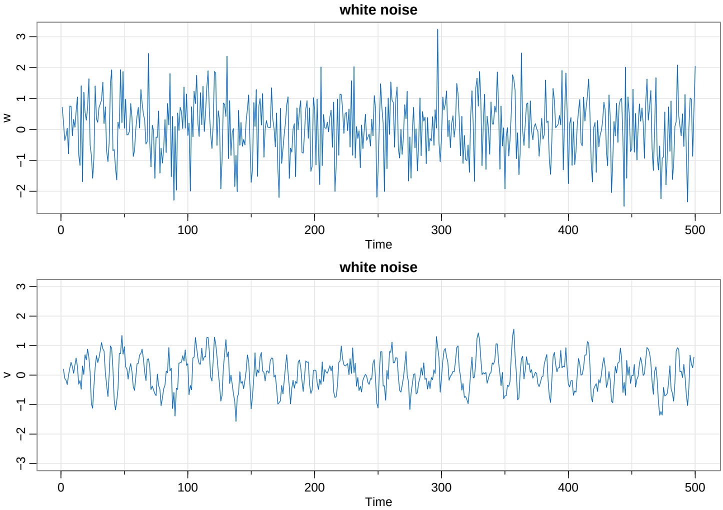
```
par(mfrow=c(3,1))
culer = c(rgb(.12,.67,.85,.7), rgb(.67,.12,.85,.7))
u = rep(c(rep(.6,16), rep(-.6,16)), 4) # stimulus signal
tsplot(fmri1[,4], ylab="BOLD", xlab="", main="Cortex",
       col=culer[1], ylim=c(-.6,.6), lwd=2)
lines(fmri1[,5], col=culer[2], lwd=2)
lines(u, type="s")
tsplot(fmri1[,6], ylab="BOLD", xlab="", main="Thalamus",
       col=culer[1], ylim=c(-.6,.6), lwd=2)
lines(fmri1[,7], col=culer[2], lwd=2)
lines(u, type="s")
tsplot(fmri1[,8], ylab="BOLD", xlab="", main="Cerebellum",
       col=culer[1], ylim=c(-.6,.6), lwd=2)
lines(fmri1[,9], col=culer[2], lwd=2)
lines(u, type="s")
mtext("Time (1 pt = 2 sec)", side=1, line=1.75)
```

Time Series Statistical Models

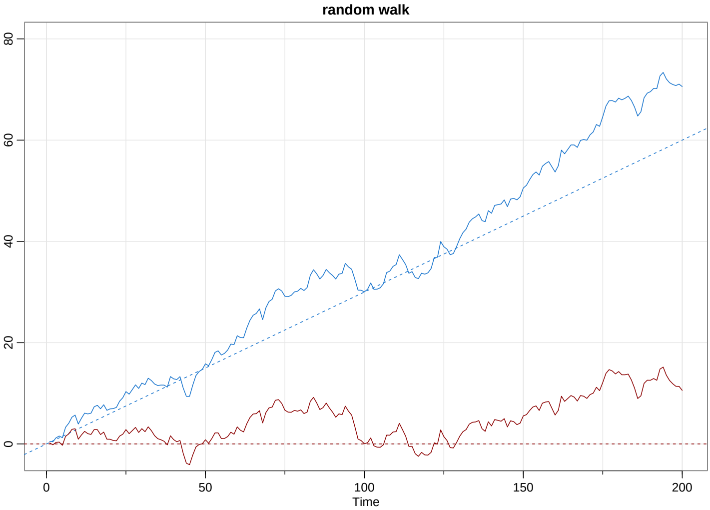
- **Time series**: a collection of random variables indexed according to the order they are obtained in time
- Time series models: provide plausible descriptions for sample data, i.e., describe the character of data that seemingly fluctuate in a random fashion over time

Note: Time series regression and Autoregressive Moving Average (ARIMA) models are basic models to analyze time series data.

White Noise


white noise
w
3
2
1
0
−1
−2
0
100
200
300
400
500
Time
white noise
v
3
2
1
0
−1
−2
−3
0
100
200
300
400
500
Time

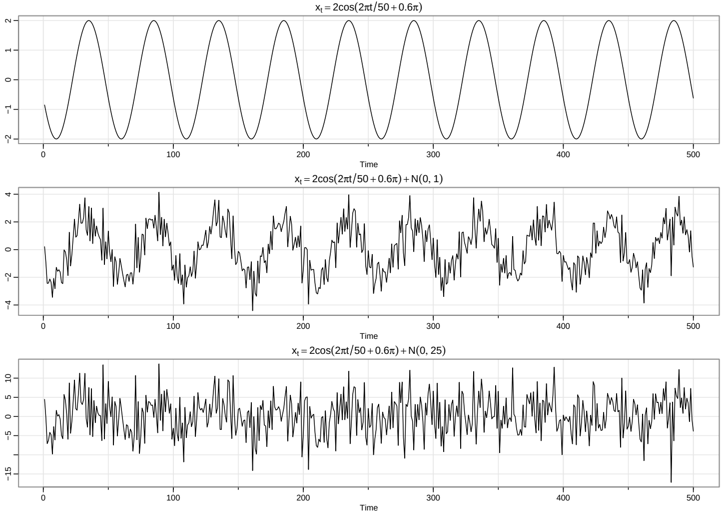
Random Walk with Drift



```r
set.seed(314159265) # so you can reproduce the results
# White noise
par(mfrow=2:1)
w = rnorm(500) # 500 N(0,1) variates
# moving average v_t = 1/3 (w_{t-1}+w_t+w_{t+1})
v = filter(w, sides=2, filter=rep(1/3,3))
tsplot(w, col=4, main="white noise")
tsplot(v, ylim=c(-3,3), col=4, main="white noise")

# Random walk with drift
w  = rnorm(200)
x  = cumsum(w)
wd = w +.3
xd = cumsum(wd)
tsplot(xd, ylim=c(-2,80), main="random walk", ylab="", col=4)
abline(a=0, b=.3, lty=2, col=4) # drift
lines(x, col="darkred")
abline(h=0, col="darkred", lty=2)
```

Signal in Noise

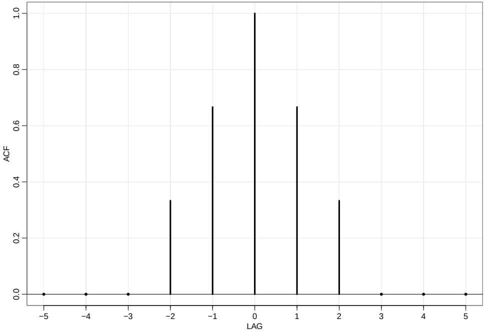


Measures of Dependence

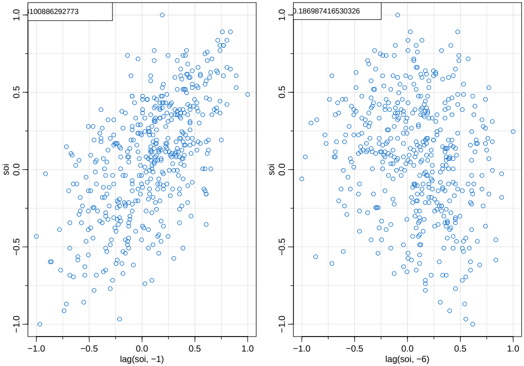
- Autocorrelation Function (ACF): measures the linear predictability of the series at time t, say x_t, using the value x_s
- Sample ACF: estimating ACF is similar to estimating of correlation in the usual setup where we have pairs of observations, (x_i, y_i), for $i = 1, \dots, n$.

  If we have time series x_t for $t = 1, \dots, n$, then the pairs of observations for estimating $\rho(h)$ are the $n - h$ pairs given by $\{(x_t, x_{t+h}); t = 1, \dots, n - h\}$.

```r
# Autocorrelation function of a three-point moving average.
# Note that ACF is symmetric about lag zero
ACF = c(0,0,0,1,2,3,2,1,0,0,0)/3
LAG = -5:5
tsplot(LAG, ACF, type="h", lwd=3, xlab="LAG")
abline(h=0)
points(LAG[-(4:8)], ACF[-(4:8)], pch=20)
axis(1, at=seq(-5, 5, by=2))
```

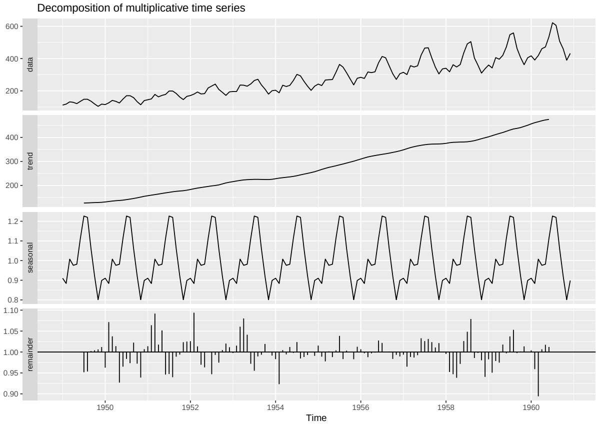


```r
# An example of using the SOI series
r = acf1(soi, 6, plot=FALSE) # sample acf values
par(mfrow=c(1,2), mar=c(2.5,2.5,0,0)+.5, mgp=c(1.6,.6,0))
plot(lag(soi,-1), soi, col="dodgerblue3", panel.first=Grid())
legend("topleft", legend=r[1], bg="white", adj=.45, cex = 0.85)
plot(lag(soi,-6), soi, col="dodgerblue3", panel.first=Grid())
legend("topleft", legend=r[6], bg="white", adj=.25, cex = 0.8)
```



Time Series Decomposition

```
#library(tseries); library(forecast)
decomposeAP <- decompose(AirPassengers,"multiplicative")
autoplot(decomposeAP)
```



Reference

- Shumway, R. H. and Stoffer, D. S. Time Series Analysis and Its Applications (4th). Springer. Chapters 1 and 2.