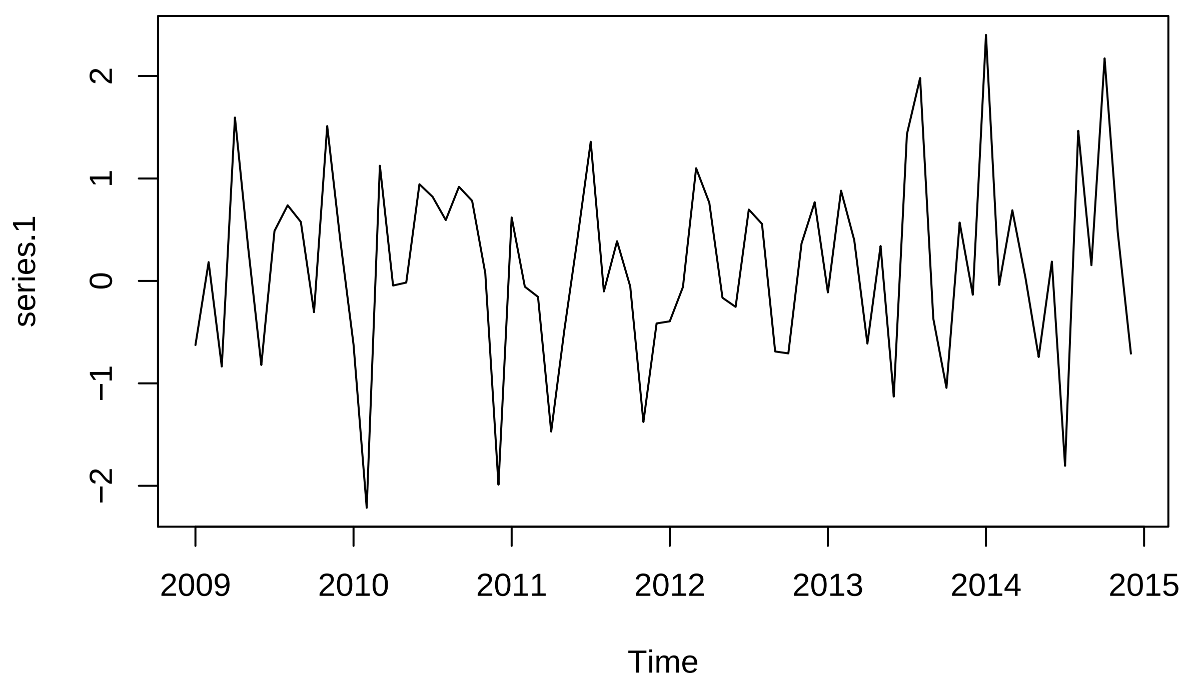
Simulations of Time Series Data

Andira Putri

Generating Time Series Data in R

The function `ts()` can convert a numeric vector into a time series object in R. The syntax is `ts(vectorname, start=, end=, frequency=)`, where start/end are the first/last time points, and frequency is the number of observations per unit in time. `frequency=1` means yearly, `frequency=12` means monthly, etc. Let's generate a time series data set here:

```
set.seed(1)
#randomly generate vector of length 72
vector=rnorm(72,0,1)
#generate time series spanning 6 years
#6 years --> 72 months, frequency=12 puts time series in terms of months
series.1=ts(vector,start=c(2009,1),end=c(2014,12),frequency=12)
plot.ts(series.1)
```

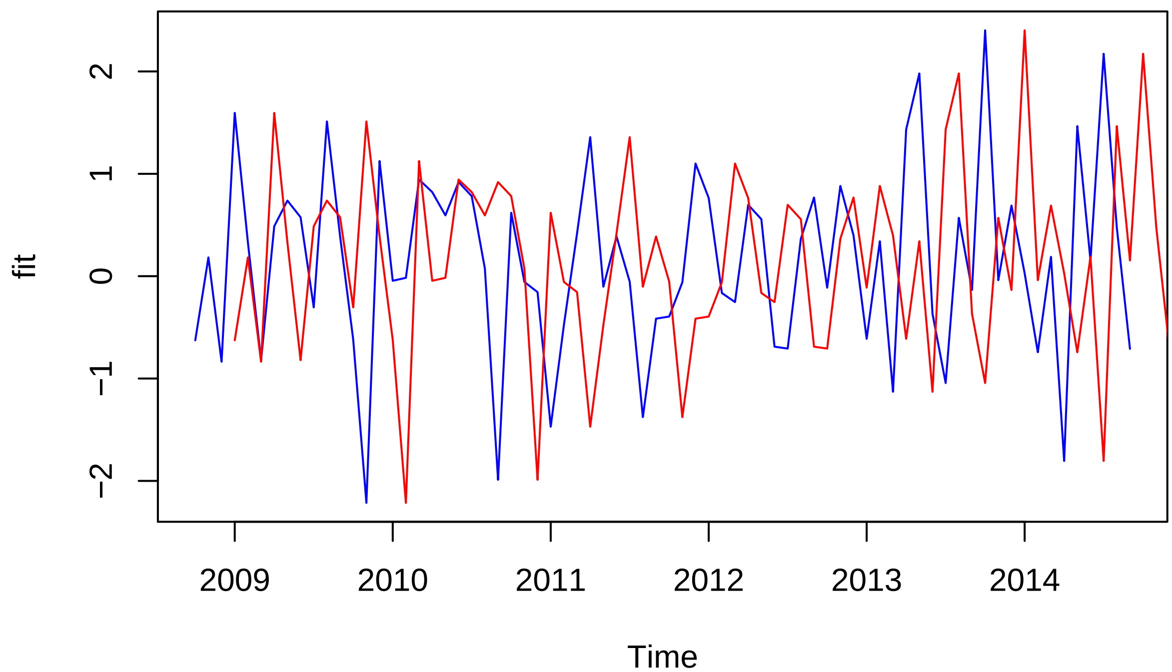


We see that the data revolves around the mean 0, and there is no random walk.

Lag

Lag causes a delay so that you can study how similar a time series is to itself. Lag is an important component of autocorrelation studies…yes, foreshadowing! I shift the time series 3 units and superimpose the two series on a plot. The red lines represent the delayed time series.

```r
fit=lag(series.1,3)
plot.ts(fit,col="blue")
lines(series.1,col="red")
library(forecast)
```



Autocorrelation

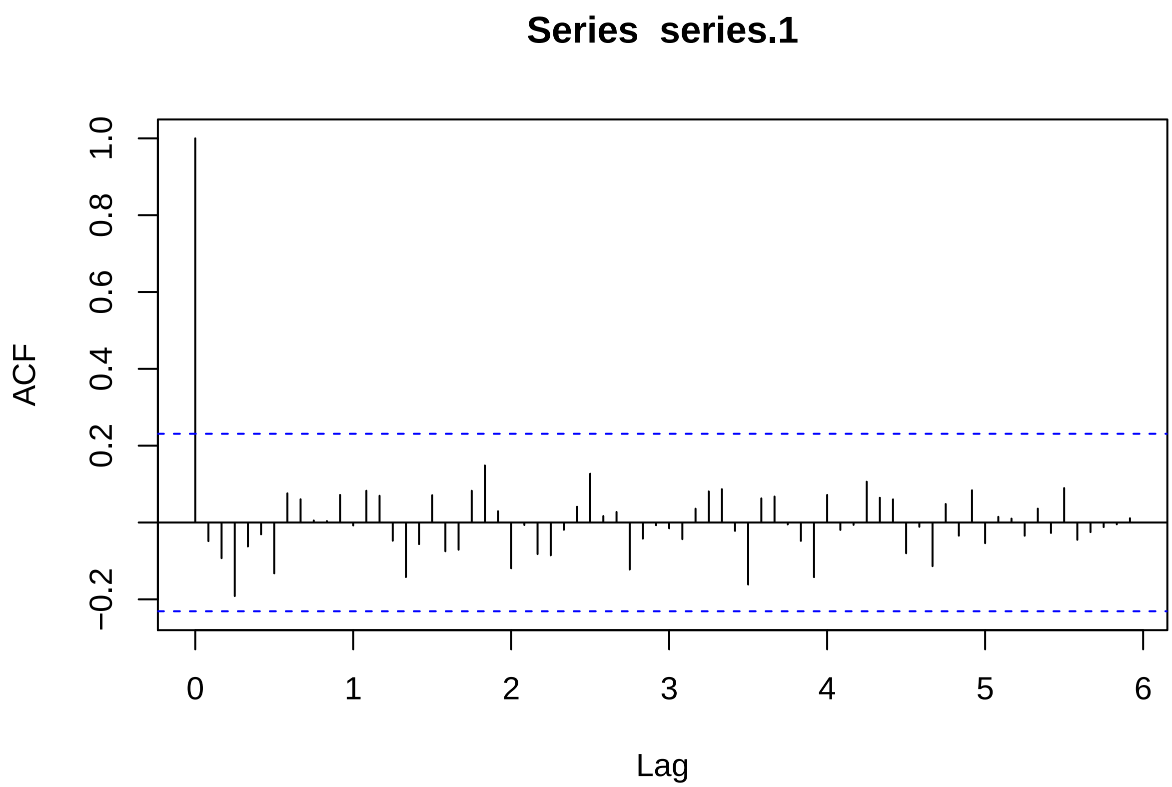
Autocorrelation functions (`acf`) are useful for measuring the linear predictability of the series at time t (x_t) using only the variable (x_s). The function is given by:

$p(s,t) = \frac{\gamma(s,t)}{\sqrt{\gamma(s,s)\gamma(t,t)}}$, where $\gamma(s,t) = cov(x_s, x_t)$ and $\gamma(t,t) = cov(x_t, x_t) = var(x_t)$. Thank you Math Stat.

Another forumalation is given measurements $Y_1, Y_2, ..., Y_N$ at time $X_1, X_2, ..., X_N$, the lag k autocorrelation function is defined as:

$r_k = \frac{\sum_{i=1}^{N-k}(Y_i - \bar{Y})(Y_{i+k} - \bar{Y})}{\sum_{i=1}^{N}(Y_i - \bar{Y})^2}$

```r
library(forecast)
#generate correlogram
acf=acf(series.1,500)
```



All correlograms start with an autocorrelation of 1; this is because when t=0, we are comparing the time series with itself. Periodicity is a good indicator of frequency in the time series data. For example, if each peak in a correlogram occurs when t is a multiple of 7, it is likely that the data is in terms of weeks and it's not just a coincidence.

When using the entirety of time series data, interpreting correlograms might not be easy. Partial autocorrelation functions (`pacf`) controls the values of the time series at shorter lags. This process removes the interference and resonance from multiple cycles and gives a more clear periodicity.

Great reads!

- https://www.alanzucconi.com/2016/06/06/autocorrelation-function/#part2
- https://www.itl.nist.gov/div898/handbook/eda/section3/eda35c.htm