First things first

ARIMA MODELS IN R

David Stoffer

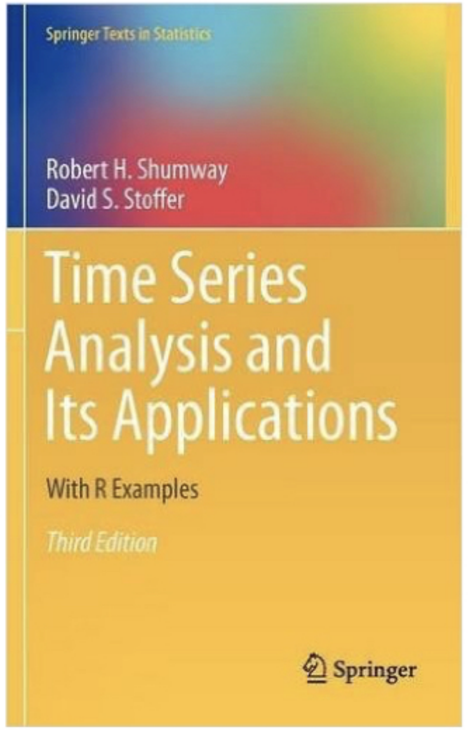
Professor of Statistics at the University of Pittsburgh

About Me

- Professor of Statistics

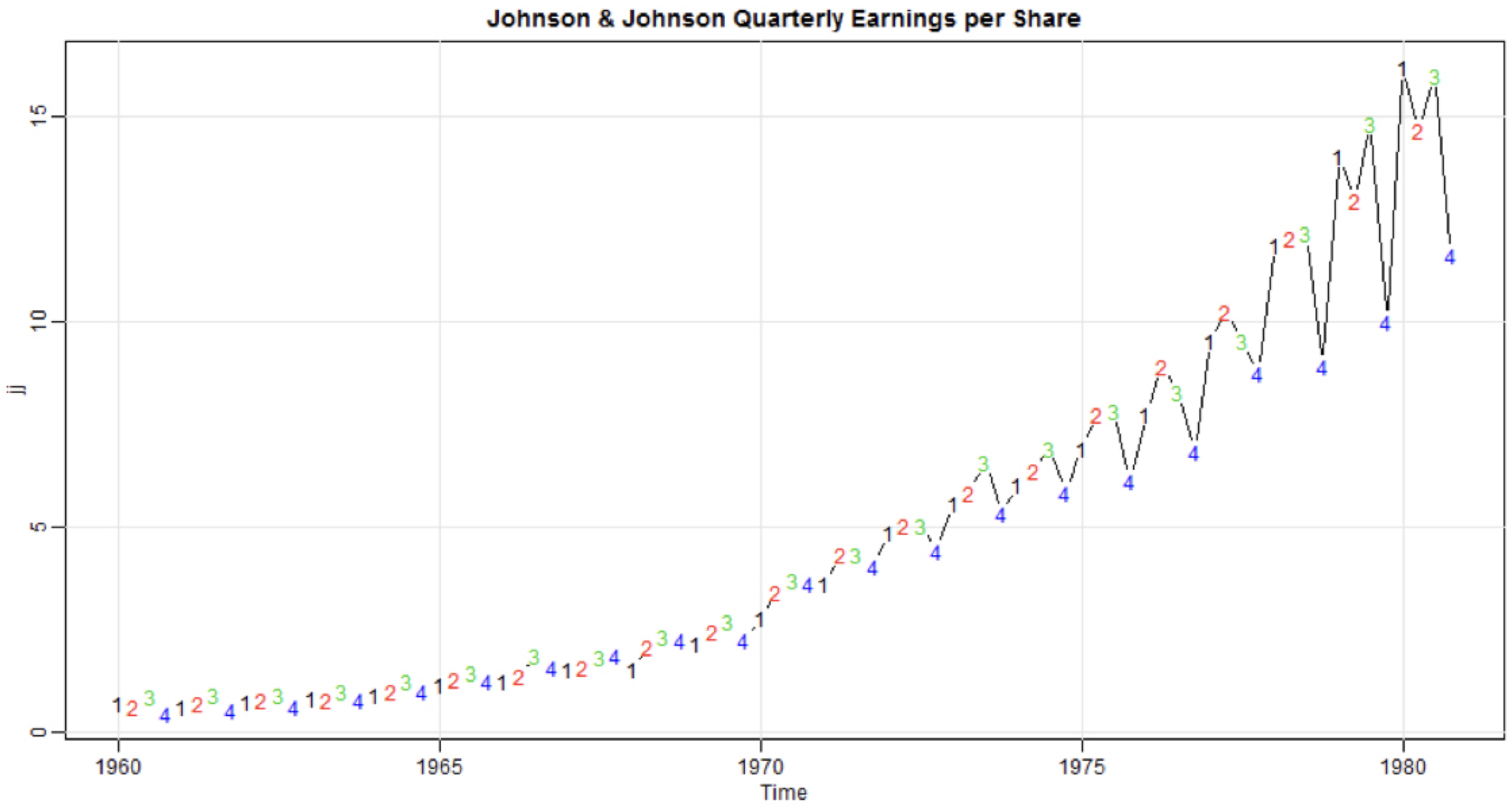
About Me

- Professor of Statistics
- Co-author of two texts on time series
- `astsa` package



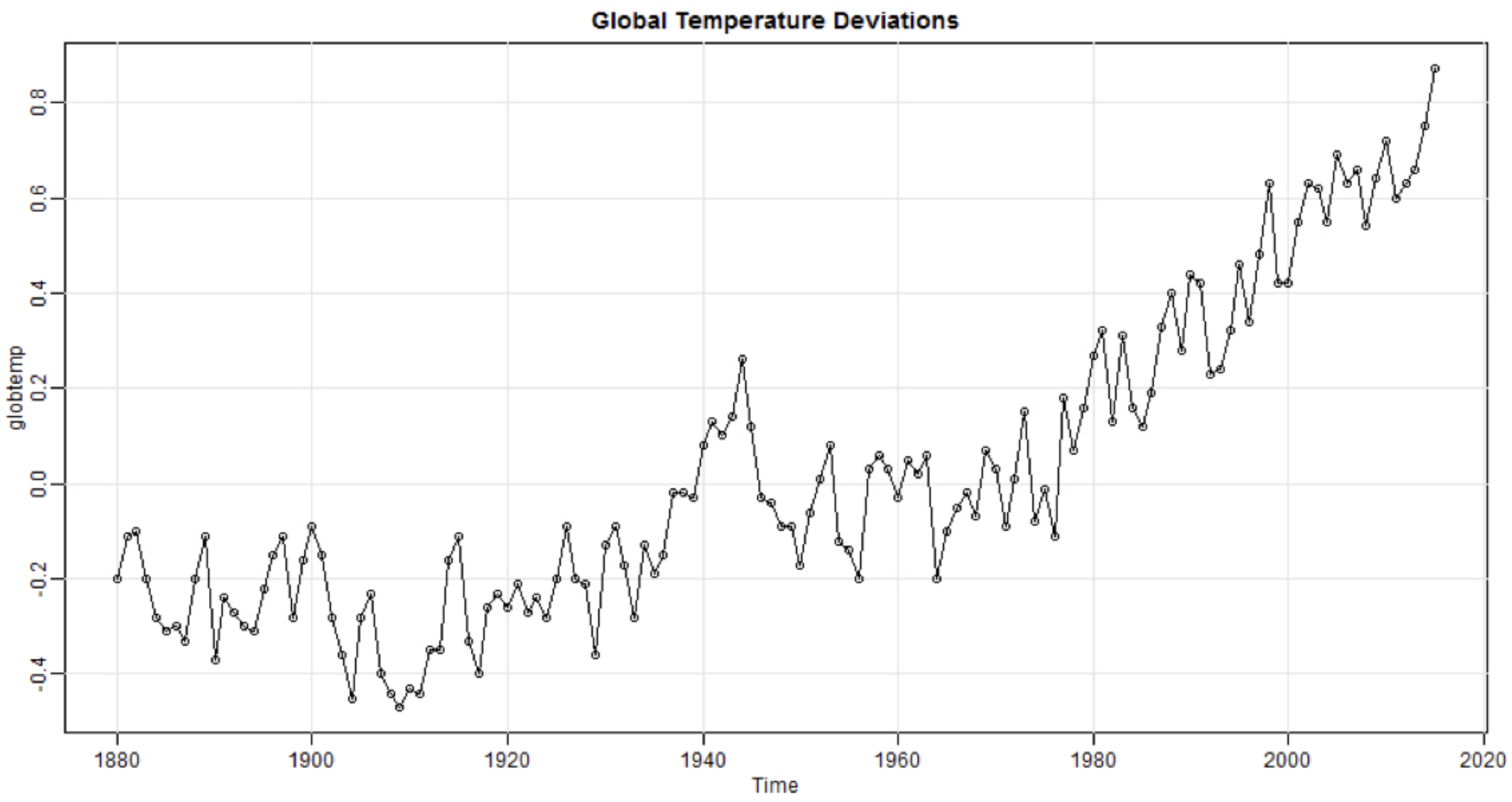
Time Series Data - I

```r
library(astsa)
plot(jj, main = "Johnson & Johnson Quarterly Earnings per Share", type = "c")
text(jj, labels = 1:4, col = 1:4)
```



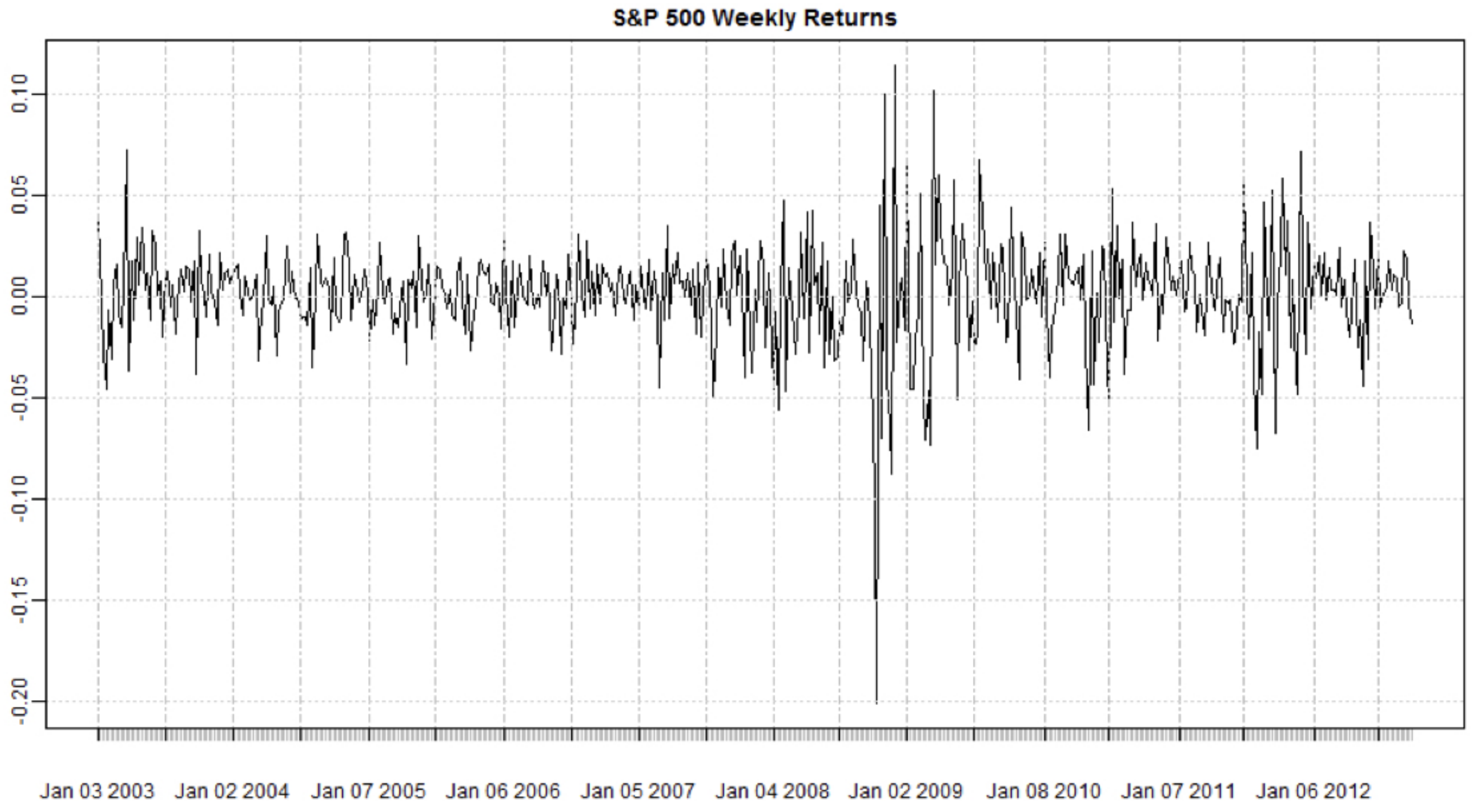
Time Series Data - II

```
library(astsa)
plot(globtemp, main = "Global Temperature Deviations", type= "o")
```



Time Series Data - III

```r
library(xts)
plot(sp500w, main = "S&P 500 Weekly Returns")
```



Time Series Regression Models

Regression: $Y_i = \beta X_i + \epsilon_i$, where ϵ_i is white noise

White Noise:

- independent normals with common variance
- is basic building block of time series

AutoRegression: $X_t = \phi X_{t-1} + \epsilon_t$ (ϵ_t is white noise)

Moving Average: $\epsilon_t = W_t + \theta W_{t-1}$ (W_t is white noise)

ARMA: $X_t = \phi X_{t-1} + W_t + \theta W_{t-1}$

Let's practice!

ARIMA MODELS IN R

Stationarity and nonstationarity

ARIMA MODELS IN R

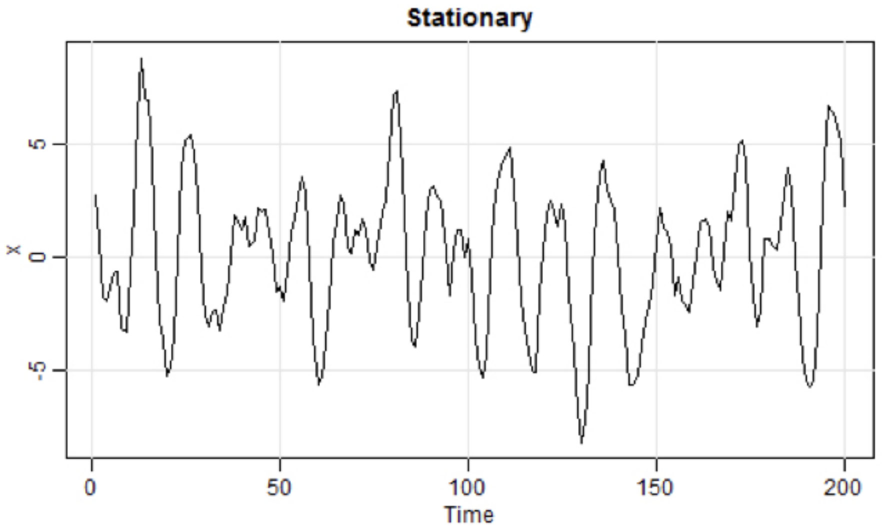
David Stoffer
Professor of Statistics at the University of Pittsburgh

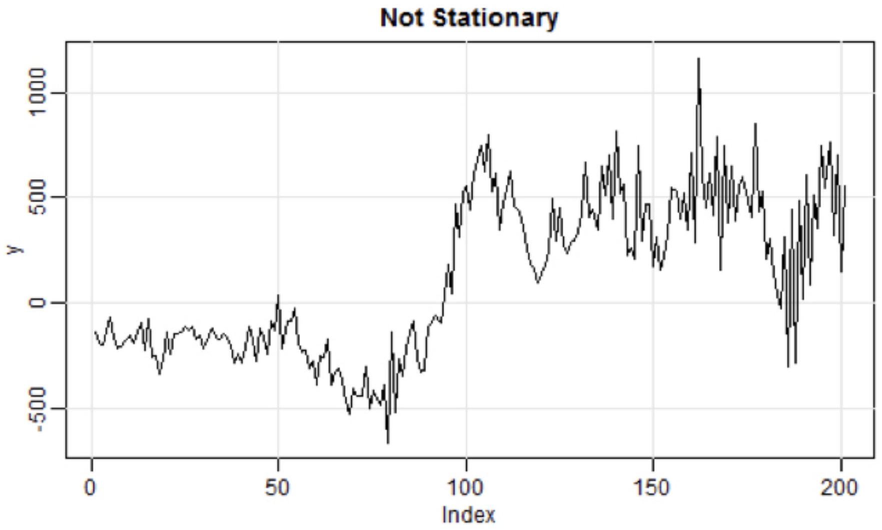
datacamp

Stationarity

A time series is stationary when it is "stable", meaning:

- the mean is constant over time (no trend)
- the correlation structure remains constant over time





Stationarity

Given data, $x_1, ..., x_n$ we can estimate by averaging

For example, if the mean is constant, we can estimate it by the sample average $\bar{x}$

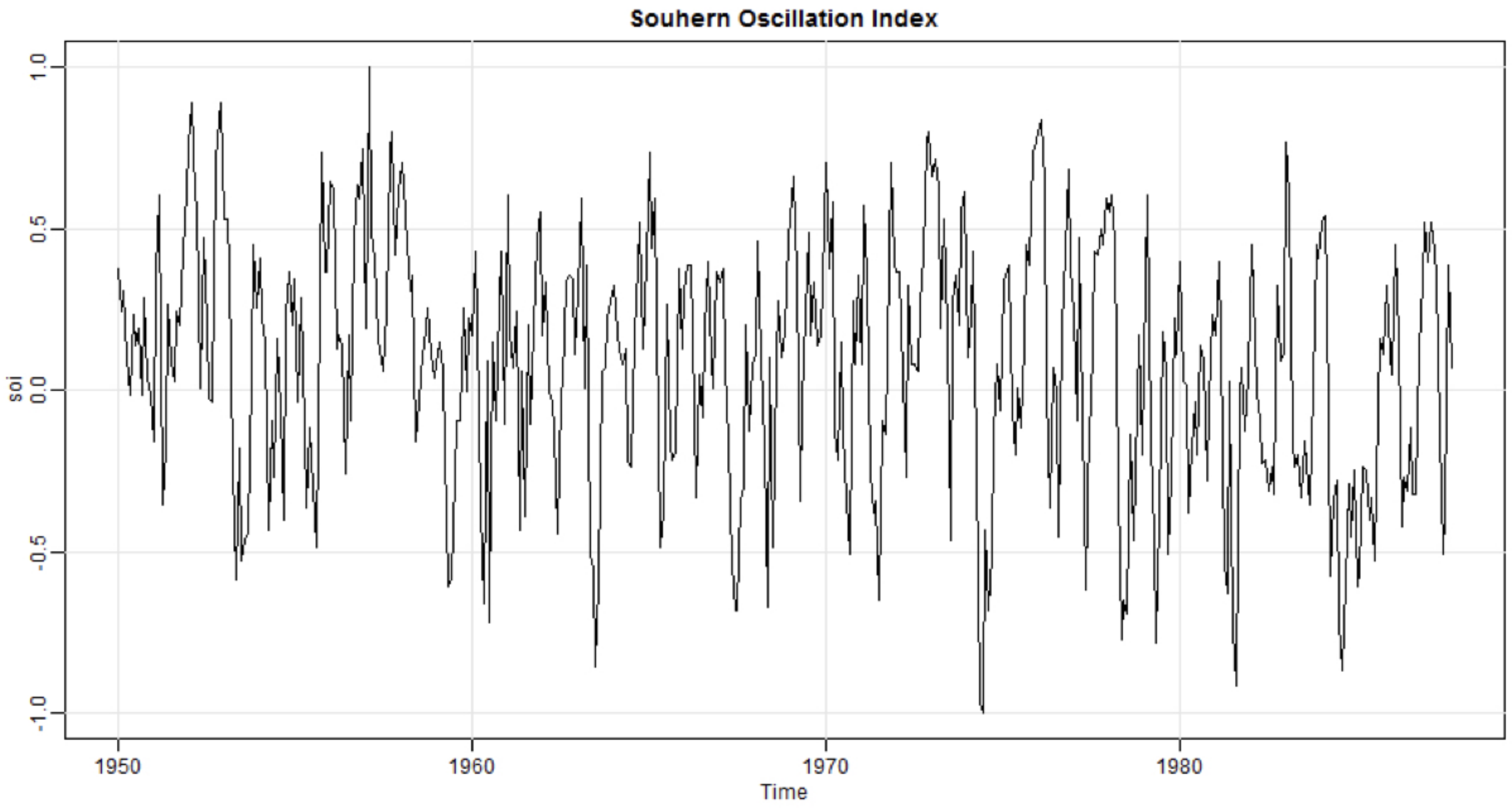
Pairs can be used to estimate **correlation** on different lags:

$(x_1, x_2), (x_2, x_3), (x_3, x_4)$, ... for lag 1

$(x_1, x_3), (x_2, x_4), (x_3, x_5)$, ... for lag 2

Southern Oscillation Index

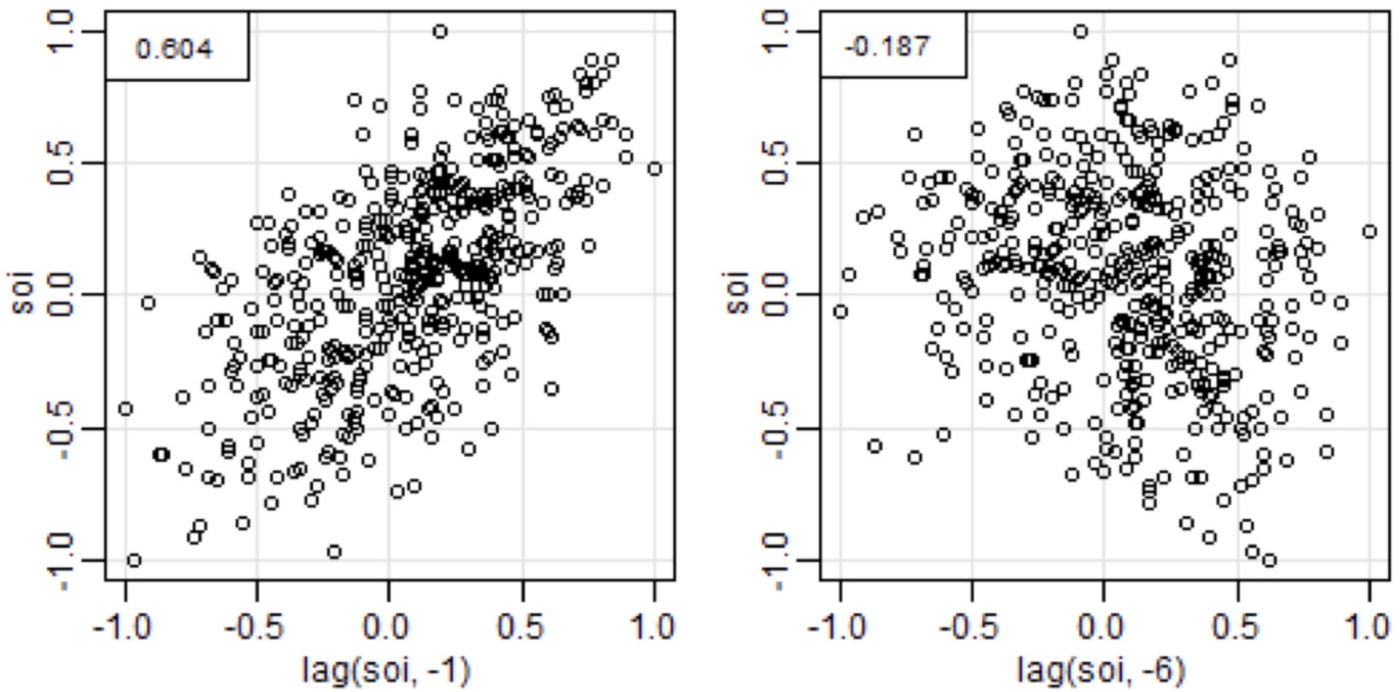
Reasonable to assume stationary, but perhaps some slight trend.



Southern Oscillation Index

To estimate autocorrelation, compute the correlation coefficient between the time series and itself at various lags.

Here you see how to get the correlation at lag 1 and lag 6.

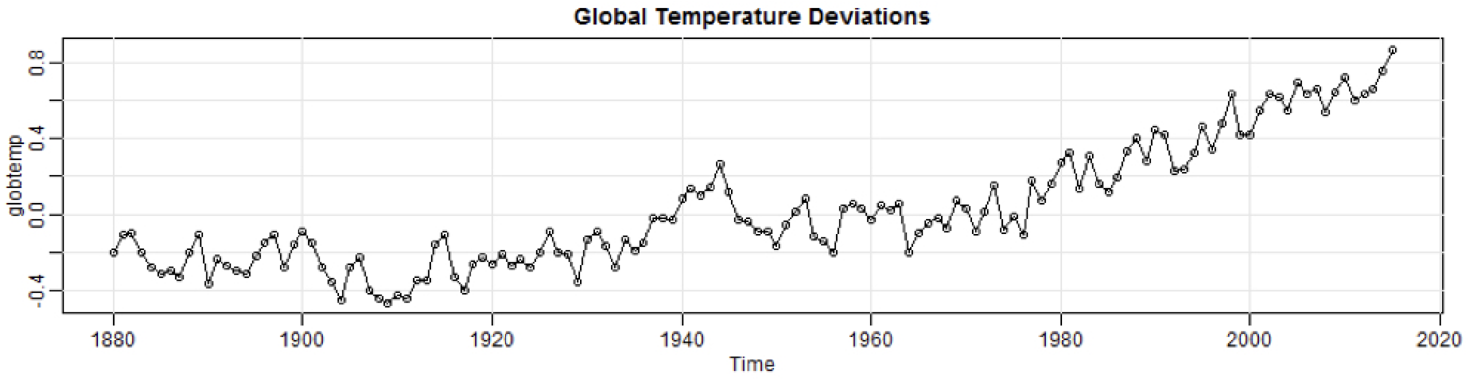


Random Walk Trend

Not stationary, but differenced data are stationary.

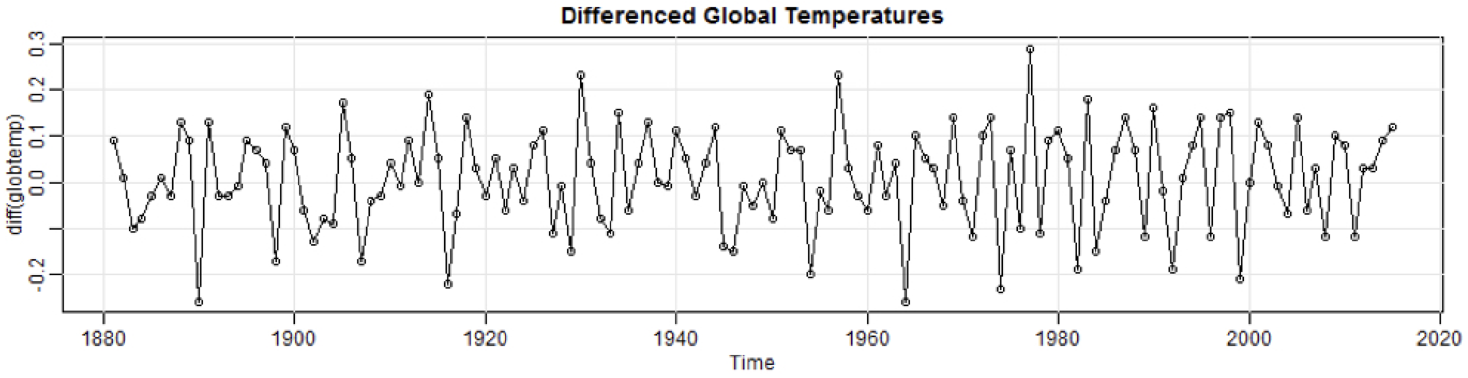
$$X_t$$

`globtemp`



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

`diff(globtemp)`



Trend Stationarity

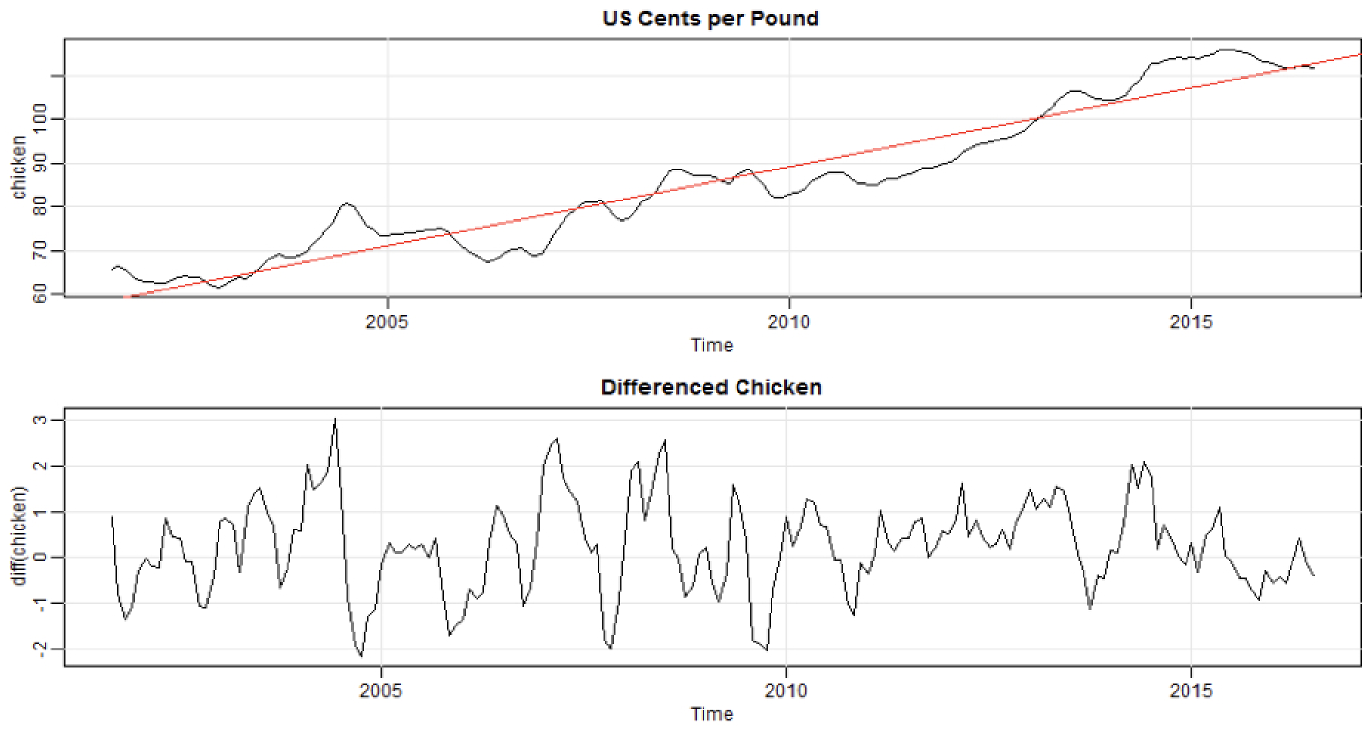
Stationarity around a trend, differencing still works!

X_t

chicken

$X_t - X_{t-1}$

diff(chicken)



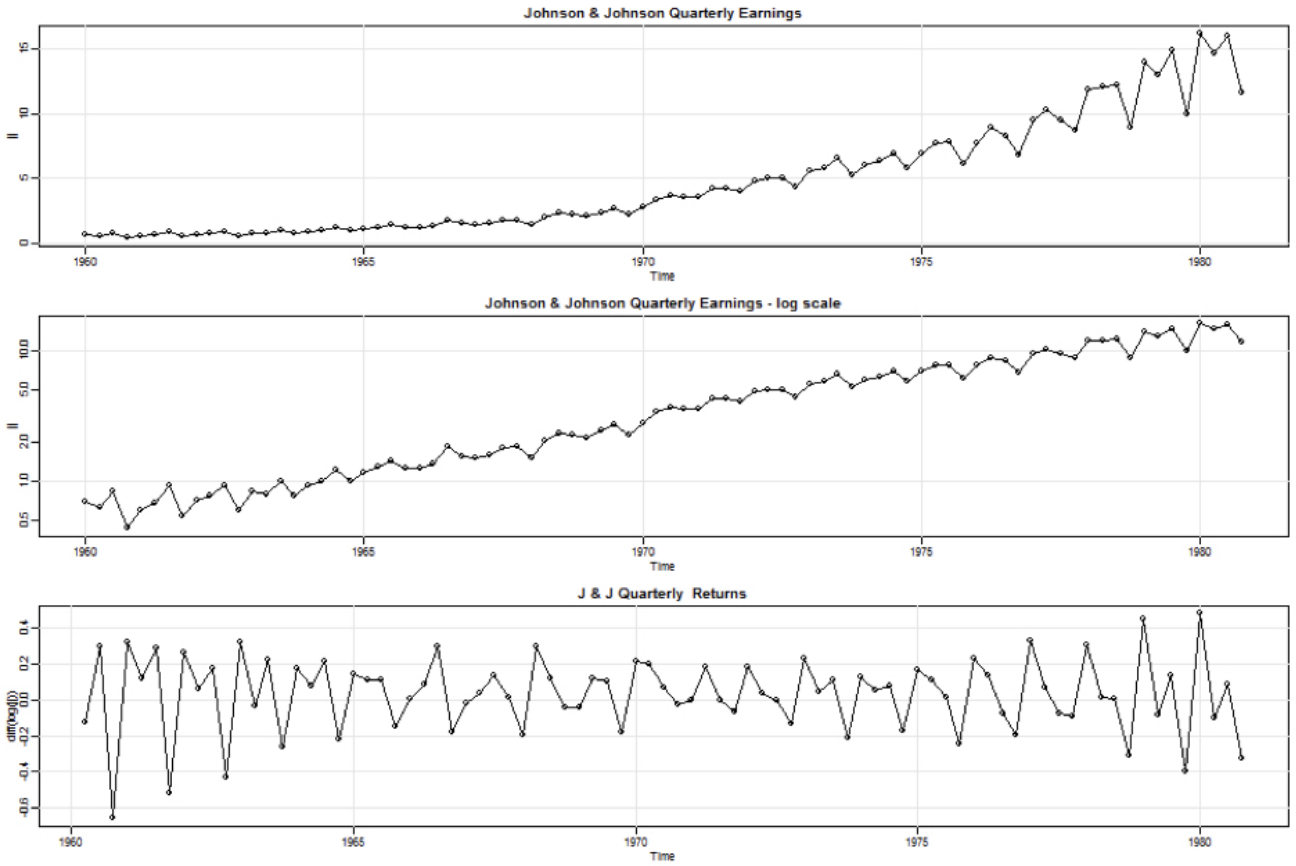
Nonstationarity in trend and variability

First log, then difference

X_t

$\log(X_t)$

$\log(X_t) - \log(X_{t-1})$



Let's practice!

ARIMA MODELS IN R

Stationary time series: ARMA

ARIMA MODELS IN R

David Stoffer
Professor of Statistics at the University of Pittsburgh

Wold Decomposition

Wold proved that any stationary time series may be represented as a linear combination of white noise:

$$X_t = W_t + a_1 W_{t-1} + a_2 W_{t-2} + \ldots$$

For constants $a_1, a_2, \ldots$

Any **ARMA** model has this form, which means they are suited to modeling time series.

Note: Special case of MA(q) is already of this form, where constants are 0 after q-th term.

Generating ARMA using arima.sim()

- Basic syntax:

```
arima.sim(model, n, ...)
```

- `model` is a list with order of the model as `c(p, d, q)` and the coefficients
- `n` is the length of the series

Generating and plotting MA(1)

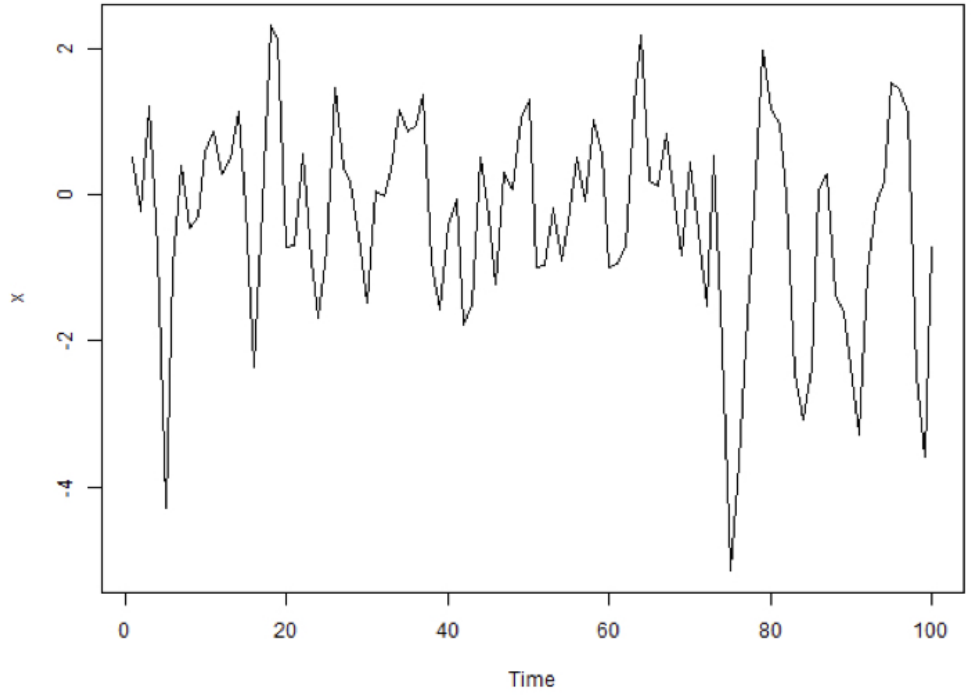
Generate MA(1) given by

$$X_t = W_t + 0.9W_{t-1}$$

Generating and plotting MA(1)

Generate MA(1) given by

$$X_t = W_t + 0.9W_{t-1}$$




```r
x <- arima.sim(list(order = c(0, 0, 1), ma = 0.9), n = 100)
plot(x)
```

Generating and plotting AR(2)

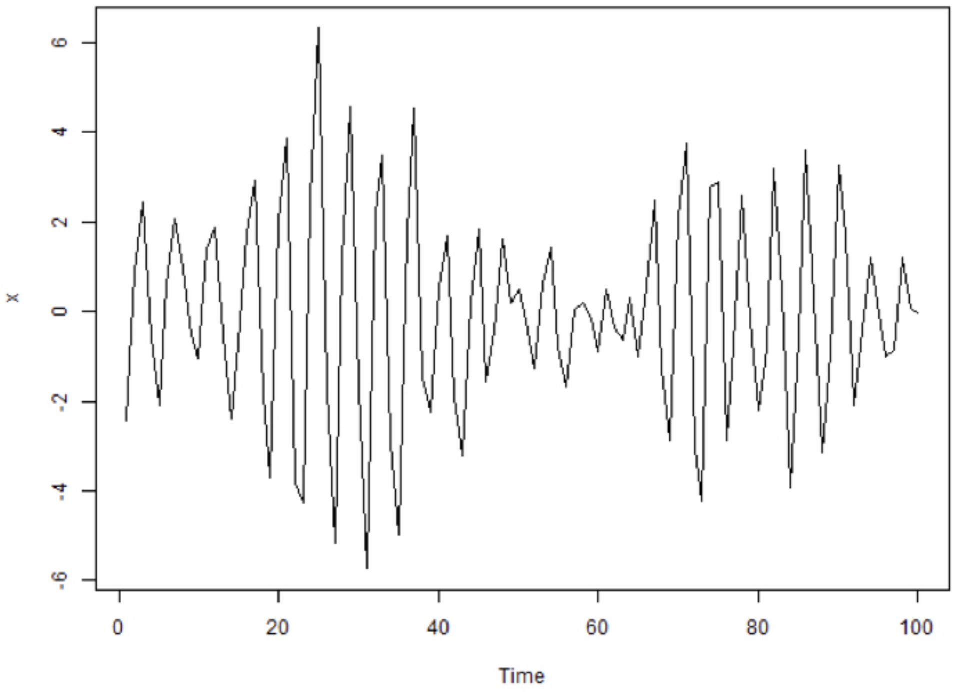
Generate AR(2) given by

$$X_t = -0.9X_{t-2} + W_t$$

Generating and plotting AR(2)

Generate AR(2) given by

$$X_t = -0.9X_{t-2} + W_t$$




```r
x <- arima.sim(list(order = c(2, 0, 0), ar = c(0, -0.9)), n = 100)
plot(x)
```

Let's practice!

ARIMA MODELS IN R