Predicting the Future

Analysing Time Series data

Avishek Sen Gupta
Bangalore
[avishek.net]

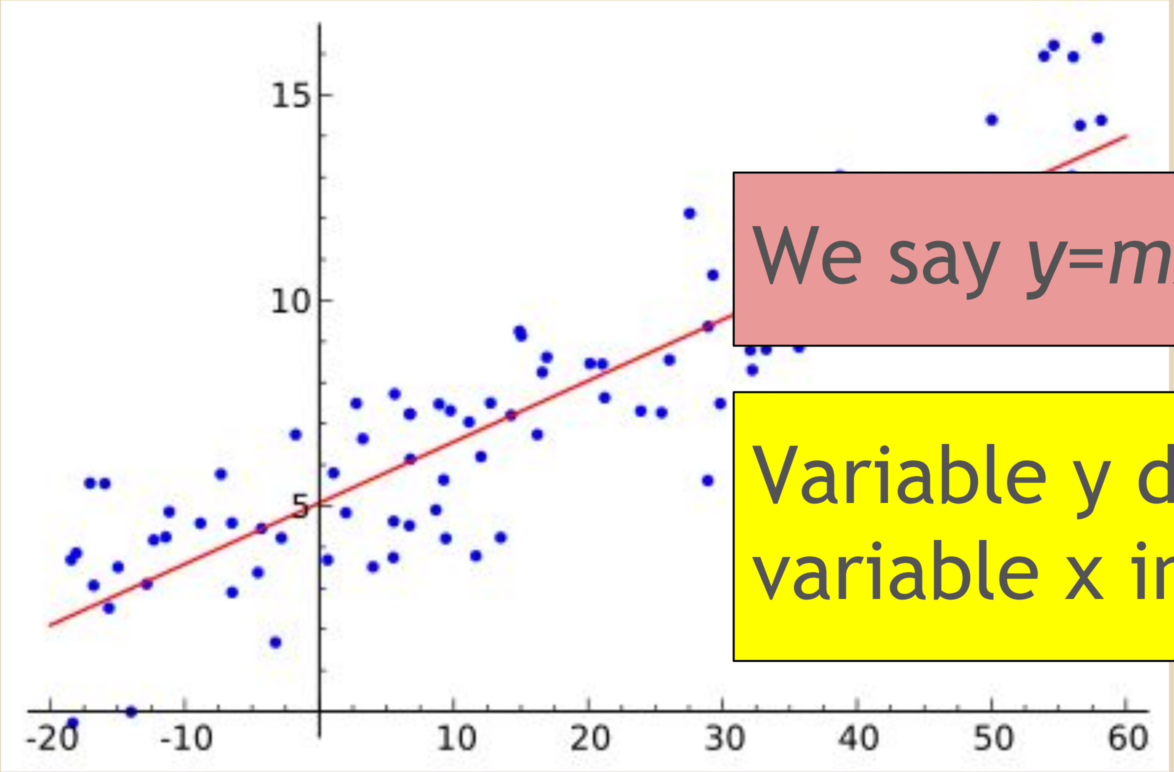
What makes Time Series analysis feasible?

What gives us the confidence that we can predict the future, when all we know is the past?

Key Insight: The future is like the past. Or, restated, future behaviour depends upon past behaviour.

Building our intuition

Remember linear regression?

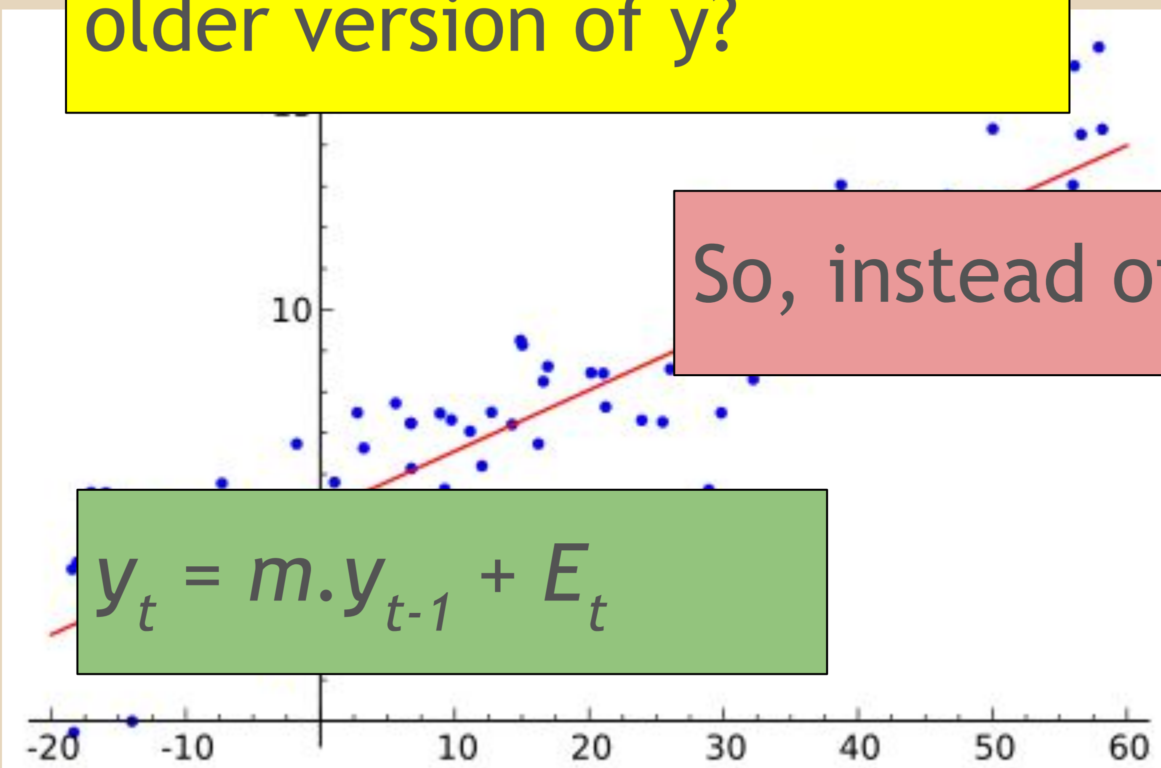


We say $y=mx + c$

Variable y depends upon variable x in some fashion

Building our intuition

What if x was simply an older version of y?



So, instead of $y = m.x + c$

$y_t = m.y_{t-1} + E_t$

Building our intuition

$y = m.x + c$ $\Rightarrow$ $y_t = m.y_{t-1} + E_t$

This is called a recurrence relation

It's like regression, except that the dependent variable depends upon a past version of itself.

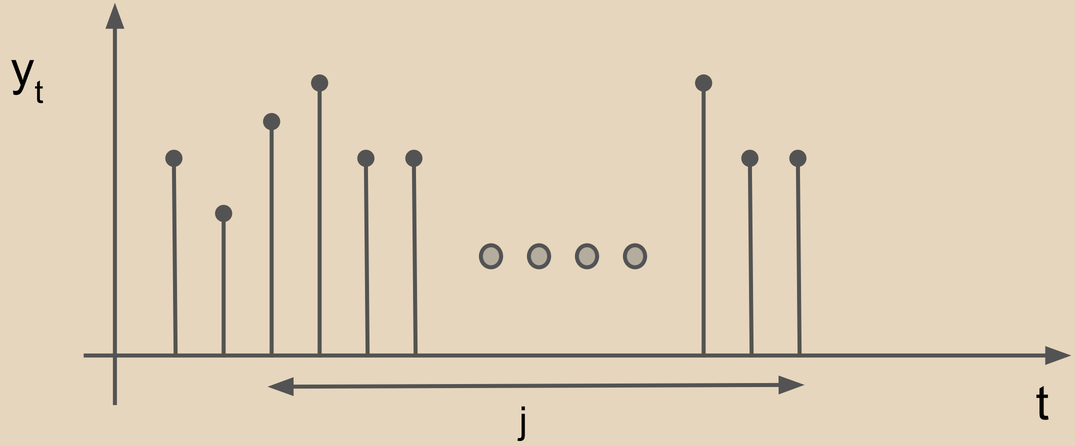
Building our intuition

Now, let's expand this a bit...

$$y_t = m.y_{t-1} + E_t$$
$$y_t = m.(m.y_{t-2} + E_{t-1}) + E_t$$
$$y_t = m^2.y_{t-2} + m.E_{t-1} + E_t$$
$$y_t = m^2.(m.y_{t-3} + E_{t-2}) + m.E_{t-1} + E_t$$
$$y_t = m^3.y_{t-3} + m^2.E_{t-2} + m.E_{t-1} + E_t$$

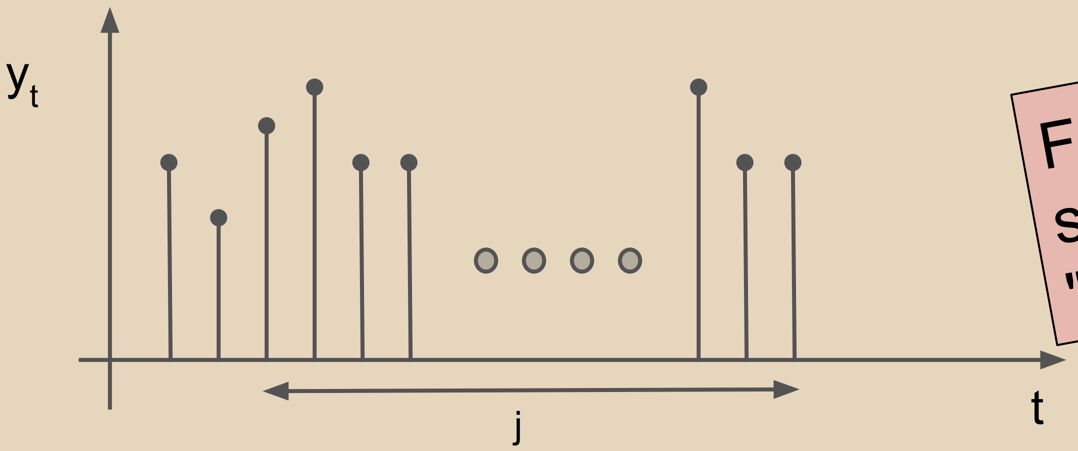
Beginning to see the pattern here?

Building our intuition



$$y_t = m^j . y_{t-j} + E_t + mE_{t-1} + m^2E_{t-2} + m^3E_{t-3} + \ldots + m^{j-1}E_{t-j+1})$$

Building our intuition



For a past value of y_t, say y_{t-j}, j is called the "lag".

$$y_t = m^j.y_{t-j} + E_t + mE_{t-1} + m^2E_{t-2} + m^3E_{t-3} + \ldots + m^{j-1}E_{t-j+1})$$

Setting t-j = 0...

$$y_t = m^t.y_0 + E_t + mE_{t-1} + m^2E_{t-2} + m^3E_{t-3} + \ldots + m^{t-1}E_1)$$

Building our intuition

$$y_t = m^t.y_0 + E_t + mE_{t-1} + m^2E_{t-2} + m^3E_{t-3} + \ldots + m^{t-1}E_1)$$

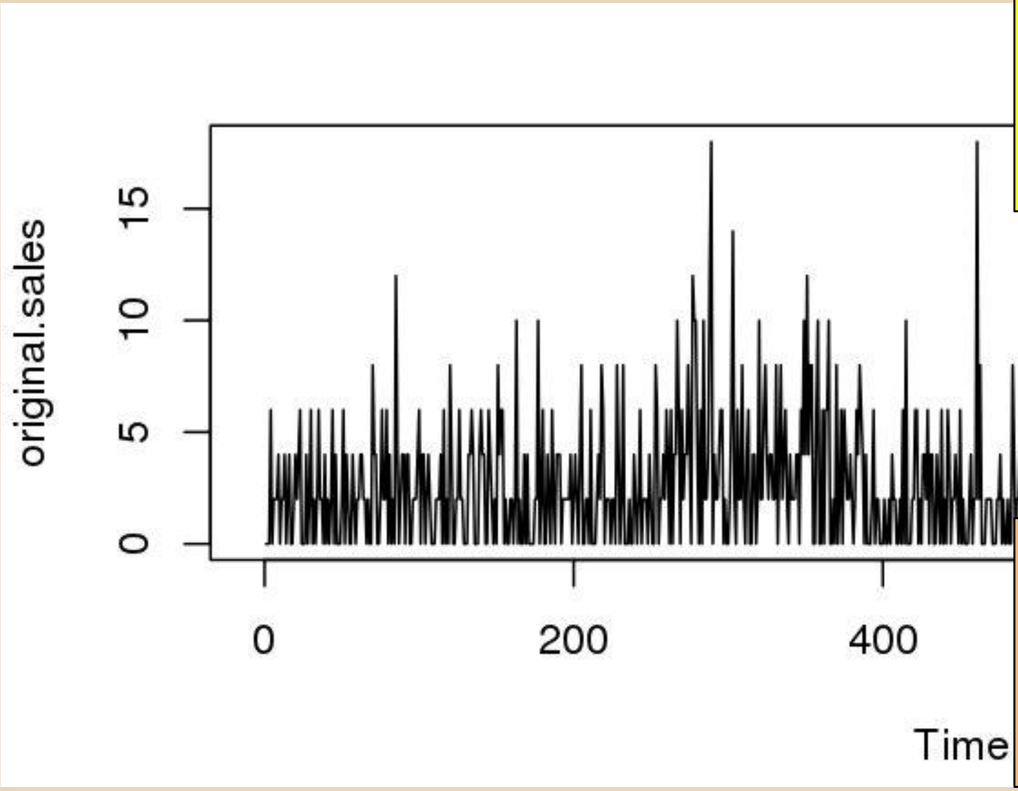
Errr, why is this important again?

This is how the current value of y (y_t) is related to the first value of y (y_0)

The equation $y_t = m.y_{t-1} + E_t$ is a specific example of an Autoregressive Model of a time series. In math, this is called a Difference Equation.

Exploring Time Series : Decomposition

When you are presented with a time series...

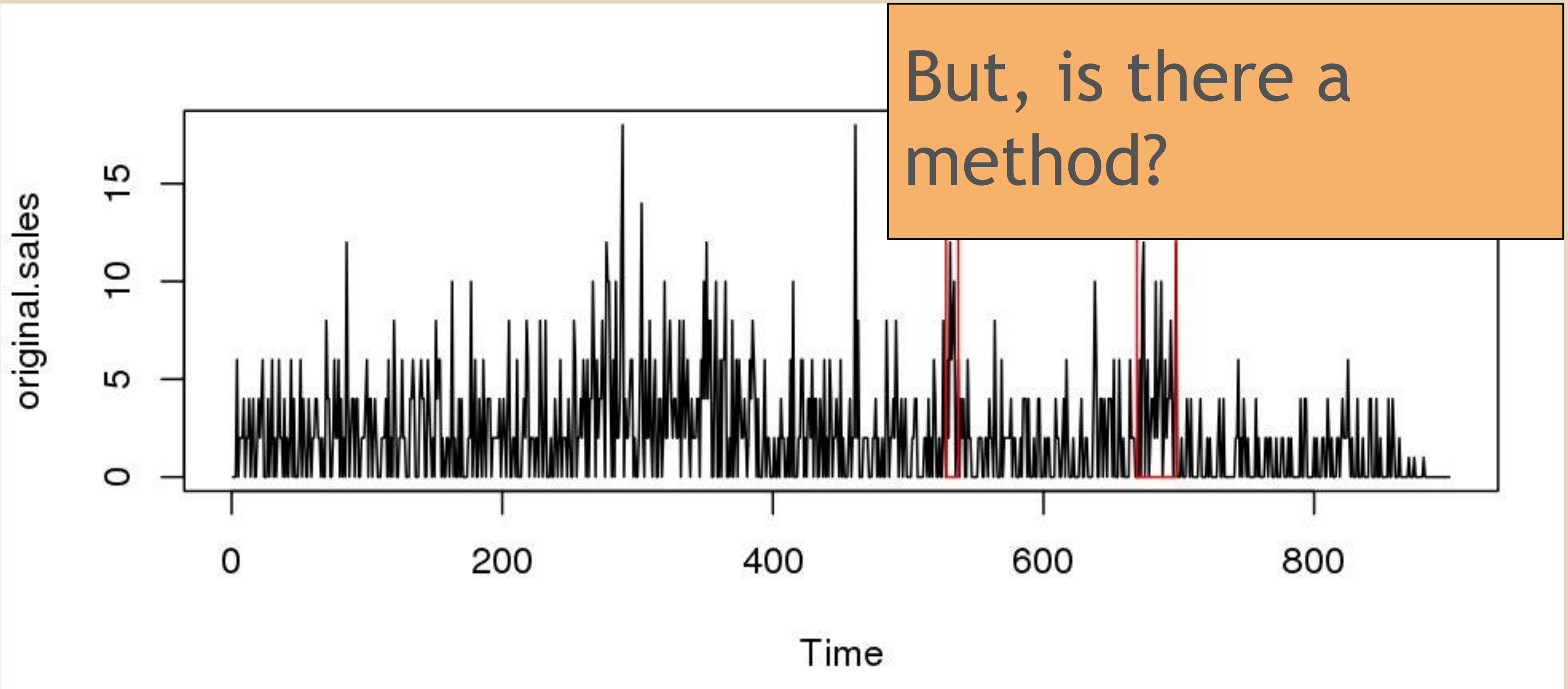


Beginning to see the pattern here?

No? Yeah, I'd agree. It's a *mess*!

Decomposition of Time Series

Time series data is amenable to exploration, just like any ordinary data set.



But, is there a method?

Decomposition of Time Series

Time series data is amenable to exploration, just like any ordinary data set.

Think of time series data as being composed of 3 components.

- Trend
- Seasonality
- Noise

Decomposition of Time Series

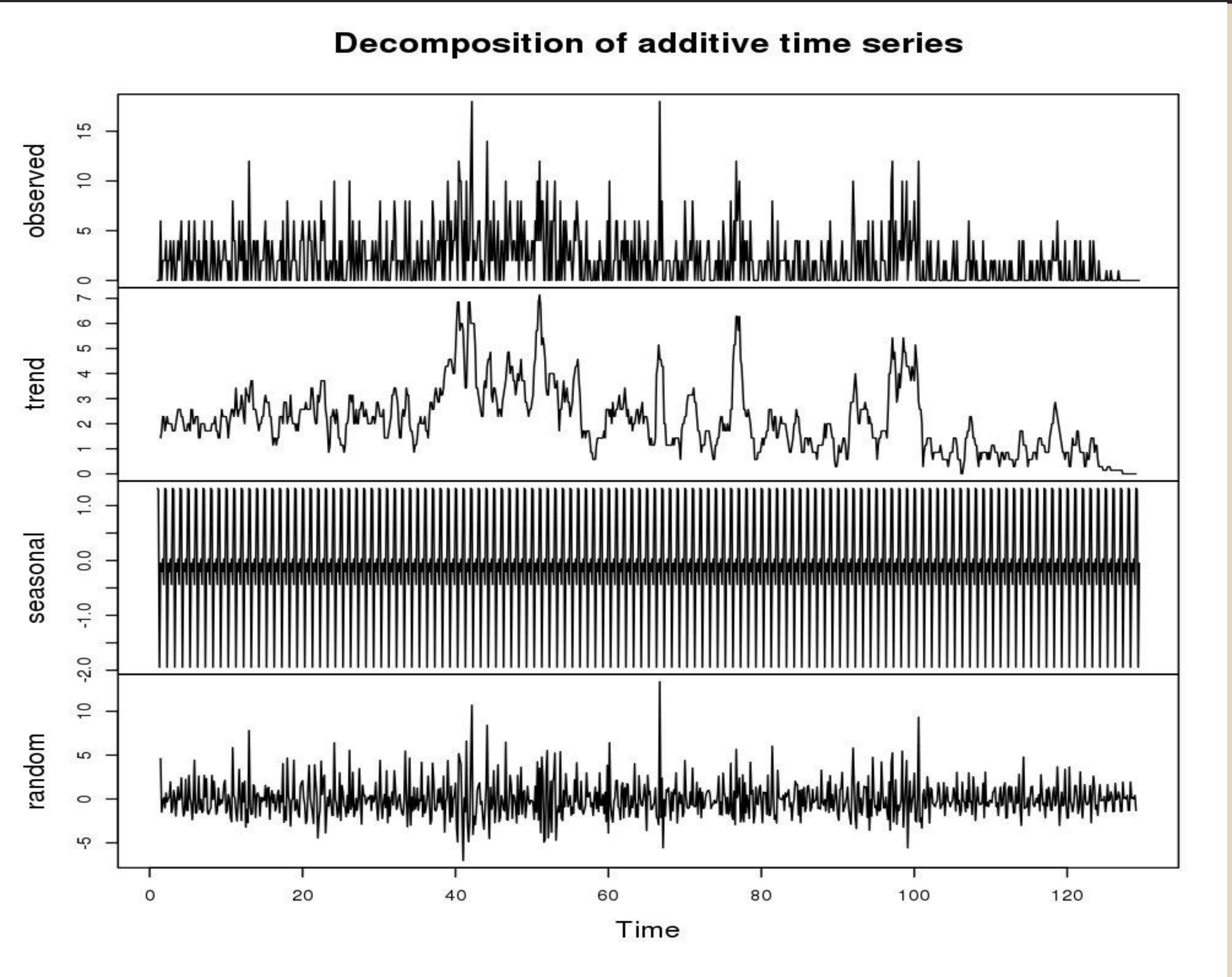
- Trend
- Seasonality
- Noise

What's the big picture like?

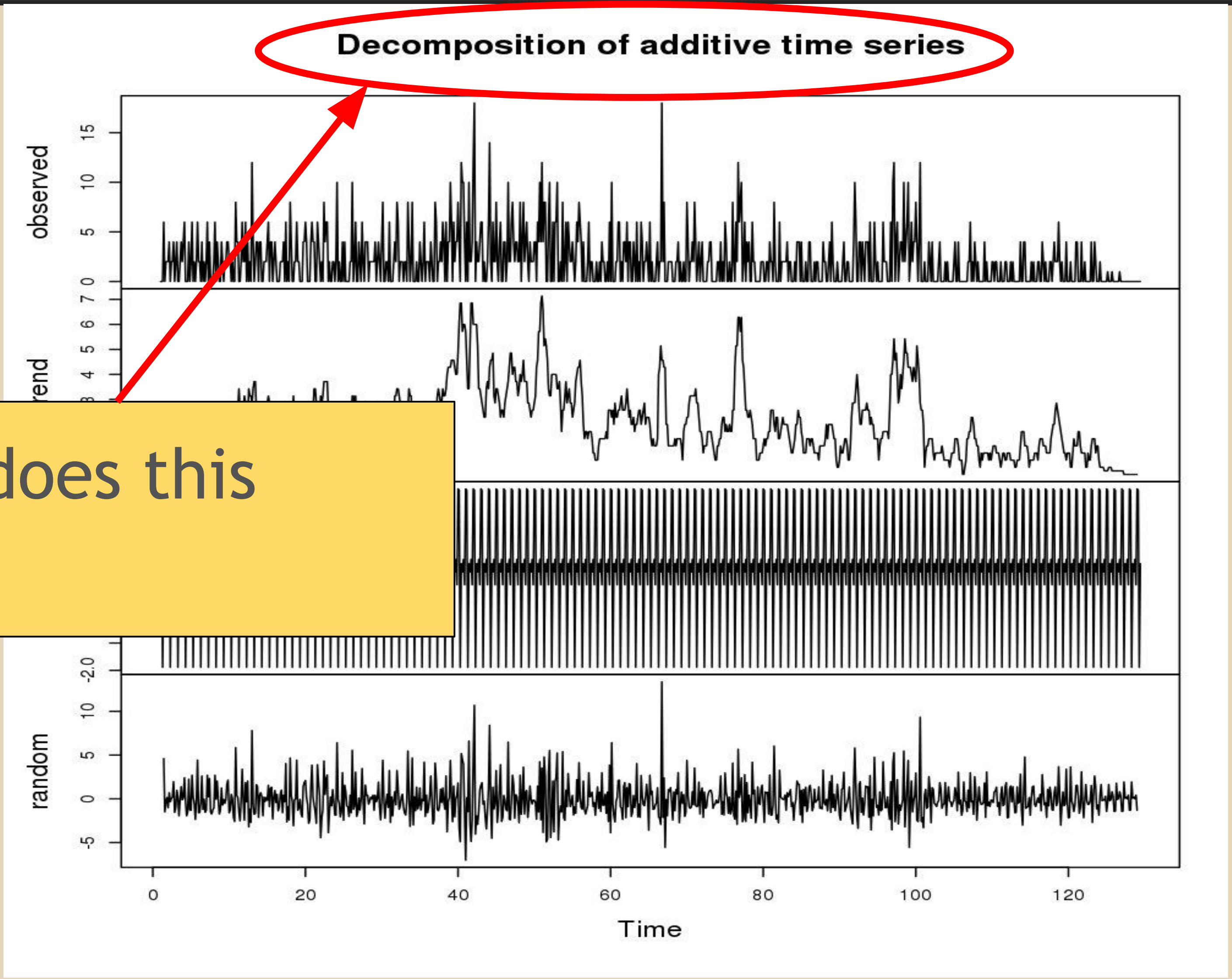
Are there any repetitive patterns?

What about randomness?

Decomposition of Time Series



Decomposition of Time Series



What does this mean?

Decomposition of Time Series

There are mainly two types of decompositions, apart from variations and hybrids.

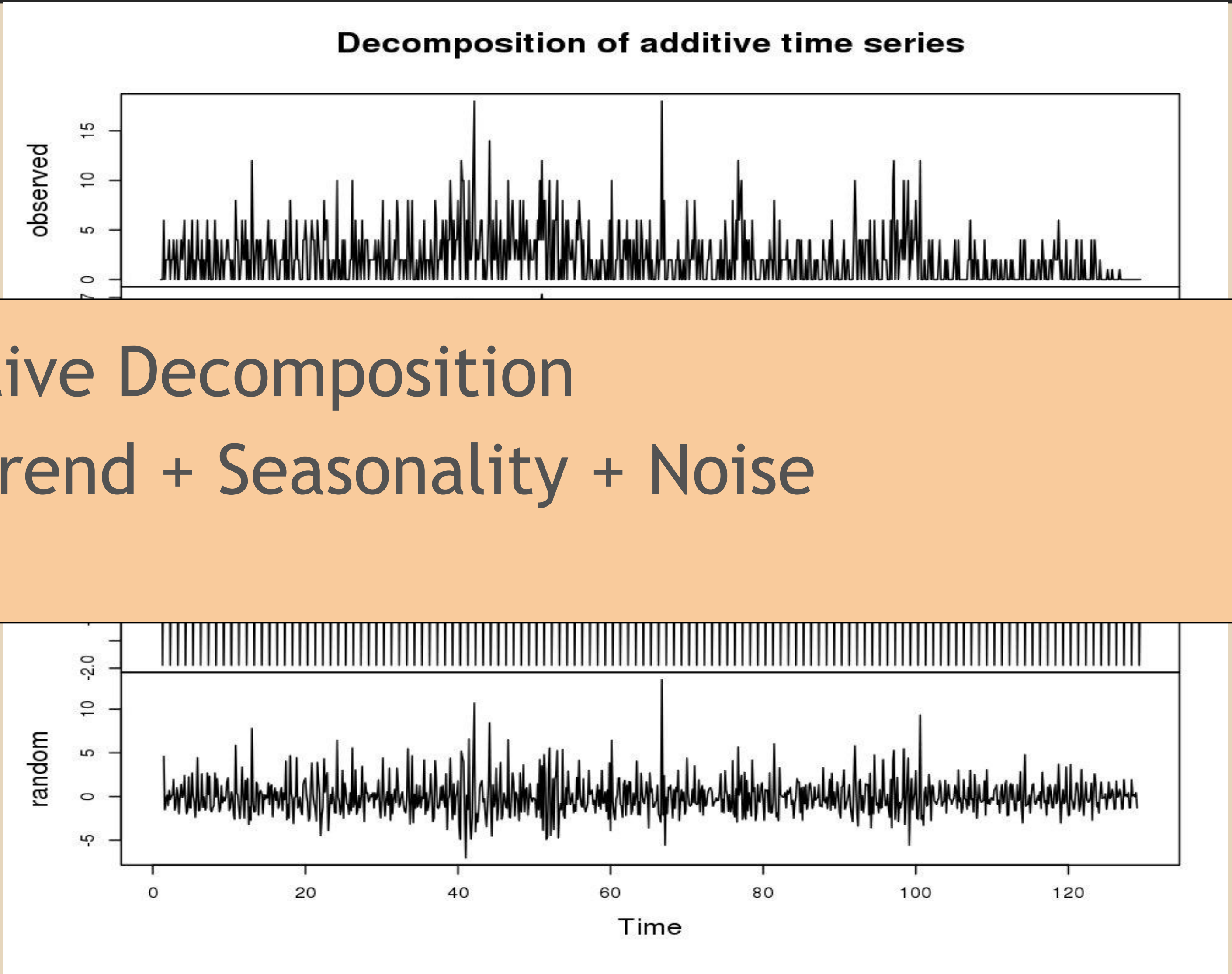
Additive Decomposition

Y = Trend + Seasonality + Noise

Multiplicative Decomposition

Y = Trend x Seasonality x Noise

Decomposition of Time Series



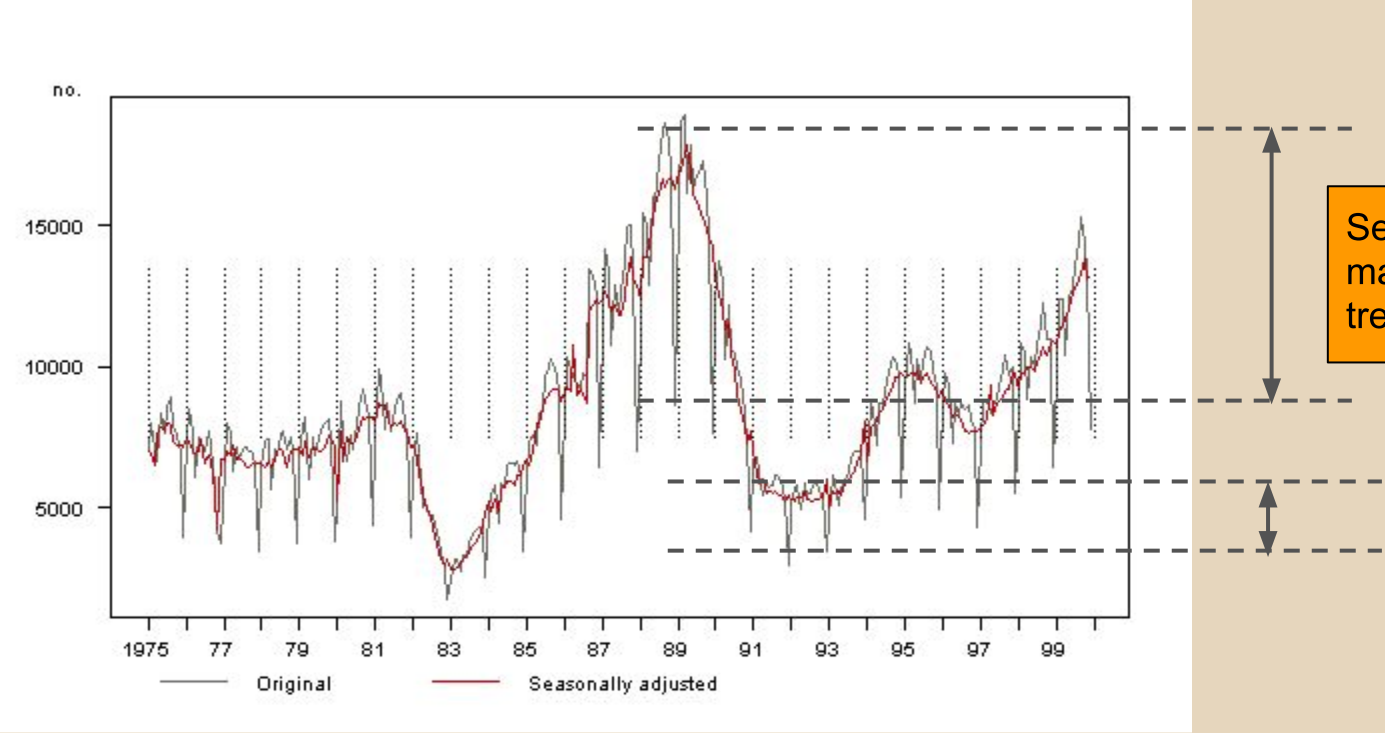
Additive Decomposition

Y = Trend + Seasonality + Noise

Decomposition of Time Series

Multiplicative Decomposition

Y = Trend x Seasonality x Noise



Seasonal fluctuations magnify with increasing trend

Why decompose time series data?

To get a sense of data which may appear chaotic at first sight.

Information gleaned at this point may be used for more formal modeling

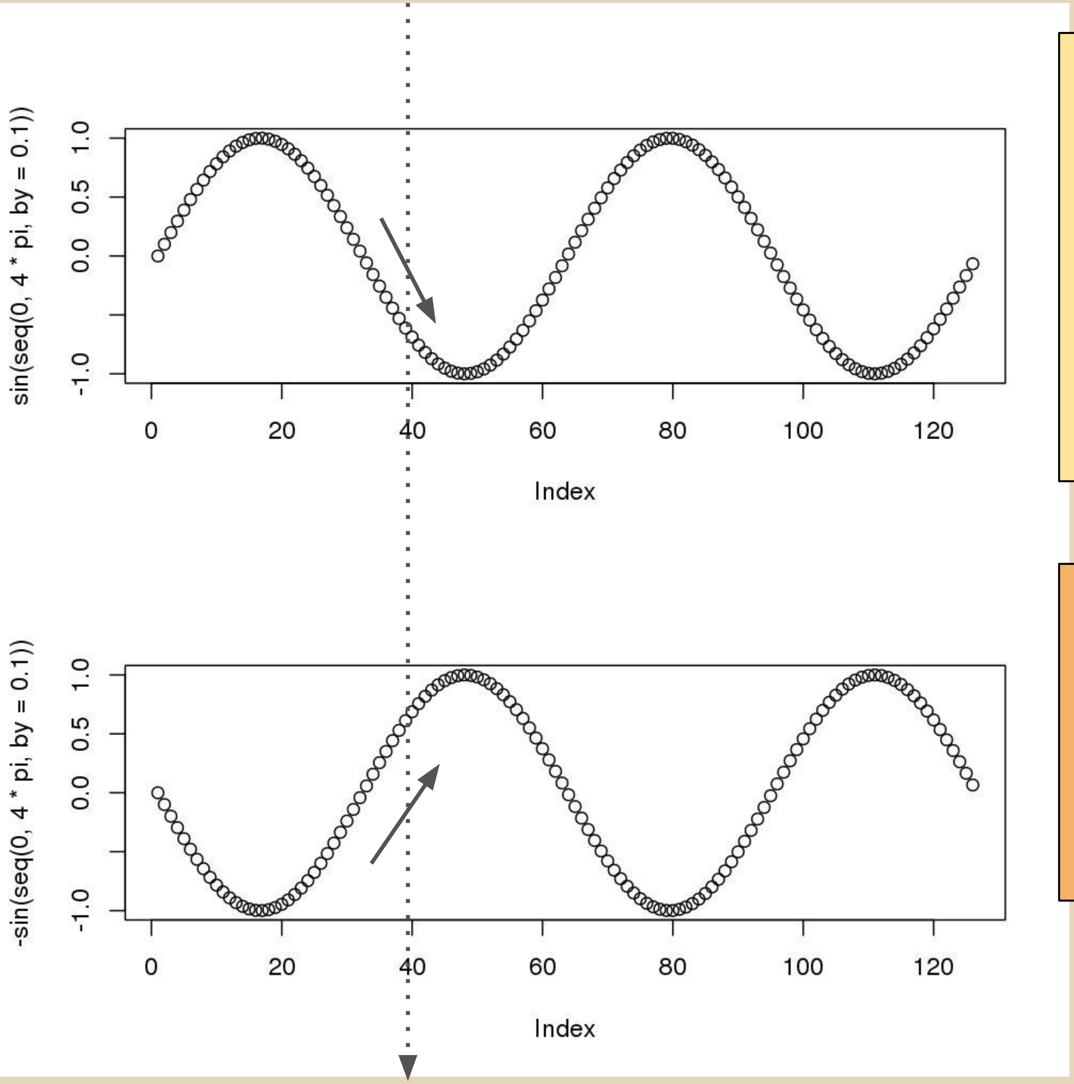
(R can do this automatically for you, btw :-)

Exploring Time Series: The Autocorrelation Function (ACF)

The Autocorrelation Function of a time series reveals important patterns which form the basis of an important class of models.

Remember correlation?
Yes? No?

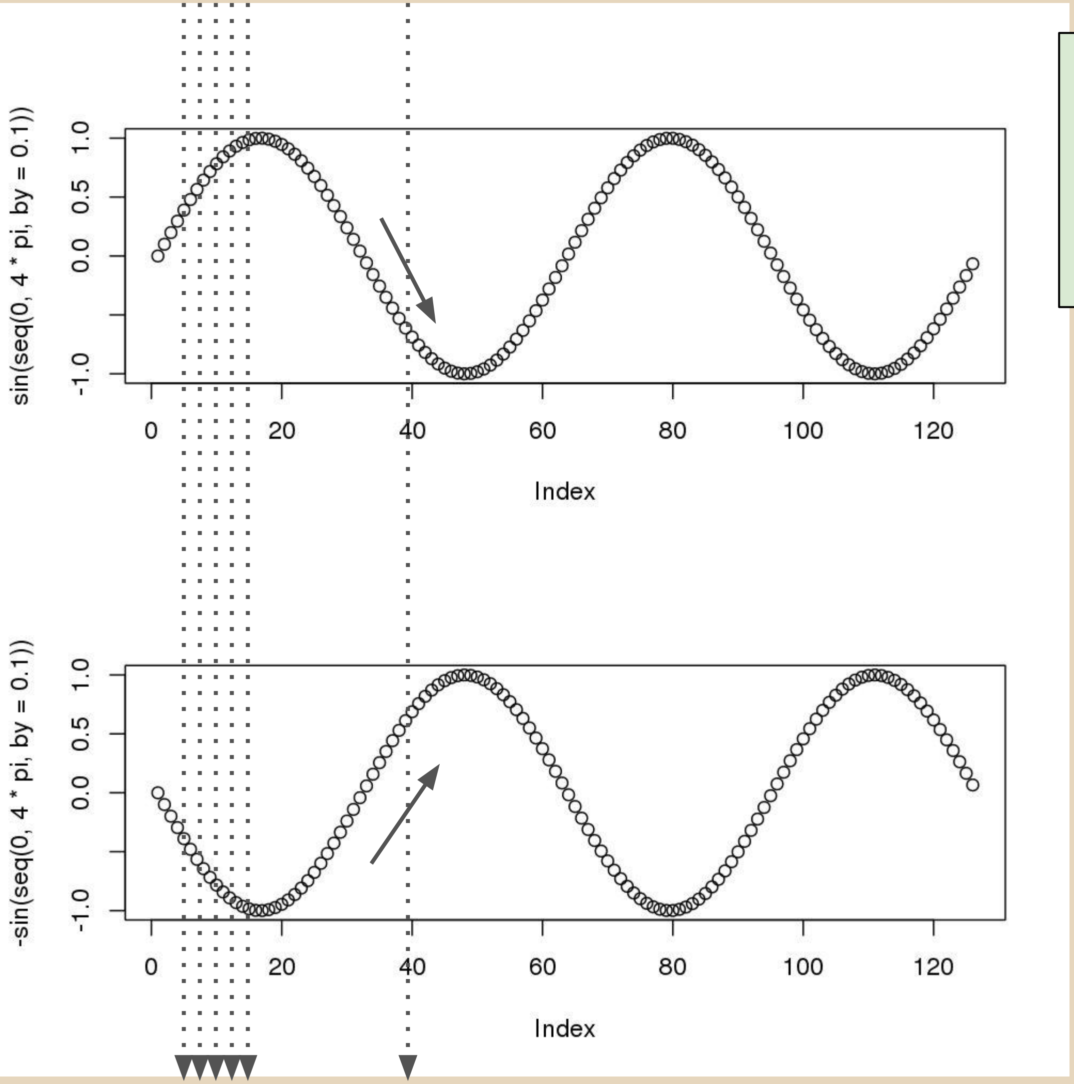
Exploring Time Series: The Autocorrelation Function (ACF)



When the sine function is decreasing, the inverse sine function is increasing (and vice versa).

Intuitively, we say that the sine and inverse sine functions are negatively correlated.

Exploring Time Series: The Autocorrelation Function (ACF)

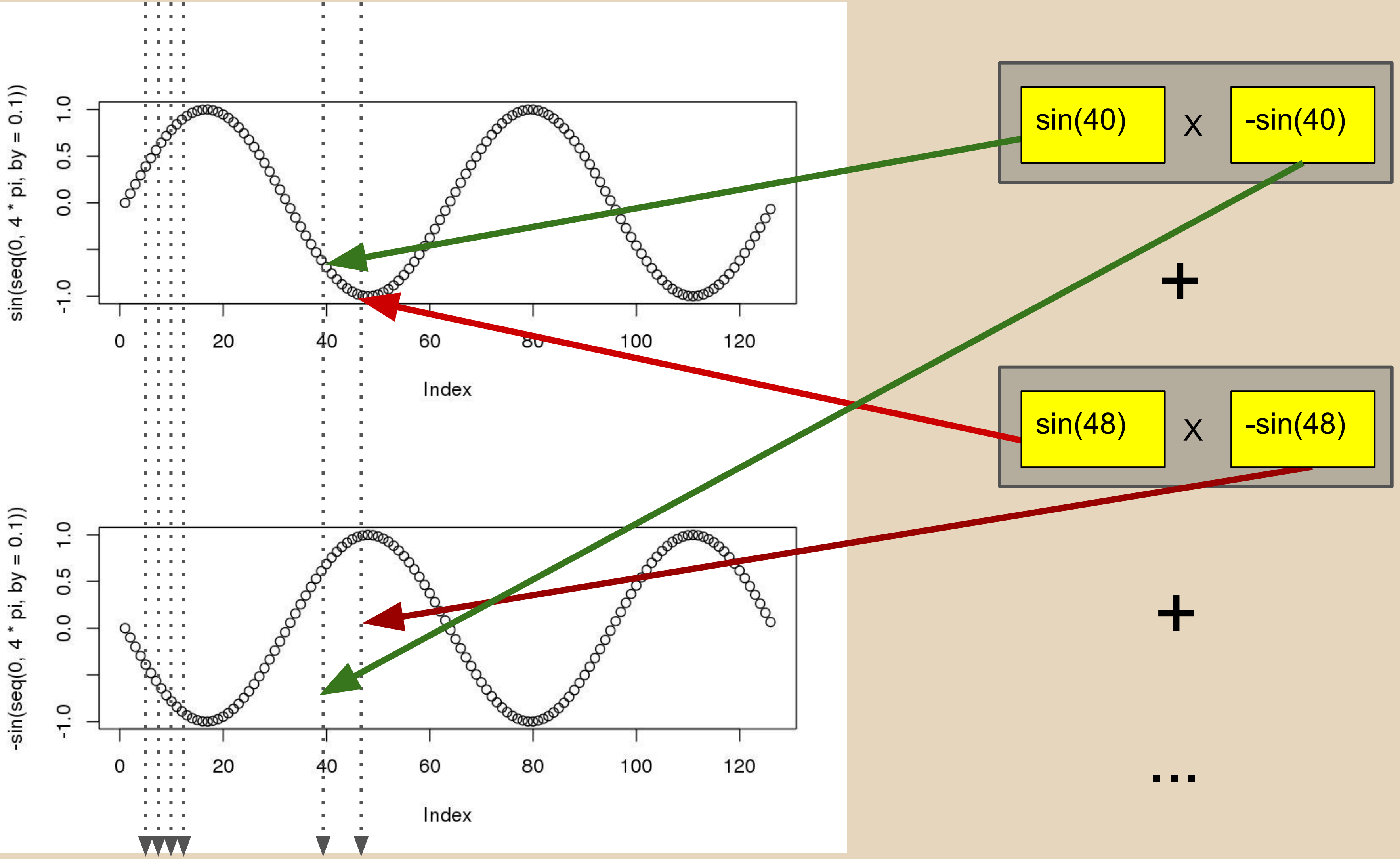


Let's take two functions, both of which have zero means.

Like our friends here, the sine and the negative sine.

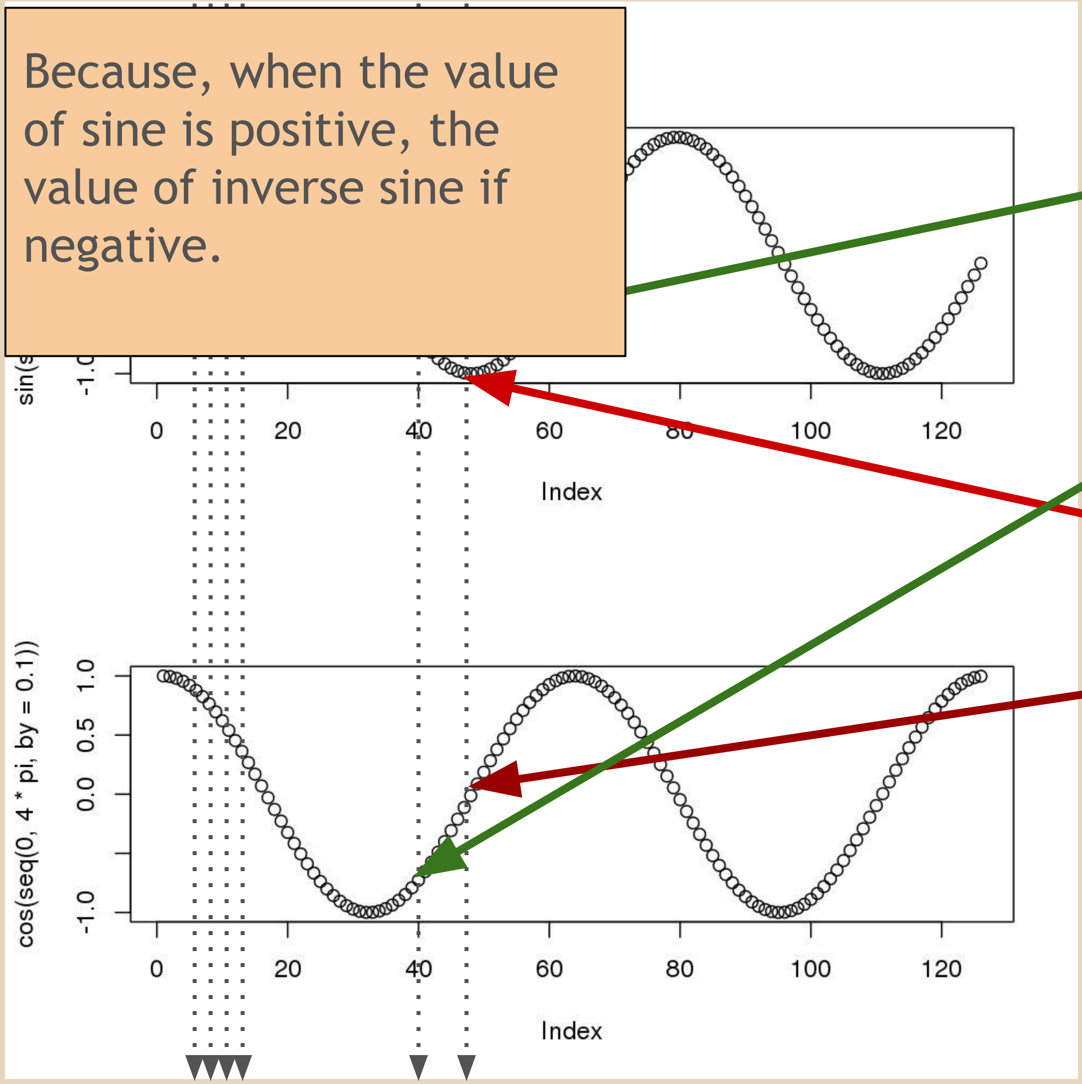
- Run through a sequence of values
- For each index x, multiply the values of the functions.
- Add up all these values

Exploring Time Series: The Autocorrelation Function (ACF)



Exploring Time Series: The Autocorrelation Function (ACF)

Because, when the value of sine is positive, the value of inverse sine if negative.



A Negative Value

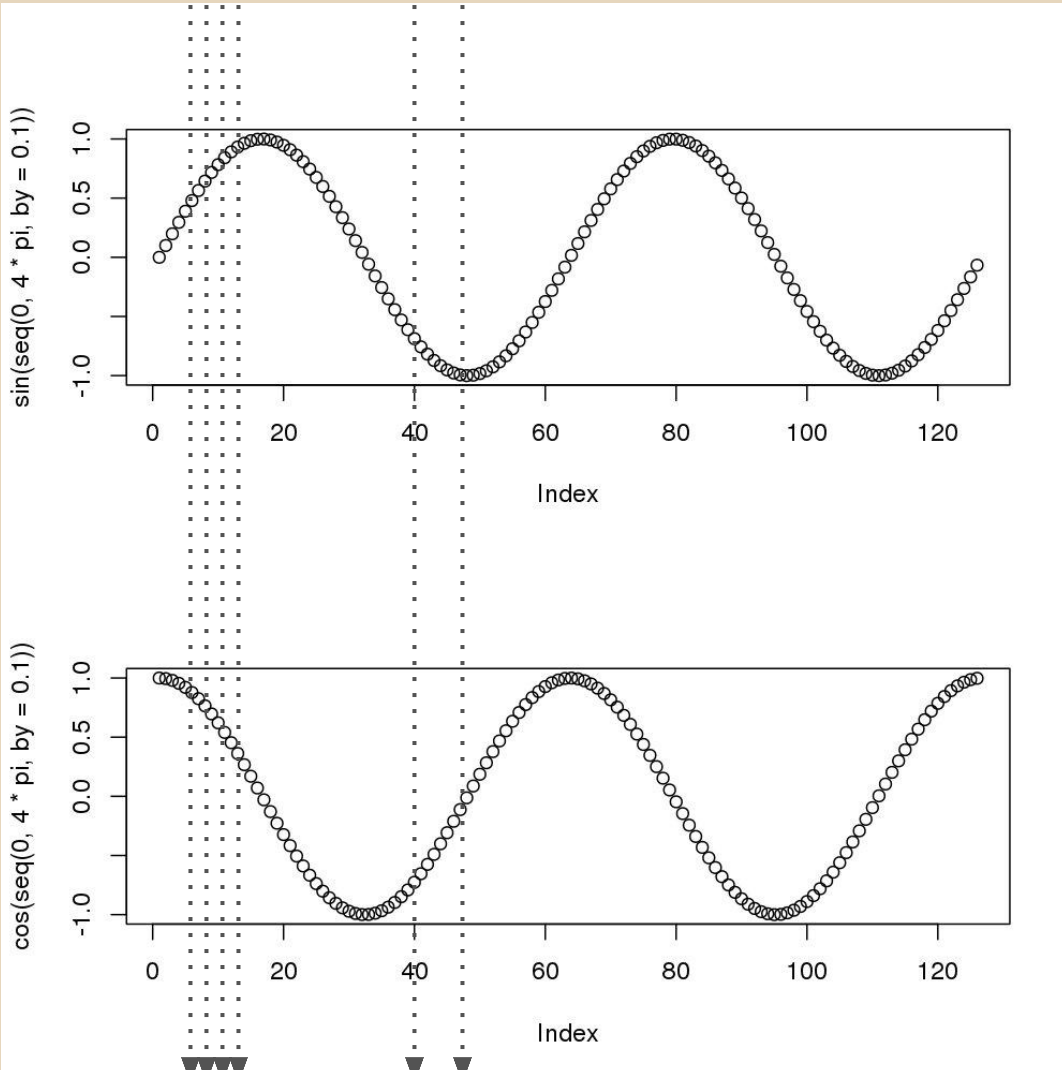
+

A Negative Value

+

...

Exploring Time Series: The Autocorrelation Function (ACF)



The result is a negative number.

By similar reasoning, the correlation between two sine functions is a positive number.

Convince yourself that sine and cosine are also positively correlated.

Exploring Time Series: The Autocorrelation Function (ACF)

Mathematically, the correlation coefficient between two variables is given as:

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

Subtract means to ensure that functions have zero means during calculations

Normalising constant to keep coefficient in [-1,1]

Exploring Time Series: The Autocorrelation Function (ACF)

You know where this is leading up to?

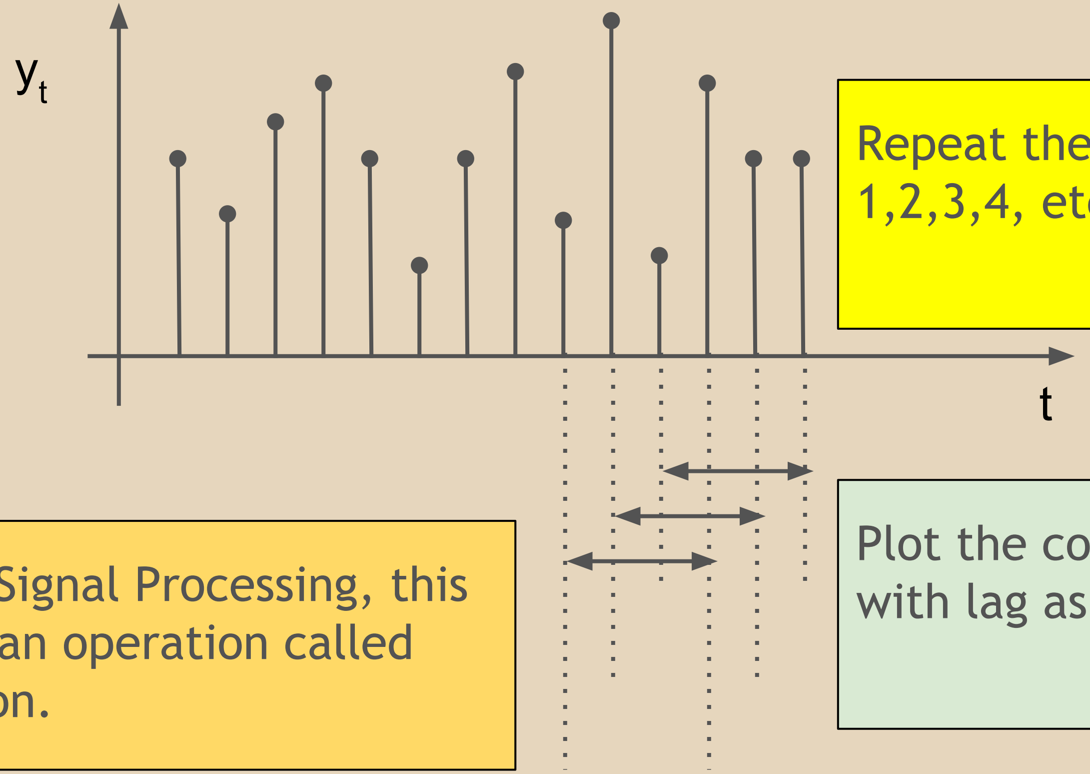
Same idea as the one behind a recurrence relation. Find the correlation coefficient between y_t and y_{t-j}. Hence the term "autocorrelation" (correlation with self).

- Assume you have 30 values of y. Pick a value of lag, say 3.
- Start with y_{30}.
- Multiply y_{30} with y_{27}.
- Multiply y_{29} with y_{26}.
- Multiply y_{28} with y_{25}.
- ... and so on.

Sum these. This is the correlation coefficient for lag 3.

Exploring Time Series: The Autocorrelation Function (ACF)

Graphical example of correlation coefficient calculation for lag 3

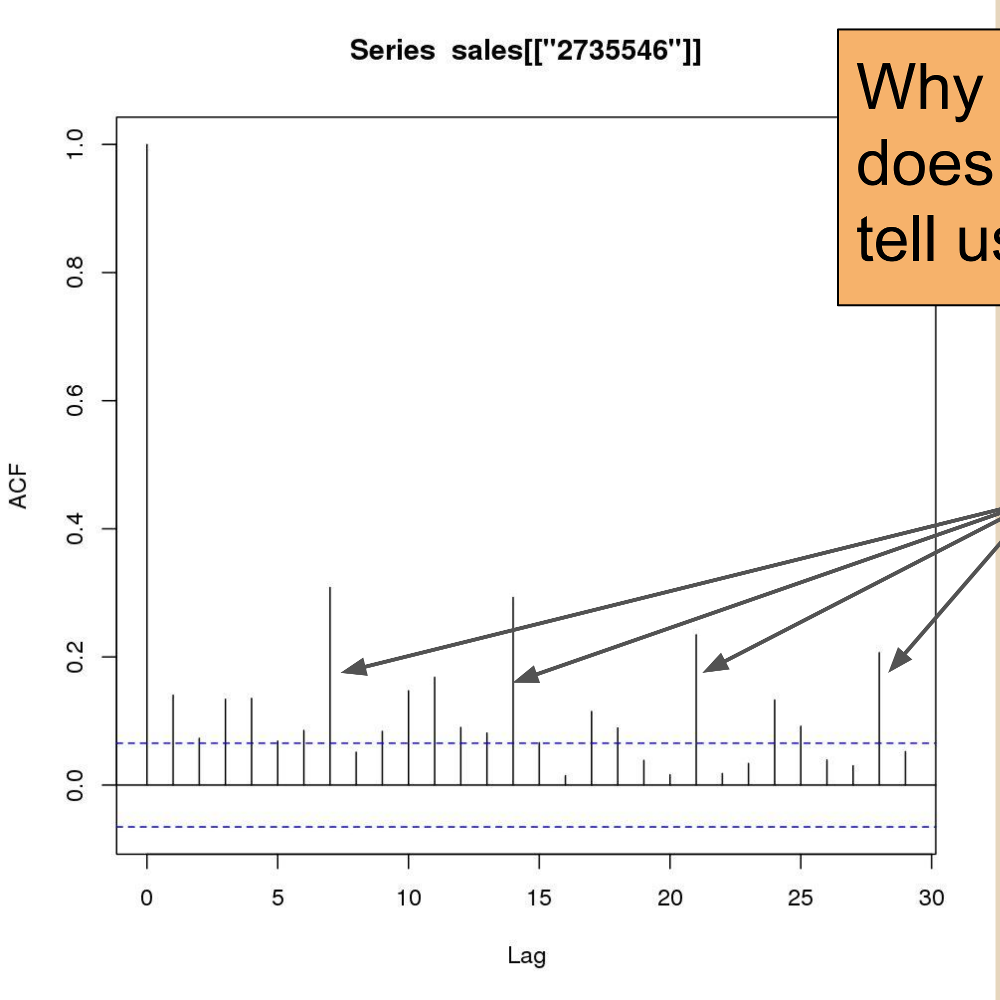


Repeat the calculation for lags 1,2,3,4, etc...

Plot the correlation coefficients with lag as the x-axis.

In Digital Signal Processing, this is akin to an operation called convolution.

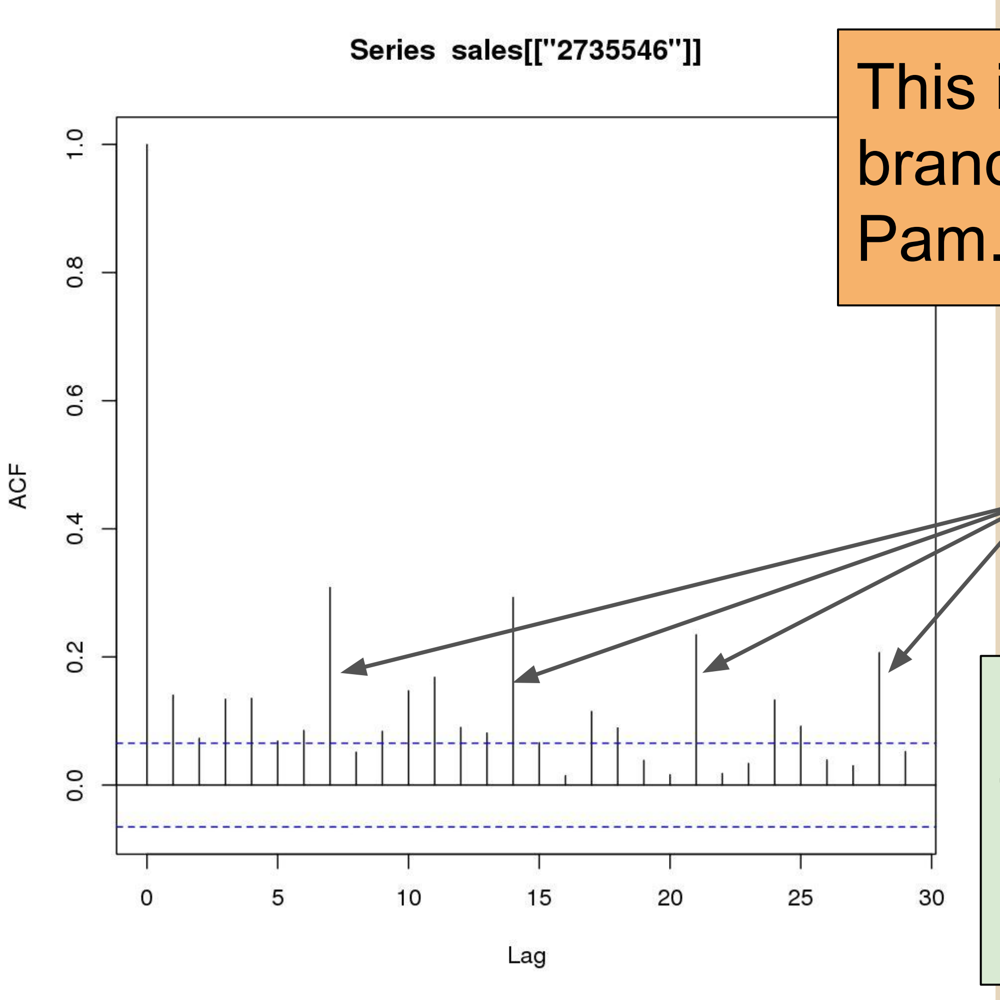
Exploring Time Series: The Autocorrelation Function (ACF)



Why go to all this trouble? What does the ACF of a time series tell us?

See these spikes at lags 7, 14, 21, 28...?

Exploring Time Series: The Autocorrelation Function (ACF)



This is sales data of a particular brand of coffee from Gruppo Pam.

See these spikes at lags 7, 14, 21, 28...?

The spikes indicate that there is a correlation in sales every 7 (and its multiple) days.

Exploring Time Series: The Autocorrelation Function (ACF)

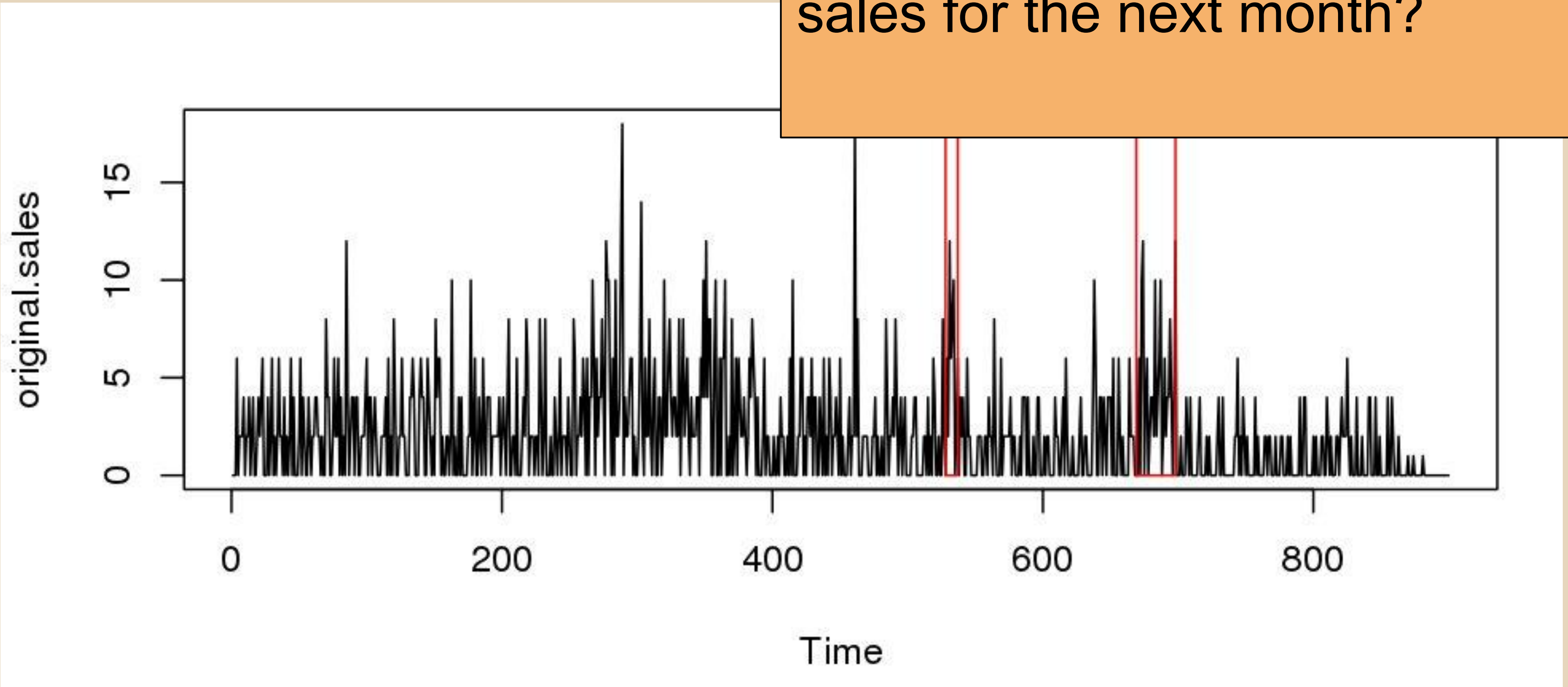
Autocorrelation Function plots also provide direct evidence of the parameters of a AR/MA, ARMA models.

Yeah, those come in a bit :-) Hold on.

Smoothing and Forecasting

Remember the spiky, chaotic time series data we saw?

What if we want to forecast sales for the next month?



Smoothing and Forecasting

Let's start simple. Our prediction for the next time point will simply be the average of all past values.

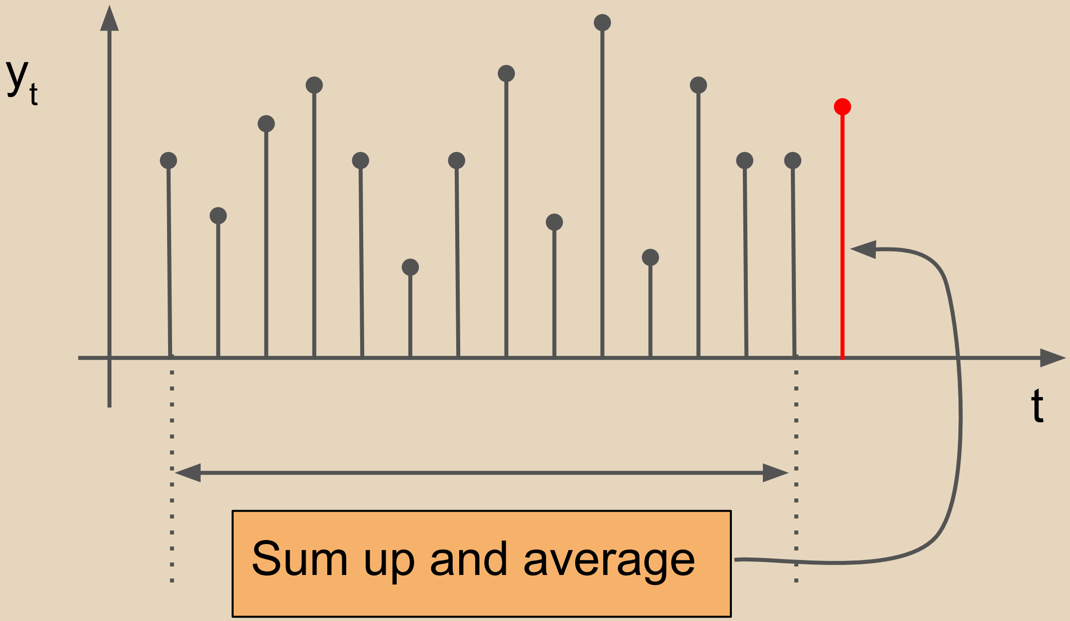
$$AM = \frac{1}{n}\sum_{i=1}^{n} a_i = \frac{a_1 + a_2 + \cdots + a_n}{n}.$$

Yuck! This is so boring. Real life is seldom average. (I don't know if that's a joke)

This does not take into account things like seasonality, trend, etc.

Smoothing and Forecasting

This is really simple. Not very useful, except as a starting point.



Smoothing and Forecasting

Let's see if we can do a bit better.

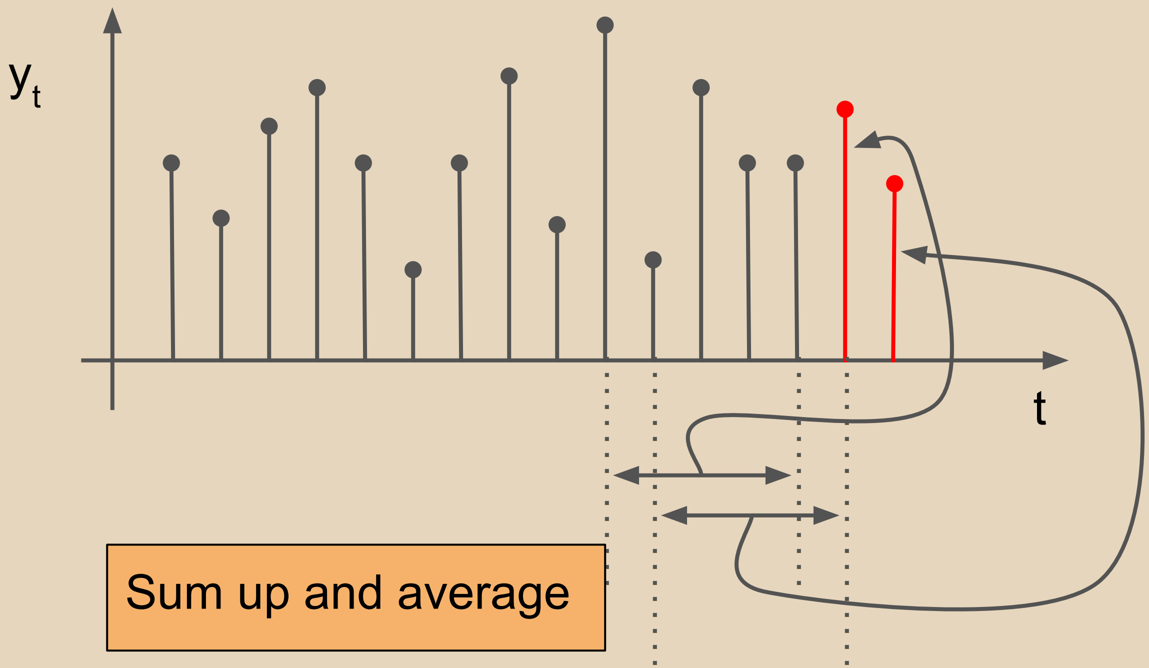
We don't have to take into account every past value. Just take the most recent n values.

$$SMA = \frac{p_M + p_{M-1} + \cdots + p_{M-(n-1)}}{n}$$

Key Idea: Values older than n recent values do not significantly contribute to the forecast.

Smoothing and Forecasting

This is also called Rolling Average.
Still a bit naive.



Sum up and average

Forecasting: Exponential Methods

Key Idea: Older values of a time series do not contribute as much to the forecast as do the more recent ones.

The exponential model also corrects for forecast errors:

$$F_{t+1} = F_t + \underbrace{B}_{\text{Parameter}} . \underbrace{(Y_t - F_t)}_{\text{Forecast Error}}$$

Forecasting: Exponential Methods

Why do we call this an exponential method?

Well, expand it!

$F_{t+1} = F_t + B.(Y_t - F_t)$

$F_{t+1} = B.Y_t + (1 - B).F_t$

$F_{t+1} = B.Y_t + (1 - B).(B.Y_{t-1} + (1 - B).F_{t-1})$

$F_{t+1} = B.Y_t + B.(1 - B).Y_{t-1} + (1 - B)^2.F_{t-1}$

$F_{t+1} = B.Y_t + B.(1 - B).Y_{t-1} + (1 - B)^2.F_{t-1} + \dots + (1 - B)^t.F_1$

The contribution of earlier forecasts decreases exponentially.

This is the Single Exponential Smoothing Method

Forecasting: Holt's/Holt-Winters'

These are extensions to the Single Exponential Smoothing technique.

- Holt's technique uses an extra parameter to track the trend; this is Double Exponential Smoothing.
- Holt-Winters' technique also estimates seasonality in a time series. Use this when you have determined that there is seasonality present in the time series.

I'm not showing the formulae. Intuition will have to do for now.

Forecasting: When to use what?

- Explore your data.
  - Decompose it.
  - Plot ACFs and PACFs.
  - Use simple regression to draw inferences about trends.
- Select an appropriate forecasting model.
  - Don't use a more complicated model when a simple one would do.
  - Use part of the data set to verify the accuracy of the model. Several criteria for verifying this exist (e.g. Akaike Information Criterion).

Some Classical Formal Models

Autoregressive (AR) Models

We have already seen an Autoregressive process.

$y_t = m.y_{t-1} + E_t$ ← AR(1) process

$y_t = m_1.y_{t-1} + m_2.y_{t-2} + E_t$ ← AR(2) process

Key Idea: The current value is the weighted sum of n past values, plus an error term.

Moving Average (MA) Models

Use Moving Average models to incorporate shocks into a time series.

It is an extension of the Simple Exponential Smoothing method.

Represent a series as the weighted sum of errors.

Remember?

$F_{t+1} = F_t + B.(Y_t - F_t) = F_{t+1} = F_t + B.E_t$

Moving Average (MA) Models

Represent a series as the weighted sum of errors.

$$Y_t = b_0 + e_t + b_1.e_{t-1} + b_2.e_{t-2} + \ldots + b_q.e_{t-q}$$

$Y_t = b_0 + e_t + b_1.e_{t-1}$ ← MA(1) process

$Y_t = b_0 + e_t + b_1.e_{t-1} + b_2.e_{t-2}$ ← MA(2) process

ARMA Models

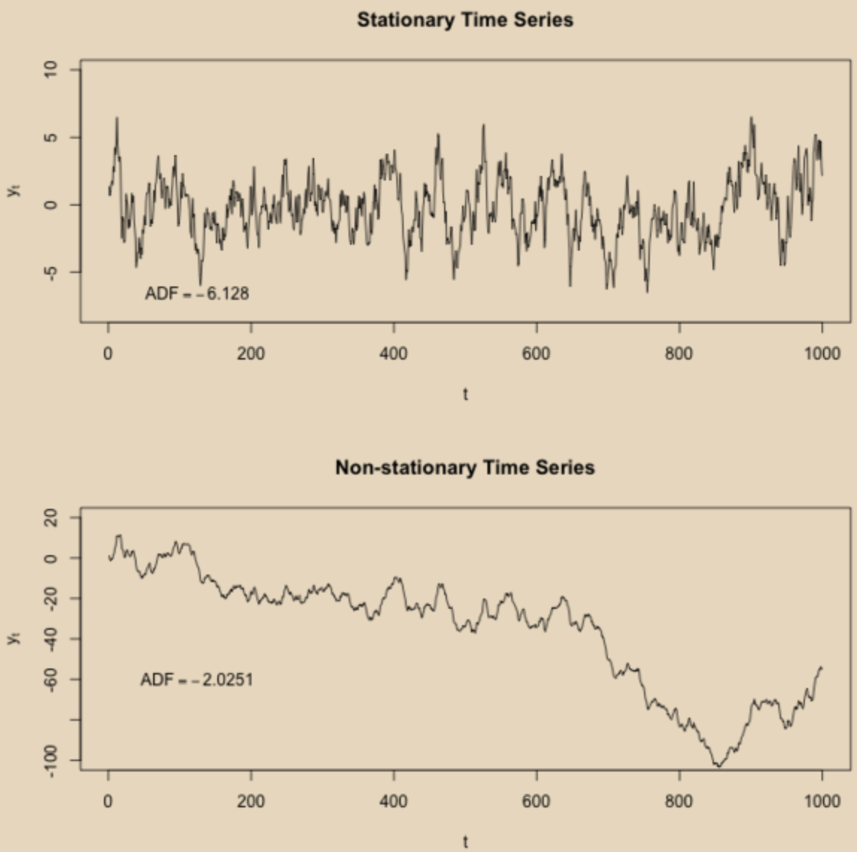
ARMA models are simply what you get if you add an AR model and an MA model.

ARMA = Autoregressive Moving Average

Very useful and general class of models, can be used to represent all sorts of time series.

Stationarity Requirements

AR/MA/ARMA models require that a series be stationary (or at least weakly stationary).

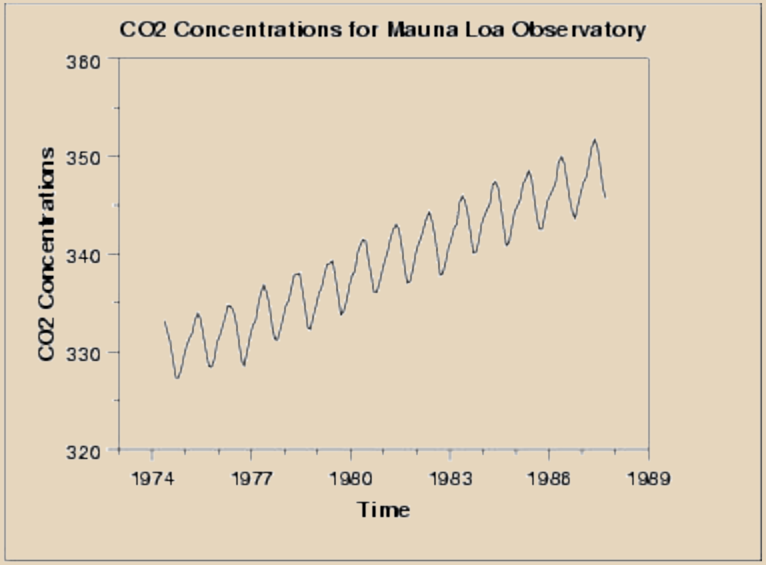


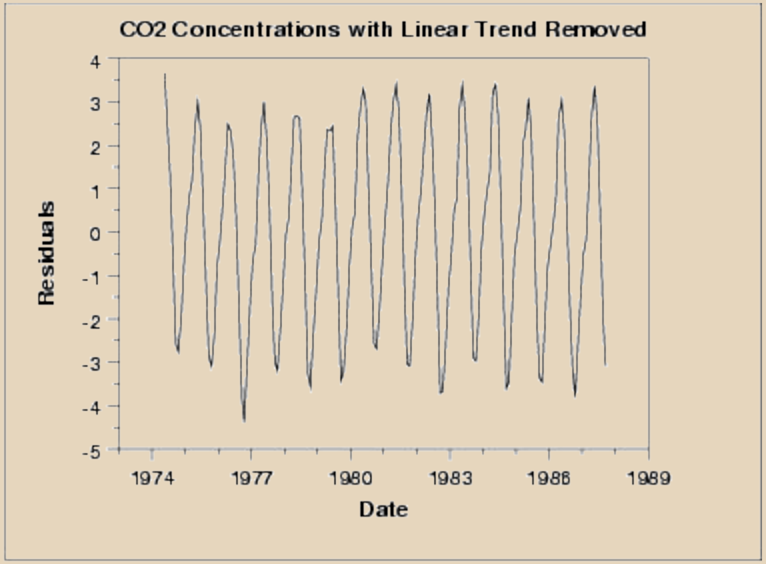
Weak stationarity:

- Fixed Mean
- Finite Variance
- Independent Covariance

Stationarity Transformations

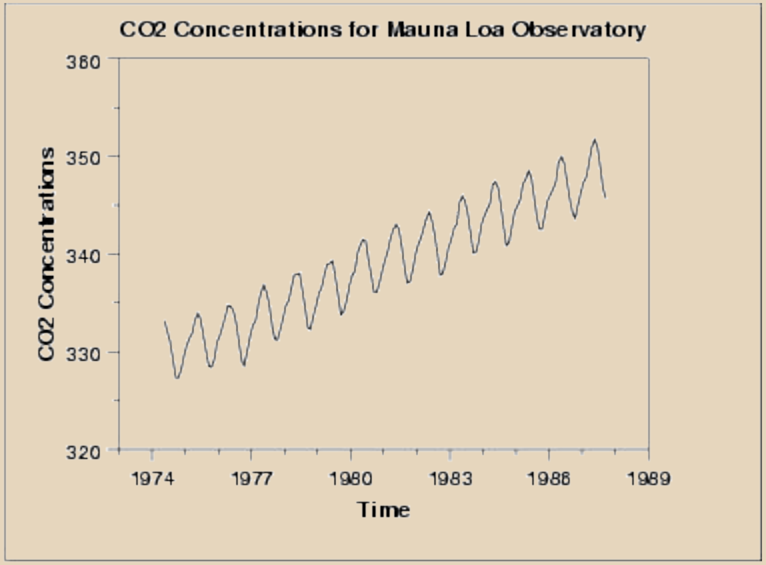
If mean is varying too much, you may use differencing to achieve stationarity.

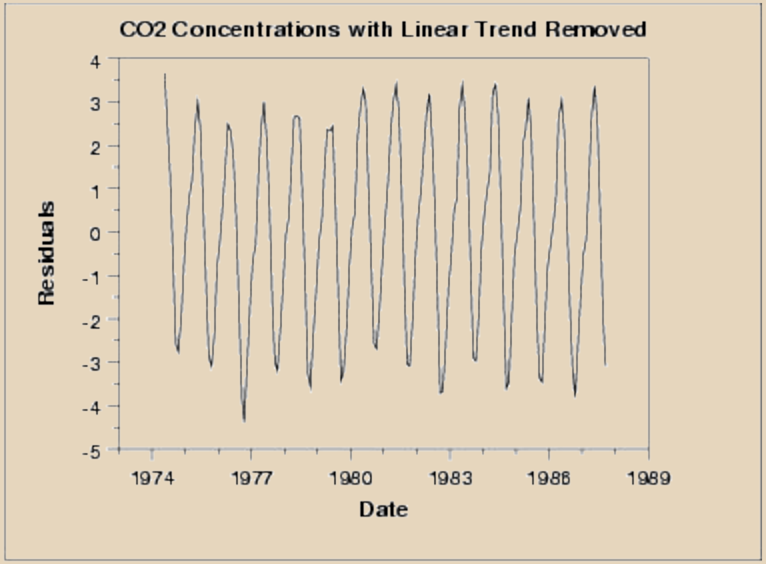




Stationarity Transformations

If mean is varying too much, you may use differencing to achieve stationarity.





For non-constant variance, taking log or square root may stabilise the time series.

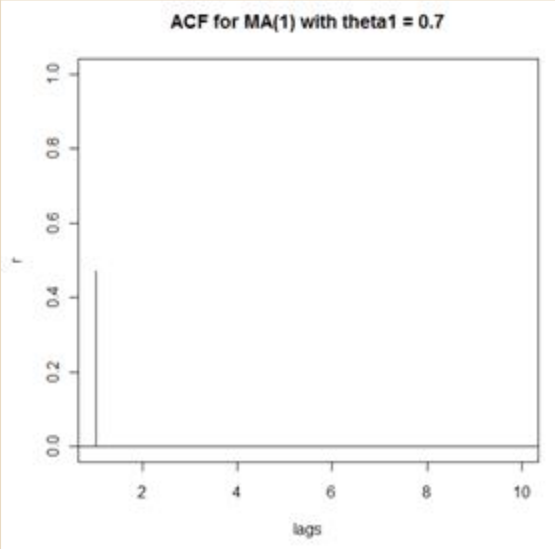
Determining the model order of an ARMA process

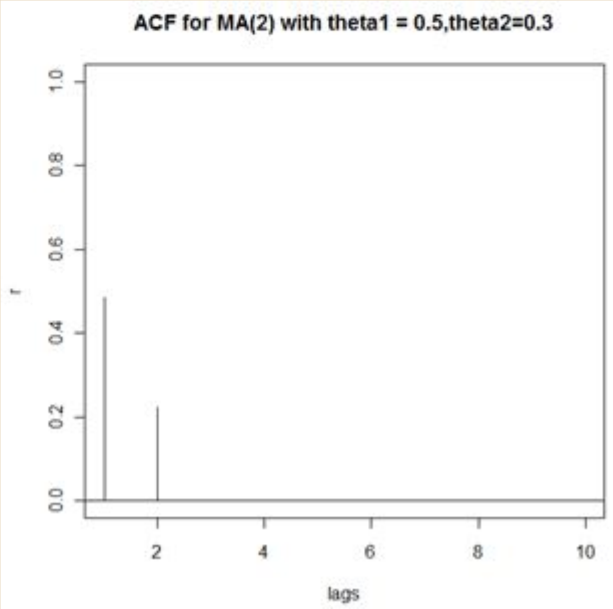
This is all well and good, but given a time series, how do we decide whether to use:

- AR process
- MA process
- ARMA process

Hint: Look at the ACF and the PACF plots :-)

Model order diagnostics for MA processes





Model order diagnostics for ARMA processes

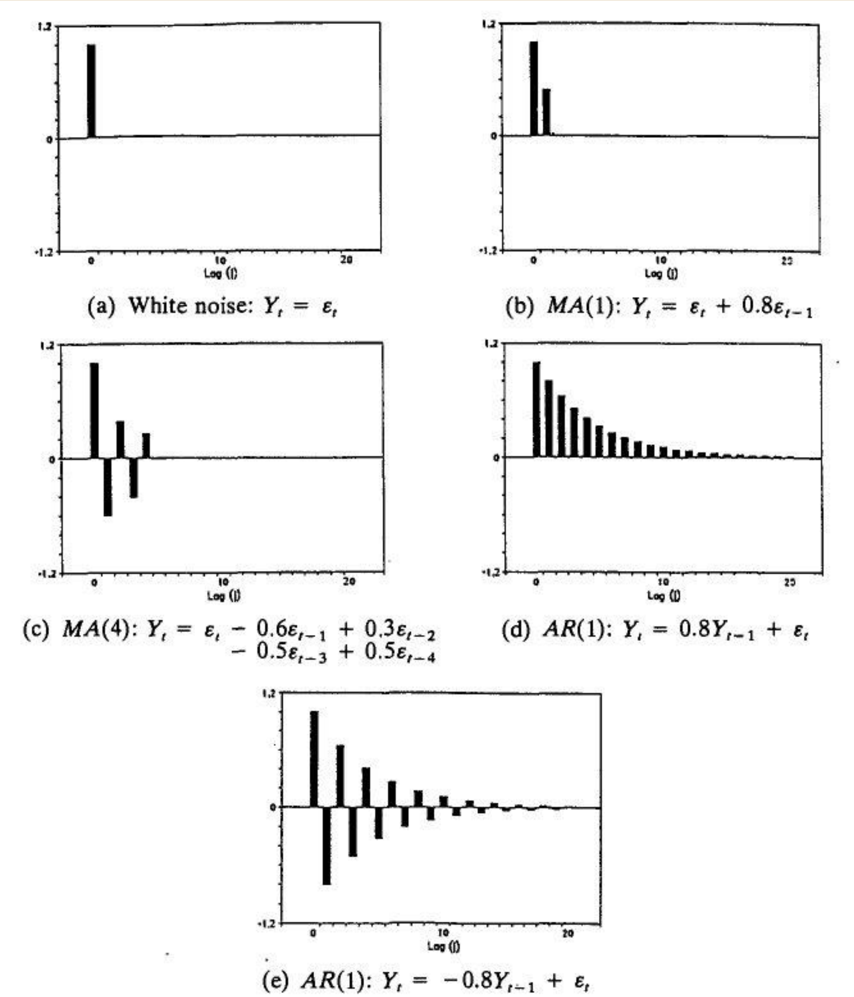


FIGURE 3.1 Autocorrelation functions for assorted *ARMA* processes.

Why do these plots arise? The theory behind them is very straightforward but out of scope for this session :-(

OK to treat these as diagnostic tools.

Intervention Analysis

What is Intervention?

Intervention happens whenever an event occurs which has a lasting/temporary impact on a time series.

Intervention Analysis

Bush's approval ratings before and after 9/11.

How do we incorporate such events in a time series model?

Intervention Analysis

How do we incorporate such events in a time series model?

Theoretically, not very difficult.
Practically, *sigh*...

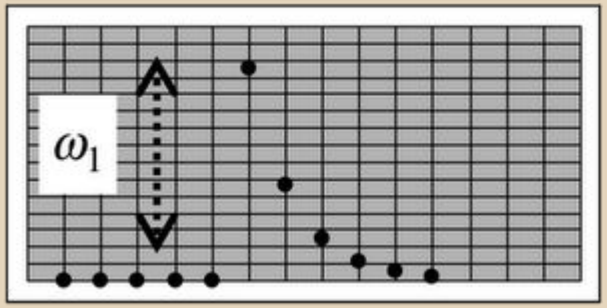
$$Y_t = a + w.X_t + N_t$$

Indicator variable (0,1)

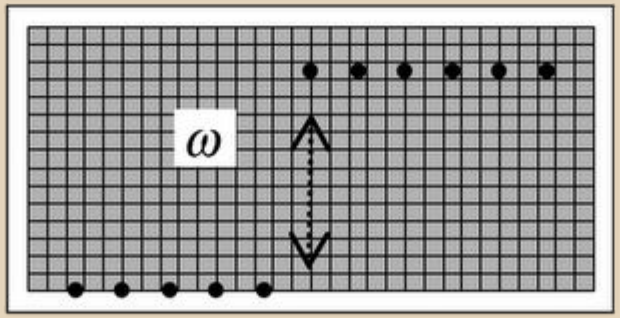
Intervention function

ARMA process

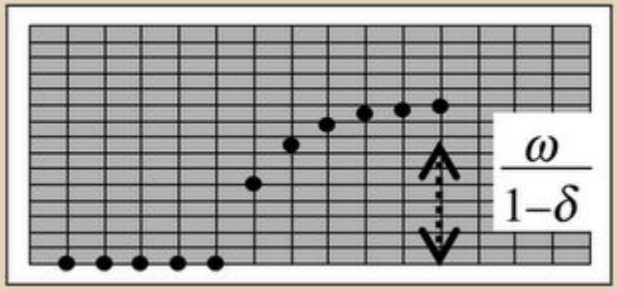
Intervention Analysis: Intervention Models



Pulse Function



Step Function



Gradual Permanent Effect Function

More advanced intervention modeling techniques include ARMA Regression Trees. Pieces of the time series are modeled using separate ARMA processes.

Things there isn't time for...

- State space models and Kalman Filters
- Parameter estimation algorithms
- Dynamic Bayesian Networks and Hidden Markov Models
- Econometric Modeling

Questions?

Thanks for Listening :-)