Data Analysis, Statistics, Machine Learning

Leland Wilkinson

Adjunct Professor
UIC Computer Science
Chief Scientist
H2O.ai

leland.wilkinson@gmail.com

Time Series

Time series statistics involve random processes over time

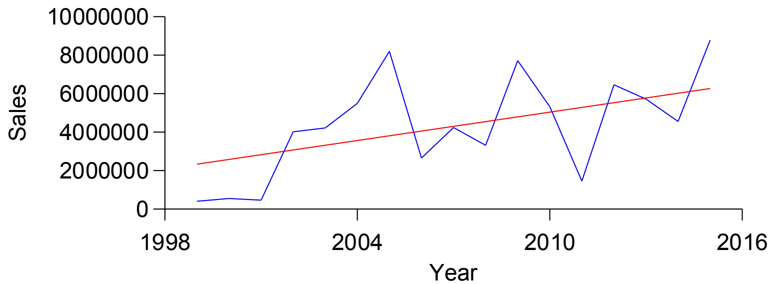
Spatial statistics involve random processes over space

Both involve similar mathematical models

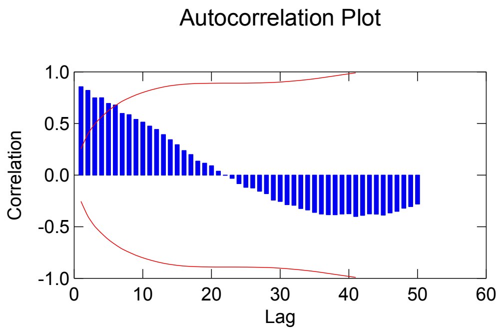
- When there is no temporal or spatial influence, these boil down to ordinary statistical methods

DO NOT USE OLS methods on temporal/spatial data

- These require stochastic models, not OLS "trend lines"
- measurements at each time/space point are not independent



Quarterly US Ecommerce Retail Sales, Seasonally Adjusted



Time and Space

Stochastic processes

Up to now, we've been dealing with i.i.d. random variables

Independent. Identically. Distributed.

We assumed there was no ordering of those random variables

Our models depended on random error plus systematic effects

Time series analytics deal with ordered random variables

We (usually) assume these variables are equally spaced across time

A variable at time t_i is predictable in part by another variable at another time

The simplest example of this type of behavior is called autoregressive (AR)

$$x_t = \phi x_{t-1} + \epsilon_t$$

In this model each observation at a given time is a function of the previous observation plus random error

$$E[\epsilon_t] = 0$$

$$E[\epsilon_t^2] = \sigma^2$$

$$E[\epsilon_s \epsilon_t] = 0 \text{ for all } s \neq t$$

Time and Space

Stochastic processes

Diagnosing a stochastic process

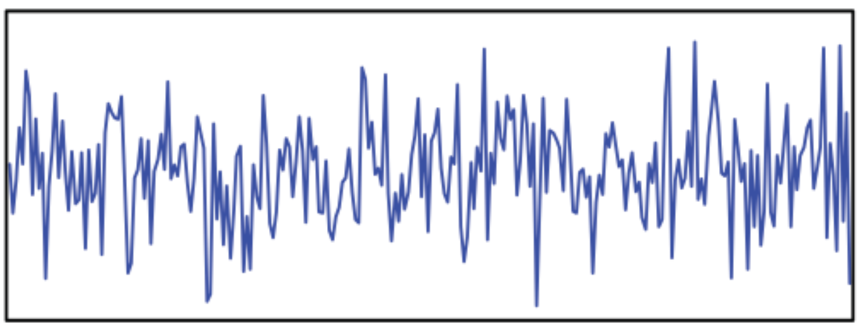
Correlate a series with itself shifted backward by one time period

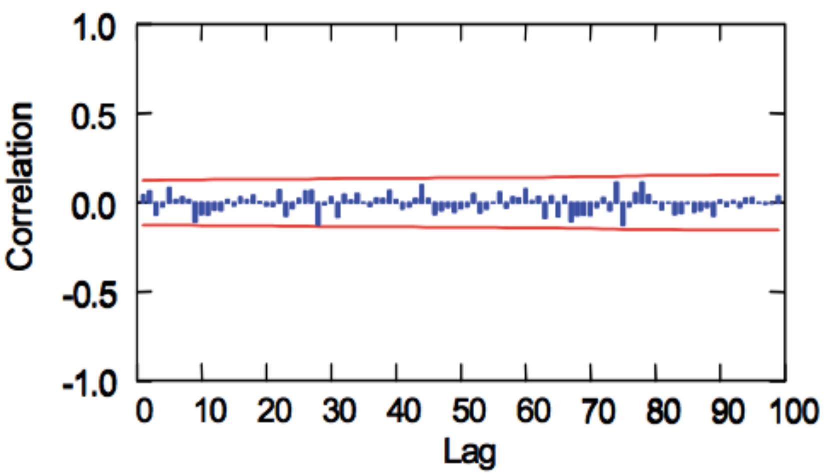
Correlate the shifted series with itself shifted backward by one time period

And so on…

Here's an Autocorrelation Function (ACF) Plot of white noise

$$x_t = \epsilon_t$$





Time and Space

Stochastic processes

Diagnosing an autoregressive process

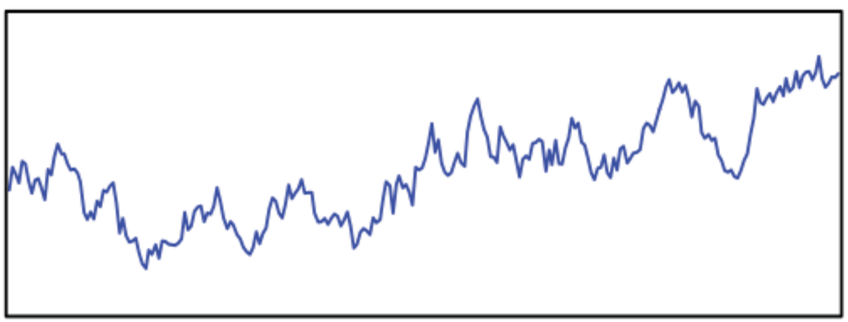
Correlate a series with itself shifted backward by one time period

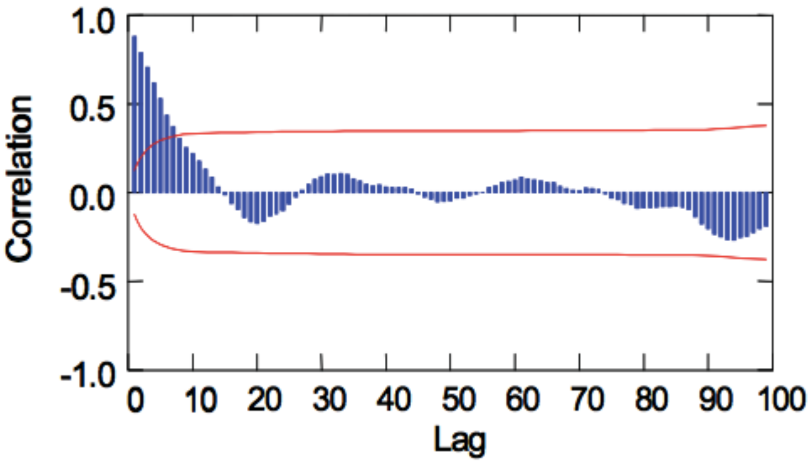
Correlate the shifted series with itself shifted backward by one time period

And so on…

Here's an Autocorrelation Function Plot of an AR(1) process

$$x_t = \phi x_{t-1} + \epsilon_t$$





Time and Space

Stochastic processes

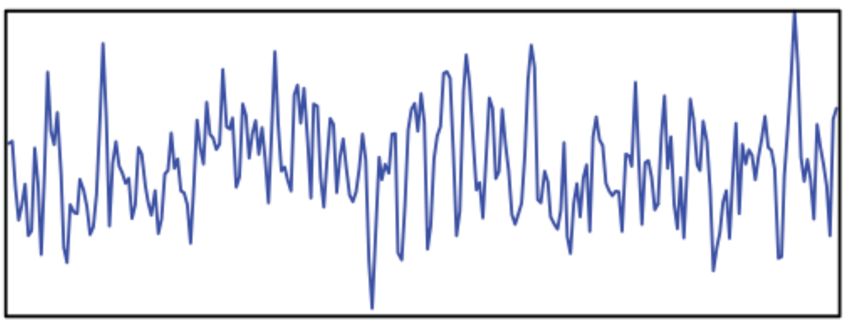
Moving Average (MA) processes

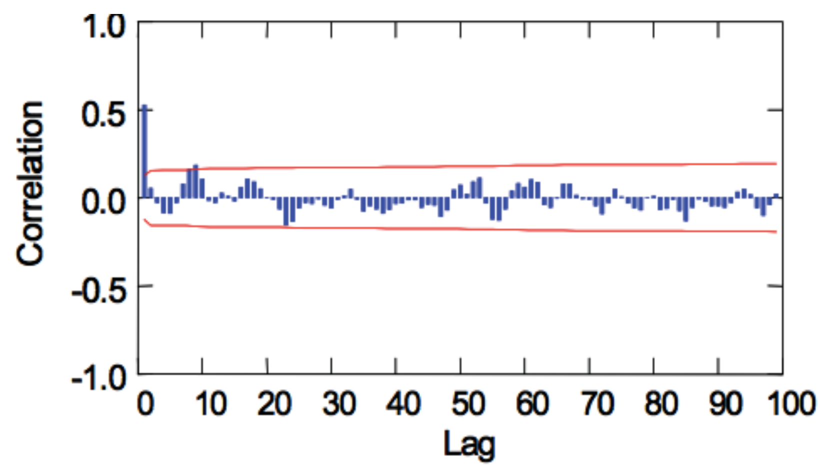
In this model each observation at a given time is a function the previous error

Plus random error

$$x_t = \theta \epsilon_{t-1} + \epsilon_t$$

Here's an Autocorrelation Function Plot of an MA(1) process





Time and Space

Stochastic processes

Moving Average (MA) processes

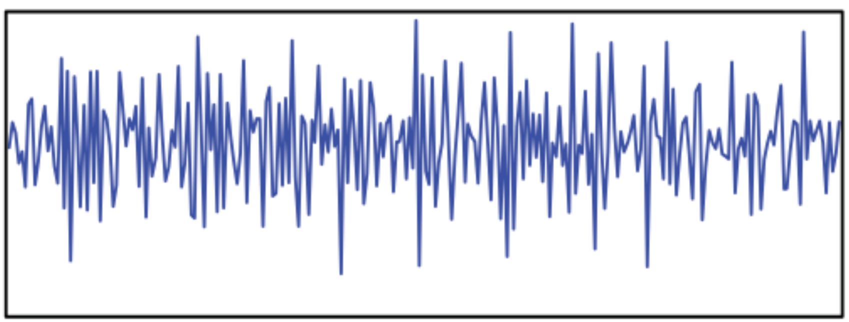
The θ parameter can be negative

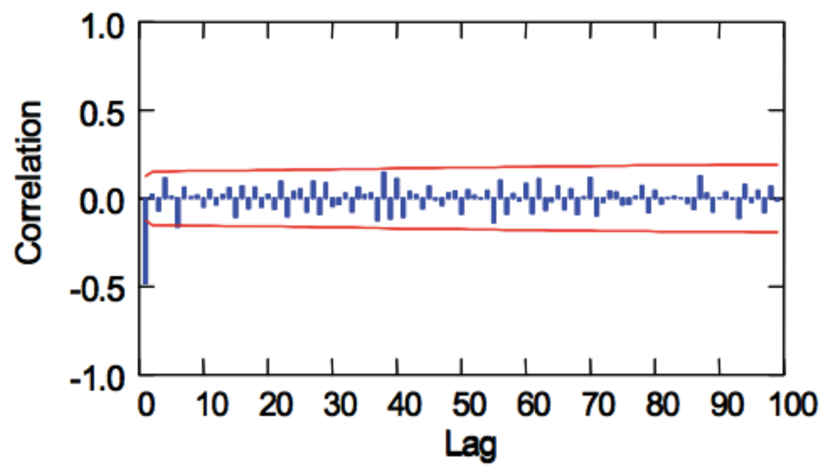
$$x_t = \theta\epsilon_{t-1} + \epsilon_t$$

Here's an Autocorrelation Function Plot of a negative MA(1) process

Negative θ enhances high frequencies

Positive θ enhances low frequencies



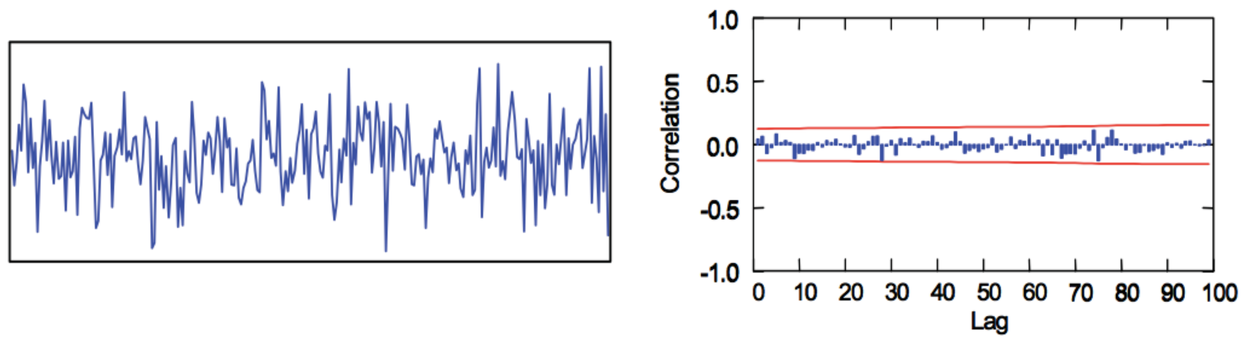


Time and Space

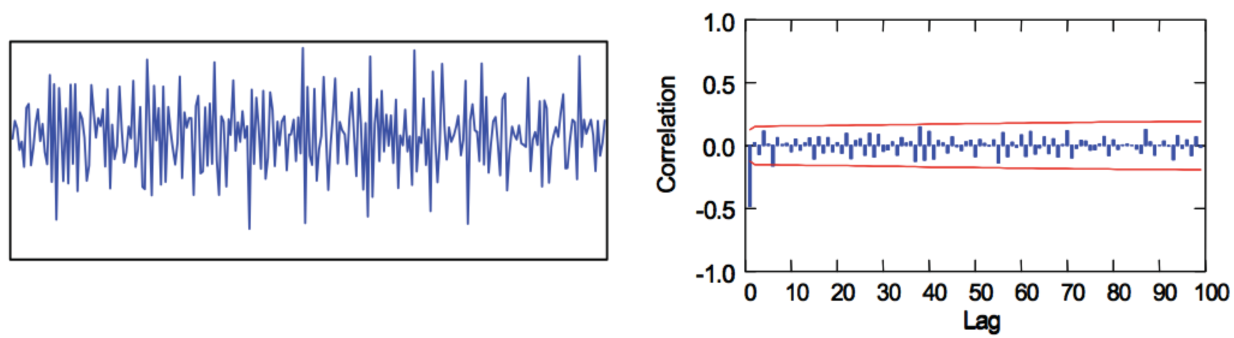
ACF Plots

Notice that without an ACF plot, diagnosis of raw series is difficult

White noise



MA(1)



Time and Space

Stochastic processes

ARMA processes (Box & Jenkins)

We can mix these models

An ARMA model looks like this

$$x_t = \sum_{i=1}^{p} \phi_i x_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t$$

In most cases, the coefficients of the terms decay exponentially

So we do not have to make p and q large for modeling most series

All the models we've seen so far can include a constant

We can also add trend to these models

$$x_t = \alpha + \beta x_t + \sum_{i=1}^{p} \phi_i x_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t$$

Time and Space

Stochastic processes

Seasonal processes

Dependencies in the model can be across seasons

A seasonal autoregressive model looks like this

$$x_t = \phi x_{t-s} + \epsilon_t$$

And a seasonal moving average model looks like this

$$x_t = \theta \epsilon_{t-s} + \epsilon_t$$

Economists love this stuff

- They even mix stochastic and classical models in the same equation
- Their goal is to account for dependencies in the residuals in regression models
- Here's an example of one of their models
  - Generalized Least Squares

$$\hat{\beta} = (\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{Y})$$

Time and Space

Stochastic processes

Estimating ARMA models

- Are you serious?
- This is a black art
- And usually you want ARIMA instead of ARMA
  - Which I haven't even told you about
- Even after a semester course in ARIMA models you won't be able to do it
- You have to learn how to diagnose ACF plots
  - And PACF plots, which I haven't even told you about
- You have to know when to difference your series to achieve stationarity
  - Which I haven't even told you about
- Leave this to the experts

That brings us to the next topic

- There's a simple model that does better than fancy ARIMA for many real forecasts
- It's called Exponential Smoothing
  - It includes seasonal effects as well

Time and Space

Stochastic processes

Exponential Smoothing

We begin with the moving average smoothing model

For a point at time t $(1 \leq t \leq n)$, a moving average smoothed value is given by

$$\hat{x}_t = \frac{1}{p} \sum_{i=1}^{p} x_{t-i}$$

Some considerations:

- Our smoothing estimate is simply the average of the p previous values.
- The first p points in the series are not smoothed.
- If each point in the series has a random component, we are averaging fixed and random components of previous points.
- In this case, the model smooths only p prior random components (not *n*).
- In other words, the model ignores any randomness before the previous *p* time points.
- If we presume only random error governs the process, we call the process a *random walk*.
- If we believe the process is a random walk, then we should set *p* = 1.
- If *p* = 1, the smooth is just the previous observation.
- If *p* = 1, we are assuming there is no more information we can get out of the data
- If *p* > 1, we are assuming we can eliminate the effects of the errors by averaging them.

Time and Space

Stochastic processes

Exponential Smoothing

Now go on to the weighted moving average smoothing model

$$\hat{x}_t = \frac{1}{p} \sum_{i=1}^{p} w_i x_{t-i}$$

Let's make these weights decline exponentially

$$w_i = p^{-i}$$

And let's normalize them to add to 1

$$w_i = \left(\frac{p-1}{1-p^{-p}} \right) p^{-i}$$

This makes the exponentially weighted smoothing model

Time and Space

Stochastic processes

The Exponentially Weighted Moving Average Model (EWMA)

Here is the recursive form of the exponentially weighted smoothing model

$$\hat{x}_t = \alpha x_{t-1} + (1 - \alpha)\hat{x}_{t-1}$$

Notice the $\hat{x}_{t-1}$ on the right

We assume $0 < \alpha < 1$ so things don't explode

This formula gives us a recursive estimation method

No fancy optimization needed

What we are doing here is projecting forward local patterns in the series

We could consider this a statistical estimation method

Or we could just think of it as a deterministic forward pattern duplicator

Time and Space

Stochastic processes

The Holt-Winters method

Now it gets powerful

Holt and Winters added trend and seasonality to EWMA

- H-W fits three types of trend models (none, linear, multiplicative)
- H-W does not fit other types of trend functions (although it could be modified to do so)
- H-W fits three types of seasonality (none, additive, multiplicative)
- H-W can fit more than simple sinusoidal seasonality functions
- H-W does not fit more than one type of seasonality in one model (but it could)
- H-W additive linear models parallel specific ARIMA models
- H-W multiplicative models do not have ARIMA parallels

Forecasting

Fit first half of series

Extrapolate to second half to get residuals

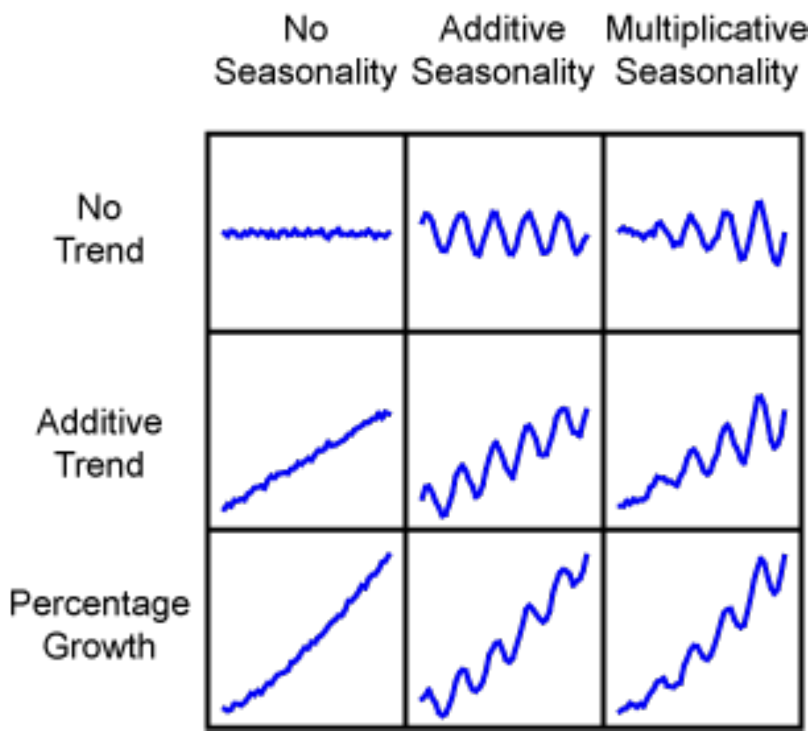
Analyze residuals for anomalies

Forecast beyond end of series

Time and Space

Stochastic processes

The Holt-Winters method

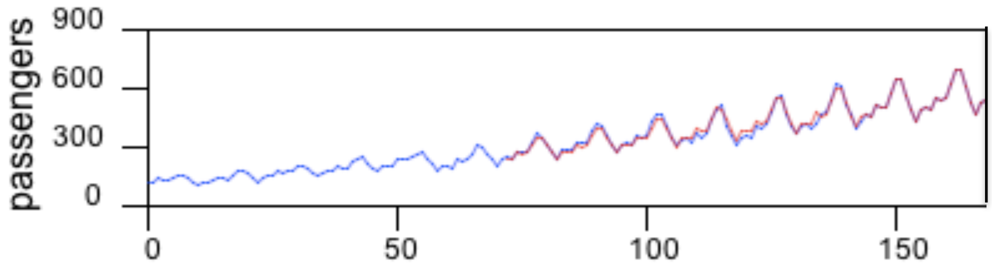


Time and Space

Stochastic processes

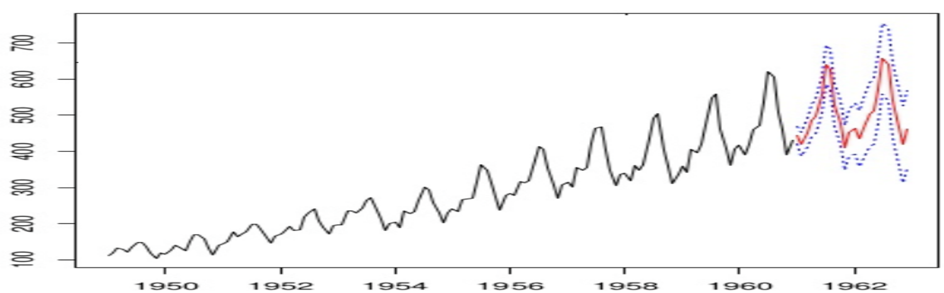
The Holt-Winters method

Here is the H-W forecast for the famous Box-Jenkins Airline dataset




```
hw passengers
  estimate /
  smooth=.3, linear=.4,season=12,
  multiplicative=.5, forecast=10
```

And here is the forecast using ARIMA (0,1,0)(0,1,0)



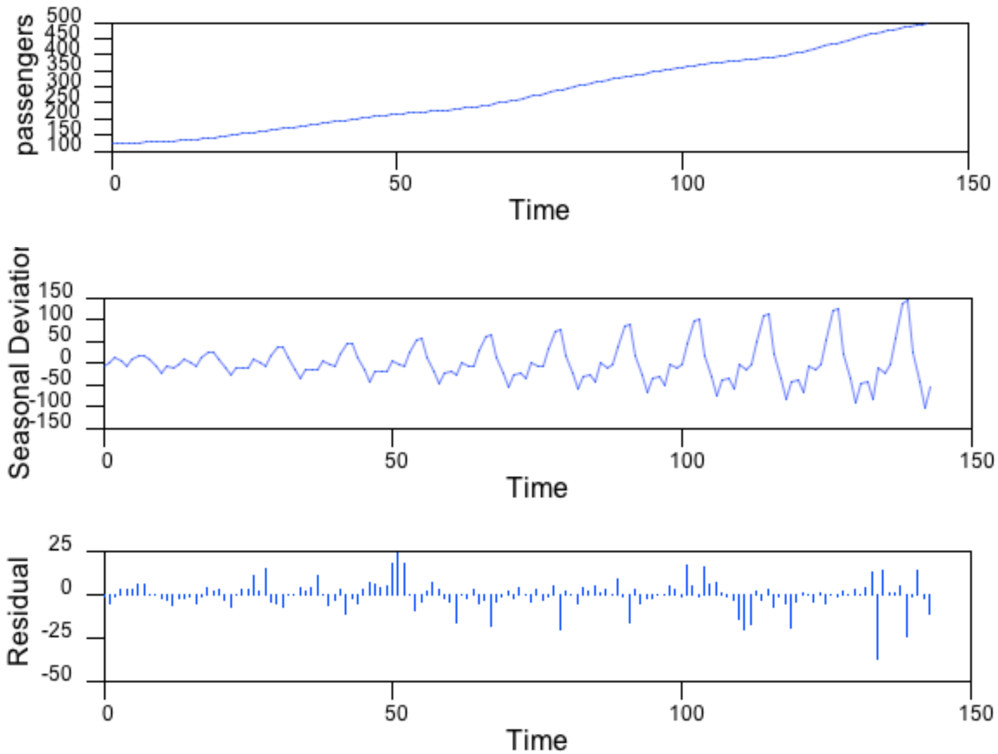

```
arima passengers
  log
  difference
  difference / lag=12
  estimate / q=1, qs = 1,
        season=12,
        backcast = 13,forecast=10
```

Time and Space

Stochastic processes

Seasonal decomposition

X11/12 (US Census), SABL (Cleveland, Bell Labs)



Trend

Season

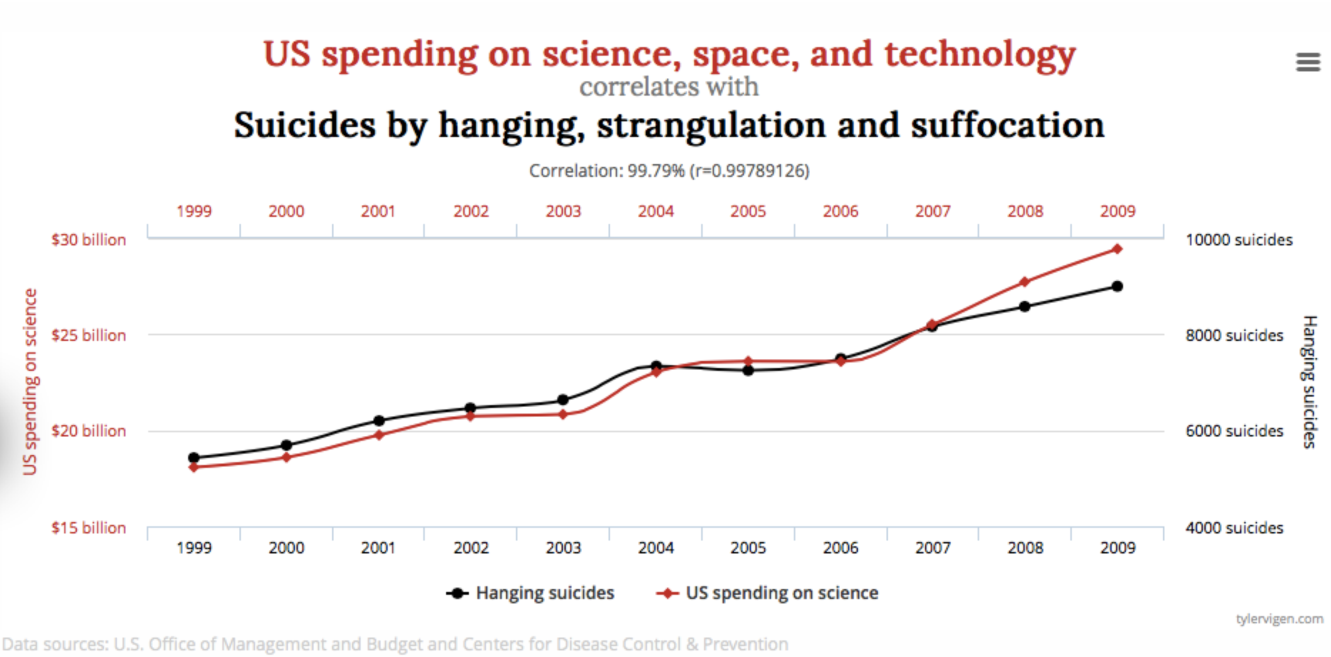
Residual

Time and Space

Correlating Time Series

Don't do this…

I love this site!

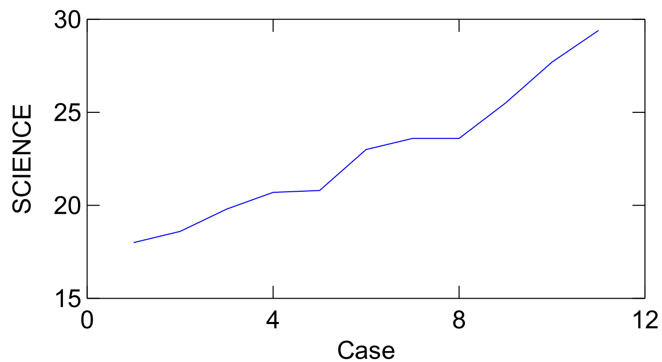


http://www.tylervigen.com/spurious-correlations

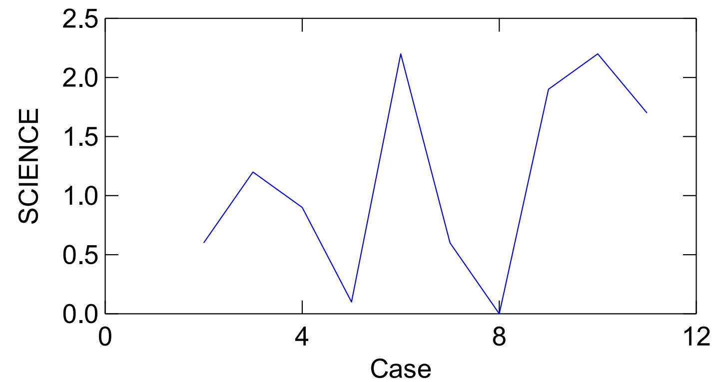
Time and Space

Correlating Time Series

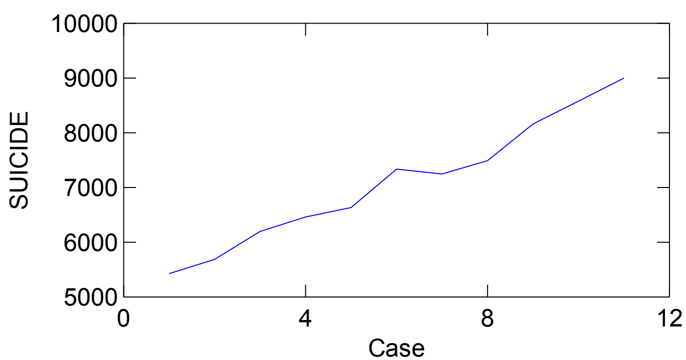
Science Raw



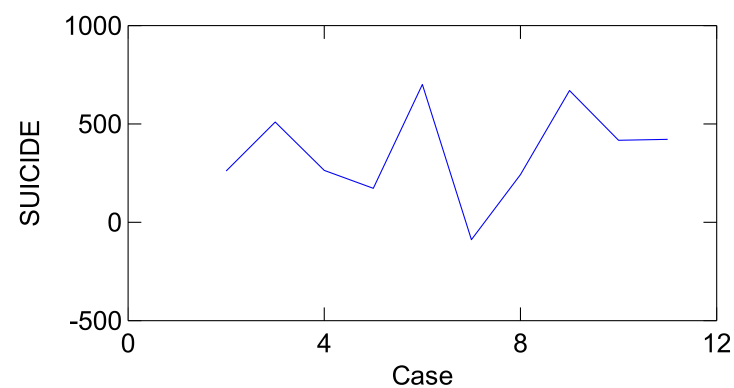
Science Detrended



Suicide Raw



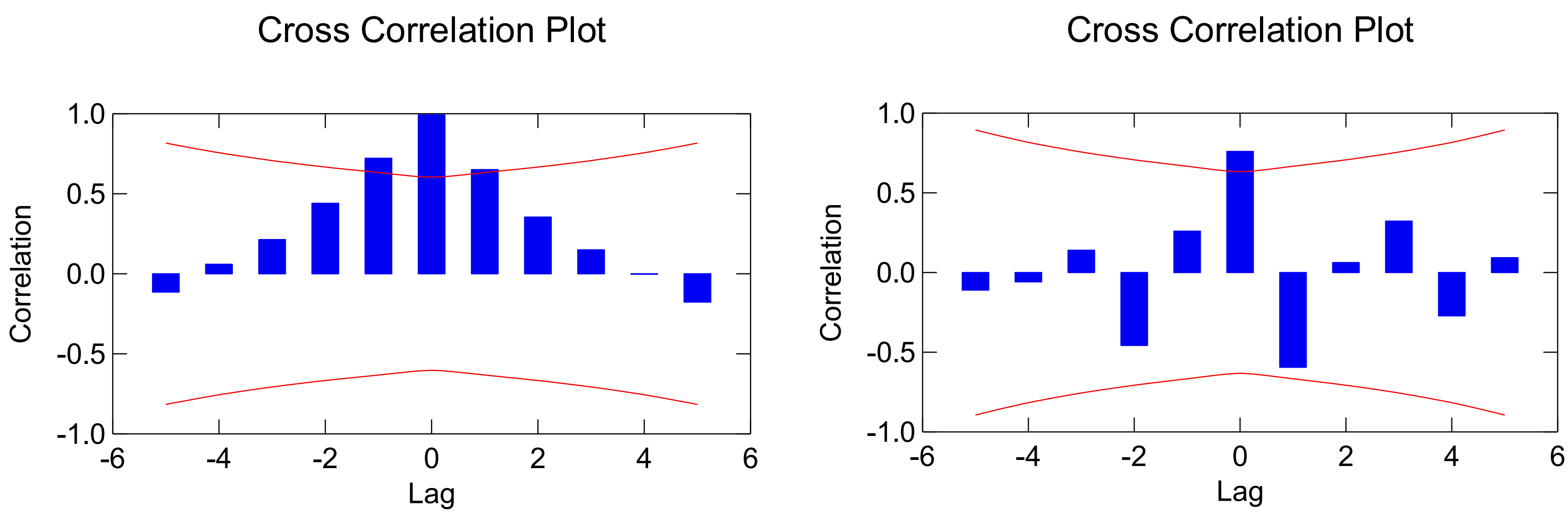
Suicide Detrended



Time and Space

Correlating Time Series

CCF of Raw series vs. CCF of detrended



Detrending doesn't always get you out of the woods

There can be second-order artifacts that influence correlation between series

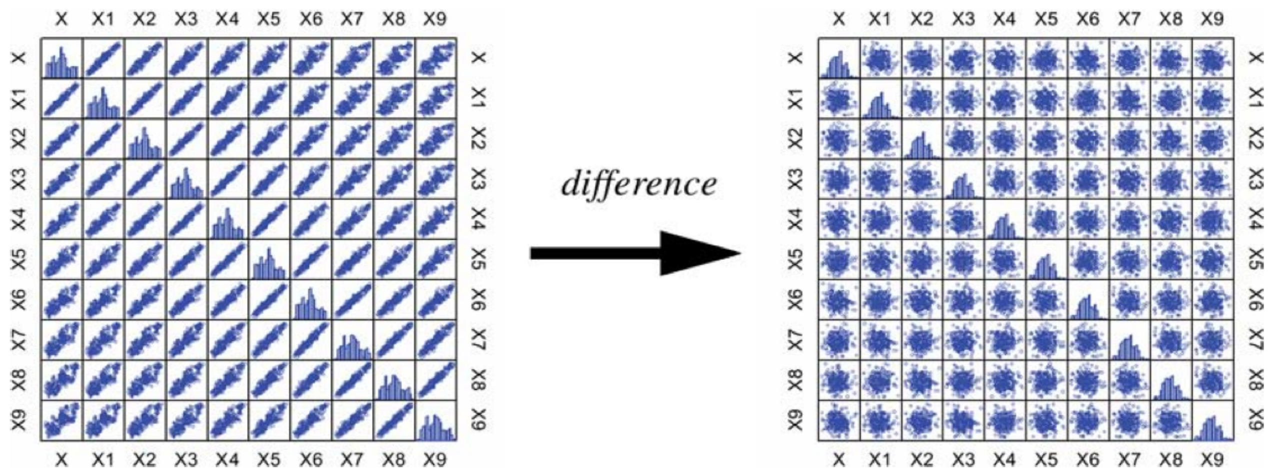
Time and Space

Multivariate analysis of time series

The cautions mentioned earlier apply to any analysis of time series

- e.g., Clustering or Principal Components of time series
  - Need to difference to achieve stationarity before clustering
  - First differences of a random walk achieves stationarity
  - Other models require more exotic measures

First 9 lags of a random walk



Time and Space

- Wait, there's more…
  - But if you insist on trying this stuff, you'd better talk to a time-series statistician or economist
    - But don't ask the economist to predict the economy!