## Autoregressive (AR) Models

$$X_t = \mathcal{E}_t + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p}$$

The above generally states that today's value is slightly different froma  combination of the last few day's values.

Autocorrelated Models are stationary, but the same cannot be said for Autoregressive Models.

The stationary nature of Autoregressive Models (AR) is dependent on the parameters:

Example Simulation:

Assuming 50,000 noise values generated from N(0,1) (normal distribtuion with mean $\mu$ of 0 and variance $\sigma^2$ of 1) to be $\mathcal{E}_1, \ldots, \mathcal{E}_{50,000}$

$X_1 = \mathcal{E}_1$

For $t = 2$ to 50,000:

$X_1 = \mathcal{E}_1 + \mathbf{0.5} X_{t-1}$

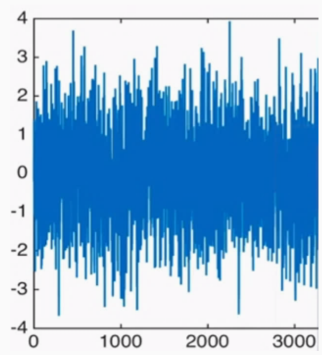
→ appears stationary

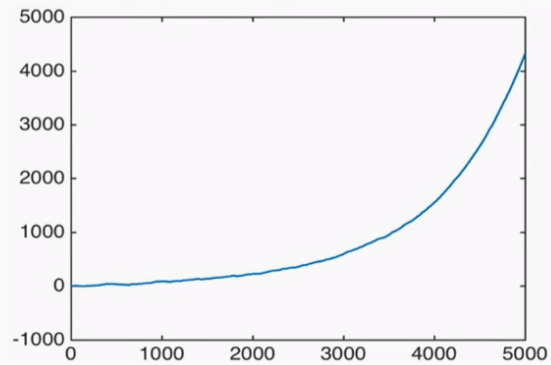
$X_1 = \mathcal{E}_1 + \mathbf{1.001} X_{t-1}$

→ appears exponential (positive)

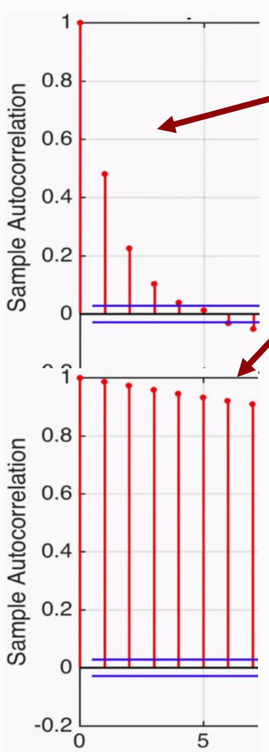
$X_1 = \mathcal{E}_1 + \mathbf{0.999} X_{t-1}$

→ appears exponential (negative)




Determining the stationary nature of an Autoregressive Model by examining Autocorrelated Models:



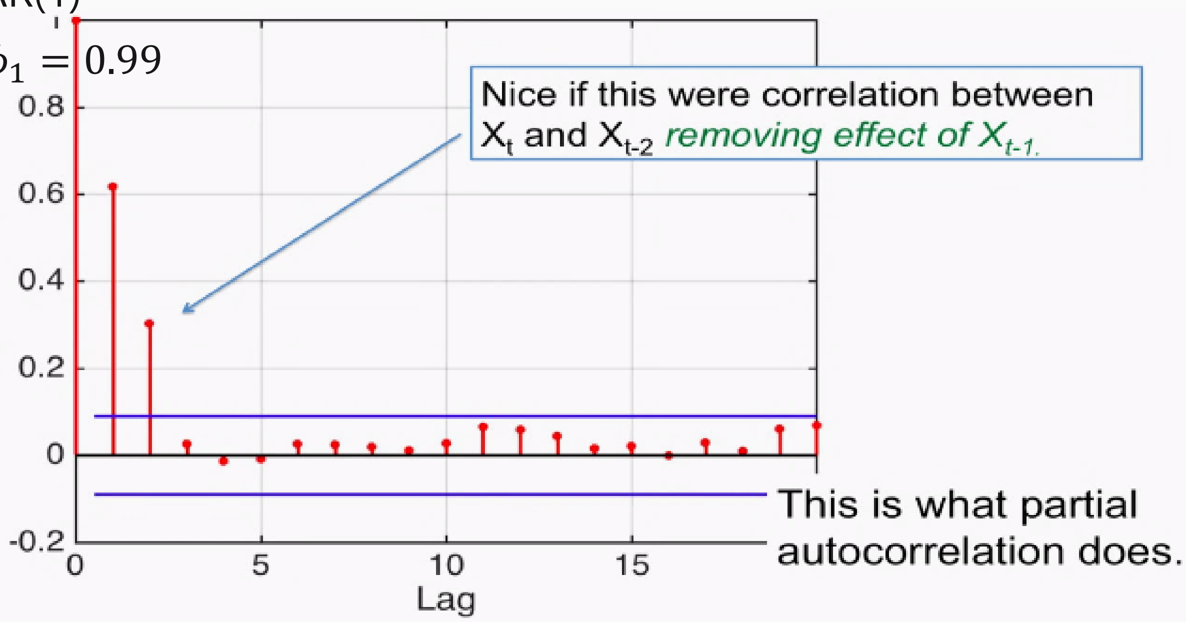

AR(1)

$\phi_1 = 0.5$

Both models exhibit exponential decay, with the second model occuring at a slower rate; meaning that the Autocorrelation Model **cannot be used** to determine the order of an Autoregressive Model:

AR(1)

$\phi_1 = 0.99$



Because Autocorrelation experiences exponential decay, the order of an Autoregressive Model cannot be determined when evaluating the stationary nature of the data. Instead, the Partial Autocorrelation function is examined for stationary inference of the data

## Partial Autocorrelation

The Partial Autocorrelation of $X_1$ and $X_{t-u}$ is the correlation not accounted for by lags $1, \dots, u-1$:

Written in terms of lag *(u)*: $\text{PartialAutoCorr}(u) = \text{Cor}\left(X_t, X_{t-u} \mid X_{t-1}, \dots, X_{t-(u+1)}\right)$

The PAC is the correlation between $X_t$ and $X_{t-u}$ given everything that happened in between the two.

Using the Sleep Model as a working example (the effect of sleep 2 nights ago on tonight's sleep):

In terms of order $2$: $\text{PartialAutoCorr}(2) = \text{Cor}(X_t, X_{t-2} \mid X_{t-1})$

→ The correlation between tonight's sleep and two nights ago that is not accounted for last night

To compute the Partial Autocorrelation: Predict correction to $X_t$ from $X_{t-1}$: $\hat{X}_t = \beta_1 X_{t-1}$

$\beta_1$ is the correlation between $X_t$ and $X_{t-1}$.

$\hat{X}_t$ is predicted based upon the sole fact that $X_t$ is correlated with $X_{t-1}$.

All predict correction to $X_{t-2}$ from $X_{t-1}$: $\hat{X}_{t-2} = \beta_1 X_{t-1}$

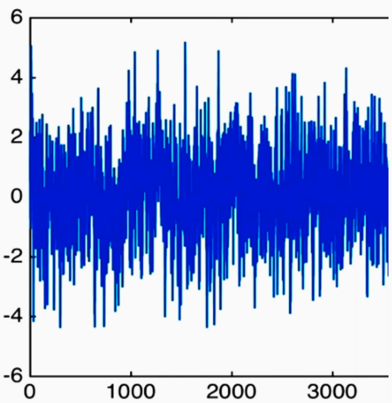
$\beta_1$ is the correlation between $X_{t-2}$ and $X_{t-1}$, which is that same as the correlation between $X_t$ and $X_{t-1}$. This is attributed to the stationary nature of the series (mean and variance constant)

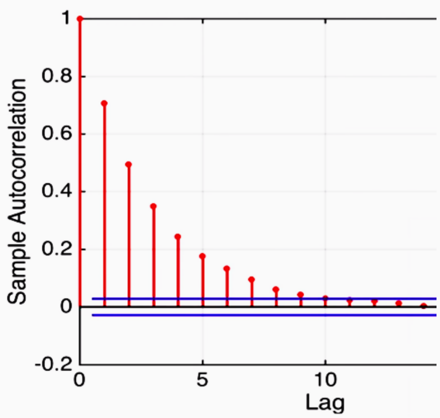
In terms of order $2$: $\text{PartialAutoCorr}(2) = \text{Cor}\left(X_t - \hat{X}_t, X_{t-2} - \hat{X}_{t-2}\right)$

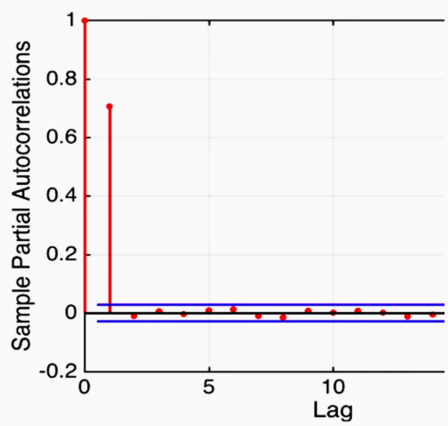
the above formulates the question that arises: Is it possible to use what cannot be explained about $X_{t-2}$ from $X_{t-1}$ to predict what cannot be explained about $X_t$ from $X_{t-1}$?

The Partial Autocorrelation computes the correlation of the prediction error at time $t$ and the prediction error at time $t-2$.

For example: assuming the model $X_1 = \mathcal{E}_1 + \mathbf{0.7} X_{t-1}$ and stationary timeseries data:




the Sample Autocorrelation Model is not useful as expected, other than proving that the model is not a true moving average model (or else the Sample Autocorrelation plot would have a cutoff). The exponential decay additionally hints the data is an Autoregressive Model (with order still unknown). In contrast, the Sample Partial Autocorrelation Model depicts an order of 1 (AR(1)) due to a single non-zero term in the Partial Autocorrelation Function.

in general terms $(u)$:

$$\text{PartialAutoCorr}(u) = \text{Cor}\left(X_t - \hat{X}_t, X_{t-u} - \hat{X}_{t-u}\right)$$

The Partial Autocorrelation for lag $u$ is the correlation between the prediction error at time $t$ and the prediction error at time t-u; where allowed to explain $X_t$ and $X_{t-u}$ using **everything** in between:

$$\hat{X}_t = \beta_1 X_{t-1} + \beta_2 X_{t-2} + \cdots + \beta_{u-1} X_{t-(u-1)}$$

$$\hat{X}_{t-u} = \beta_1 X_{t-(u+1)} + \beta_2 X_{t-(u+2)} + \cdots + \beta_{u-1} X_{t-1)}$$

The least squares method will be used to determine the value of the $\beta$'s

Example: use a linear model to predict from the lags:

$$\hat{X}_t = \beta_1 X_{t-1} + \beta_2 X_{t-2} + \cdots + \beta_{u-1} X_{t-(u-1)}$$

use least squares to get the coefficients in a minimized error fashion: $\min_{\beta_0, \beta_1, \ldots, \beta_{u-1}} \left(\hat{X}_t - X_t\right)^2$

then use a linear model to predict $X_{t-u}$ from the lags:

$$\hat{X}_{t-u} = \beta_1 X_{t-(u+1)} + \beta_2 X_{t-(u+2)} + \cdots + \beta_{u-1} X_{t-1)}$$

the coeficients are the same in both models because least squares is about correlations (which are not directed forward or backwards in time; no directionality):

$$\text{PartialAutoCorr}(u) = \text{Cor}\left(X_t - \hat{X}_t, X_{t-u} - \hat{X}_{t-u}\right)$$

The Partial Autocorrelation describes the unexplained part of $X_t$ from the unexplained part of $X_{t-u}$; where explinations are derived from all of the lags in between.

## Theorhetical Summary of Autoregressive Models:

- AR models are not always stationary; this depends on the parameters
- The autocorrelation functions for stationart AR(p) models experiences exponential decay and thus, order cannot be determined
- The partial autocorrelation for stationary (AR(p) models experiences a strict cutoff at (p) and thus, the order can effectively be determined