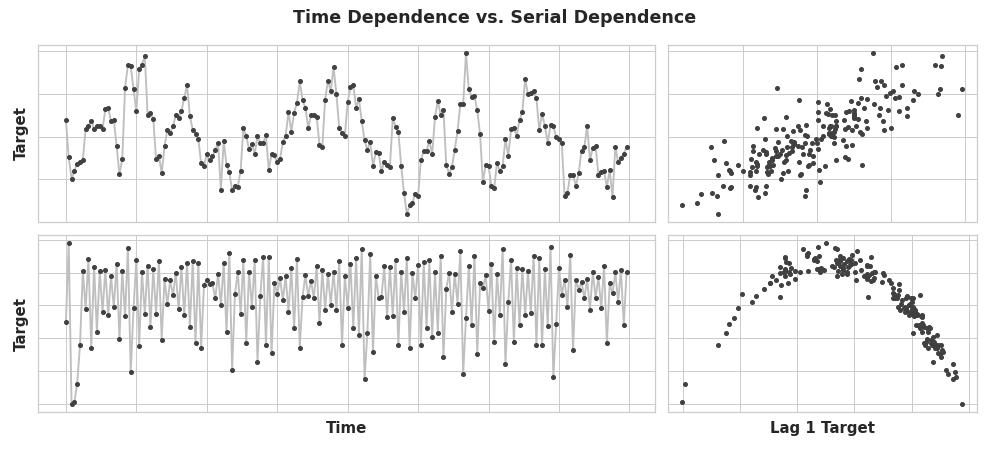
Time Series: 4th lesson – Time Series As Features

What is serial dependence?

In earlier lessons, we investigated properties of time series that were most easily modeled as *time dependent* properties, that is, with features we could derive directly from the time index. Some time series properties, however, can only be modeled as *serially dependent* properties, that is, using as features past values of the target series. The structure of these time series may not be apparent from a plot over time; plotted against past values, however, the structure becomes clear -- as we see in the figure below.

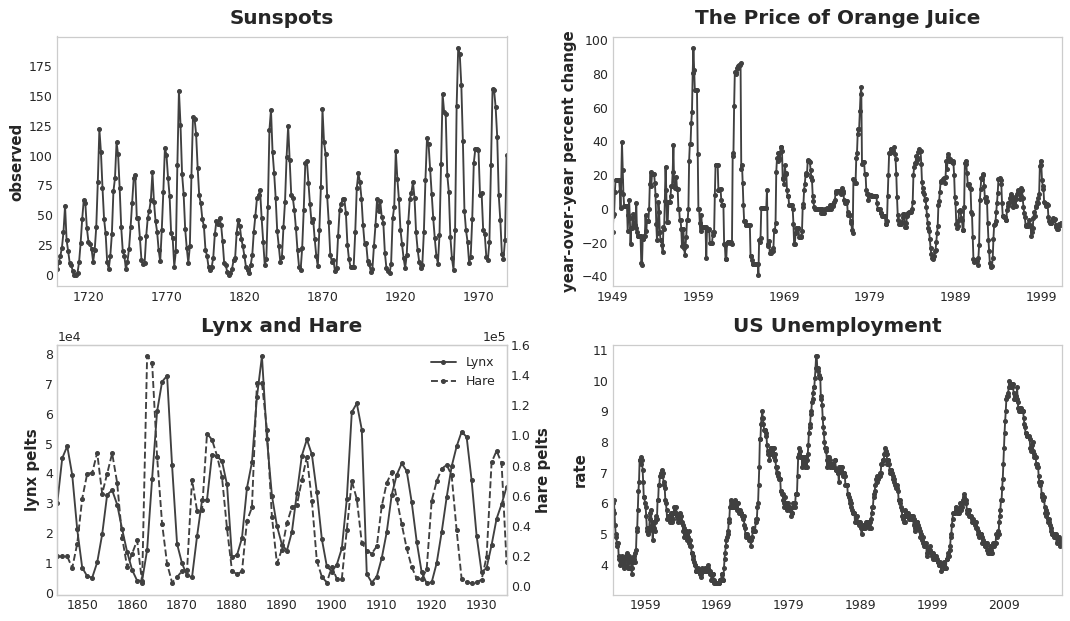


\*these two series have serial dependence, but not time dependence. Points on the right have coordinates: value at time t-1, value at time t

With trend and seasonality, we trained models to fit curves to plots like those on the left in the figure above -- the models were learning time dependence. The goal in this lesson is to train models to fit curves to plots like those on the right -- we want them to learn serial dependence.

Cycles:

Patterns of growth and decay in a time series associated with how the value in a series at one time depends on values at previous times, but not necessarily on the time step itself. Cyclic behavior is characteristic of systems that can affect themselves or whose reactions persist over time, economics, epidemics, animal populations, volcano eruptions, and similar natural phenomena often display cyclic behavior.



What distinguishes cyclic behavior from seasonality is that cycles are not necessarily time dependent, as seasons are. What happens in a cycle is less about the particular date of occurence, and more about what has happened in the recent past. At least, relative independence from time means that cyclic behavior can be much more irregular than seasonality.

Lagged series and lag plots:

To investigate possible serial dependence (like cycles) in a time series, we need to create "lagged" copies of the series. Lagging a time series means to shift its values forward one or more time steps, or equivalently, to shift the times in its index backward one or more steps. In either case, the effect is that the observations in the lagged series will appear to have happened later in time.

This shows the monthly unemployment rate in the US (y) together with its first and second lagged series (y\_lag\_1 and y\_lag\_2, respectively). Notice how the values of the lagged series are shifted forward in time.

import pandas as pd

*# Federal Reserve dataset: https://www.kaggle.com/federalreserve/interest-rates*

reserve = pd.read\_csv(

"../input/ts-course-data/reserve.csv",

parse\_dates={'Date': ['Year', 'Month', 'Day']},

index\_col='Date',

)

y = reserve.loc[:, 'Unemployment Rate'].dropna().to\_period('M')

df = pd.DataFrame({

'y': y,

'y\_lag\_1': y.shift(1),

'y\_lag\_2': y.shift(2),

})

df.head()

y y\_lag\_1 y\_lag\_2

Date

1954-07 5.8 NaN NaN

1954-08 6.0 5.8 NaN

1954-09 6.1 6.0 5.8

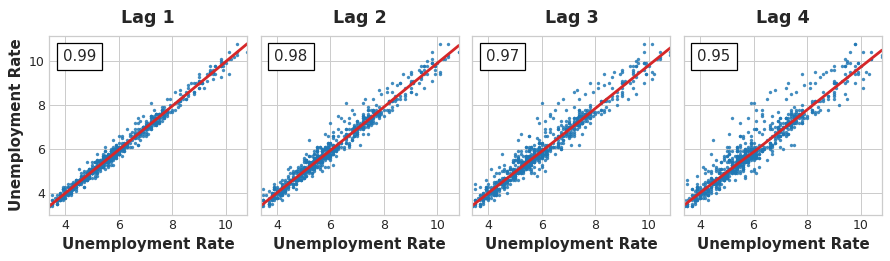
1954-10 5.7 6.1 6.0

1954-11 5.3 5.7 6.1

By lagging a time series, we can make its past values appear contemporaneous with the values we are trying to predict (in the same row, in other words). This makes lagged series useful as features for modeling serial dependence. To forecast the US unemployment rate series, we could use y\_lag\_1 and y\_lag\_2 as features to predict the target y. This would forecast the future unemployment rate as a function of the unemployment rate in the prior two months.

Lag plots:

A lag plot of a time series shows its values plotted against its lags. Serial dependence in a time series will often become apparent by looking at a lag plot. We can see from this lag plot of US Unemployment that there is a strong and apparently linear relationship between the current unemployment rate and past rates.

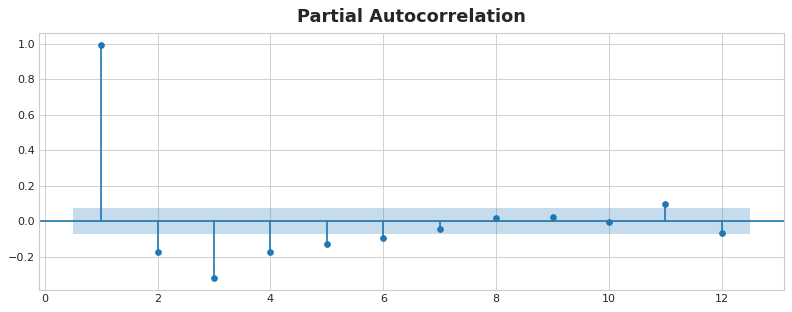


The most commonly used measure of serial dependence is known as autocorrelation, which issimply the correlation a time series has with one of its lags. US Unemployment has an autocorrelation of 0.99 at lag 1, 0.98 at lag 2, and so on.

Choosing lags:

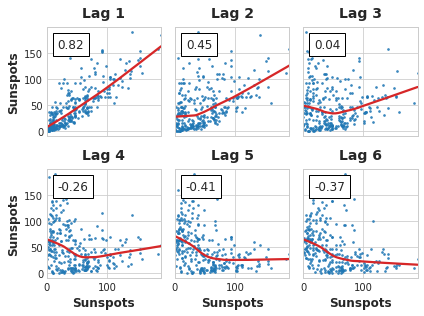
When choosing lags to use as features, it generally won't be useful to include every lag with alarge autocorrelation. In *US Unemployment*, for instance, the autocorrelation at lag 2 might result entirely from "decayed" information from lag 1 -- just correlation that's carried over fromthe previous step. If lag 2 doesn't contain anything new, there would be no reason to include it if we already have lag 1.

The partial autocorrelation tells you the correlation of a lag accounting for all of the previous lags -- the amount of "new" correlation the lag contributes, so to speak. Plotting the partial autocorrelation can help you choose which lag features to use. In the figure below, lag 1 through lag 6 fall outside the intervals of "no correlation" (in blue), so we might choose lags 1 through lag 6 as features for *US Unemployment* (lag 11 is likely a false positive).



A plot like that above is known as a correlogram. The correlogram is for lag features essentially what the periodogram is for Fourier features.

Finally, we need to be mindful that autocorrelation and partial autocorrelation are measures of *linear dependence*. Because real-world time series often have substantial non-linear dependences, it's best to look at a lag plot (or use some more general measure of dependence, like mutual information) when choosing lag features. *The Sunspots* series has lags with non-linear dependence which we might overlook with autocorrelation.



Non-linear relationships like these can either be transformed to be linear or else learned by anappropriate algorithm.