

## Lecture 19

### Lags, Autocorrelation, ACF methods

In a regular time series, autocorrelation measures the linear relationship between observations at different time points within the same time series. Specifically, autocorrelation at lag  $k$  ( $ACF(k)$ ) measures the correlation between the observations at time  $t$  and the observations at time  $t - k$ , where  $k$  is a positive integer representing the time lag.

For example, in a monthly time series of sales data, the autocorrelation at lag 1 would measure the correlation between the sales in January and the sales in February, while the autocorrelation at lag 12 would measure the correlation between the sales in January and the sales in January of the previous year.

**Autocorrelation** is important in regular time series analysis because it helps identify the presence of trends, seasonality, and other patterns in the data. Autocorrelation can be used to test for the presence of serial correlation (the dependence of observations on previous observations), which is a violation of the assumption of independence required by many statistical models. Autocorrelation can also be used to help identify the appropriate lag order for autoregressive and moving average models, and to help diagnose model misspecification or inadequate modeling assumptions.

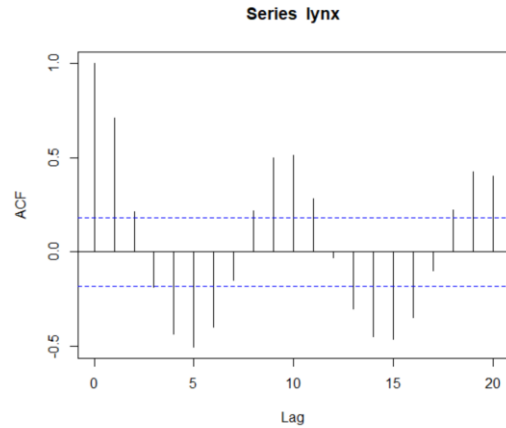
In time series analysis, a **lag** refers to the number of time periods between two consecutive observations in a time series. For example, in a monthly time series of sales data, the lag between January and February is one  $y_{t-1}$ , and the lag between January and March is two  $y_{t-2}$ .

Lags are often used in autocorrelation analysis to measure the linear relationship between observations at different points in time. For example, the autocorrelation at lag  $k$  measures the correlation between the observations at time  $t$  and the observations at time  $t-k$ . By examining the autocorrelation at different lags, analysts can identify the presence of patterns in the data such as trends, seasonality, or cyclical behavior.

Lags are also used in time series modeling, such as in autoregressive integrated moving average (ARIMA) models, where the lag order is an important parameter to be specified. The lag order represents the number of past observations used in the model to predict future values of the time series. The choice of the appropriate lag order can have a significant impact on the accuracy of the model's predictions.

In time series analysis, **ACF** stands for "**autocorrelation function**". The ACF is a mathematical tool used to measure the correlation between observations at different lags within a time series.

The ACF at lag  $k$  ( $ACF(k)$ ) measures the correlation between the observations at time  $t$  and the observations at time  $t - k$ . By calculating the ACF at different lags, analysts can identify the presence of patterns in the data such as trends, seasonality, or cyclical behavior.

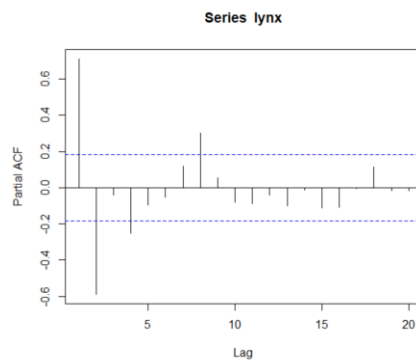


The ACF is typically visualized using a plot called the autocorrelation plot, which displays the ACF values for different lags. The autocorrelation plot can help analysts identify the lag at which the autocorrelation is strongest, which can provide insights into the underlying patterns in the data.

The ACF is an important tool in time series analysis and modeling, as it is used to diagnose model adequacy and to help identify the appropriate lag order for autoregressive and moving average models.

In time series analysis, **PACF** stands for "**partial autocorrelation function**". The PACF is a mathematical tool used to measure the correlation between observations at different lags within a time series, after removing the effects of the intervening lags.

The PACF at lag  $k$  ( $PACF(k)$ ) measures the correlation between the observations at time  $t$  and the observations at time  $t - k$ , after removing the effects of the intervening lags. This is different from the ACF, which measures the correlation at lag  $k$  without accounting for the effects of the intervening lags.



The PACF is typically visualized using a plot called the partial autocorrelation plot, which displays the PACF values for different lags. The partial autocorrelation plot can help analysts identify the lag at which the partial autocorrelation is strongest, which can provide insights into the underlying patterns in the data. It looks similar to the ACF graph, but the correlations plotted will not be the same.

The PACF is an important tool in time series analysis and modeling, as it is used to diagnose model adequacy and to help identify the appropriate lag order for autoregressive models.

Both the ACF and PACF are important tools in time series analysis and modeling, and they are used for different purposes.

The ACF is primarily used to identify the presence of autocorrelation in a time series, which is a measure of the relationship between observations at different lags. By analyzing the ACF plot, analysts can identify the presence of patterns such as seasonality, cyclical behavior, and trends. The ACF is also used to identify the appropriate lag order for moving average (MA) models.

The PACF, on the other hand, is primarily used to identify the appropriate lag order for autoregressive (AR) models. The PACF measures the correlation between observations at different lags, while controlling for the effects of intervening lags. By analyzing the PACF plot, analysts can identify the lag order for AR models.

An **autoregressive (AR) model** is a type of statistical model used in time series analysis to describe the relationship between an observation and a lagged value of the same time series. In an AR model, the current value of a variable is modeled as a linear combination of its past values, with the coefficients of the past values estimated from the data.

The order of an AR model (denoted as " $p$ ") indicates the number of past values of the variable that are used to model the current value. For example, an  $AR(1)$  model uses the lagged value from the previous time step to predict the current value, while an  $AR(2)$  model uses the values from the two previous time steps.

The general form of an  $AR(p)$  model can be written as:

$$y_t = c + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \dots + \Phi_p y_{t-p} + \varepsilon_t$$

where:

$y_t$  is the value of the variable at time  $t$

$c$  is a constant term

$\Phi_1, \Phi_2, \dots, \Phi_p$  are the coefficients of the lagged values of the variable

$\varepsilon_t$  is the error term, which is assumed to be white noise

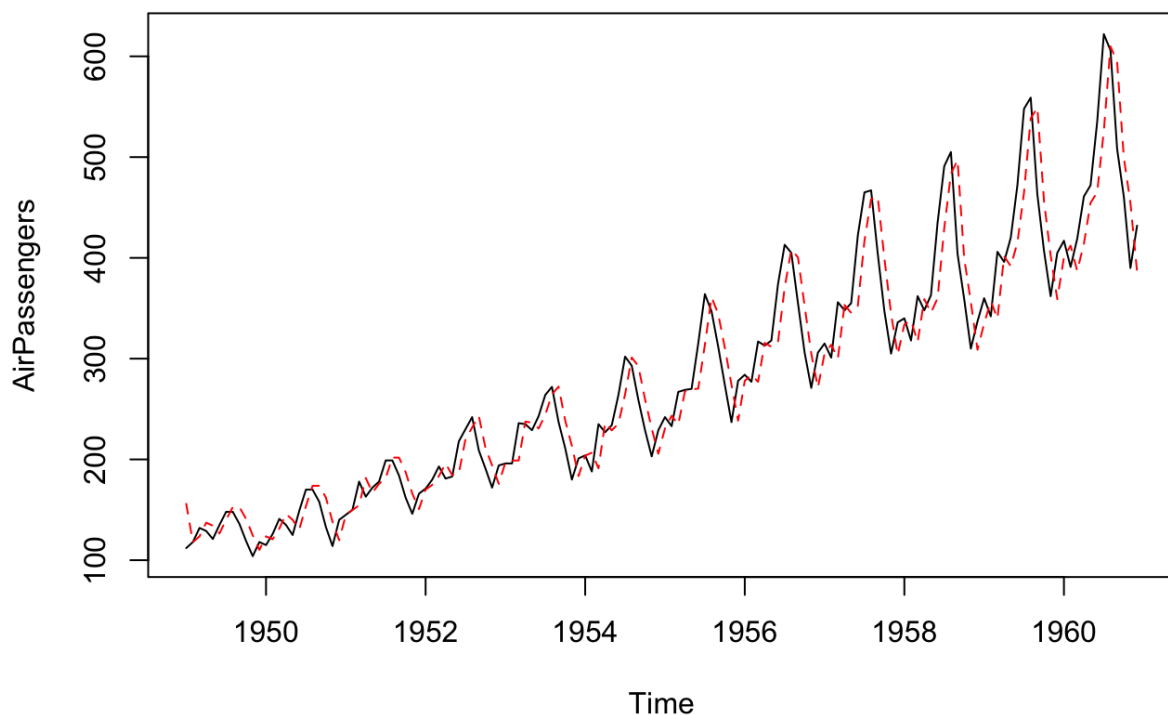
AR models are useful for modeling stationary time series data that exhibit autocorrelation, which means that the current value of the variable is correlated with its past values. AR models are commonly used in financial forecasting, weather forecasting, and other applications in which past values of a variable can be used to predict future values.

The partial autocorrelation function (PACF) can be used to identify the appropriate number of lags to include in an autoregressive (AR) model. The PACF measures the correlation between the series and its lags at different time lags.

To choose the number of lags in an AR model using the PACF, one can look for significant correlation coefficients at different lags. If the PACF shows a significant correlation at a certain lag and then gradually decreases to zero as the lags increase, this suggests that the AR model with that lag is appropriate.

One common approach is to use the partial autocorrelation function (PACF) to identify the lag structure of the AR model. The PACF measures the correlation between the series and its lags after adjusting for the effect of the intermediate lags. Significant coefficients in the PACF at a certain lag indicate that this lag should be included in the model.

If the ACF and PACF do not provide clear guidance on the number of lags to include in the model, other methods such as the Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC) can be used to select the best model. These methods compare the goodness-of-fit of models with different numbers of lags and choose the one that balances the fit with the complexity of the model.



Resources:

1. <https://vitalflux.com/different-types-of-time-series-forecasting-models/>
2. <https://neptune.ai/blog/select-model-for-time-series-prediction-task>
3. <https://statisticsbyjim.com/time-series/autocorrelation-partial-autocorrelation/>
4. <https://online.stat.psu.edu/stat462/node/188/>
5. <https://towardsdatascience.com/interpreting-acf-and-pacf-plots-for-time-series-forecasting-af0d6db4061c>
6. <https://www.linkedin.com/pulse/time-series-analysis-short-introduction-/?trk=pulse-article>
7. <https://machinelearningmastery.com/gentle-introduction-autocorrelation-partial-autocorrelation/>
8. <https://otexts.com/fpp2/AR.html>
9. <https://rpubs.com/JSHAH/481706>