Autocorrelation

Autocorrelation is the correlation between observations that are *n* time periods apart. It measures the relationship between *lagged values* of a time series, just as Pearson's correlation measures the degree of a linear relationship between two variables. Autocorrelation function is used in econometric modeling to determine stationarity and seasonality.

How To

- ✓ Run the **STATISTICS->TIME SERIES ->AUTOCORRELATION AND PARTIAL AUTOCORRELATION** command.

- ✓ Select a variable containing a time series x_i.

- ✓ Enter the value for the lag to the **LAG LENGTH** field. The magnitude of the time lag determines the order of autocorrelation coefficient.

- ✓ To plot a correlogram check the **PLOT ACF** option. The **PLOT PARTIAL CORRELATIONS** option adds partial correlogram to the report.

- ✓ Check the **REMOVE MEAN** option to prepare time series with removed mean.

- ✓ Use the **COMPUTE DIFFERENCED SERIES** option to apply the differencing operator to the time series. *Differencing can help stabilize the mean of a time series by removing changes in the level of a time series, and so eliminating trend and seasonality* [HYN].

    - o To apply the operator more than once – change the **DIFFERENCES OF ORDER (REPEAT N TIMES)** option value.
    - o Optionally, change the **DIFFERENCING LAG** value (default value is 1).

- ✓ Optionally, select the standard error calculation algorithm. There are two ways to calculate the standard error of the sample autocorrelation:

    - o By default we assume that the true MA order of the process is *k-1*, so the variance of r_k is
      $$SE(r_k) \cong \sqrt{\frac{1}{n}\left(1 + 2\sum_{j=1}^{k-1} r_j\right)}.$$

    - o If we assume that the process is white noise (check the **WHITE NOISE STANDARD ERROR** option), the standard error is approximated by the square root (Box and Jenkins, [BOX]):
      $$SE(r_k) \cong \sqrt{\frac{1}{n}\left(\frac{n-k}{n+2}\right)}.$$

Assumptions

The time series observations are equally spaced.

Results

The report includes the overall mean, variance, and a table, showing following statistics for each lag value: autocorrelation coefficient r_k, lower and upper confidence limits for r_k, standard error, partial R, Box-Ljung Q.

Autocorrelation

Overall mean	81.60702					
Variance	93,309.08767					
Partial R Standard Error	0.13245					
Lag	Correlation	LCL	UCL	Standard Error	Partial R	Box-Ljung Q
1	-0.20543	-0.27212	-0.13678	0.13245	-0.20543	2.53436
2	0.17318	0.10381	0.24087	0.13793	0.13675	4.36820
3	-0.26770	-0.33206	-0.20086	0.14169	-0.22226	8.83125

OVERALL MEAN – the mean of all observations in the time series.

VARIANCE – the variance of the time series.

PACF (PARTIAL R) STANDARD ERROR – the standard error of the estimated partial autocorrelation function (PACF) φ_k, calculated as $SE(\varphi_k) \cong 1/\sqrt{n}$.

Lags table

CORRELATION r_k – k^{th} lag autocorrelation. For example, r_1 measures the relationship between x_t and x_{t-1}, r_2 measures the relationship between x_t and x_{t-2}, and so on. For a discrete process with known mean and variance an estimate of the autocorrelation may be obtained as $r_k = \frac{1}{(n-k)\partial^2}\sum_{t=1}^{n-k}(x_t - \mu)(x_{t+k} - \mu)$, where $(x_1, x_2, ...x_n)$ are observations, μ - mean.

LCL, UCL are the lower and upper confidence limits for r_k at selected alpha.

R STANDARD ERROR – standard error for r_k.

PARTIAL R (PACF) – estimated partial autocorrelation φ_k. Partial autocorrelation is the correlation between a time series and its lags with the effects of lower order lags held constant, and so it further removes the linear ties between the lagged series. Partial autocorrelation function (PACF) is calculated using the Durbin-Levinson algorithm [QEN]: $\varphi_{k+1,k+1} = \frac{r_{k+1}-\sum_{j=1}^{k}\varphi_{k,j}r_{k-j}}{1-\sum_{j=1}^{k}\varphi_{k,j}r_j}$.

PACF can reveal the presence of the autoregressive process in the time series.

BOX-LJUNG Q – a measure of autocorrelation. Box-Ljung Q statistic is chi-squared distributed with degrees of freedom k-p-q, where p and q are autoregressive and moving average orders, respectively.

At lag k, the Box-Ljung statistic is defined as: $Q_k = n(n + 2)\sum_{k=1}^{h}\frac{\rho_k^2}{n-k}$.
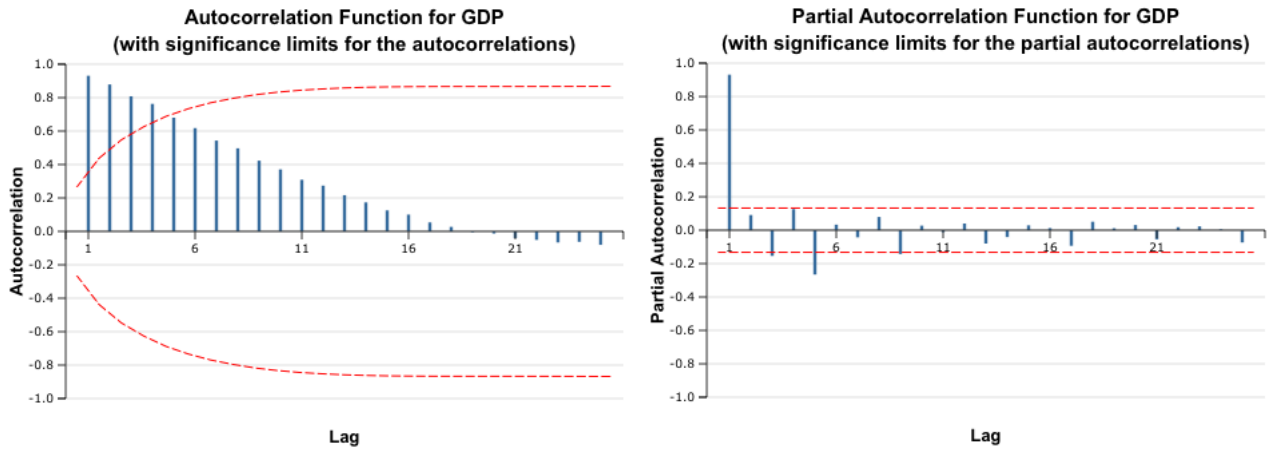
Correlograms

Correlograms (autocorrelation and partial autocorrelation plots) are useful for preliminary identification of an ARIMA model.

- Compare the sample correlogram to the theoretical correlogram for a stationary process.
- If the time series is stationary – the ACF declines to zero almost immediately.

- If the plot looks non-stationary, try to identify the trend and apply differencing to remove it. If the series is not seasonally adjusted, it may need special treatment.
- Positive autocorrelations are often found among the economic times series due to the persistence associated with business cycles and periods of expansionor recession.
- Moving average processes have the ACF with spikes for the first lags and the PACF shows exponential decay. The number of spikes indicates the order of the moving average.

ACF shows exponential decay	Time series is stationary. Autoregressive model.
Long runs of positive autocorrelations	Apply extra differencing.
ACF slowly declines, there are significant values at large lags or noteworthy spikes	Include further moving average terms. Moving average model with a trend.
Lag 1 is coming to the value of -0.5	Could be the result of overdifferencing. Decrease the order of differencing. Autoregressive model.
The PACF correlogram has noteworthy spikes	Include extra autoregressive terms.

Autocorrelation Function for GDP
(with significance limits for the autocorrelations)

Partial Autocorrelation Function for GDP
(with significance limits for the partial autocorrelations)

*The correlograms show that the autocorrelation function for the **GDP** time series remains significant for the first five lags. Even without additional indicators, we can conclude that the given time series is non-stationary. However, later we will see that the first order differencing restores stationarity.*
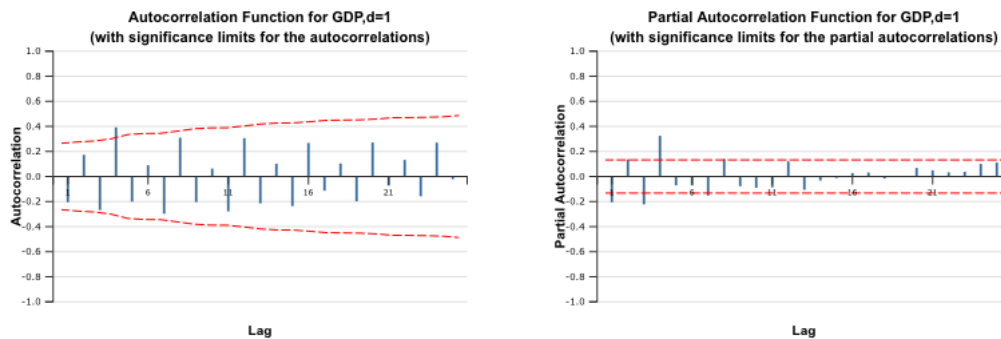
Example

1. Open the **TIME SERIES - AUTOCORRELATION** dataset**.**
2. **GDP** variable values are the Estonia gross domestic product (GDP) in millions of euros for 2001 – 2015 years (on a quarterly basis).

3. Run the **STATISTICS->TIME SERIES->AUTOCORRELATION** command.

4. Select the **GDP** variable as time series, enter **25** (half the number of observations) as lags count. Check the **PLOT ACF** and **PLOT PARTIAL** options and leave default values for other options.

Check the **DIFFERENCE LAG VALUES** option and leave the **DIFFERENCES OF ORDER** field equal to 1, to get first order differences.

4. Run the command.

ACF and PACF plots for differenced (d=1) GDP time series

5. The plots above show that the ACF for the GDP remains significant and high, fluctuating about zero because the GDP has a trend due to its economic nature. As the ACF correlogram shows alternating positive and negative values (an indicator of a stationary series), we can assume that the differenced time series is stationary and use it for further modeling. The decay is not exponential, so it is recommended to run tests for stationarity.

REFERENCES

[BAR] Bartlett, M. S. 1946. On the theoretical specification of sampling properties of autocorrelated time series. Journal of Royal Statistical Society, Series B, 8: 27.

[BOX] Box, G. E. P., and Jenkins, G. M. 1976. Time series analysis: Forecasting and control. San Francisco: Holden-Day.

[QEN] Time series analysis. Boston: Duxbury Press. Quenouville, M. H. 1949. Approximate tests of correlation in time series. Journal of the Royal Statistical Society, Series B, 11: 68.

[HYN] Hyndman R., Athanasopoulos G. (2014) *Forecasting: principles and practice.* Published by Otexts.com. Available online at otexts.com/fpp/

[ORD] Principles of Business Forecasting, Keith Ord, Robert Fildes. South-Western College Publishing, New York, NY (2012)