Autocorrelation and Partial Autocorrelation

What Are Autocorrelation and Partial Autocorrelation?

Autocorrelation is the linear dependence of a variable with itself at two points in time. For stationary processes, autocorrelation between any two observations only depends on the time lag h between them. Define $Cov(y_t, y_{t-h}) = \gamma_h$. Lag-$h$ autocorrelation is given by

$$\rho_h = Corr(y_t, y_{t-h}) = \frac{\gamma_h}{\gamma_0}.$$

The denominator γ_0 is the lag 0 covariance, i.e., the unconditional variance of the process.

Correlation between two variables can result from a mutual linear dependence on other variables (confounding). *Partial autocorrelation* is the autocorrelation between y_t and y_{t-h} after removing any linear dependence on $y_1, y_2, \ldots, y_{t-h+1}$. The partial lag-$h$ autocorrelation is denoted $\phi_{h,h}$.

Theoretical ACF and PACF

The autocorrelation function (ACF) for a time series y_t, $t = 1,\ldots,N$, is the sequence ρ_h, $h = 1, 2,\ldots,N - 1$. The partial autocorrelation function (PACF) is the sequence $\phi_{h,h}$, $h = 1, 2,\ldots,N - 1$.

The theoretical ACF and PACF for the AR, MA, and ARMA conditional mean models are known, and quite different for each model. The differences in ACF and PACF among models are useful when selecting models. The following summarizes the ACF and PACF behavior for these models.

Conditional Mean Model	ACF	PACF
AR(p)	Tails off gradually	Cuts off after p lags
MA(q)	Cuts off after q lags	Tails off gradually
ARMA(p,q)	Tails off gradually	Tails off gradually

Sample ACF and PACF

Sample autocorrelation and sample partial autocorrelation are statistics that estimate the theoretical autocorrelation and partial autocorrelation. As a qualitative model selection tool, you can compare the sample ACF and PACF of your data against known theoretical autocorrelation functions [1].

For an observed series $y_1, y_2,\ldots,y_T$, denote the sample mean $\bar{y}$. The sample lag-h autocorrelation is given by

$$\hat{\rho}_h = \frac{\sum_{t=h+1}^{T}(y_t - \bar{y})(y_{t-h} - \bar{y})}{\sum_{t=1}^{T}(y_t - \bar{y})^2}.$$

The standard error for testing the significance of a single lag-h autocorrelation, $\hat{\rho}_h$, is approximately

$$SE_\rho = \sqrt{(1 + 2\sum_{i=1}^{h-1}\hat{\rho}_i^2)/N}.$$

When you use `autocorr` to plot the sample autocorrelation function (also known as the correlogram), approximate 95% confidence intervals are drawn at $\pm 2SE\rho$ by default. Optional input arguments let you modify the calculation of the confidence bounds.

The sample lag-*h* partial autocorrelation is the estimated lag-*h* coefficient in an AR model containing *h* lags, $\hat{\phi}_{h,h}$. The standard error for testing the significance of a single lag-*h* partial autocorrelation is approximately $1/\sqrt{N-1}$. When you use `parcorr` to plot the sample partial autocorrelation function, approximate 95% confidence intervals are drawn at $\pm 2/\sqrt{N-1}$ by default. Optional input arguments let you modify the calculation of the confidence bounds.

References

[1] Box, G. E. P., G. M. Jenkins, and G. C. Reinsel. *Time Series Analysis: Forecasting and Control*. 3rd ed. Englewood Cliffs, NJ: Prentice Hall, 1994.

See Also

`autocorr` | `parcorr`

Related Examples

- Detect Autocorrelation
- Detect ARCH Effects
- Box-Jenkins Model Selection

More About

- Ljung-Box Q-Test
- Autoregressive Model
- Moving Average Model
- Autoregressive Moving Average Model