Cross Validated

How does ACF & PACF identify the order of MA and AR terms?

Asked 4 years, 5 months ago Active 3 months ago Viewed 32k times

24

19

It's been more than 2 years that I am working on different time series. I have read on many articles that ACF is used to identify order of MA term, and PACF for AR. There is a thumb rule that for MA, the lag where ACF shuts off suddenly is the order of MA and similarly for PACF and AR.

Here is one of the articles I followed from PennState Eberly College of Science.

My Question is why is it so? For me even ACF can give AR term. I need explanation of thumb rule mentioned above. I am not able to understand thumb rule intuitively/mathematically that why -

Identification of an AR model is often best done with the PACF.
Identification of an MA model is often best done with the ACF rather than the PACF

Please note:- I don't need how but "WHY". :)

time-series arima autoregressive moving-average

Share Cite Edit Follow Flag

edited May 26 '17 at 0:58
Antoni Parellada
23.1k 15 96 194

asked May 25 '17 at 14:35
Arpit Sisodia
1,039 2 7 23

3 Answers

Active Oldest Votes

19

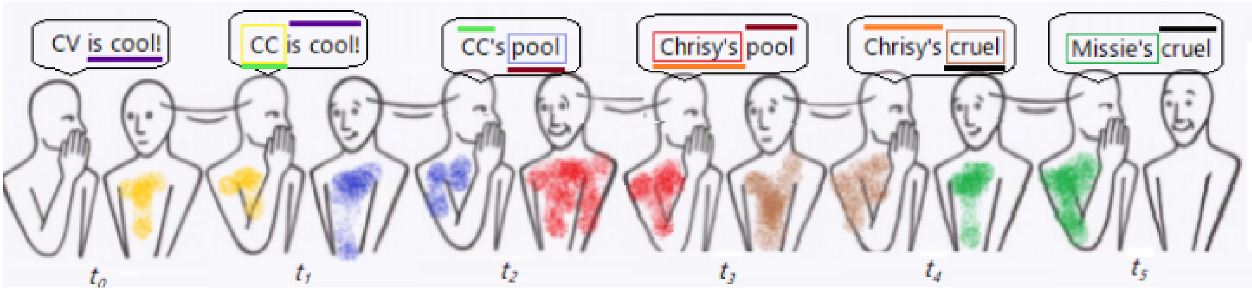
The quotes are from the link in the OP:

> **Identification of an AR model is often best done with the PACF.**
>
> For an AR model, the theoretical PACF "shuts off" past the order of the model. The phrase "shuts off" means that in theory the partial autocorrelations are equal to 0 beyond that point. Put another way, the number of non-zero partial autocorrelations gives the order of the AR model. By the "order of the model" we mean the most extreme lag of x that is used as a predictor.
>
> ... a k^{th} order autoregression, written as AR(k), is a multiple linear regression in which the value of the series at any time t is a (linear) function of the values at times $t-1, t-2, \ldots, t-k:$

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \cdots + \beta_2 y_{t-k} + \epsilon_t.$$

This equation looks like a regression model, as indicated on the linked paged... So what is a **possible** intuition...

In Chinese whispers or the telephone game as illustrated here

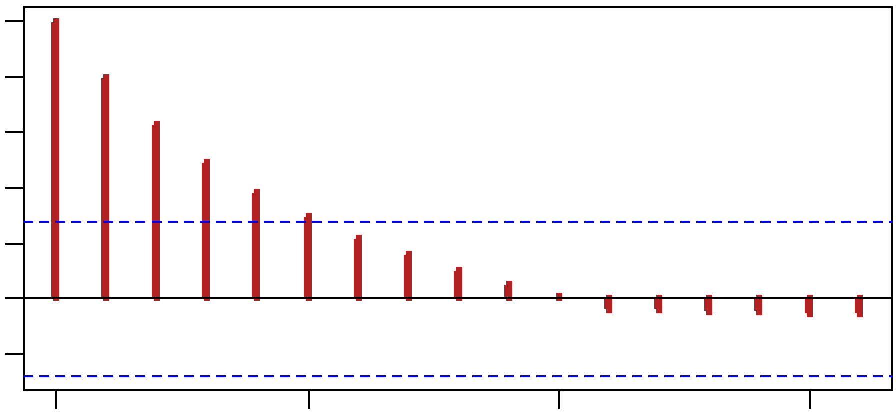


the message gets distorted as it is whispered from person to person, and the sentence is completely new after passing through two people. For instance, at time t_2 the message, i.e. " CC `'s pool` ", is completely different in meaning from that at t_o, i.e. " `CV is cool!` " The "*correlation*" that existed with t_1 ("CC is cool!") in the word "CC" is gone; there are no remaining identical words, and even the intonation ("!") has changed.

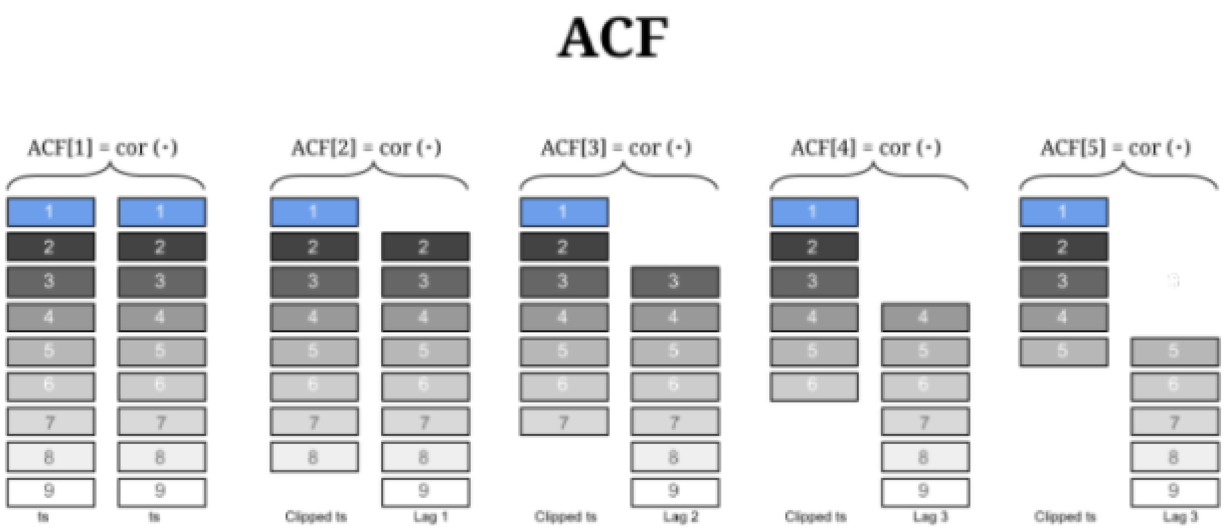
This pattern repeats itself: there is a word shared at any given two consecutive time stamps, which goes away if t_k is compared to t_{k-2}.

However, in this process of introducing errors at each step there is a similarity that spans further than just one single step: Although `Chrisy's pool` is different in meaning to `CC is cool!`, there is no denying their phonetic similarities or the rhyming of "pool" and "cool". Therefore it wouldn't be true that the correlation stops at t_{k-1}. It does decay (exponentially) but it can be traced downstream for a long time: compare t_5 (`Missi's cruel`) to t_0 (`CV is cool!`) - there are still similarities.

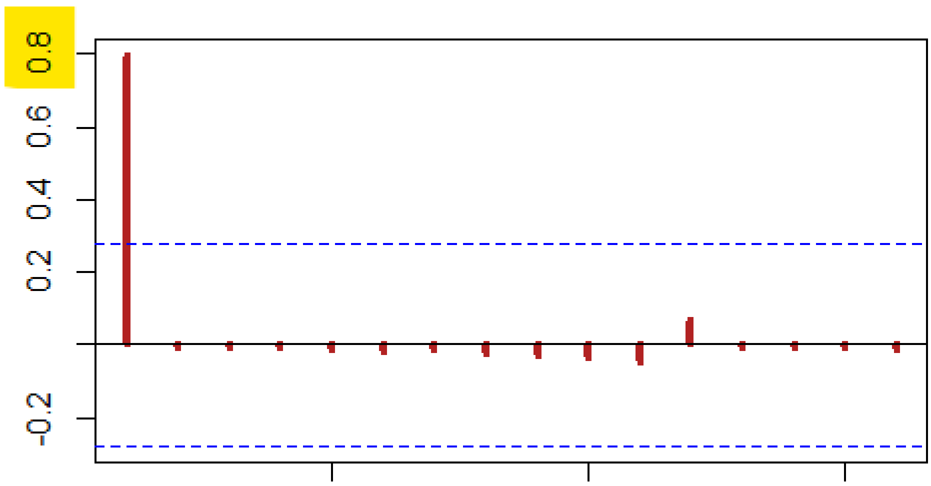
This explains the correlogram (ACF) in an AR(1) processes (e.g. with coefficient 0.8):



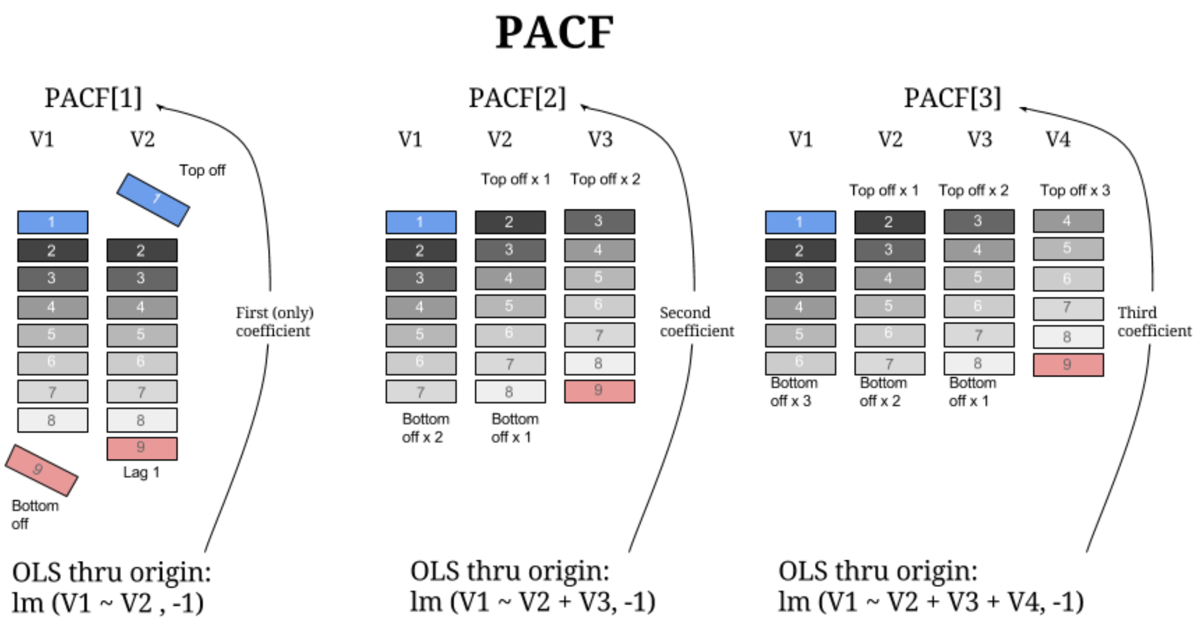
Multiple, progressively offset sequences are correlated, discarding any contribution of the intermediate steps. This would be the graph of the operations involved:



In this setting the PACF is useful in showing that once the effect of t_{k-1} is controlled for, older timestamps than t_{k-1} do not explain any of the remaining variance: all that remains is white noise:



It is not difficult to come very close to the actual output of the R function by actually obtaining consecutive OLS regressions through the origin of farther lagged sequences, and collecting the coefficients into a vector. Schematically,



Identification of an MA model is often best done with the ACF rather than the PACF.

For an MA model, the theoretical PACF does not shut off, but instead tapers toward 0 in some manner. A clearer pattern for an MA model is in the ACF. The ACF will have non-zero autocorrelations only at lags involved in the model.

A moving average term in a time series model is a past error (multiplied by a coefficient).

The q^{th}-order moving average model, denoted by MA(q) is

$$x_t = \mu + w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \cdots + \theta_q w_{t-q}$$

with $w_t \overset{\text{iid}}{\sim} N(0, \sigma_w^2)$.

It turns out that the behavior of the ACF and the PACF are flipped compared to AR processes:

Terms	ACF	PACF
AR	Geometric	p significant lags
MA	q significant lags	Geometric
ARMA	Geometric	Geometric

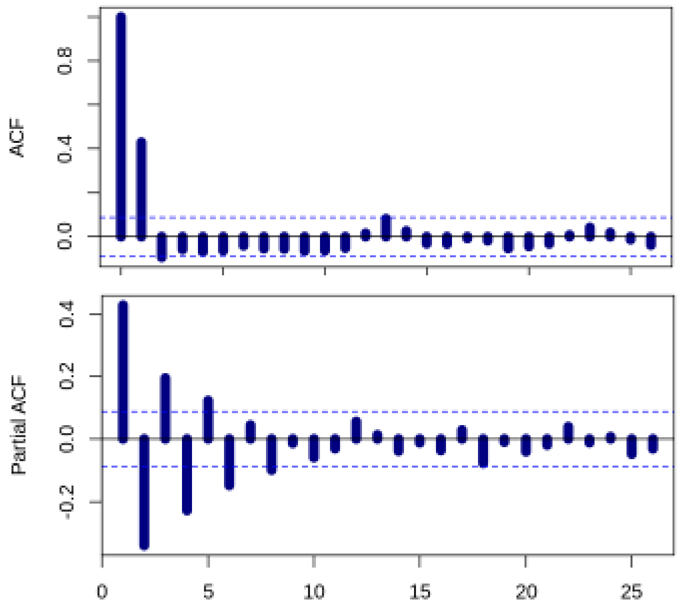
In the game above, t_{k-1} was enough to explain all prior errors in transmitting the message (single significan bar in PACF plot), absorbing all prior errors, which had shaped the final message one error at a time. An alternative view of that AR(1) process is as the addition of a long series of correlated mistakes (Koyck transformation), an MA(∞). Likewise, with some conditions, an MA(1) process can be inverted into an AR(∞) process.

conditions, an MA(1) process can be inverted into an AR(∞) process.

$$x_t = -\theta x_{t-1} - \theta^2 x_{t-2} - \theta^3 x_{t-3} + \cdots + \epsilon_t$$

The confusing part then is why the significant spikes in the ACF stop after the number of lags in MA(q). But in an MA(1) process the covariance is different from zero only at consecutive times $\text{Cov}(X_t, X_{t-1}) = \theta\sigma^2$, because only then the expansion $\text{Cov}(\epsilon_t + \theta\epsilon_{t-1}, \epsilon_{t-1} + \theta\epsilon_{t_2}) = \theta\text{Cov}(\epsilon_{t-1}, \epsilon_{t-1})$ will result in a match in time stamps - all other combinations will be zero due to iid condition.

This is the reason why the ACF plot is helpful in indicating the number of lags, as in this MA(1) process $\epsilon_t + 0.8\epsilon_{t-1}$, in which only one lag shows significant correlation, and the PACF shows typical oscillating values that progressively decay:



In the game of whispers, the error at t_2 (`pool`) is "correlated" with the value at t_3 (`Chrissy's pool`); however, there is no "correlation" between t_3 and the error at t_1 (`cc`).

Applying a PACF to a MA process will not result in "shut offs", but rather a progressive decay: controlling for the explanatory contribution of later random variables in the process does not render more distant ones insignificant as it was the case in AR processes.

Share Cite Edit Follow Flag

edited Jul 12 at 15:09

answered May 25 '17 at 19:51

Antoni Parellada

23.1k 15 96 194

This is a really cool answer, good job (+1) – Firebug Nov 15 '17 at 12:25

Rob Hyndman suggests this strategy for ARIMA, which uses both pacf and acf to determine the orders. Do we need to know beforehand what series we have to use the strategy described in your answer? thanks! – stucash May 8 '19 at 11:51

Please take my answer as a didactic exercise. I am no expert in the topic. – Antoni Parellada May 8 '19 at 11:53

6

Robert Nau from Duke's Fuqua School of Business gives a detailed and somewhat intuitive explanation of how ACF and PACF plots can be used to choose AR and MA orders here and here. I give a brief summary of his arguments below.

A simple explanation of why PACF identifies the AR order

The partial autocorrelations can be computed by fitting a sequence of AR models starting with the first lag only and progressively adding more lags. The coefficient of lag k in an AR(k) model gives the partial autocorrelation at lag k. Given this, if the partial autocorrelation "cuts off"/ceases to be significant at a certain lag (as seen in an ACF plot) this indicates that that lag does not add explanatory power to a model and therefore that the AR order should be the previous lag.

A more complete explanation which also addresses the use of ACF to identify the MA order

Time series can have AR or MA signatures:

- An AR signature corresponds to a PACF plot displaying a sharp cut-off and a more slowly decaying ACF;
- An MA signature corresponds to an ACF plot displaying a sharp cut-off and a PACF plot that decays more slowly.

AR signatures are often associated with positive autocorrelation at lag 1, suggesting that the series is slightly "underdifferenced" (this means that further differencing is necessary to completely eliminate autocorrelation). Since AR terms achieve partial differencing (see below), this can be fixed by adding an AR term to the model (hence the name of this signature). Therefore a PACF plot with a sharp cut-off (accompanied by a slowly decaying ACF plot with a positive first lag) can indicate the order of the AR term. Nau puts it as folows:

> If the PACF of the differenced series displays a sharp cutoff and/or the lag-1 autocorrelation is positive--i.e., if the series appears slightly "underdifferenced"--then consider adding an AR term to the model. The lag at which the PACF cuts off is the indicated number of AR terms.

MA signatures, on the other hand, are commonly associated with negative first lags, suggesting that the series is "overdifferenced" (i.e. it is necessary to partially cancel out the differencing to obtain a stationary series). Since MA terms can cancel an order of differencing (see below), the ACF plot of a series with an MA signature indicates the necessary MA order:

> If the ACF of the differenced series displays a sharp cutoff and/or the lag-1 autocorrelation is negative--i.e., if the series appears slightly "overdifferenced"--then consider adding an MA term to the model. The lag at which the ACF cuts off is the indicated number of MA terms.

Why AR terms achieve partial differencing and MA terms partially cancel previous differencing

Take a basic ARIMA(1,1,1) model, presented without the constant for simplicity:

$$y_t = Y_t - Y_{t-1}$$

$$y_t = \phi y_{t-1} + e_t - \theta e_{t-1}$$

Defining B as the lag/backshift operator, this can be written as follows:

$$y_t = (1 - B)Y_t$$

$$y_t = \phi B y_t + e_t - \theta B e_t$$

which can be further simplified to give:

$$(1 - \phi B)y_t = (1 - \theta B)e_t$$

or equivalently:

$$(1 - \phi B)(1 - B)Y_t = (1 - \theta B)e_t.$$

We can see that the AR(1) term gave us the $(1 - \phi B)$ term, thus partially (if $\phi \in (0, 1)$) increasing the order of differencing. Moreover, if we manipulate B as a numeric variable (which we can do because it is a linear operator), we can see that the MA(1) term gave us the $(1 - \theta B)$ term, thus partially cancelling out the original differencing term—$(1 - B)$—in the left-hand-side.

Share Cite Edit Follow Flag

answered Aug 8 '18 at 11:20
Lino Ferreira
273 2 9

3

On a higher level, here is how to understand it. (If you need a more mathematical approach, I can gladly go after some of my notes on time series analysis)

ACF and PACF are theoretical statistical constructs just like an expected value or variance, but on different domains. The same way that Expected values come up when studying random variables, ACF and PACF come up when studying time series.

When studying random variables, there is the question of how to estimate their parameters, which is where the method of moments, MLE and other procedures and constructs come in, as well as inspecting the estimates, their standard errors and etc.

Inspecting the estimated ACF and PACF come is from the same idea, estimating the parameters of a random time series process. Get the idea?

If you think you need a more mathematically inclined answer, please let me know, and I'll try and see if I can craft something by the end of the day.

Share Cite Edit Follow Flag

answered May 25 '17 at 14:49
Guilherme Marthe
1,209 6 12