**Compare true and estimated ACFs and PACFs**

We can compare the true ACF and PACF for an ARMA model to the estimated ACF and PACF obtained from an observed time series. Values and patterns that tend to match suggest a particular model to use for the data.

We have the following for some ARMA models:

|  |  |  |
| --- | --- | --- |
| AR(1):xt=ϕ1xt-1+wt | ρ(h) = |  |
| ϕhh = |  |
| AR(2): xt=ϕ1xt-1+ϕ2xt-2+wt | ρ(h) = | ϕ1ρ(h-1)+ϕ2ρ(h-2) for h1 |
| ϕhh = | ϕ11=ρ(1)ϕ22=ϕ2ϕhh=0 for h3 |
| MA(1): xt = θ1wt-1 + wt | ρ(h) = |  |
| ϕhh = |  for h1  |

|  |  |  |
| --- | --- | --- |
| MA(2): xt = θ1wt-1 +θ2wt-2+wt | ρ(h) = |  |
| ϕhh = |  |

Remember: “+” signs are used in the moving average operator, θ(B). **Many textbooks use “-“ signs so be careful when you examine these book!!!!** Thus, ACFs and PACFs may be slightly different.

In summary,

|  |  |  |  |
| --- | --- | --- | --- |
|  | **AR(p)** | **MA(q)** | **ARMA(p,q)** |
| ACF | Tails off to 0 | Cuts off to 0 after lag q | Tails off to 0 after lag q |
| PACF | Cuts off to 0 after lag p | Tails off to 0 | Tails off to 0 after lag p |

Intuitive explanation for the ARMA(p,q) result:

Consider an ARMA(2,1) process.

* The AR(2) in the above table says the ACF tails off and the PACF cuts off to 0 after lag 2.
* The MA(1) in the above table says the ACF cuts off to 0 after lag 1 and the PACF tails off.

Putting this information together, we obtain:

* The ACF for the ARMA(2,1) process will look like the ACF from an AR(2) model after lag 1 (because the ACF for MA(1) cuts off to 0 after lag 1).
* The PACF for the ARMA(2,1) will look like the PACF from a MA(1) model after lag 2 (because the PACF for AR(2) cuts off to 0 after lag 2).
* One can think of overlaying the two ACF plots and the two PACF plots from both the AR(2) and MA(1) models.

Identifying ARMA models can be difficult. Often, it is hard to determine when the “tails off” part begins.

How can you use this information???

Examine the ACF and PACF of an observed time series. If it matches any of the above behaviors, choose the corresponding model for the observed time series.

For example, suppose a hypothesis test for Ho:ρ(1) = 0 vs. Ha:ρ(1) ≠ 0 rejects Ho for h = 1 and does not reject Ho for h ≥ 2. Also, the PACF tends to “tail off”. Then a MA(1) may be appropriate for the time series data.

Example: ACF and PACF plots of ARMA models for simulated data (ARMA\_sample.R).

I used arima.sim() to simulate observations from an ARMA process. The estimated ACF and PACF are calculated and then compared to the true ACF and PACF. For all of the simulated data sets,  and n = 1,000. The horizontal lines for the ACF and PACF plots are drawn at ±1.96n-1/2=±1.96/.



 Estimate True









Overlay of AR(1) (black) and MA(1) (red) ACFs and PACFs. Compare to the ARMA(1,1) ACF and PACF.











Below is my R code for the first set of plots. Note that I used the same seed each time for each simulated time series data set (usually, you would want to change it):

 list.ar <- c(0.8); list.ma <- c(0.5);

 plot.title <- "AR par = 0.8, MA par 0.5"

 #Simulate the data

 set.seed(1788)

 x <- arima.sim(model = list(ar = list.ar, ma = list.ma),

 n = 1000, rand.gen = rnorm, sd = 1)

 #Plot of the data

 dev.new(width = 8, height = 6, pointsize = 10)

 plot(x = x, ylab = expression(x[t]), xlab = "t", type =

 "l", col = c("red"), main = paste("ARMA model:",

 plot.title), panel.first=grid(col = "gray", lty =

 "dotted"))

 points(x = x, pch = 20, col = "blue")

 #Estimated ACF and PACF

 dev.new(width = 8, height = 6, pointsize = 10)

 par(mfcol = c(2,2))

 acf(x = x, type = "correlation", lag.max = 20, ylim = c(-

 1,1), xlab = "h", main = paste("ARMA model:",

 plot.title))

 abline(h = c(-0.5, 0.5), lty = "dotted", col = "gray")

 pacf(x = x, lag.max = 20, ylim = c(-1,1), xlab = "h",

 main = paste("ARMA model:", plot.title))

 abline(h = c(-0.5, 0.5), lty = "dotted", col = "gray")

 #Note: acf(x = x, type = "partial") also gives the PACF

 plot

 #True ACF and PACF

 plot(y = ARMAacf(ar = list.ar, ma = list.ma, lag.max =

 20), x = 0:20, type = "h", ylim = c(-1,1), xlab =

 "h", ylab = expression(rho(h)), main = paste("ARMA

 model:", plot.title))

 abline(h = 0)

 abline(h = c(-0.5, 0.5), lty = "dotted", col = "gray")

 plot(x = ARMAacf(ar = list.ar, ma = list.ma, lag.max =

 20, pacf = TRUE), type = "h", ylim = c(-1,1), xlab =

 "h", ylab = expression(phi1[hh]), main = paste("ARMA

 model:", plot.title))

 abline(h = 0)

 abline(h = c(-0.5, 0.5), lty = "dotted", col = "gray")

Notes:

* Run the program a few times on your own with different seeds, samples sizes, and ARMA models to see what changes occur. For example, see what happens with the agreement of the estimated and true ACFs and PACFs as the sample size decreases.
* It is not as easy to see where the “cuts off to 0 after lag…” occurs! We will use additional methods to help us decide if a model fits the data well.
* Examining the ACF and PACF is one of the first steps to the constructing an ARMA model for a time series data set. This examination allows one to pick a few different candidate models to examine more closely!
* On a test, I may give you some of the ACF and PACF plots above (without the model stated) and have you suggest a possible ARMA model.