PACF, in depth

Jared Fisher

Lecture 8a

Announcements

Announcements

- Project Checkpoint 3 is due tomorrow, Wednesday March 31, by 11:50pm MDT
- Homework 5 will be posted today and is due Wednesday April 7
- Midterm 2 will be held on April 15

Transition in Course Topics

Transition

- We have been discussing the theory of ARMA: given $\phi(B), \theta(B)$, what is X_t?
- We will now embark on the methodology/application: given X_t, what $\phi(B), \theta(B)$ make sense?

Big Picture

$$f(Y_t) = m_t + s_t + X_t$$

- Pursue stationarity: model trend and seasonality, stabilize variance
- Model stationary $\{X_t\}$ (e.g. ARMA model)
- What empirical tools have we discussed for doing this?

Empirical Tools

- Pursuing stationarity
  - Model trend: parametric, smoothing, etc.
  - Model seasonality: sinusoids, indicators, etc.
  - Remove trend/seasonality: differencing
  - Variance stabilization: log(), sqrt()
- Stationary process modeling: lots of theory! but short on empirical tools
  - Sample autocovariance, sample autocorrelation
  - Sample autocorrelation plot: correlogram from acf()

One Approach to Empirical Modeling

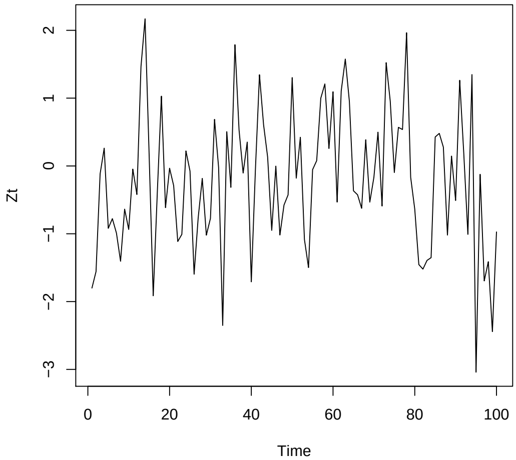
- Address trend/seasonality/variance to obtain a stationary process (i.e. residuals)
- Analyze stationary residuals to determine values of p and q to be used in ARMA(p,q) model.
- Once p and q are chosen, we can estimate the values of the ϕ, θ parameters.

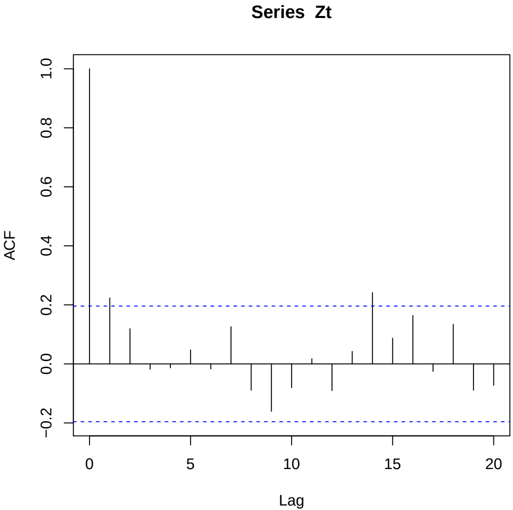
Sample ACF Plot

- Before last lecture, we had essentially only one tool for analyzing a stationary time series: the sample ACF plot/correlogram from R's acf(). Recall what it can/cannot show us:

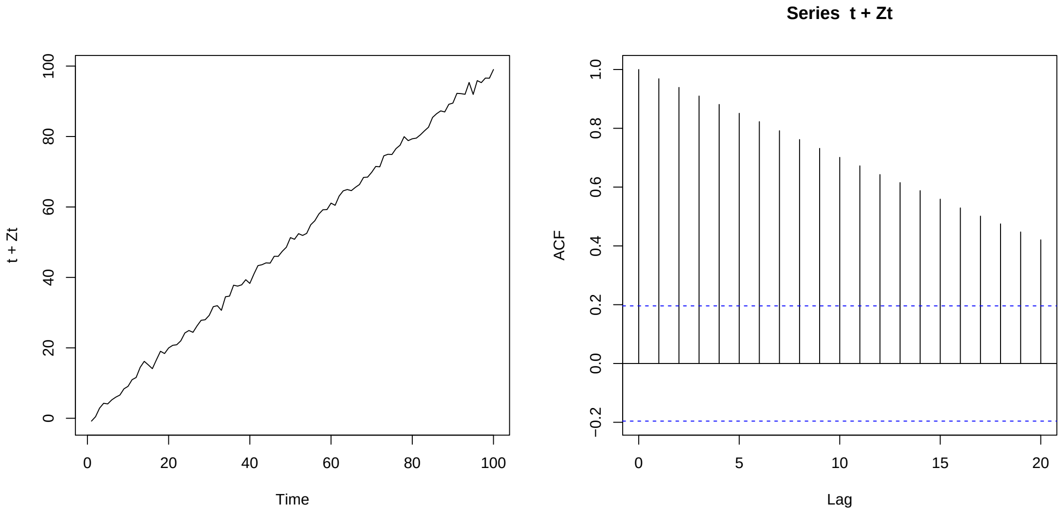
Y_t	Stationary?	Time series plot	Sample ACF
White Noise	Yes	"stable"	Zero's
Trend	No	Clearly not stable	Slow decay
Seasonality	No	Probably not stable	Periodic
MA(q)	Yes	"stable"	Zero's for $\|h\| > q$
AR(p)	Yes	"stable"	Usually exponential decay
ARMA(p,q)	Yes	"stable"	Exponential decay, and more

White noise


Zt
Time
0
20
40
60
80
100
−3
−2
−1
0
1
2

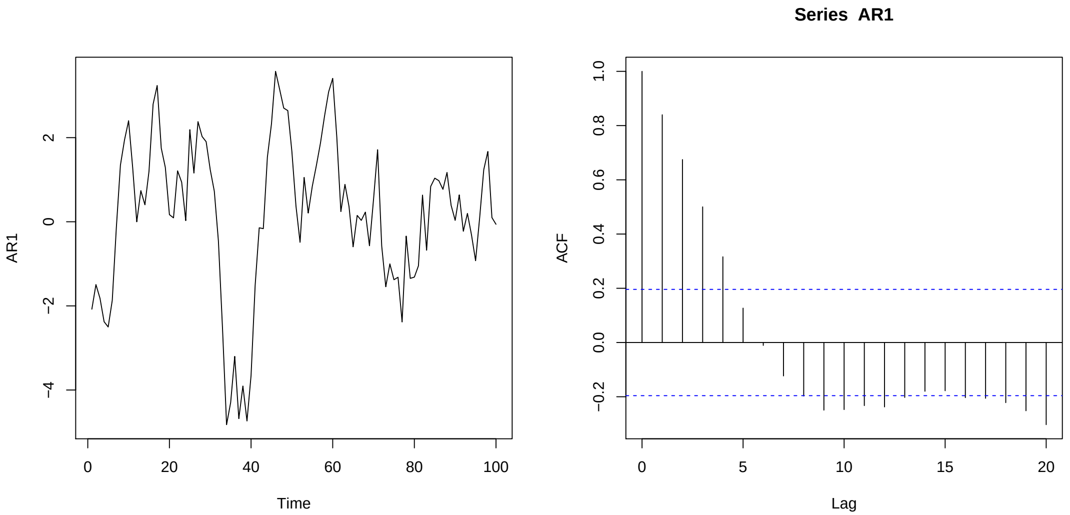

Series Zt
ACF
Lag
0
5
10
15
20
−0.2
0.0
0.2
0.4
0.6
0.8
1.0

Trend

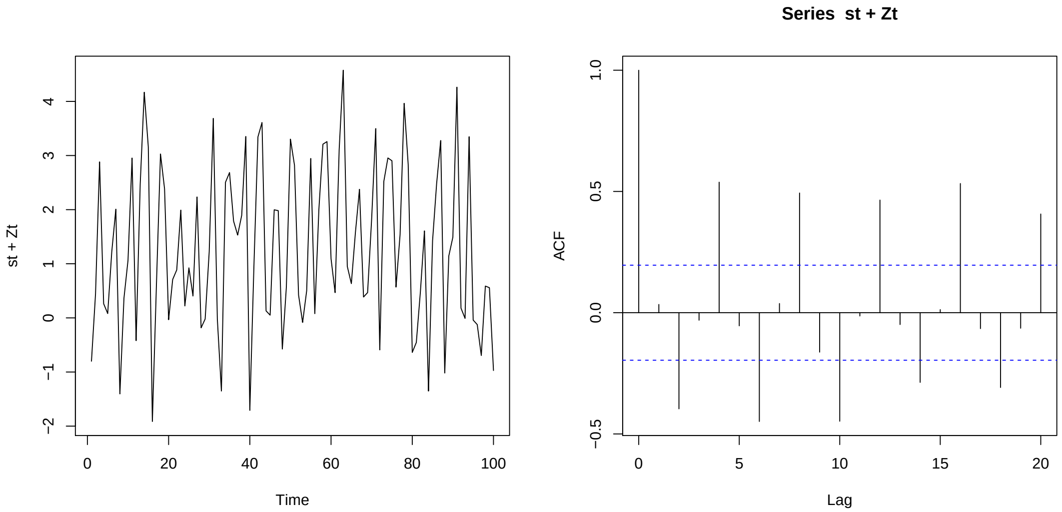


Series t + Zt
t + Zt
0
20
40
60
80
100
Time
ACF
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
Lag

AR(1), $\phi = 0.9$

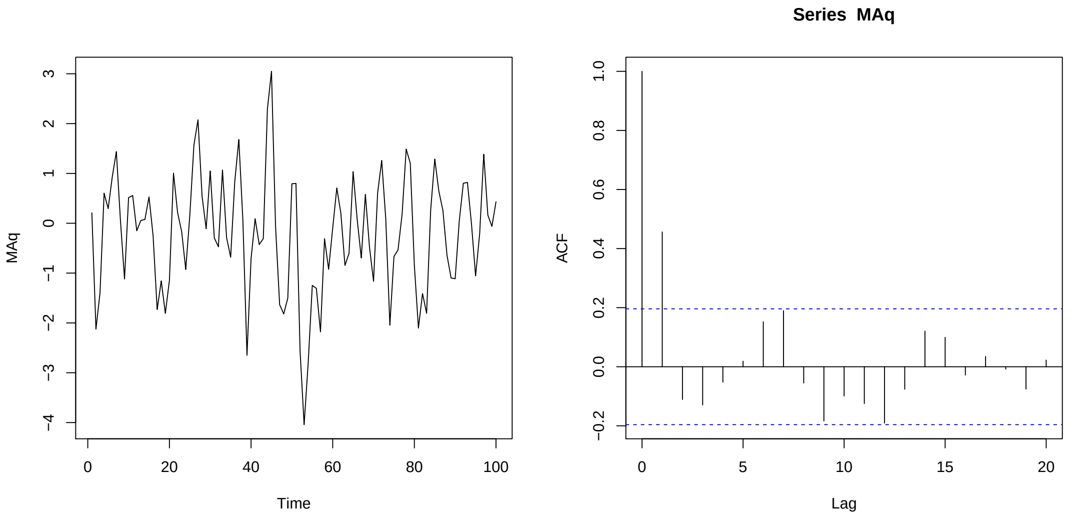


Seasonality

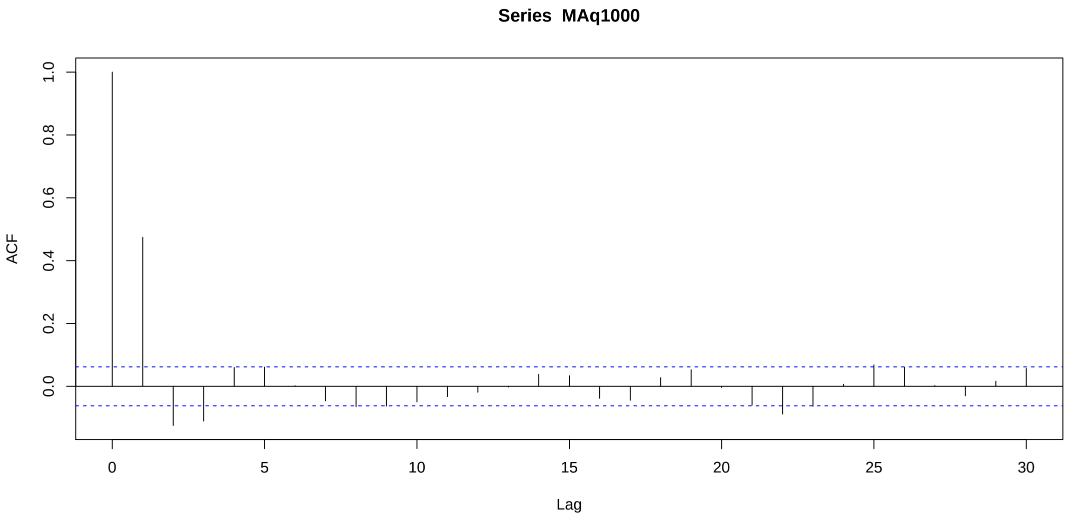


Note: next week we'll see Seasonal ARMA processes that have similar ACFs

MA(3)

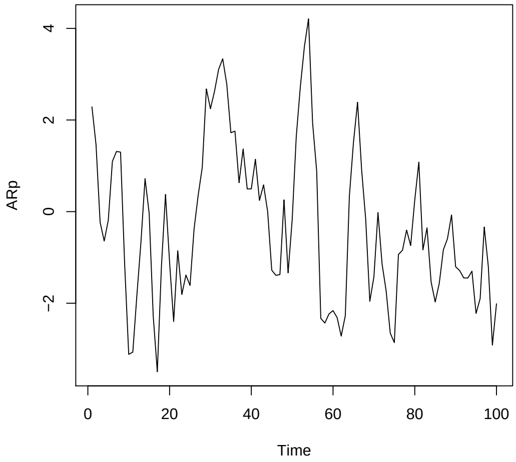


MA(3) increase sample size from n=100 to n=1000

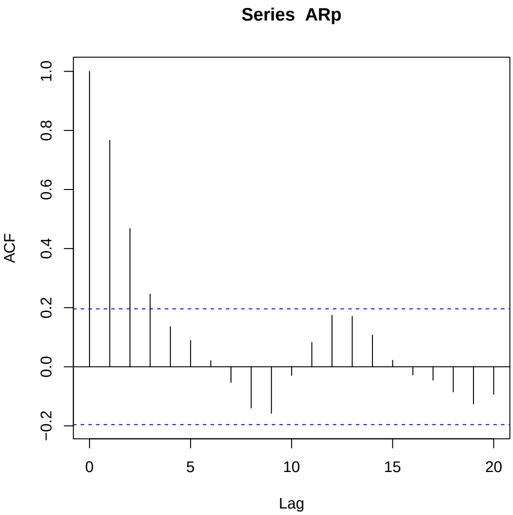


q=3 is visible!

AR(3)

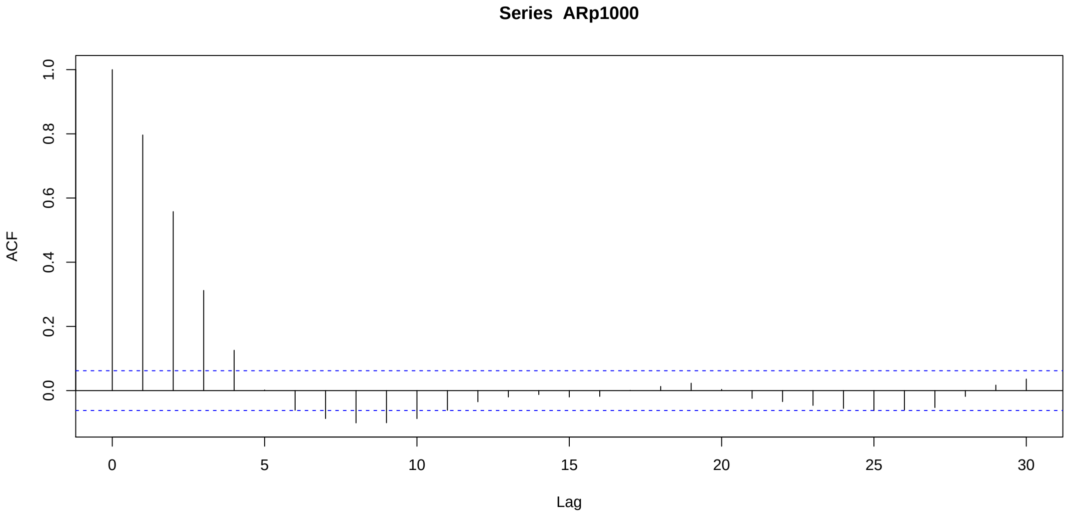


ARp
4
2
0
−2
0
20
40
60
80
100
Time



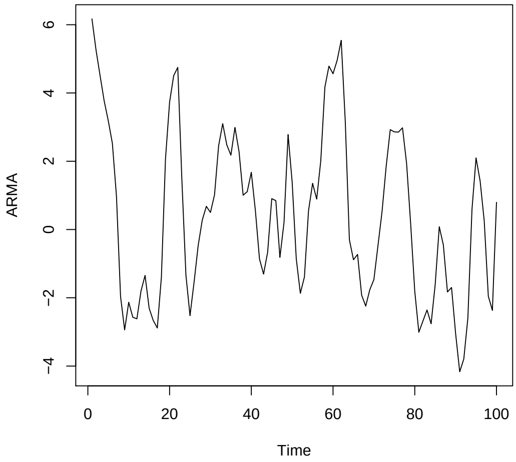
Series ARp
ACF
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0
5
10
15
20
Lag

AR(3), n=1000

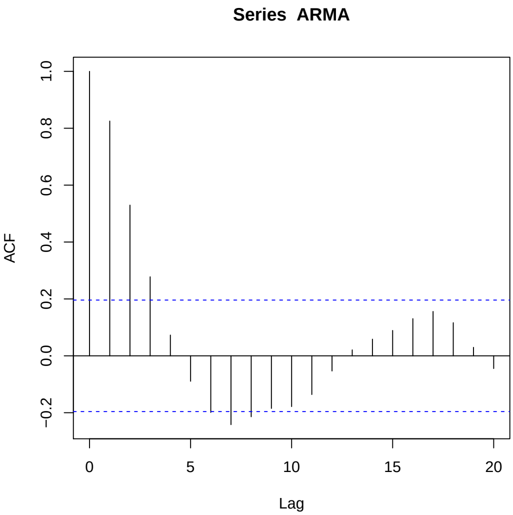


p=???

ARMA(3,3)

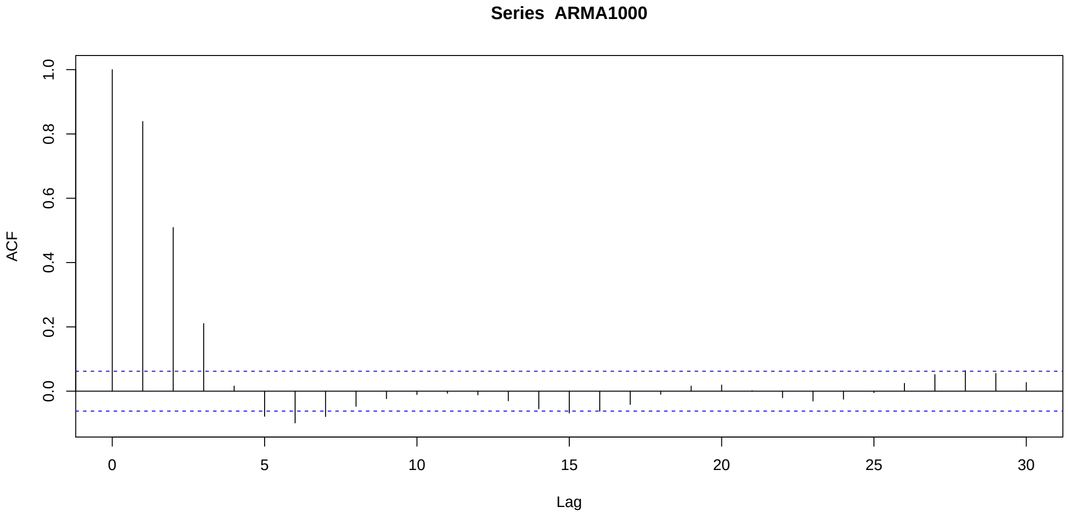


ARMA
6
4
2
0
−2
−4
0
20
40
60
80
100
Time



Series ARMA
ACF
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0
5
10
15
20
Lag

ARMA(3,3), n=1000



p=???, q=???

What's up with Autoregressive ACF's?

- Let p=1, such that: $X_t = \phi X_{t-1} + W_t$
- $$\begin{aligned}\gamma(1) &= cov(X_t, X_{t-1}) \\ &= cov(\phi X_{t-1} + W_t, X_{t-1}) \\ &= \phi\gamma(0)\end{aligned}$$
- $$\begin{aligned}\gamma(2) &= cov(X_t, X_{t-2}) \\ &= cov(\phi X_{t-1} + W_t, X_{t-2}) \\ &= cov(\phi^2 X_{t-2} + \phi W_{t-1} + W_t, X_{t-2}) \\ &= \phi^2\gamma(0)\end{aligned}$$
- Such that $\rho(1) = \phi$, $\rho(2) = \phi^2$, ... $\rho(j) = \phi^j$ (as we divide $\gamma(j)$ by $\gamma(0) = var(X_t)$)
- All are non-zero correlations, but p shouldn't be infinite...

What's up with Autoregressive ACF's?

- But! Note that X_t and X_{t-2} are only connected through their mutual relationship with X_{t-1}:

$$X_t = \phi X_{t-1} + W_t$$
$$X_{t-1} = \phi X_{t-2} + W_{t-1}$$

- What if we could measure the correlation between X_t and X_{t-2} while accounting for X_{t-1}?

How to diagnose p for AR's?

- We'll need a new tool to diagnose p
- For this tool we'll need to review prediction

Recap: Prediction

Theorem - Best Prediction

Let $Y, W_1, \ldots, W_n$ be random variables. Then for the best **mean squared error** prediction $f^\star(W_1, \ldots, W_n)$ of Y, that is

$$E(Y - f^\star(W_1, \ldots, W_n))^2 = \min_f E(Y - f(W_1, \ldots, W_n))^2,$$

it holds that

$$f^*(W_1, \ldots, W_n) := E\left(Y | W_1, \ldots, W_n\right).$$

Best (Linear?) Prediction

- Problem: in general, we'd need to know the entire joint distribution of $Y, W_1, \ldots, W_n$ in order to compute it.
- On the other hand, it is much easier to compute the best **linear** prediction of Y in terms of $W_1, \ldots, W_n$.
- Assume that W_i and Y all have finite second moments
- Let Δ denote the covariance matrix of $W = (W_1, \ldots, W_n)$
- Assume Δ invertible (this just excludes the situation that a linear combination of the W_i's has variance zero), that is

$$\Delta_{ij} = cov(W_i, W_j) \quad \text{and} \quad \zeta_i = cov(Y, W_i).$$

Thoerem: Best Linear Prediction ("BLP")

Let $Y, W_1, \ldots, W_n$ be zero mean random variables with finite second moments. Then for the best mean squared error linear prediction $a_1 W_1 + \ldots + a_n W_n$ of Y, that is

$$E(Y - (a_1^\star W_1 + \ldots + a_n^\star W_n))^2 = \min_a E(Y - (a_1 W_1 + \ldots + a_n W_n))^2,$$

it holds that

$$(a_1^\star, \ldots, a_n^\star)^\top = \Delta^{-1}\zeta.$$

Theorem: Characterization of the Best Linear Prediction (CBLP)

The best linear predictor $(a_1^\star, \dots, a_n^\star)^\top$ in the BLP Theorem is uniquely characterized by the property that

$$cov(Y - a_1 W_1 - \cdots - a_n W_n, W_i) = 0 \text{ for all } i = 1, \dots, n.$$

PACF : Partial Autocorrelation Function

Definition

Let $\{X_t\}$ be a mean zero stationary process. The **Partial Autocorrelation** at lag h, denoted by pacf(h) is defined as the coefficient of X_{t-h} in the best linear predictor for X_t in terms of $X_{t-1}, \ldots, X_{t-h}$.

h=1

- If h=1:

$$\begin{aligned} pacf(1) &= \Delta^{-1}\zeta \\ &= cov(X_{t-1}, X_{t-1})^{-1} cov(X_t, X_{t-1}) \\ &= \gamma(0)^{-1}\gamma(1) \\ &= \rho(1) \end{aligned}$$

- But $pacf(h)$ for $h > 1$ can be quite different from $\rho(h)$.

AR

Recall that for an AR(p) model we have $X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = W_t$ and hence by Theorem CBLP we immediately get the following theorem:

For the partial autocorrelation function of a causal AR(p) model $\phi(B)X_t = W_t$ it holds that $pacf(p) = \phi_p$ and $pacf(h) = 0$ for $h > p$.

Best Linear Prediction

- From the definition, it is not quite clear why this is called a correlation. We make this more clear in the following.
- pacf(h) is the correlation between X_t and X_{t-h} "with the linear effect of everything 'in the middle' removed" (TSA4e)
- corr$(X_t - \hat{X}_t \ , X_{t-h} - \hat{X}_{t-h})$, where $\hat{X}$'s are the best linear predictor of $X_{t-1}, \ldots, X_{t-h+1}$.

Best Linear Prediction

- Let $a_1 X_{t-1} + \cdots + a_{h-1} X_{t-h+1}$ denote the best linear predictor of X_t in terms of $X_{t-1}, \ldots, X_{t-h+1}$.
- By stationarity, the two sequences

  $$X_t, X_{t-1}, \ldots, X_{t-h+1}$$

  and

  $$X_{t-h}, X_{t-h+1}, \ldots, X_{t-1}$$

  have the same covariance matrix.
- Therefore, the best linear prediction of X_{t-h} in terms of $X_{t-h+1}, \ldots, X_{t-1}$ equals $a_1 X_{t-h+1} + \cdots + a_{h-1} X_{t-1}$.

In Other Words

- $$pacf(h) = \text{corr}(X_t - a_1 X_{t-1} - \cdots - a_{h-1} X_{t-h+1},$$
  $$X_{t-h} - a_1 X_{t-h+1} - \cdots - a_{h-1} X_{t-1}).$$
- $pacf(h)$ is the correlation between the errors in the best linear predictions of X_t and X_{t-h} in terms of the intervening variables $X_{t-1}, \ldots, X_{t-h+1}$.
- That is, the correlation between X_t and X_{t-h} with the effect of the intervening variables $X_{t-1}, X_{t-2}, \ldots, X_{t-h+1}$ removed.

h>p?

- *pacf*(h) equals zero for lags $h > p$ for an AR(p) model
- Note that for $h > p$, the best linear predictor for X_t in terms of $X_{t-1}, \ldots, X_{t-h+1}$ equals $\phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p}$.
- In other words, $a_1 = \phi_1, \ldots, a_p = \phi_p$ and $a_i = 0$ for $i > p$.
- Therefore for $h > p$, we have by causality

$$
\begin{aligned}
pacf(h) &= \text{corr}(X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p}, \\
&\qquad X_{t-h} - \phi_1 X_{t-h+1} - \cdots - \phi_p X_{t-h+p}) \\
&= \text{corr}\left(W_t, X_{t-h} - \phi_1 X_{t-h+1} - \cdots - \phi_p X_{t-h+p}\right) \\
&= 0.
\end{aligned}
$$

So does it work???

- Recall $\rho(2) = \phi^2$ for AR(1).
- We previously showed that pacf(2) $= 0$ for AR(1), (this is recap :)

Estimating the PACF with data

- Estimate the entries in Δ and ζ by the respective sample autocorrelations
- To choose p for AR(p), a natural approach is to plot the sample pacf.
- As the true pacf for an AR(p) model is zero for lags larger than p, the sample pacf should be close to zero for lags larger than p.
- Just as with Bartlett's Theorem for the autocorrelation function, one can quantify the variability of the pacf function precisely, as the following theorem shows

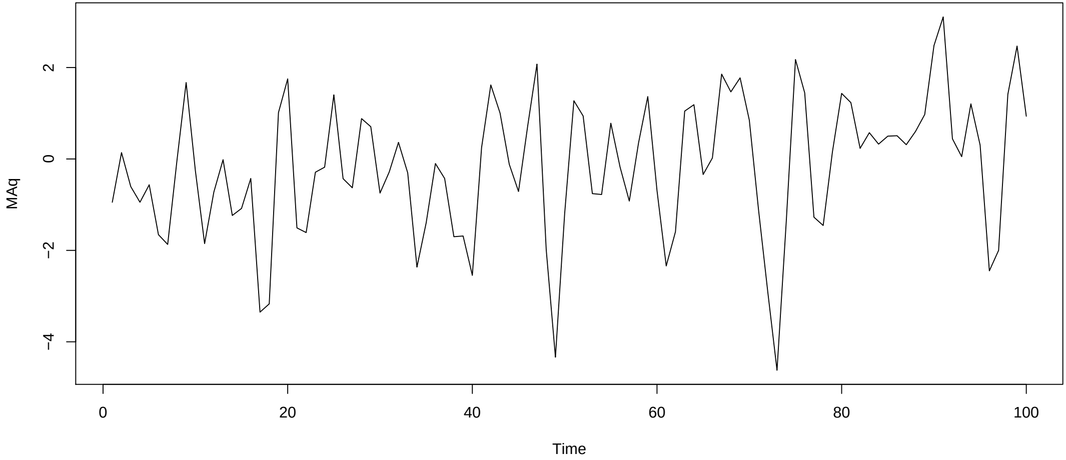
PACF Approximate Distribution

- Theorem: Let $\{X_t\}$ a causal AR(p) process with i.i.d. noise $\{W_t\}$. Let p_k denote the sample pacf at lag k defined above. Then for $k > p$ we have that the p_k's are approximately independent normally distributed with mean zero and variance $1/n$.
- Thus for $h > p$, the pacf() plot bands at $\pm 1.96 n^{-1/2}$ can be used for checking if an AR(p) model is appropriate.

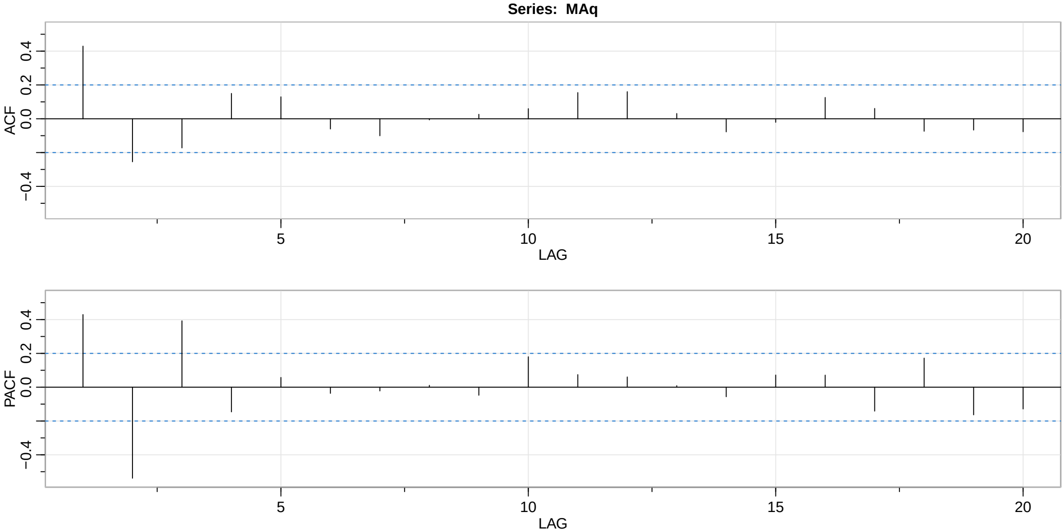
Returning to our example

MA(3)

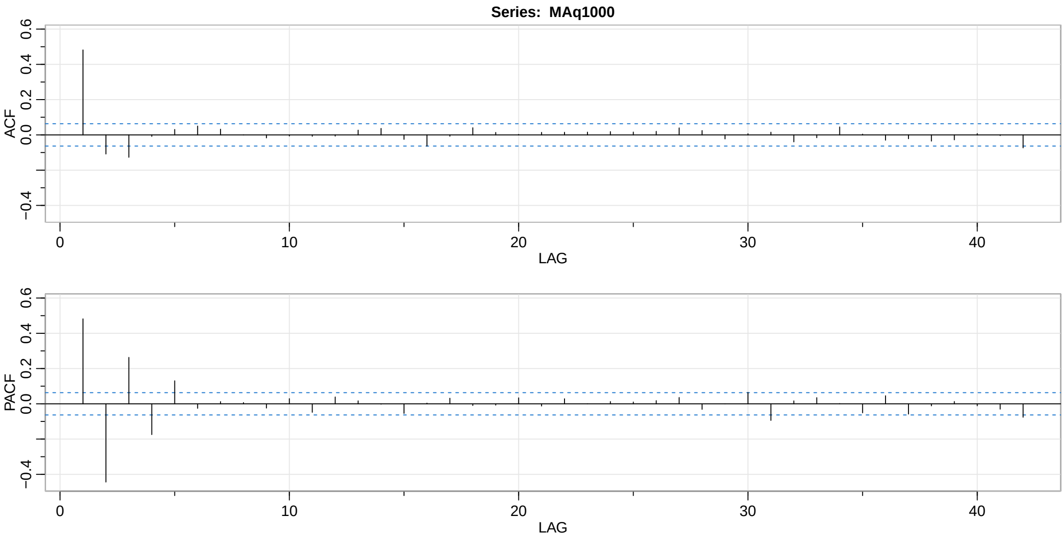
```
MAq = arima.sim(n=100,model=list(ma=c(.9,0,-.2)))
plot.ts(MAq)
```



MA(3), using acf2()



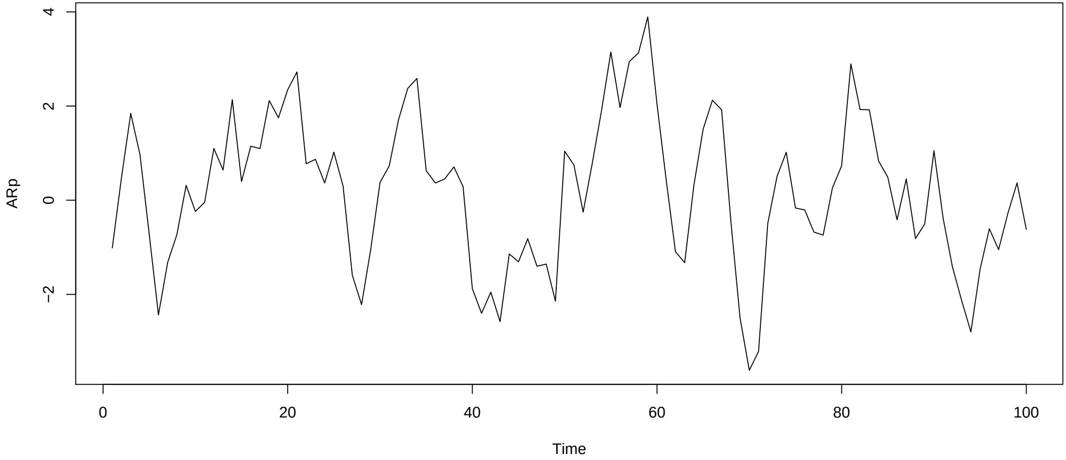
MA(3), n=1000



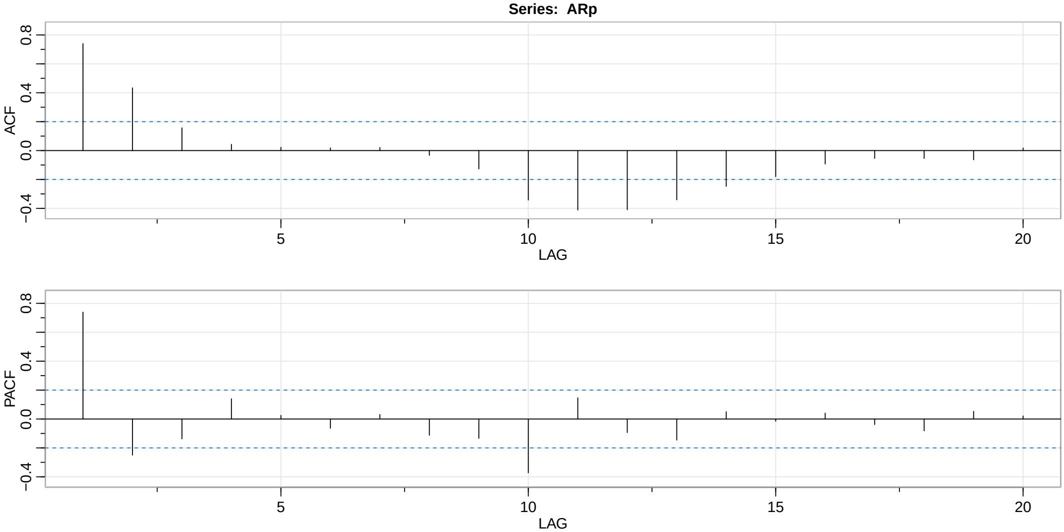
Series: MAq1000
ACF
−0.4
0.0
0.2
0.4
0.6
0
10
20
30
40
LAG
PACF
−0.4
0.0
0.2
0.4
0.6
0
10
20
30
40
LAG

AR(p)

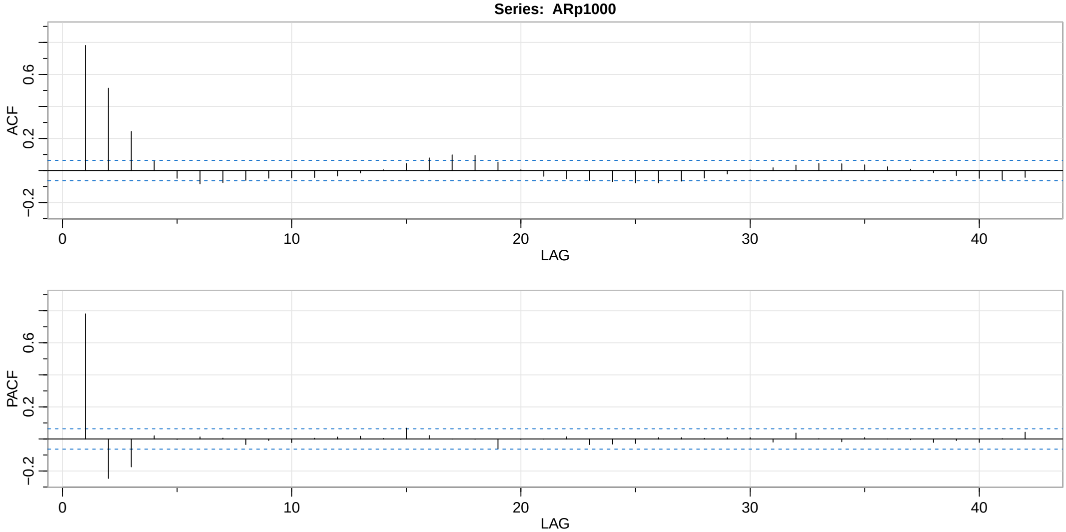
```r
ARp = arima.sim(n=100,model=list(ar=c(.9,0,-.2)))
plot.ts(ARp)
```



AR(p)


Series: ARp
ACF
0.8
0.4
0.0
−0.4
5
10
15
20
LAG
PACF
0.8
0.4
0.0
−0.4
5
10
15
20
LAG

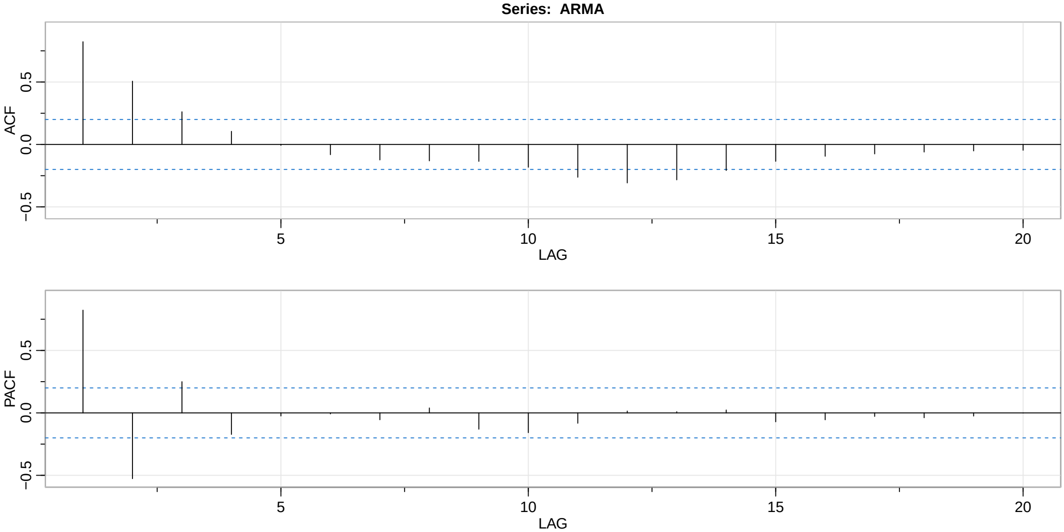
AR(p), n=1000



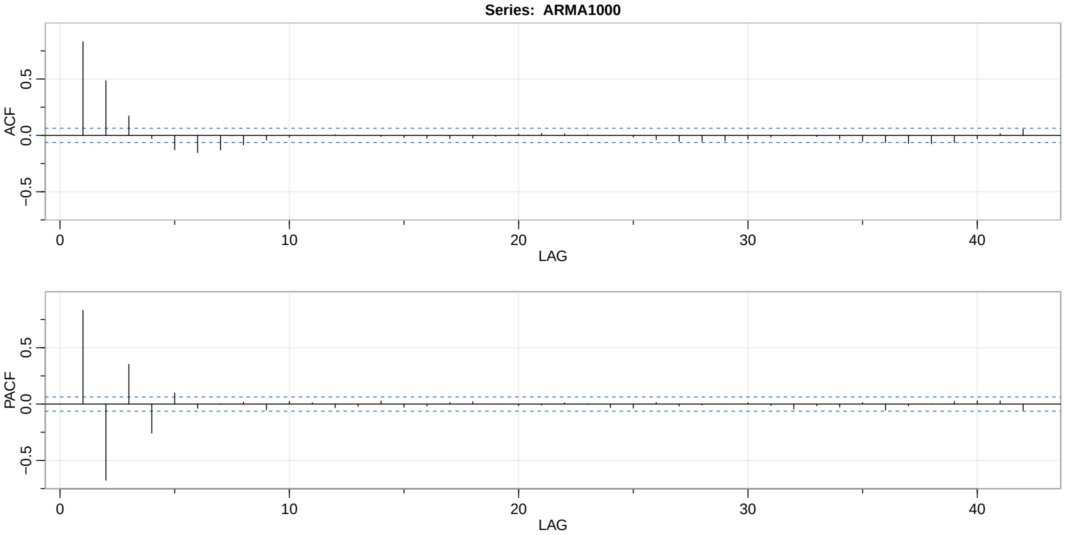
Series: ARp1000
ACF
0.6
0.2
−0.2
0
10
20
30
40
LAG
PACF
0.6
0.2
−0.2
0
10
20
30
40
LAG

p=3 (!)

ARMA(p,q)



ARMA(p,q), n=1000



p=???, q=???

Conclusion

- For an $MA(q)$ model,
  - The autocorrelation function $\rho_X(h)$ equals zero for $h > q$.
  - Also for $h > q$, the sample autocorrelation functions r_h are approximately normal with mean 0 and variance w_{hh}/n where $w_{hh} := 1 + 2\rho^2(1) + \cdots + 2\rho^2(q)$ by Bartlett's Theorem.
- For an $AR(p)$ model,
  - The partial autocorrelation function $pacf(h)$ equals zero for $h > p$
  - For $h > p$, the sample partial autocorrelations are approximately normal with mean 0 and variance $1/n$.

Conclusion

- If the sample ACF for a data set cuts off at some lag, we use an MA model (i.e. p=0). If the sample PACF cuts off at some lag, we use an AR model (i.e. q=0).
- In other words, if the sample ACF has no reasonable cutoff, then we have evidence that p>0 is reasonable.
- If the sample PACF has no reasonable cutoff, then we have evidence that q>0 is plausible.
- What if neither has a reasonable cutoff? Then we probably have p>0 and q>0, and in principle this is a model selection problem! To be continued.
- To the code!

More on Prediction

Prediction

- We've discussed prediction in general and prediction of a stationary process (X_t).
- However, you likely want to predict time series data (Y_t). Let's talk about that.
- Consider a smoother/filter estimate of the signal model, such that

$$Y_t = \left(\sum_{j=1}^{b} a_j Y_{t-j} \right) + X_t$$

- How can we forecast Y_{n+2} and Y_{n+1} given $Y_1, ..., Y_n$? We'll see that it's easiest to start with Y_{n+1}.

Prediction of Y_t

- Allowing for some shorter notation in the conditional:

$$\begin{aligned} E(Y_{n+1}|Y_{1,\ldots,n}) &= \left(\sum_{j=1}^{b} a_j E(Y_{n-j}|Y_{1,\ldots,n})\right) + E(X_{n+1}|Y_{1,\ldots,n}) \\ &= \sum_{j=1}^{b} a_j Y_{n-j} + E(X_{n+1}|Y_{1,\ldots,n}) \end{aligned}$$