ARMA models

Financial time series – lecture four

Outline

- ACF and PACF. Detailed accounts for ACF and PACF of ARMA models. Need to understand basic concepts and definitions of ACF and PACF. Know how to compute ACF for ARMA models.

- MLE and LS estimations for ARMA models. Understand the standard MLE and LS approach. Able to distinguish exact and conditional estimates. Able to write out the optimization form of these two approach for ARMA models.

- ARMA model prediction. Understand the general form of prediction for linear time series models. Able to conduct multiple step ahead forecast for AR, MA and simple ARMA models. Able to compute the prediction error and it's variance. (2.4.4., 2.5.4 and 2.6.4 of Tsay book)

The autoregressive model AR(p)

- First order AR

$$X_t = \phi_0 + \phi_1 X_{t-1} + w_t$$

where $w_t \sim WN(0, \sigma^2)$.

- Properties of AR(1) process:

$$\begin{aligned} E(X_t) &= \phi_0 \\ Var(X_t) &= \frac{\sigma^2}{1-\phi_1^2}\ if\ |\phi_1| < 1 \\ \rho(h) &= \phi_1^h \end{aligned}$$

- ACF examples of AR(1) processes.

```
ts.sim = arima.sim(list(order = c(1,0,0), ar = 0.8),
n = 200);
acf(ts.sim,lag.max=12);
```

AR(1) with positive ϕ_1.

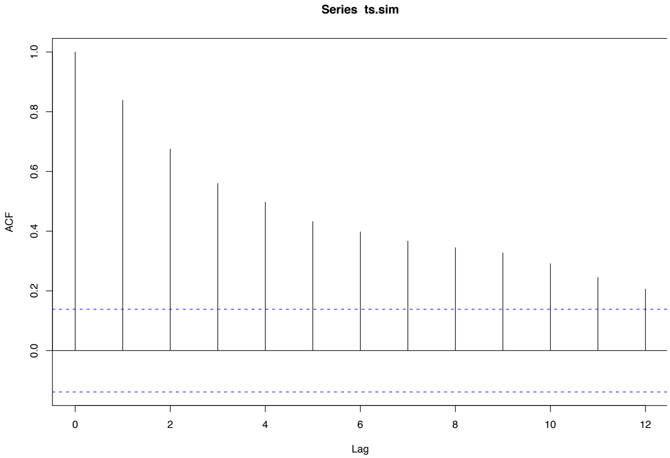
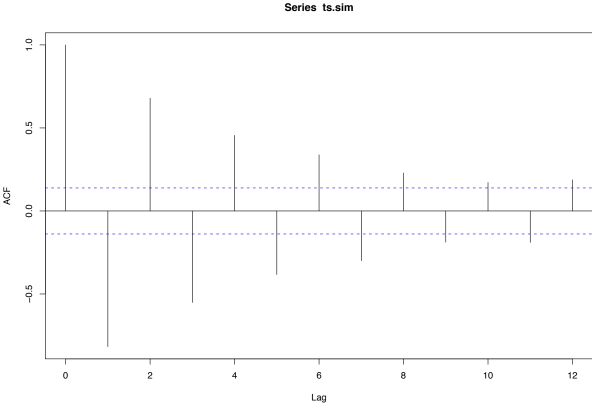


Figure: ACF when ϕ_1=0.8

AR(1) with negative ϕ_1.

```
ts.sim = arima.sim(list(order = c(1,0,0), ar = -0.8),
n = 200);
acf(ts.sim,lag.max=12);
```



AR(1) with small ϕ_1.

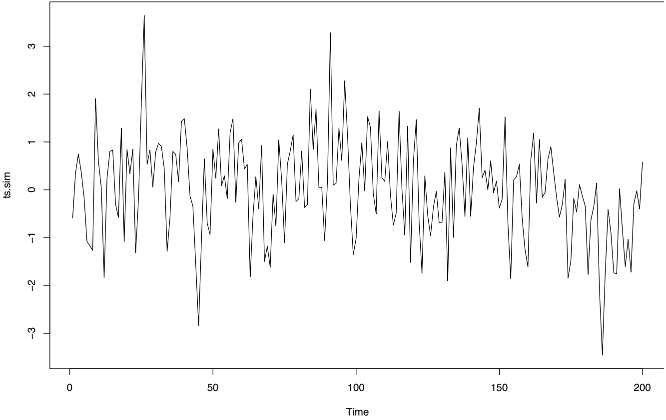


Figure: ACF when ϕ_1=0.2

AR(1) with large ϕ_1. What are the differences?

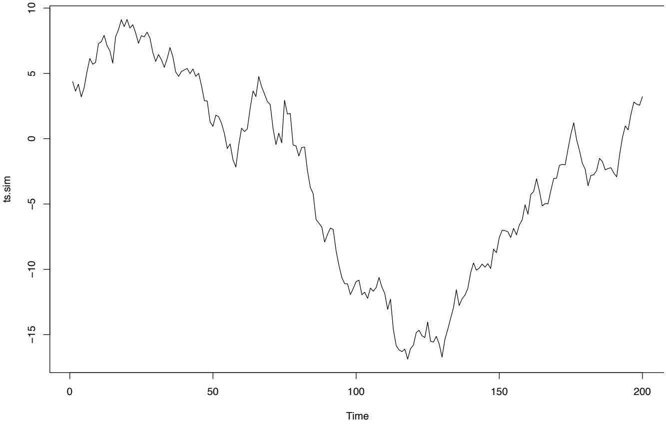


Figure: ACF when ϕ_1=0.99

Properties of AR(1)

- The stationary condition for AR(1) process is $|\phi_1| < 1$.

$$X_t = \phi_0 + \phi_1 X_{t-1} + w_t$$
$$Var(X_t) = \phi_1^2 Var(X_{t-1}) + \sigma^2$$

  If X_t is stationary,

$$(1 - \phi_1^2) Var(X_t) = \sigma^2,$$

  thus $|\phi_1| < 1$.

- Condition on the past, we have

$$E(X_t|X_{t-1}) = \phi_0 + \phi_1 X_{t-1}, \quad Var(X_t|X_{t-1}) = Var(w_t) = \sigma^2.$$

MA representation of AR(1)

- For simplicity, consider the mean zero series

$$\begin{aligned} X_t &= \phi_1 X_{t-1} + w_t \\ &= \phi_1(\phi_1 X_{t-2} + w_{t-1}) + w_t \\ &= \phi_1^2 X_{t-2} + \phi_1 w_{t-1} + w_t \\ &= \cdots \\ &= w_t + \phi_1 w_{t-1} + \phi_1^2 w_{t-2} + \cdots \end{aligned}$$

- The shock at $t-k$ has an exponentially decaying influence on X_t. The associated coefficients are called the ψ weights.

Higher order AR models

- The zero mean AR(p) model is

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + w_t,$$

  where w_t is uncorrelated with X_{t-j} for any j. Note that mean is zero here.

- Include the mean, the model is

$$X_t = \phi_0 + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + w_t,$$

  where $\phi_0 = \mu(1 - \phi_1 - \cdots - \phi_p)$, and $\mu = E(X_t)$.

- Stationarity condition: all roots of

$$1 - \phi_1 B - \cdots - \phi_p B^p = 0$$

  are outside the unit disc. B is the backshift operator.

AR(2) with complex roots

- Consider an AR(2) $X_t = 1.5X_{t-1} - 0.75X_{t-2} + w_t$, the roots of $1 - 1.5B + 0.75B^2$ are outside of unit circle. This will ensure a stationary AR(2) process.

```
ts.sim = arima.sim(list(ar = c(1.5,-0.75),
 n = 200);
acf(ts.sim);
```

- However, the roots are complex numbers. The acf plot will show damping sine and cosine waves, with the length of cycle

$$k = \frac{2\pi}{\cos^{-1}[\phi_1/2\sqrt{-\phi_2}]}.$$

- When some of the roots are complex numbers, the series will show pseudo-cyclic behavior. In this example, the cycle is 12.

ACF for an AR(2) with complex root

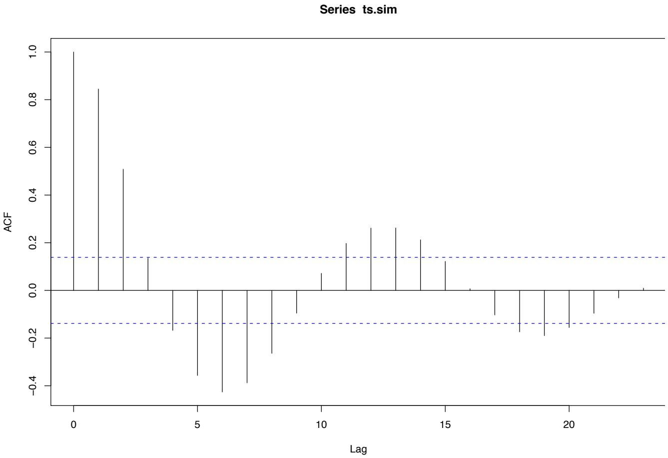


Figure: ACF when for AR(2) with complex root

The ACF of AR

- The ACF of a zero mean AR(p) process

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + w_t$$

$$\gamma(h) = \mathbb{E}\left(\sum_{i=1}^{p} \phi_i X_{t-i} + w_t\right) X_{t-h}.$$

- This can be further simplified to

$$\sum_{i=1}^{p} \phi_i \gamma(h-i) + COV(w_t, X_{t-h}).$$

- The $COV(w_t, X_{t-h}) = \sigma^2$ when $h = 0$, and 0 when $h > 0$.

The ACF of AR

- we have

$$\gamma(h) = \sum_{i=1}^{p} \phi_i \gamma(h-i), \quad h > 0 \text{ and}$$

$$\gamma(0) = \sum_{i=1}^{p} \phi_i \gamma(-i) + \sigma^2.$$

- For $h = 0, 1, 2, \cdots, p$, if we know the parameters $\phi_1, \cdots, \phi_p$ and σ^2, then we can solve for γ.
- Or if we know the γ, we will have $p+1$ linear equations to estimate ϕ and σ.

Operations with autoregressive operators

- Use the backshift operator B,

$$X_t = \phi_1 B X_t + \phi_2 B^2 X_t + \cdots + \phi_p B^p X_t + w_t$$

- We define the AR operator

$$\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$$

- Order p AR model is $\Phi(B) X_t = w_t$.
- Consider the first order model $X_t = \phi X_{t-1} + w_t$. As shown previously, when $|\phi| < 1$

$$X_t = \sum_{j=0}^{\infty} \phi^j w_{t-j}.$$

Operations with autoregressive operators

- This is now a MA(∞) process. It can be shown that $E(X_t) = 0$, $VAR(X_t) = \sigma^2/(1-\phi^2)$, $\gamma(h) = \sigma^2\phi^h/((1-\phi^2)$, and $\rho(h) = \phi^h$.
- Assuming that the inverse operation $\Phi^{-1}(B)$ exists, $\Phi(B)X_t = w_t$ implies

$$\Phi^{-1}(B)\Phi(B)X_t = \Phi^{-1}(B)w_t$$

thus

$$X_t = \Phi^{-1}(B)w_t.$$

This is similar to polynomial operations on B,

$$\Phi^{-1}(B) = 1 + \phi B + \phi^2 B^2 + \cdots + \phi^j B^j + \cdots .$$

Model forecast

- The underlying governing rule does not change
- Future behavior can be induced from the past
- We have sufficient record of the past to understand the underlying nature
- Assume we have the observed series $x_1, \cdots, x_n$, where x_n is the current state. Our **Goal** is to forecast h steps ahead. Namely, estimate the random variable X_{t+h}.
  Denote $X_n(h)$ as forecast of X_{n+h} made at time n.
  Examples: $X_n(1)$, $X_n(5)$, $X_{n+2}(2)$.
- What's a good forecast? One that has the smallest mean square error:

$$\text{Loss} = E(X_n(h) - X_{n+h})^2.$$

  The minimizer of the loss function is $X_n(h) = E(X_{n+h}|\mathcal{F}_n)$.
  Conditional expectation given past events.

Prediction for AR(1)

- Consider an AR(1) model with mean $E(X_t) = \mu$

$$X_t = \phi_0 + \phi_1 X_{t-1} + w_t,$$

where $w_t \sim N(0, \sigma^2)$ and $\phi_0 = (1 - \phi_1)\mu$.
Since $X_{n+1} = \phi_0 + \phi_1 X_n + w_{n+1}$, the one step ahead forecast is

$$\begin{aligned} X_n(1) &= E(X_{n+1}|x_1, \cdots, x_n) = E(\phi_0 + \phi_1 X_n + w_{n+1}|x_1, \cdots, x_n) \\ &= E(\phi_0 + \phi_1 x_n + w_{n+1}) = \phi_0 + \phi_1 x_n. \end{aligned}$$

Thus the estimated forecast is $\hat{\phi}_0 + \hat{\phi}_1 x_n$.

- The prediction error

$$U(1) = X_n(1) - X_{n+1} = \phi_0 + \phi_1 x_n - (\phi_0 + \phi_1 x_n + w_{n+1}) = -w_{n+1},$$

which follows a normal distribution with std σ and the 95% prediction interval is $(X_n(1) - 2\hat{\sigma}, X_n(1) + 2\hat{\sigma})$.

Prediction for AR(1)

- Two steps forecast.

  $X_{n+2} = \phi_0 + \phi_1 X_{n+1} + w_{n+2}$.

  $$\begin{aligned} X_n(2) &= E(X_{n+2}|x_1, \cdots, x_n) = E(\phi_0 + \phi_1 X_{n+1} + w_{n+2}|x_1, \cdots, x_n \\ &= \phi_0 + \phi_1 E(X_{n+1}|x_1, \cdots, x_n) \\ &= \phi_0 + \phi_1 X_n(1). \end{aligned}$$

  In general $X_n(h) = \phi_0 + \phi_1 X_n(h-1)$.

- The prediction error

  $$U(2) = X_n(2) - X_{n+2} = \phi_0 + \phi_1 X_n(1) - (\phi_0 + \phi_1 X_{n+1} + w_{n+2})$$

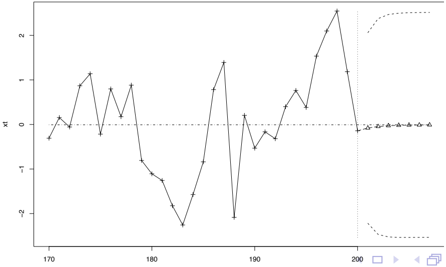
  which is

  $$-w_{n+2} + \phi_1(X_n(1) - X_{n+1}) = -w_{n+2} - \phi_1 w_{n+1} \sim N(0, (1+\phi_1^2)\sigma^2).$$

  Exercise. Show that

  $$U(h) \sim N(0, (1 + \phi_1^2 + \cdots + \phi_1^{2(h-1)}\sigma^2)$$

```
x=arima.sim(n = 200, list(ar = c(0.5)), sd = 1)
out=arima(x,c(1,0,0))
pp=predict(out,7)
print(t(cbind(pp$pred,pp$se)))
               [,1]       [,2]       [,3]       [,4]       [,5]
pp$pred -0.847106 -0.3918024 -0.2172091 -0.1502585 -0.1245852
pp$se    1.016656  1.0888409  1.0990556  1.1005497  1.1007692
                [,6]       [,7]
pp$pred -0.1147404 -0.1109652
pp$se    1.1008015  1.1008062
```



Prediction for AR(2) models

- The AR(2) model $X_t = \phi_0 + \phi_1 X_{t-1} + \phi_2 X_{t-2} + w_t$ where $\phi_0 = (1 - \phi_1 - \phi_2)\mu$ and $E(X_t) = \mu$.
- We have $X_n(1) = E(X_{n+1} | x_1, \cdots, x_n)$ $= E(\phi_0 + \phi_1 x_n + \phi_2 x_{n-1} + w_{n+1}) = \phi_0 + \phi_1 x_n + \phi_2 x_{n-1}$. Thus the estimate is $X_n(1) = \hat{\phi}_0 + \hat{\phi}_1 x_n + \hat{\phi}_2 x_{n-1}$.
- The prediction error is $-w_{n+1} \sim N(0, \sigma^2)$.
- Two step prediction. $X_{n+2} = \phi_0 + \phi_1 X_{n+1} + \phi_2 X_n + w_{n+2}$.

$$\begin{aligned}
X_n(2) &= E(X_{n+2} | x_1, \cdots, x_n) \\
&= E(\phi_0 + \phi_1 X_{n+1} + \phi_2 x_n + w_{n+2} | x_1, \cdots, x_n) \\
&= \phi_0 + \phi_1 E(X_{n+1} | x_1, \cdots, x_n) + \phi_2 x_n \\
&= \phi_0 + \phi_1 X_n(1) + \phi_2 x_n.
\end{aligned}$$

Prediction for AR(2) models

- In general $X_n(h) = \phi_0 + \phi_1 X_n(h-1) + \phi_2 X_n(h-2)$.
- The prediction error

  $$U(2) = X_n(2) - X_{n+2} = \phi_0 + \phi_1 X_n(1) + \phi_2 x_n - (\phi_0 + \phi_1 X_{n+1} + \phi_2 x_n + w_n$$

  which is

  $$-w_{n+2} + \phi_1 (X_n(1) - X_{n+1}) = -w_{n+2} - \phi_1 w_{n+1} \sim N(0, (1 + \phi_1^2)\sigma^2).$$

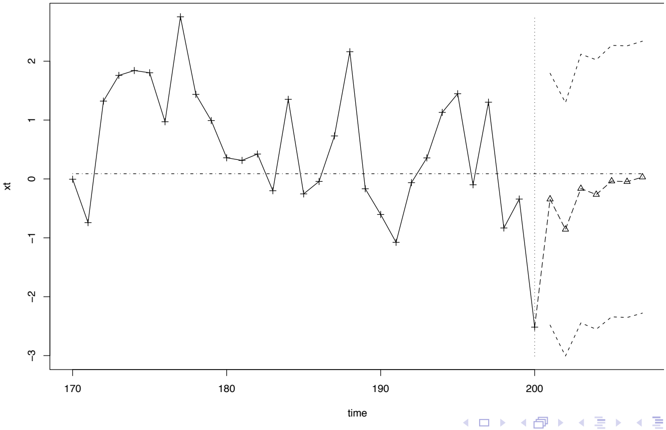
  In general the prediction error

  $$U(h) \sim N(0, (1 + \sum_{i=1}^{h} \psi_i^2)\sigma^2)$$

  where ψ_i are the coefficients in the MA presentation of the AR model.

- Example for predicting AR(2) model. Data generated using

```
x=arima.sim(n = 200, list(ar = c(0.2,0.3)), sd = 1)
```



The moving average model

- Let $\{w_t\}$ be WN$(0, \sigma^2)$ process. Let $\theta_0 = 1$. Let q be any integer, and consider the real constants $\theta_1, \theta_2, \cdots, \theta_q$ with $\theta_q \neq 0$. Define

$$\begin{aligned} X_t &= w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \cdots + \theta_q w_{t-q} \\ &= \theta_0 w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \cdots + \theta_q w_{t-q} \\ &= \sum_{j=0}^{q} \theta_j w_{t-j} \end{aligned}$$

- $\{X_t\}$ is the moving-average of order q, or the MA(q) process.
- X_t is a linear combination of $q + 1$ white noise variables.

The moving average model

- The MA(q) process is q-correlated, since X_t and X_{t+h} are uncorrelated for all lags $h > q$. What about w_t and X_{t-j}?
- Using the backward shift operator $B^j X_t = X_{t-j}$ we can write

$$X_t = \theta(B) w_t$$

where

$$\Theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q$$

- The MA process is stationary for any values of θ.

Moments

- Mean: $E(X_t) = 0$.
- Autocovariance:

$$\begin{aligned}
\gamma(h) &= cov(X_{t+h}, X_t) \\
&= E\left[\left(\sum_j \theta_j w_{t+h-j}\right)\left(\sum_k \theta_k w_{t-k}\right)\right] \\
&= \sigma^2 \sum_{k+h\leq q} \theta_k \theta_{k+h} \\
&= 0 \text{ if } h > q.
\end{aligned}$$

$\gamma(q) = \sigma^2 \theta_q$.

- The ACF of a moving average model is zero except for a finite number of lags h and the ACF of an autoregressive model goes to zero geometrically.

An MA(2) processes

- An MA(2) with $\theta_1 = 0.45$ and $\theta_2 = -0.45$
  ts.sim = arima.sim(n=200, list(ma=c(0.45, -0.45), sd = 1))

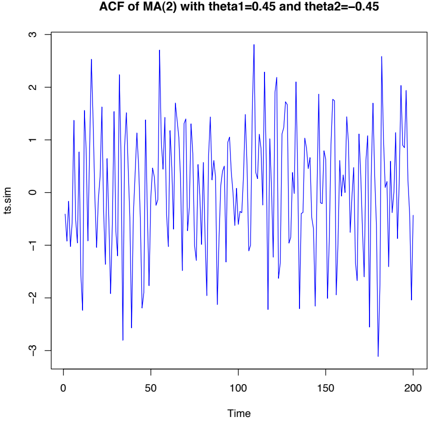


Figure: MA(2) process

The acf of an MA(2) processes

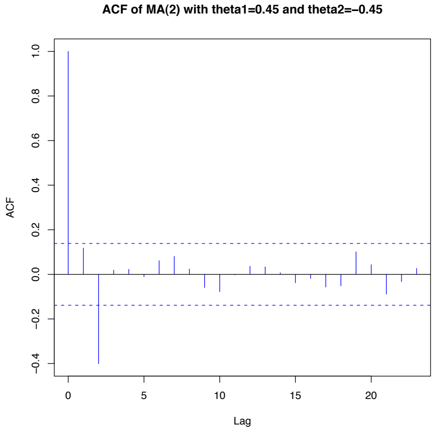


Figure: ACF of the MA(2) process

- Note the sharp cut-off at lag 2.

Identifiability and Invertibility of MA models

- Consider first order model $X_t = w_t + \theta w_{t-1}$, the autocorrelation $\rho(h)$ is the same for $\theta = 5$ and $\theta = 1/5$. Thus they are not identifiable.
- Invertible process: a process that has an infinite AR representation.
- Choose parameters that makes the process invertible.
- Write $X_t = \Theta(B) w_t = (1 + \theta B) w_t$. If $|\theta| < 1$ we have

$$w_t = (1 + \theta B)^{-1} X_t = \pi(B) X_t,$$

  where $\pi(B) = \sum_{j=0}^{\infty} (-\theta)^j B^j$.
- Thus X_t satisfies an infinite autoregression

$$X_t = -\sum_{j=1}^{\infty} (-\theta)^j X_{t-j} + w_t.$$

  This is the desired AR infinity representation if $|\theta| < 1$.

Prediction for MA(1) models

- MA(1) model with $E(X_t) = \mu$ and $w_t \sim N(0, \sigma^2)$.

$$X_t = c + w_t + \theta_1 w_{t-1}, \quad \text{where} \quad c = \mu.$$

$$X_n(1) = E(c + w_{n+1} + \theta_1 w_n | x_1, \cdots, x_n) = c + \theta_1 E(w_n | x_1, \cdots, x_n).$$

Replaced by estimated quantities, $x_n(1) = \hat{c} + \hat{\theta}_1 \hat{w}_n$.

- Prediction error

$$U(1) = \theta_1 E(w_n | x_1, \cdots, x_n) - (w_{n+1} + \theta_1 w_n) \approx -w_{n+1} \sim N(0, \sigma^2)$$

Excecise: $x_n(2) =$

$U(2) =$

Prediction for MA(2) models

- Prediction for MA(2)

$$x_n(1) = \hat{c} + \hat{\theta}_1 \hat{w}_n + \hat{\theta}_2 \hat{w}_{n-1}$$

$$x_n(2) = \hat{c} + \hat{\theta}_2 \hat{w}_n$$

$$x_n(h) = \hat{c}, \;\; h > 2$$

- Prediction error

$$U(1) \approx -w_{n+1} \sim N(0, \sigma^2)$$

$$U(2) \approx -w_{n+2} - \theta_1 w_{n+1} \sim N(0, (1 + \theta_1^2)\sigma^2)$$

$$U(h) \approx -w_{n+3} - \theta_1 w_{n+2} - \theta_2 w_{n+1} \sim N(0, (1 + \theta_1^2 + \theta_2^2)\sigma^2), \;\; h > 2$$

Prediction for MA models

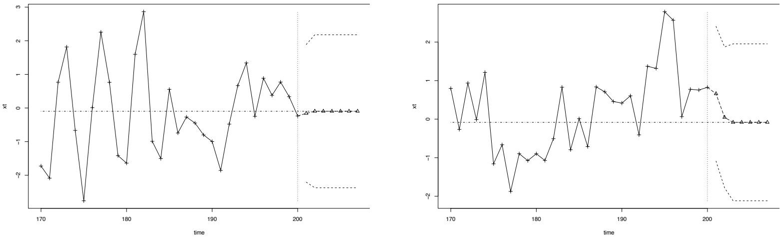


Figure: Prediction and prediction interval for MA(1) and MA(2)

- **ARMA** A time series $\{X_t : t = ..., -2, -1, 0, 1, 2, ...\}$ is ARMA(p,q) if it is stationary and

$$X_t = \phi_1 X_{t-1} + ... + \phi_p X_{t-p} + w_t + \theta_1 w_{t-1} + ... + \theta_q w_{t-q}$$

- The current observation X_t is a linear combination of previous observations $X_{t-j}, j = 1, \ldots, p$, plus a random term. The random term is a weighted average of $w_{t-j}, j = 1, \ldots, q$. If X_t has a nonzero mean μ, then

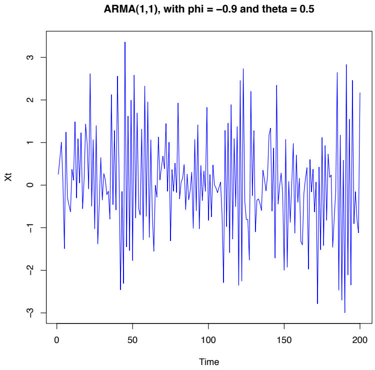
$$X_t = \phi_0 + \phi_1 X_{t-1} + ... + \phi_p X_{t-p} + w_t + \theta_1 w_{t-1} + ... + \theta_q w_{t-q}$$

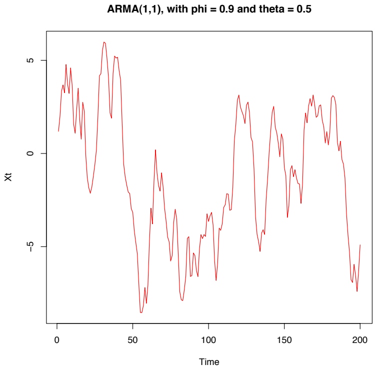
where $\phi_0 = \mu(1 - \phi_1 - ... - \phi_p)$.

- Using backshift operator, we have $\Phi(B)X_t = \Theta(B)w_t$.

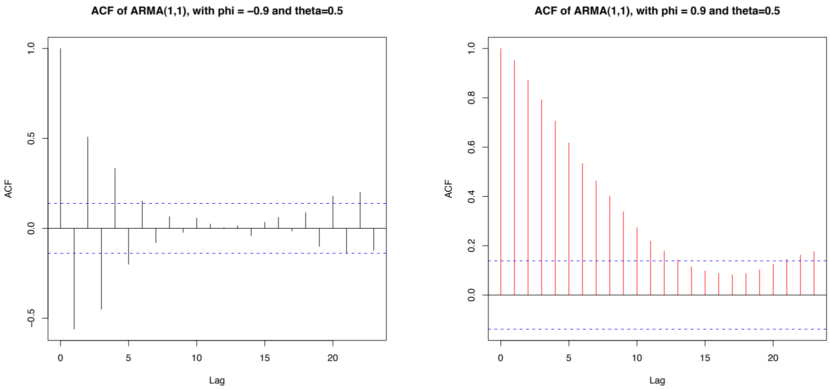
Simulated ARMA(1,1)

- To simulate an ARMA(1,1) with $\phi = -0.9$ and $\theta = 0.5$, use `ts.sim<- arima.sim(n=200,list(ar=-0.9,ma=0.5),sd=1)`
- Below are plots of four different ARMA(1,1) with various values of ϕ and θ. Note the distinct qualitative differences.





ACF plots of the simulated ARMA(1,1)



Parameter Redundancy

- Consider a white noise time series

$$X_t = w_t$$

for all t. Thus we have

$$X_{t-1} = w_{t-1}$$

Thus

$$0 = 0.5X_{t-1} - 0.5w_{t-1}.$$

- Add these two questions, we have

$$X_t = 0.5X_{t-1} - 0.5w_{t-1} + w_t$$

Seems to be an ARMA(1,1) model, but in reality this is, of course, still white noise!!

Parameter Redundancy

- Thus, it is important to check AR and MA polynomial for redundancies.
- When we write the ARMA(1,1) in backshift notation we will see the problem.

$$(1 - 0.5B)X_t = (1 - 0.5B)w_t$$

  There is a common factor in the AR and MA polynomials. We could simply divide out.
- In reality, we could easily mis-specify the model and fit an ARMA(1,1) to white noise data and find that the parameter estimates are significant. If we were unaware of parameter redundancy, we might claim the data are correlated when they are not. This is another reason to check the data with diagnostics.

Invertibility

- An ARMA model is said to be invertible if the time series can be written as

$$\pi(B)X_t = \sum_{j=0}^{\infty} \pi_j X_{t-j} = w_t$$

where $\sum_{j=0}^{\infty} |\pi_j| < \infty$ and $\pi_0 = 1$.

- **Condition for invertiblity.** A process is called invertible iff the roots of the MA polynomial $\Theta(B)$ are all outside the unit circle. With that, we have

$$\pi(B)\Theta(B) = \Phi(B)$$

The ACF of ARMA

- A stationary ARMA(p,q) process, $\Phi(B)X_t = \Theta(B)w_t$ can be written as
$$X_t = \sum_{j=0}^{\infty} \psi_j w_{t-j}$$
Define the AR and MA polynomial as
$$\Phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p, \quad \phi_i \neq 0,$$
$$\Theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q, \quad \theta_j \neq 0.$$
ψ_j can be obtained by matching the coefficients in
$$(\psi_0+\psi_1 B+\psi_2 B^2+\cdots)(1-\phi_1 B-\phi_2 B^2-\cdots) = (1+\theta_1 B+\theta_2 B^2+\cdots)$$

The Partial ACF

- The MA(q) can be identified from its ACF: cut off at lag q.
- Partial ACF enables similar property for AR(p).
- Motivating example:
  In an AR(1) model $X_t = \phi X_{t-1} + w_t$, where X_t and X_{t-2} are correlated with $\rho_2 = \phi^2$.
  $X_t = \phi X_{t-1} + w_t$ and $X_{t-1} = \phi X_{t-2} + w_{t-1}$. The connection

$$X_t \leftrightarrow X_{t-1} \leftrightarrow X_{t-2}$$

  Thus if one takes out the common knowledge on X_{t-1}, then X_t and X_{t-2} are uncorrelated.

The Partial ACF

- Markov property:

$$P(X_{t+1}|X_t, ...) = P(X_{t+1}|X_t).$$

- P-th order Markov property:

$$P(X_{t+1}|X_t, ...) = P(X_{t+1}|X_t, \cdots, X_{t-p}).$$

  Clearly AR(p) is p-th order Markovian.
- Partial correlation between random variables: Given X_1, X_2 and X_3, the partial correlations of X_1 and X_3 given X_2 are the correlations between
  - Residuals of X_1 from its regression on X_2;
  - Residuals of X_3 from its regression on X_2;

The Partial ACF

- Partial autocorrelation ϕ_{hh} = PACF is the correlation between X_{t+h} and X_t with linear effects of $X_{t+h-1}, \cdots, X_{t+1}$ removed.
- Formally, the PACF of a stationary X_t, denoted ϕ_{hh}, is

$$\phi_{11} = \rho(1)$$

$$\phi_{hh} = \mathrm{corr}(X_h - X_h^{h-1}, X_0 - X_0^{h-1}), \quad h > 1$$

where

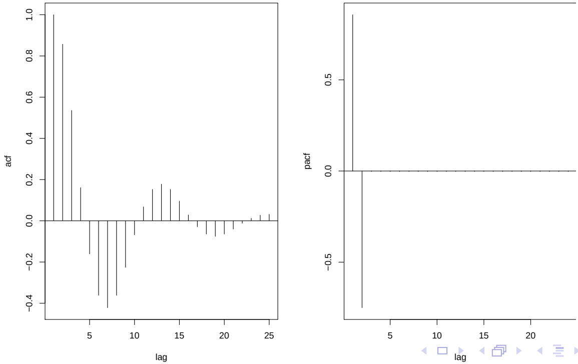
$$X_h^{h-1} = \beta_1 X_{h-1} + \beta_2 X_{h-2} + \cdots + \beta_{h-1} X_1;$$

$$X_0^{h-1} = \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_{h-1} X_{h-1}.$$

- For AR(p) process, we should see the cut off effect at lag p.
- We can use the command ARMAacf to obtain theoretical values for ACF and PACF.

The Partial ACF

```
acf=ARMAacf(ar=c(1.5,-0.75),ma=0,24);
pacf=ARMAacf(ar=c(1.5,-0.75),ma=0,24,pacf=T);
par(mfrow=c(1,2));
plot(acf,type="h",xlab="lag");  abline(h=0);
plot(pacf,type="h",xlab="lag");abline(h=0);
```



ACF and PACF

- What about the PACF of an invertible ARMA process? Think about it's AR representation.

	AR(p)	MA(q)	ARMA(p,q)
ACF	Tails off	Cut off after q	Tails off
PACF	Cut off after p	Tails off	Tails off

Diagnosis-check estimated coefficients

```
> out=arima(x,c(3,0,2))
> out
Call:
arima(x = x, order = c(3, 0, 2))
Coefficients:
         ar1     ar2     ar3     ma1     ma2  intercept
      0.0302  0.8767  0.0296  0.6531  0.4869     0.8687
s.e.  0.1520  0.0373  0.1440  0.1333  0.0722     2.0760
sigma^2 estimated as 0.9767:  log likelihood = -283.91,  a
```

We can roughly check the t-stat = estimate/s.e. for the estimated coefficients. For this case, the 3rd lag coefficient is insignificant. So we should think about fitting a ARMA(2,2) model.

Diagnosis-invertibility and stationarity

Check that the roots of the AR and MA polynomial are outside of the unit circle.

```
> #AR roots
> polyroot(c(1, -0.0302,-0.8767, -0.0296))
[1]   1.033340+0i  -1.106575+0i -29.545008-0i
> a=polyroot(c(1, -0.0302,-0.8767, -0.0296))
> Mod(a)
[1]  1.033340  1.106575 29.545008
```

Diagnosis

- Polynomial factorization: if one of the roots is close to 1, then differencing should be consider. If there are four roots close to $(1, -1, i, -i)$ respectively, then seasonal (quarterly) differencing

$$1 - B^4 = (1 - B)(1 + B)(1 - iB)(1 + iB)$$

  should be considered.
- Model redundancy: consider a ARMA(p,q) model $\Phi(B)X_t = \Theta(B)w_t$, then it also follows ARMA(p+k,q+k) model, since

$$\beta(B)\Phi(B)X_t = \beta(B)\Theta(B)w_t$$

- Underspecified model: increase model size. Add seasonal terms. If residual shows an isolated significance at lag k, adding X_{t-k} orw_{t-k} usually helps.

Diagnosis- Residual analysis

Check the following things: Time plot of the series, ACF, PACF, Box-Ljung test and outliers.

- If time plot shows non-zero mean, include a constant term in the model.
- If non-constant variance, transformation is needed, try log and cox transformation.
- If ACF PACF are not clean, probably the model is underspecified.
- Outliers: check data, treat as missing and replaced using mean or interpolation ect.

Use model selection criteria for order specification

- Akaike Information Criterion (AIC)

$$AIC = ln\hat{\sigma}^2 + \frac{2(p+q)}{T}.$$

- Bayesian Information Criterion (BIC)

$$BIC = ln\hat{\sigma}^2 + \frac{(p+q)ln(T)}{T}.$$

- AICc, DIC, MDL ect.
- Regularization methods: Lasso, group Lasso, Elastic Net, SCAD ect.

Prediction for ARMA(1,1) models

- The ARMA(1,1)

$$X_t = c + \phi_1 X_{t-1} + w_t + \theta_1 w_{t-1}, \quad c = (1 - \phi_1)\mu.$$

The 1 step prediction is

$$X_n(1) = E(X_{n+1}|x_1, \cdots, x_n) = c + \phi_1 x_n + \theta_1 E(w_n|x_1, \cdots, x_n).$$

With estimated parameters

$$X_n(1) = \hat{c} + \hat{\phi}_1 x_n + \hat{\theta}_1 \hat{w}_n.$$

Prediction error $U(1) = X_n(1) - X_{n+1}$

$$= c + \phi_1 x_n + \theta_1 E(w_n|x_1, \cdots, x_n) - (c + \phi_1 x_n + w_{n+1} + \theta_1 w_n)$$

Which is $-w_{n+1} \sim N(0, \sigma^2)$.

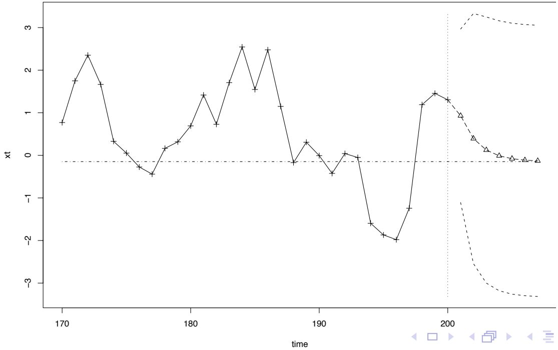
Prediction for ARMA(1,1) models

In general we have

$$X_n(h) = c + \phi_1 X_n(h-1).$$

Prediction error

$$U(h) \sim N(0, (1 + \sum_{i=1}^{h} \psi_i^2)\sigma^2)$$



Prediction for general ARMA(p,q) models

$$X_{n+h} = c+\phi_1 X_{n+h-1}+\cdots+\phi_p X_{n+h-p}+w_{n+h}+\theta_1 w_{n+h-1}+\cdots+\theta_q w_{n+h-q}$$

Steps for obtaining $x_n(h)$:

1. Replace all parameters with their estimates
2. Replace the current and past errors w_{n-j}, $j \geq 0$ with residuals.
3. Replace the current and past observations X_{n-j}, $j > 0$ with actual observations x_{n-j}.
4. Replace each future error w_{n+j}, $j > 0$ with its expectation, which is 0.
5. Replace each future value of X_{n+j}, $j > 0$ with its prediction $X_n(j)$. We need to predict $X_{n+1}, \cdots, X_{n+h-1}$ to obtain $X_n(1), \cdots, X_n(h-1)$ first.

Prediction for general ARMA(p,q) models

- Forecasting error:
  Assume that X_t (a causal ARMA(p,q)) has an MA(∞) representation:

$$X_t = w_t + \psi_1 w_{t-1} + \ldots$$

  Then

$$\begin{aligned} X_n(h) - X_{n+h} &= -(w_{n+h} + \psi_1 w_{n+h-1} + \cdots, \psi_{h-1} w_{n+1}) \\ &\sim N(0, (1 + \sum_{j=1}^{h-1} \psi_j^2)\sigma^2) \end{aligned}$$

Estimation

For an ARMA model

$$\phi(B)X_t = \theta(B)w_t,$$

we discuss estimation using conditional LS and maximum likelihood.

- Conditional LS:

$$\hat{\Theta} = \arg\min \sum_{t=1}^{T} \hat{w}_t^2$$

- MLE:

$$\hat{\Theta} = \arg\max \log P(X_1, \cdots, X_t | \Theta)$$

- Let's look at AR(1) model first.

Conditional LS

- Given $x_1, \cdots, x_n$, $x_t - \phi_1 x_{t-1} = w_t$, where $w_t \sim N(0, \sigma^2)$.
- Given $\hat{\phi}_1$, we have

$$\hat{w}_2 = x_2 - \hat{\phi}_1 x_1, \hat{w}_3 = x_3 - \hat{\phi}_1 x_2, \cdots$$

- We are finding $\hat{\phi}_1$ that minimizes

$$\sum_{t=2}^{n} (x_t - \phi_1 x_{t-1})^2$$

which gives

$$\hat{\phi}_1 = \frac{\sum_{t=2}^{n} x_t x_{t-1}}{\sum_{t=2}^{n} x_{t-1}^2}$$

- Here we ignored $\hat{w}_1$, since it's evaluation requires x_0. We can suggest a value for it, but usually it's ignored, since the impact is minimum usually.

MLE

$$
\begin{aligned}
l(x^n|\phi_1) &= P(x_1, \cdots, x_n|\phi_1) \\
&= P(x_n|x_1, \cdots, x_{n-1}, \phi_1)P(x_1, \cdots, x_{n-1}|\phi_1) \\
&= P(x_n|x_1, \cdots, x_{n-1}, \phi_1)P(x_{n-1}|x_1, \cdots, x_{n-2}, \phi_1) \\
&\quad P(x_1, \cdots, x_{n-2}|\phi_1) \\
&= P(x_1|\phi_1)\Pi_{t=2}^n P(x_t|x_1, \cdots, x_{t-1}, \phi_1) \\
&= P(x_1|\phi_1)\Pi_{t=2}^n P(x_t|x_{t-1}, \phi_1) \\
&= P(x_1|\phi_1)\Pi_{t=2}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{(x_t - \phi_1 x_{t-1})^2}{2\sigma^2}\},
\end{aligned}
$$

assuming $w_t \sim N(0, \sigma^2)$.
If we ignore the first term $P(x_1|\phi_1)$, we will get the same solution as in the LS case. If we keep the first term, we will need to resort to some convex optimization schemes.

AR(p) estimation

- For the AR(p) cases, LS becomes minimizing

$$\sum_{t=p+1}^{n} (x_t - \phi_1 x_{t-1} - \cdots - \phi_p x_{t-p})^2.$$

  And MLE reduces to similar form.
- For non-zero mean cases, we can
  1. Estimate the mean with $\bar{x}$, then perform the same analysis using the demeaned series.
  2. Or, estimate the model with constant, i.e. minimizing

$$\sum_{t=p+1}^{n} (x_t - c - \phi_1 x_{t-1} - \cdots - \phi_p x_{t-p})^2.$$

MA(1) estimation

- MA(1) model: $X_t = w_t + \theta w_{t-1}$. Given $\hat{\theta}$, we have

$$\hat{w}_1 = x_1 - \hat{\theta}_1 w_0, \hat{w}_2 = x_2 - \hat{\theta}_1 w_1, \cdots .$$

- Since the effect of w_0 propagates, we need to have a starting value for it.
- Set $w = 0$. This is the conditional estimation.
- Treat it as a parameter and estimate it with θ_1 jointly. This is call exact estimation.

$$\min_{\theta_1, w_0} \sum_{t=1}^{n} \hat{w}_t^2.$$

- Similarly for MLE

Details of arima in R

```
arima(x, order = c(0, 0, 0), seasonal = list(order = c(0, 0, 0),
period = NA), xreg = NULL, include.mean = TRUE,
method = c(CSS-ML, ML, CSS))
```

- x: a univariate time series
- order: A specification of the non-seasonal part of the ARIMA model: the three components (p, d, q) are the AR order, the degree of differencing, and the MA order.
- seasonal: A specification of the seasonal part of the ARIMA model, plus the period (which defaults to frequency(x)). This should be a list with components order and period, but a specification of just a numeric vector of length 3 will be turned into a suitable list with the specification as the order.
- xreg: Optionally, a vector or matrix of external regressors, which must have the same number of rows as x.

Details of arima in R

- include.mean: Should the ARMA model include a mean/intercept term? The default is TRUE for undifferenced series, and it is ignored for ARIMA models with differencing.
- method: Fitting method: maximum likelihood or minimize conditional sum-of-squares. The default (unless there are missing values) is to use conditional-sum-of-squares to find starting values, then maximum likelihood.

```
> out=arima(x,c(2,0,2))
> out
Call:
arima(x = x, order = c(2, 0, 2))
Coefficients:
         ar1     ar2     ma1     ma2  intercept
      0.0604  0.8760  0.6296  0.4787     0.8578
s.e.  0.0372  0.0374  0.0662  0.0627     2.0488
sigma^2 estimated as 0.9769:  log likelihood = -283.93,  a
```