Chapter 3. Autoregressive Integrated Moving Average (ARIMA) Models

Dexter Cahoy

October 31, 2018

Outline

Autoregressive Models

Moving Average Process

Autoregressive Moving Average (ARMA) Model

Autocorrelation and Partial Autocorrelation

Basics

- Classical regression is often insufficient for explaining all of the interesting dynamics of a time series.
- Chapter 2 are for the static case, namely, we only allow the dependent variable to be influenced by current values of the independent variables.
- In the time series case, it is desirable to allow the dependent variable to be influenced by the past values of the independent variables and possibly by its own past values.

Autoregressive Process

- **Def'n.** An **autoregressive process of order p (AR(p))** is of the form
$$x_t = \phi_1 x_{t-1} + \ldots + \phi_p x_{t-p} + w_t$$
where x_t is stationary, w_t is $wn(0, \sigma_w^2)$, and ϕ's are constants ($\phi_p \neq 0$).
- Above is standard form used by `R's arima.sim, ARMAacf, arima, etc.`
- The variables $x_{t-1}, \ldots,$ and x_{t-p} are random covariates.
- x_t is a weighted linear combination of the values of the previous p time points plus a *shock* or *innovation* w_t at time t.
- We assume that w_t is independent of $x_{t-1}, \ldots, x_{t-p}$.

Autoregressive Process

- **Def'n.** Let $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \phi_p B^p$ be an **autoregressive operator**. It is a polynomial in B of order p. Then **AR(p)** can be written concisely as

$$\phi(B)x_t = w_t.$$

- For simplicity, we continue to assume that $E(x_t) = 0$.

- **AR(1):** $x_t = \phi x_{t-1} + w_t$. We get the random walk if $\phi = 1$. The autocovariance function is

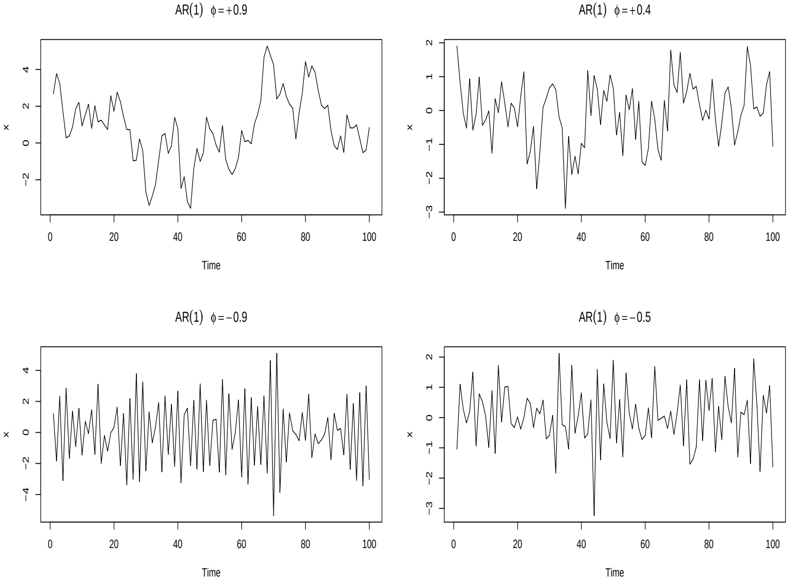
$$\gamma_x(h) = \phi^h \left(\frac{\sigma_w^2}{1 - \phi^2} \right), \qquad h \geq 0.$$

- Note that the **characteristic equation** for AR(1) is

$$\phi(B) = 1 - \phi B = 0.$$

AR(1)

- Use `arima.sim(p=1,d=0,q=0, ...)` in R.



AR(1)

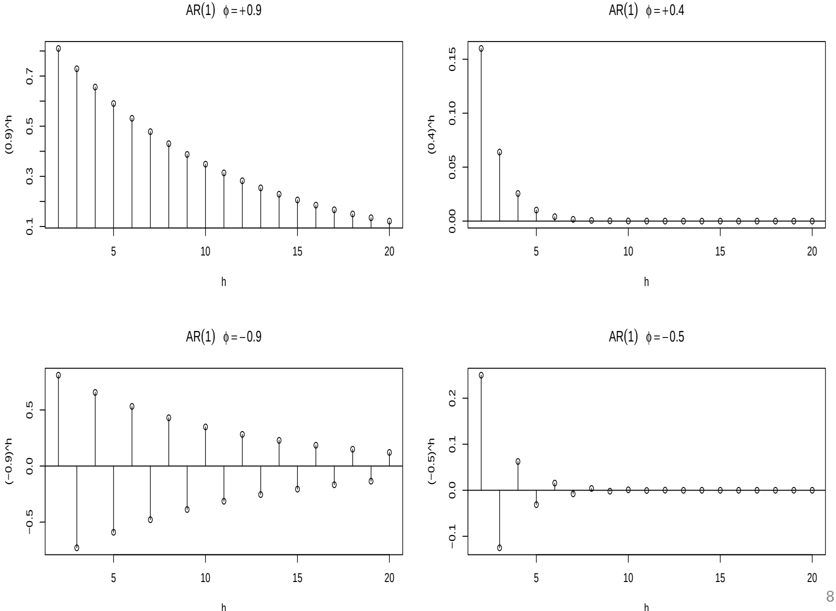
- The ACF of an AR(1) process is

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \frac{\phi^h \left(\frac{\sigma_w^2}{1-\phi^2}\right)}{\frac{\sigma_w^2}{1-\phi^2}} = \phi^h, \qquad h \geq 0.$$

- Since $-1 < \phi < 1$ (the stationarity condition), the (true/population) ACF decays exponentially with h.
- If $\phi \sim \pm 1$ then ACF decays slowly. Neighboring (small h) observations are highly correlated.
- If $\phi \sim 0$ then ACF decays rapidly .
- If $\phi > 0$ then ACF is positive.
- If $\phi < 0$ then ACF has alternating signs (more oscillations).
- AR(1) becomes a *stationary* $\iff$ the root of $\phi(B) = 0$ is such that $|B| > 1 (|\phi| < 1)$.

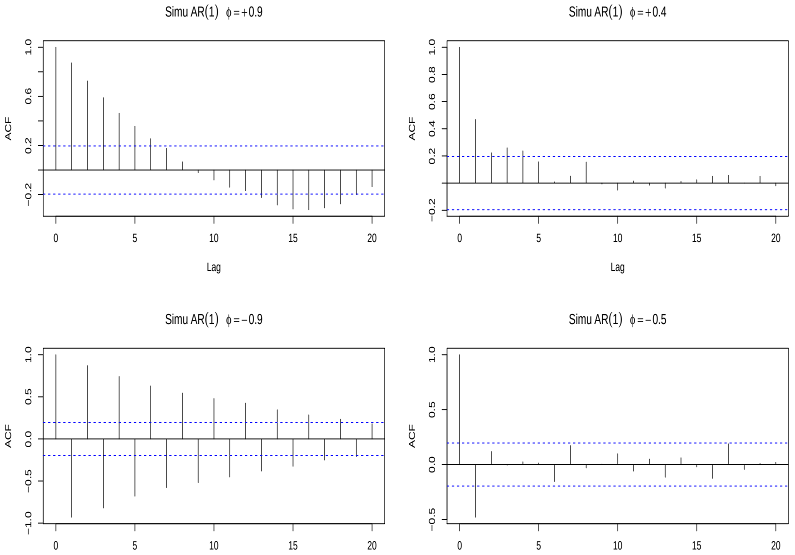
AR(1)

- True ACF's.



AR(1)

- Sample ACF's of simulated `arima.sim(list(order=c(1,0,0), ar=.9), n=100)`.



AR(2)

- **Model:** $x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + w_t$.
- The **AR(2) characteristic polynomial** is

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2$$

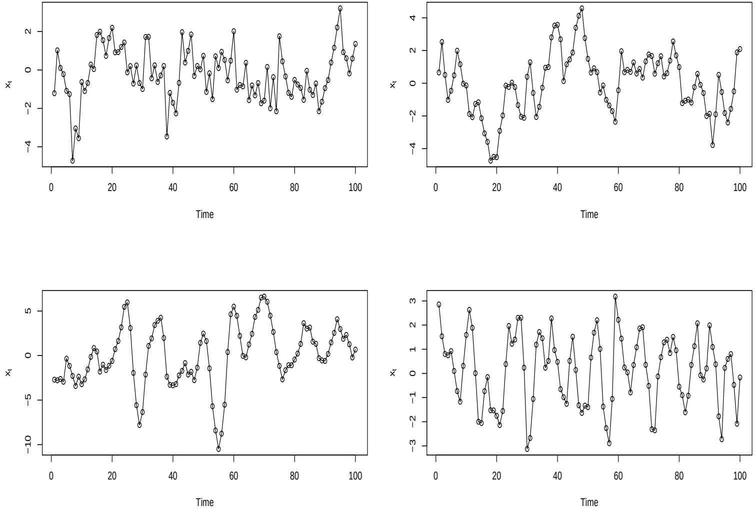
and **the AR(2) characteristic equation** is

$$1 - \phi_1 B - \phi_2 B^2 = 0.$$

- The equation has two (possibly complex) roots $B_1, B_2 = \frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{-2\phi_2}$.
- An AR(2) is **stationary** $\iff |B_i| > 1$.
- If $z = a + bi, i = \sqrt{-1}$ then `Modulus(z)=Mod(z) = |z|` $=\sqrt{a^2 + b^2}$
- Use `arima.sim(list(order=c(2,0,0), ar=c(`ϕ_1`, `ϕ_2`), n=1000)` to simulate AR(2).

AR(2)

- Sample trajectories of AR(2) using $i)(\phi_1, \phi_2) = (0.5, 0.25)$, $ii)(\phi_1, \phi_2) = (1, -0.25)$, $iii)(\phi_1, \phi_2) = (1.5, -0.75)$, $iv)(\phi_1, \phi_2) = (1, -0.6)$. All these are stationary.



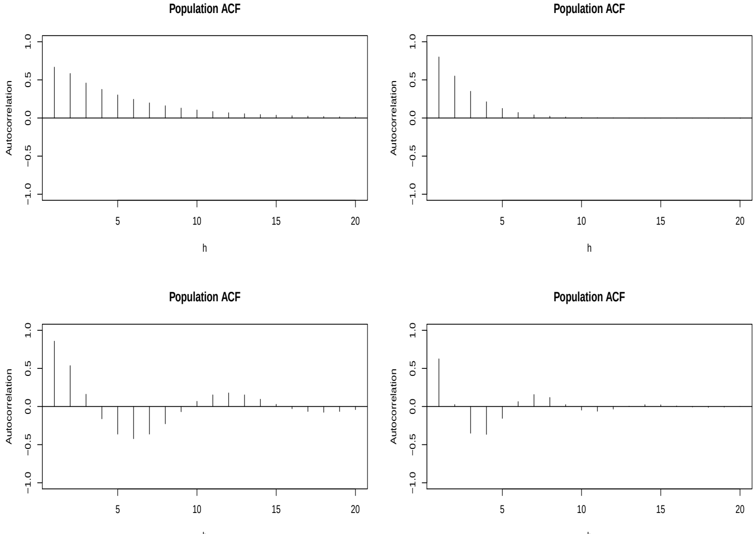
AR(2)

- The explicit formulas of ρ_h are on page 73 (CaC). These population or true results for AR(2) are needed to guide us in identifying a model for the data.

- Use `ARMAacf(ar = c(`ϕ_1`, `ϕ_2`), lag.max = m)` in R.

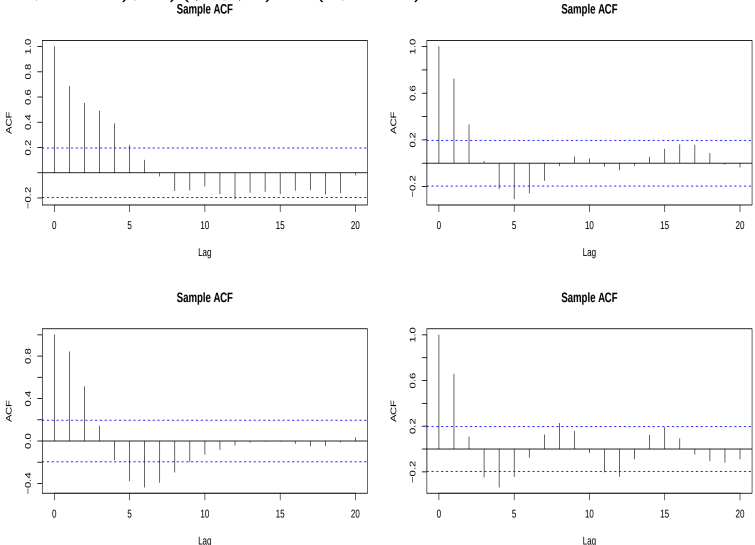
AR(2)

- Below are true ACF's of AR(2) using $i)(\phi_1, \phi_2) = (0.5, 0.25)$, $ii)(\phi_1, \phi_2) = (1, -0.25)$, $iii)(\phi_1, \phi_2) = (1.5, -0.75)$, $iv)(\phi_1, \phi_2) = (1, -0.6)$. Stationary.



AR(2)

- Below are sample ACF's of the simulated of AR(2) using $i)(\phi_1, \phi_2) = (0.5, 0.25)$, $ii)(\phi_1, \phi_2) = (1, -0.25)$, $iii)(\phi_1, \phi_2) = (1.5, -0.75)$, $iv)(\phi_1, \phi_2) = (1, -0.6)$.



AR(p)

- Recall an **AR(p)** process:

$$x_t = \sum_{j=1}^{p} \phi_j x_{t-j} + w_t. \tag{1}$$

- The **AR(p)** characteristic equation is

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p = 0.$$

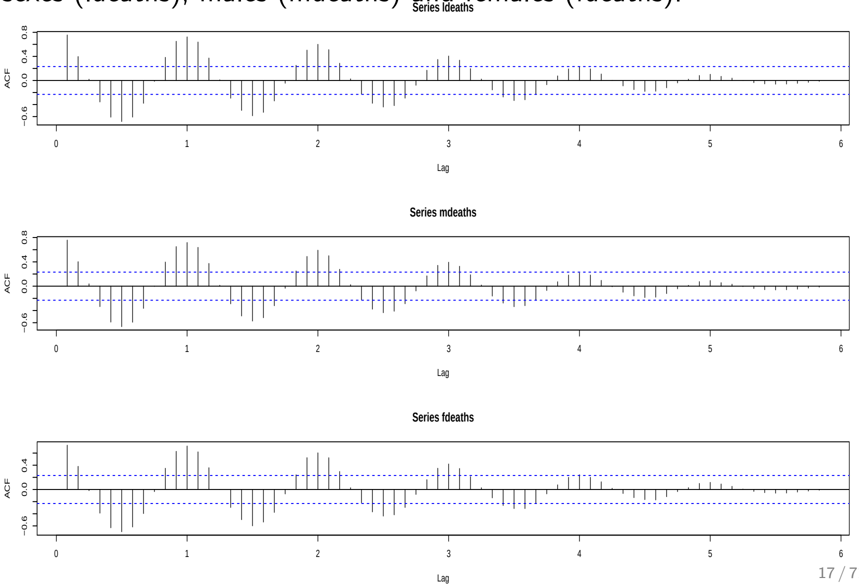
- An **AR(p)** process is *stationary (linear process)* $\iff$ all roots of the above equation are larger than 1 in modulus, i.e, $|B| > 1$ (they all lie outside the unit circle).
- The ACF of **AR(p) tails off** or **dies off** or **decreases** or **decays** to zero.
- Use `polyroot()` and `Mod()` in R to test stationarity of the model.

AR(p)

- Recall AR(1): $x_t = \phi x_{t-1} + w_t$, with
$$\phi(B) = 1 - \phi B = 0 \implies B = \frac{1}{\phi}.$$
Hence, the stationarity condition for AR(1) is
$$|B| > 1 \iff |\phi| < 1.$$
- Consider AR(2): $x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + w_t$, with
$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 = 0$$
whose roots are
$$B = \frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{-2\phi_2}.$$
- An AR(2) is *stationary* $\iff$ both roots are larger than 1 in modulus.

AR(p) Application

- Monthly deaths from lung diseases in the UK, 1974-1979, both sexes (ldeaths), males (mdeaths) and females (fdeaths).



Box-Jenkins Method

1) Determine the theoretical ACF and PACF for known classes of time series models (e.g., ARMA(p,q)). Use the sample ACF/PACF to match the data to a possible known model.

2) Estimate parameters using a method appropriate to the chosen model and assess the fit, primarily by studying the residuals. The residuals should then "look like" white noise.

3) If the fit is inadequate go back to steps 1 and 2. Otherwise, use the "final" model to forecast.

Lab for Today

- Consider the luteinizing hormone (`data(lh)` from `library(datasets)`) in blood samples at 10-min intervals from a human female, 48 samples. Is the pattern of the ACF consistent with a stationary AR model?

- Simulate an AR(3) using $n = 50$ and the coefficients$=c(0.64, -0.06, -0.22)$. Using the coefficients, is this a stationary AR(3)? Compare the theoretical ACF, the sample ACF of the simulated data, and the ACF of the luteinizing hormone data. Observations? Write down the algebraic expression of the AR(3) model.

Moving Average Process

- **Def'n.** The **moving average model of order q or MA(q)** is

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

  where $\theta_q \neq 0$ and w_t's are white noise.

- Let the **MA operator** $\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q$. The concise form of MA(q) is

$$x_t = \theta(B) w_t.$$

MA(1)

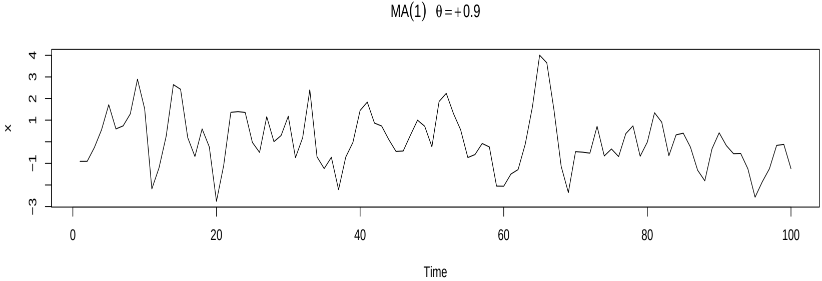
- Consider $x_t = w_t + \theta w_{t-1}$. Clearly, $E(x_t) = 0$.
- The ACF is

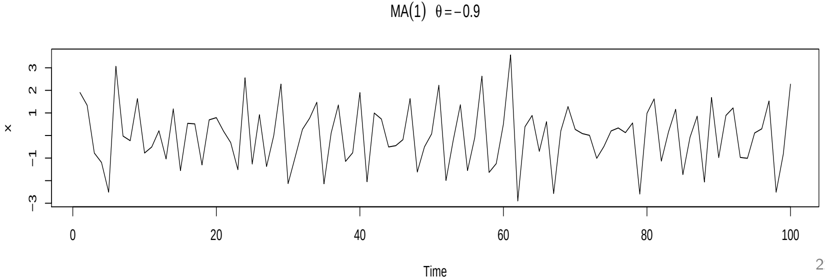
$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \begin{cases} 1, & h = 0 \\ \theta/(1+\theta^2), & h = 1 \\ 0, & h > 1. \end{cases}$$

- The ACF **cuts off** after lag $h = 1$, i.e., two observations that are more than one time unit apart are not correlated.
- $\theta > 0 \implies +\rho(1)$ while $\theta < 0 \implies -\rho(1)$.
- When $\theta = 0$, MA(1) process becomes a white noise process.
- As θ ranges from -1 to 1, the (true) lag 1 autocorrelation $\rho(1)$ ranges from -0.5 to 0.5; see R.

MA(1)

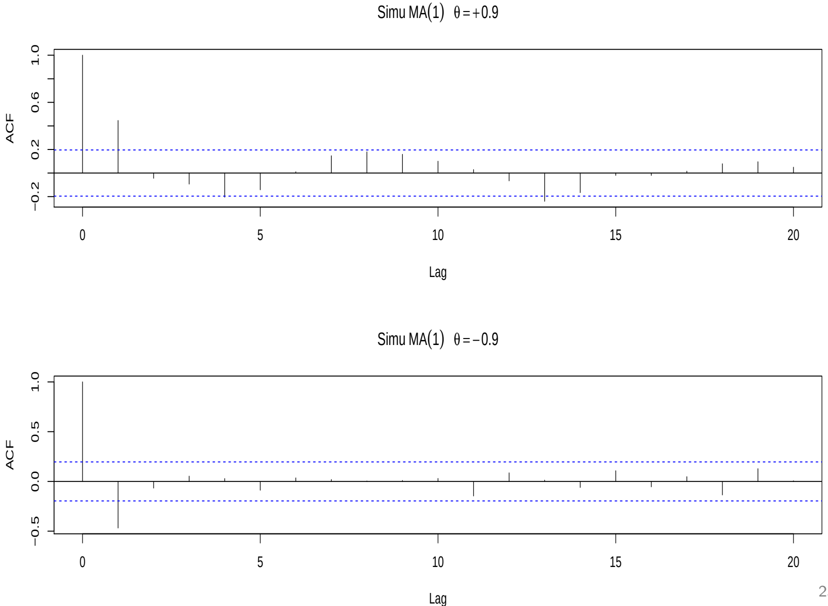
- Use `arima.sim(list(order=c(0,0,1), ma=`θ`), n=100)` .





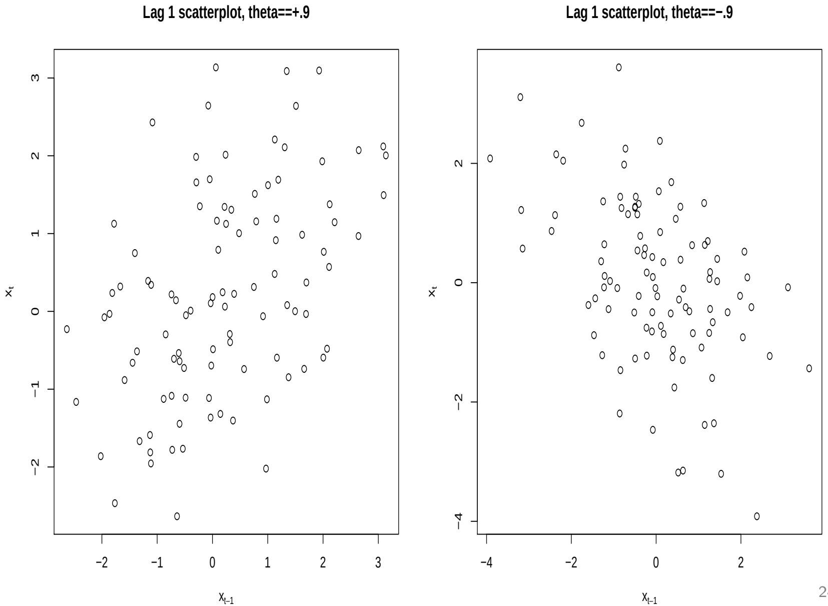
MA(1)

- ACF's of simulated MA(1).



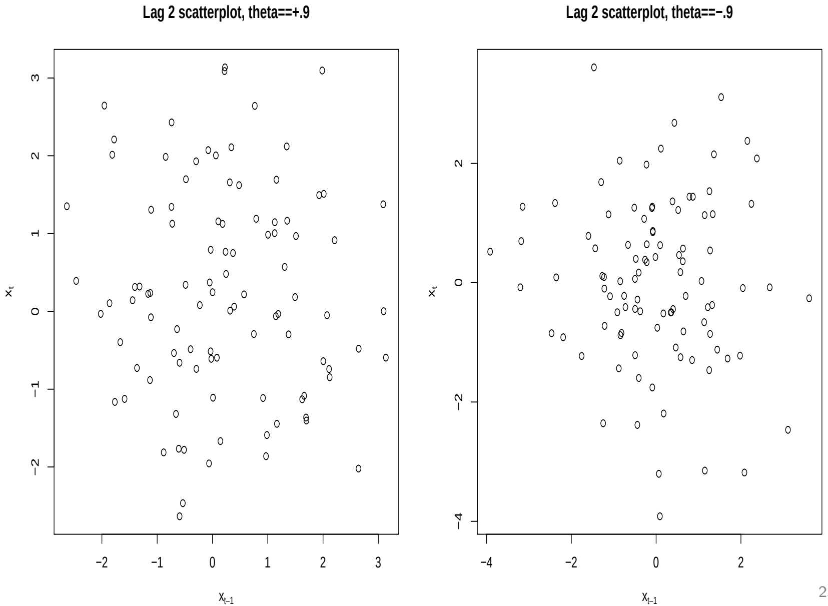
MA(1)

- Lag 1 scatter plots for simulated MA(1); $\rho_1 \approx \pm 0.497$.



MA(1)

- Lag 2 scatter plots for simulated MA(1); $\rho_2 = 0$.



MA(1)

- The ACF $\rho(h)$ is the same for θ and $1/\theta$, hence, non-unique.
- Consider

$$\texttt{Model 1:}\ x_t = w_t + 0.2w_{t-1}, \qquad w_t \overset{iid}{\sim} N(0,25)$$

  and

$$\texttt{Model 2:}\ y_t = v_t + 5v_{t-1}, \qquad v_t \overset{iid}{\sim} N(0,1).$$

- Both MA(1) models above give

$$\rho(h) = \begin{cases} 1, h = 0 \\ 5/26, h = 1 \\ 0, h > 1. \end{cases}$$

- Which of the two models should we choose?

MA(1)

- We want to choose the *invertible* one, which is Model 1.
- An **MA model is invertible** if it can be written as an **AR**(∞) model with coefficients converging to zero.
- `Model 1:` $x_t = w_t + 0.2w_{t-1}, \qquad w_t \overset{iid}{\sim} N(0, 25)$ can be written as
$$x_t = -\sum_{j=1}^{\infty}(-0.2B)^j x_t + w_t.$$
- MA(1) becomes *invertible* $\iff$ the root B of $\theta(B) = 1 + \theta B = 0$ lies outside the unit circle, i.e., $|B| > 1$ ($|\theta| < 1$).
- `Model 2` is not invertible as $1 + 5B = 0 \implies |B| = 1/5 < 1$.

MA(2)

- MA(2) Process: $x_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2}$. We can compute

$$\begin{aligned}
\gamma_0 &= var\{x_t\} = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} = (1 + \theta_1^2 + \theta_2^2)\sigma_w^2, \\
\gamma_1 &= cov\{x_t, x_{t-1}\} = (\theta_1 + \theta_1\theta_2)\sigma_w^2, \\
\gamma_2 &= cov\{x_t, x_{t-2}\} = \theta_2\sigma_w^2, \\
\gamma_h &= 0 \quad \text{for } h > 2.
\end{aligned}$$

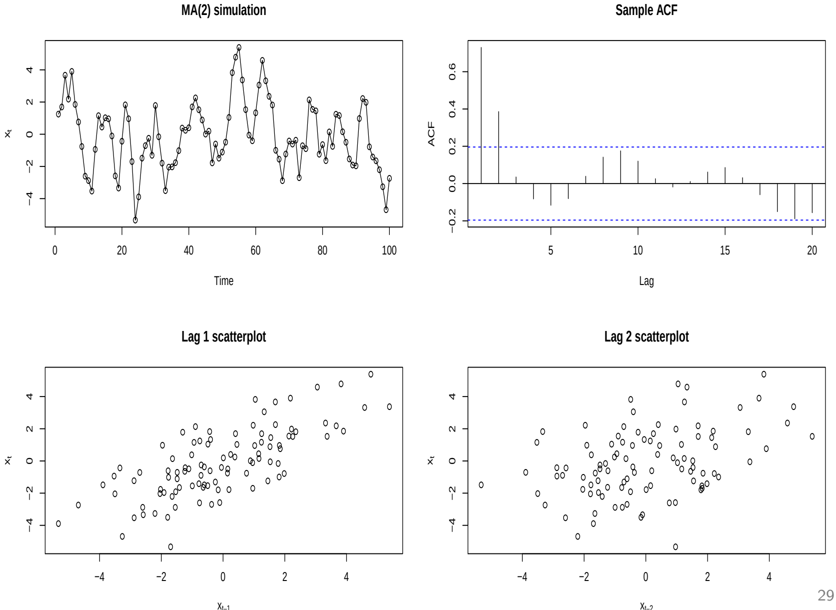
- Thus,

$$\rho_1 = \frac{\theta_1 + \theta_1\theta_2}{1 + \theta_1^2 + \theta_2^2}, \qquad \rho_2 = \frac{\theta_2}{1 + \theta_1^2 + \theta_2^2}, \qquad \rho_3 = \rho_4 = \cdots = 0.$$

- Use `arima.sim(list(order=c(0,0,2), ma= c(`θ_1 , θ_2`),` `n=100)` to simulate MA(2).

MA(2)

- We simulate $x_t = w_t + w_{t-1} + 0.9w_{t-2}$; $\rho_1 \approx 0.676$, $\rho_2 \approx 0.320$



MA(2)

- MA(2) is invertible $\iff$ all roots of $\theta(B) = 0$ lie outside the unit circle, i.e., $|B| > 1$.
- MA(2) : $x_t = w_t + w_{t-1} + 0.9w_{t-2}$. The characteristic equation is

  $$\theta(B) = 1 + B + 0.9B^2 = 0$$

  which has roots

  $$B = -0.5555556 \pm 0.8958064i \qquad (\textit{see } R)$$

  and $|B| = \sqrt{(-0.5555556)^2 + (\pm 0.8958064)^2} = 1.054 > 1$. Hence, invertible.
- MA(2) : $x_t = w_t - 2w_{t-1} + 2w_{t-2}$. The characteristic equation is

  $$\theta(B) = 1 - 2B + 2B^2 = 0$$

  which has roots

  $$B = 0.5 \pm 0.5i \qquad (\textit{see } R)$$

  and $|B| = \sqrt{(0.5)^2 + (\pm 0.5)^2} = 0.707 < 1$. Thus, non-invertible.

MA(q)

- The **moving average model** of order q or MA(q) is defined as

$$x_t = w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q}, \qquad w_t \sim wn(0, \sigma_w^2).$$

- More concisely,

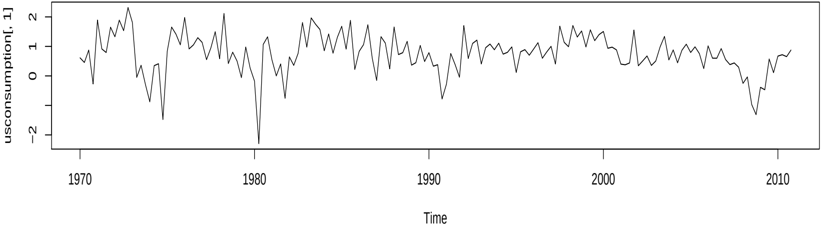
$$x_t = \theta(B) w_t, \qquad \theta(B) = 1 + \theta_1 B + \ldots + \theta_q B^q.$$

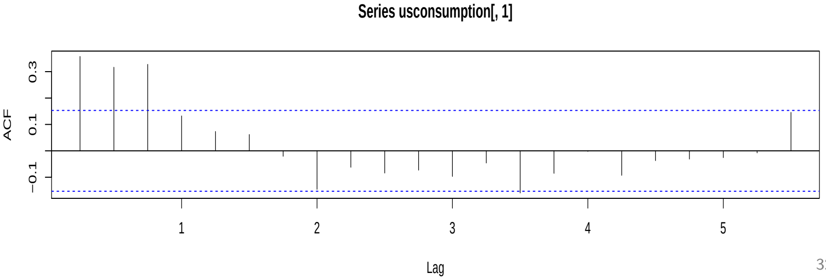
- An *MA(q) process is invertible* $\iff$ all roots of $\theta(B) = 0$ lie outside the unit circle, i.e., $|B| > 1$.

- An MA(q) process is *stationary* but not necessarily invertible.

MA(q) Application

- Percentage changes in quarterly personal consumption expenditure and personal disposable income for the US, 1970-2010.





Summary

- An MA(q) process is **invertible** if all the roots of

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q = 0$$

  exceed 1 in modulus.
- An invertible MA process is stationary.
- An invertible MA(q) process is an AR process of infinite order.
- The autocorrelation function of MA(q) is

$$\rho(h) = \frac{\sigma_w^2 \sum_{j=0}^{q-h} \theta_j \theta_{j+h}}{1 + \theta_1^2 + \ldots + \theta_q^2}. \qquad \text{(Verify)}$$

  The autocovariance(autocorrelation) **cuts off** after lag q.
- Use `ARMAacf(ma= numeric(), lag.max = r)` for true ACF.

Lab for Today

1. Simulate an MA(3). Write down the algebraic expression of your model. Is it invertible? What do you notice about the pattern of the ACF?

2. Analyze the percentage changes in quarterly personal disposable income for the US, 1970 to 2010. What model can you propose?

ARMA(p,q)

- **Def'n.** $\{x_t\}$ is an **autoregressive moving average** process of orders $\boldsymbol{p}$ and $\boldsymbol{q}$, denoted as $\boldsymbol{ARMA(p,q)}$ if it is *stationary* and $x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q}$.

- Above is standard form used by `R's arima.sim, ARMAacf, arima, etc.`; the concise form is $\phi(B)x_t = \theta(B)w_t$, $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p \quad \text{and} \quad \theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$.

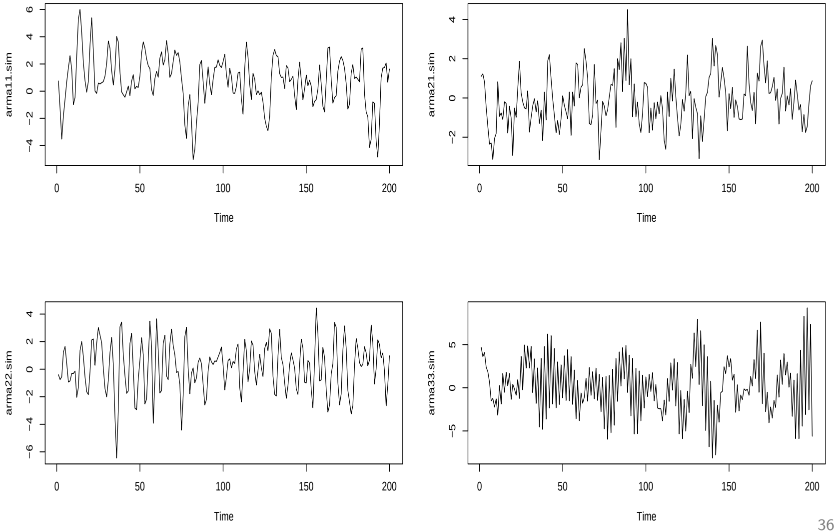
- The parameters p and q are the AR and MA orders, respectively.

- ARMA$(p, 0)$ = AR(p); *ARMA*(0, q)= *MA*(q).

- If $\mu_t = \mu \neq 0$ then simply add $\alpha = \mu(1 - \phi_1 - \cdots - \phi_p)$ on the RHS of the first equation above and write $x_t = \alpha + \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q}$.

ARMA(p,q)

- Sample ARMA(p,q) paths: *i*) ARMA(1,1), *ii*) ARMA(2,1), *iii*) ARMA(2,2), *iv*) ARMA(3,3) using `arima.sim()`.



ARMA(p,q)

- It is **always** assumed that $\phi(B)$ and $\theta(B)$ do not have common factors
- **ARMA(1,1)** : $x_t = 0.5x_{t-1} + w_t - 0.5w_{t-1}$. It looks like an ARMA(1,1) but rewriting,

$$(1 - 0.5B)x_t = (1 - 0.5B)w_t \implies x_t = w_t.$$

- Model above is over-parameterized (with redundant parameters)
- Let $x_t = 0.4x_{t-1} + 0.45x_{t-2} + w_t + w_{t-1} + 0.25w_{t-2}$. ARMA(2,2)? Rewriting,

$$\begin{aligned} (1 - 0.4B - 0.45B^2)x_t &= (1 + B + 0.25B^2)w_t \\ \implies (1 + 0.5B)(1 - 0.9B)x_t &= (1 + 0.5B)^2 w_t. \\ \implies (1 - 0.9B)x_t &= (1 + 0.5B)w_t. \end{aligned}$$

- It is ARMA(1,1) indeed which can also be written as

$$x_t = 0.9x_{t-1} + 0.5w_{t-1} + w_t.$$

ARMA(p,q)

- An ARMA (p,q) process x_t that satisfies

$$\phi(B)x_t = \theta(B)w_t$$

  is said to be **causal** if

$$x_t = \sum_{j=0}^{\infty} \psi_j w_{t-j} = \psi(B)w_t$$

  where $\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$.

- An **ARMA**(p, q) **is causal** if and only if all of the roots of its AR polynomial $\phi(B)$ are outside the unit circle: $\phi(B) = 0$ only if $|B| > 1$.

ARMA(p,q)

- An ARMA (p,q) x_t process which satisfies

$$\phi(B)x_t = \theta(B)w_t$$

  is said to be **invertible** if

$$\pi(B)x_t = w_t$$

  where $\pi(B) = \sum_{j=0}^{\infty} \pi_j B^j$.

- An **ARMA(p,q) is invertible** if and only if all of the roots of its MA polynomial $\theta(B)$ are outside the unit circle: $\theta(B) = 0$ only if $|B| > 1$.

ARMA(p,q)

- Recall ARMA(1,1)

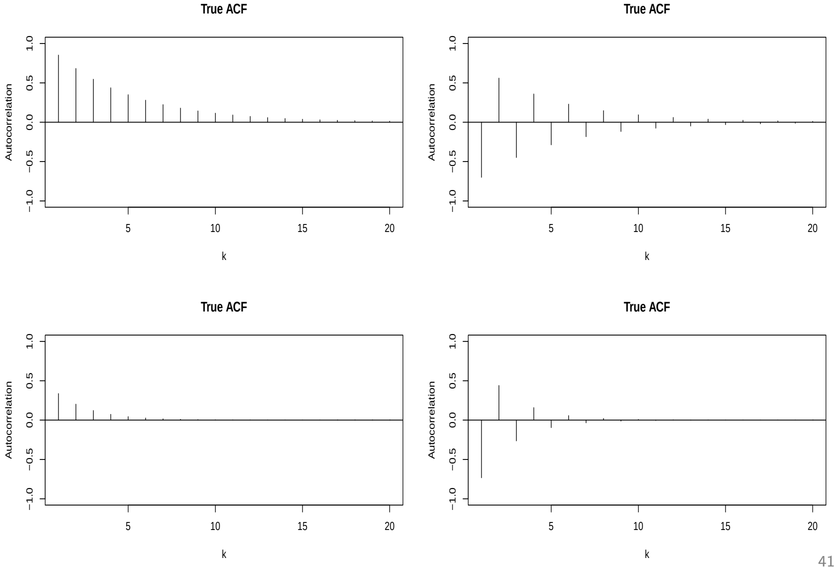
$$x_t = 0.9x_{t-1} + 0.5w_{t-1} + w_t.$$

- It is *causal* as $\phi(B) = (1 - 0.9B) = 0 \implies B = 10/9 > 1.$
- The above *causal* or MA(∞) representation only exists if $|\phi| < 1$ or the root of $(1 - \phi B) = 0$ exceeds unity in modulus.
- It is also *invertible* as

$$\theta(B) = (1 + 0.5B) = 0 \implies |B| = |-2| > 1.$$

ARMA(p,q)

- Below are true ACF's of ARMA(1,1) using $(\phi, \theta) = i)(0.8, 0.2), ii)(-0.8, .2), iii)(0.6, -0.3), iv)(-0.6, -0.3)$.



ACF of ARMA(p,q)

- The general ARMA(1,1) can be defined as

$$x_t = \phi x_{t-1} + w_t + \theta w_{t-1}. \tag{2}$$

- It can be shown that

$$\rho_h \quad = \quad \frac{(1+\theta\phi)(\phi+\theta)}{1+2\theta\phi+\theta^2}\phi^{h-1} \qquad \text{for } h \geq 1. \tag{3}$$

- The ACF of an ARMA(1,1) decays exponentially with the damping factor ϕ as h increases.
- The pattern is not that different from an AR(1). It is unlikely to tell the difference between ARMA$(1,1)$ and AR(1).
- In general, the true/theoretical/population ACF of ARMA(p,q) **tails/tapers/dies off.**

Summary

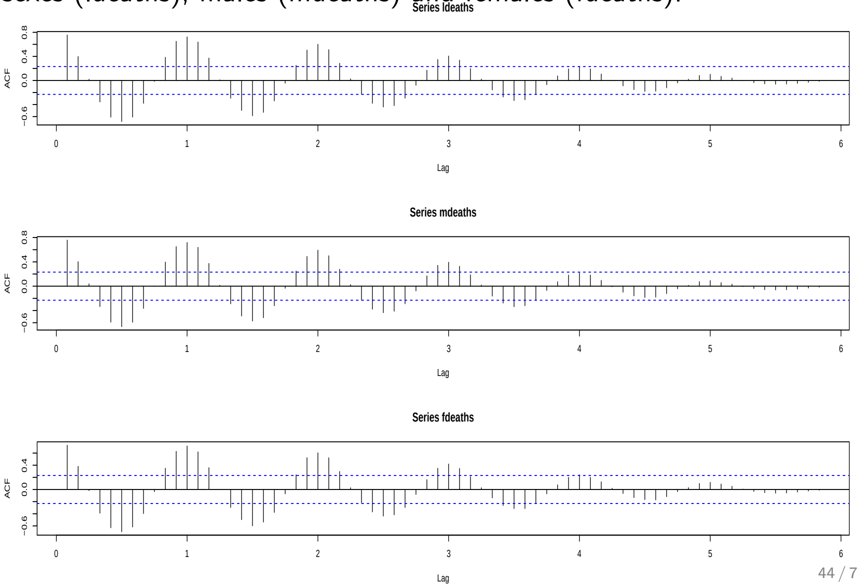
- The true/theoretical/population ACF of MA(q) **cuts/drops off after lag** q**.**

- The true/theoretical/population ACF of AR(p) **tails/tapers/dies off**.

- The true/theoretical/population ACF of ARMA(p,q) **tails/tapers/dies off.**

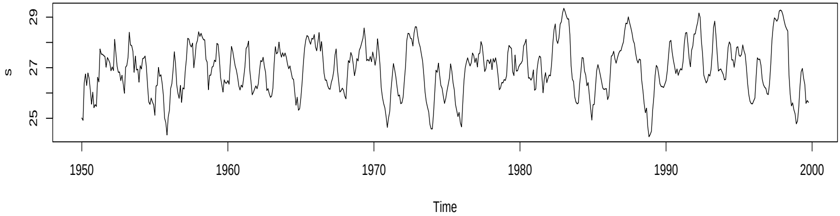
ARMA(p,q) Application

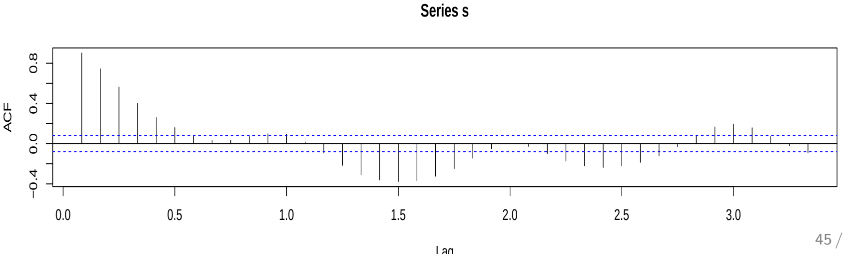
- Monthly deaths from lung diseases in the UK, 1974-1979, both sexes (ldeaths), males (mdeaths) and females (fdeaths).



ARMA(p,q) Application

- Sea Surface Temperature (SST) Nino 3.4 Indices: These are 598 measurements in degrees Celsius. The Nino 3.4 Region is bounded by 120W-170W and 5S-5N.





Partial Autocorrelation Function (PACF)

- Consider a mean-zero stationary time series. Let the predictors (as a linear function of the $h-1$ variables)

$$\hat{x}_t = \beta_1 x_{t-1} + \beta_2 x_{t-2} + \cdots + \beta_{h-1} x_{t-(h-1)}$$

and (using the same regressors)

$$\hat{x}_{t-h} = \beta_1 x_{t-(h-1)} + \beta_2 x_{t-(h-2)} + \cdots + \beta_{h-1} x_{t-1}.$$

- **Def'n.**The **partial autocorrelation function (PACF)** of a stationary process x_t is

$$\phi_{11} = corr(x_{t-1}, x_t) = \rho(1), \qquad h = 1,$$

and

$$\phi_{hh} = corr(x_{t-h} - \hat{x}_{t-h}, x_t - \hat{x}_t), \qquad h \geq 2.$$

Partial Autocorrelation Function (PACF)

- **AR(1):** $x_t = \phi x_{t-1} + w_t, \qquad \phi_{11} = \rho(1) = \phi,$

$$\phi_{22} = \frac{\rho(2) - \rho^2(1)}{1 - \rho^2(1)} = \frac{\phi^2 - \phi^2}{1 - \phi^2} = \phi_{33} = \cdots = 0.$$

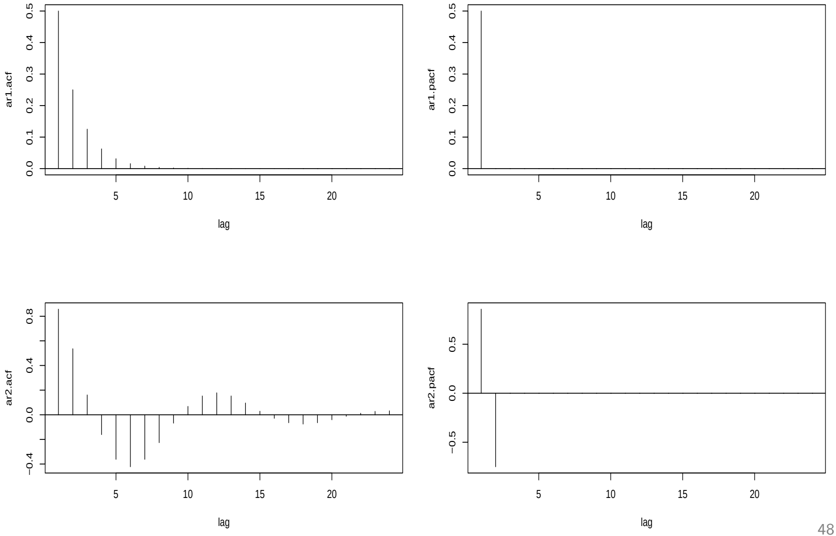
- **AR(2):** $x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + w_t, \qquad \phi_{11} = \rho(1) = \frac{\phi_1}{1-\phi_2} \neq 0,$

$$\phi_{22} = \phi_2 \neq 0, \qquad \text{and} \qquad \phi_{33} = 0. \qquad \text{(see pp 105, SaS)}$$

- **IMPORTANT:** If $\{x_t\}$ is a stationary **AR(p)** then $\phi_{pp} = \phi_p \neq 0$ and $\phi_{hh} = 0$ for all $h > p$.
- Use `ARMAacf (ar=coef, ma=0, 24, pacf=TRUE)` in R.

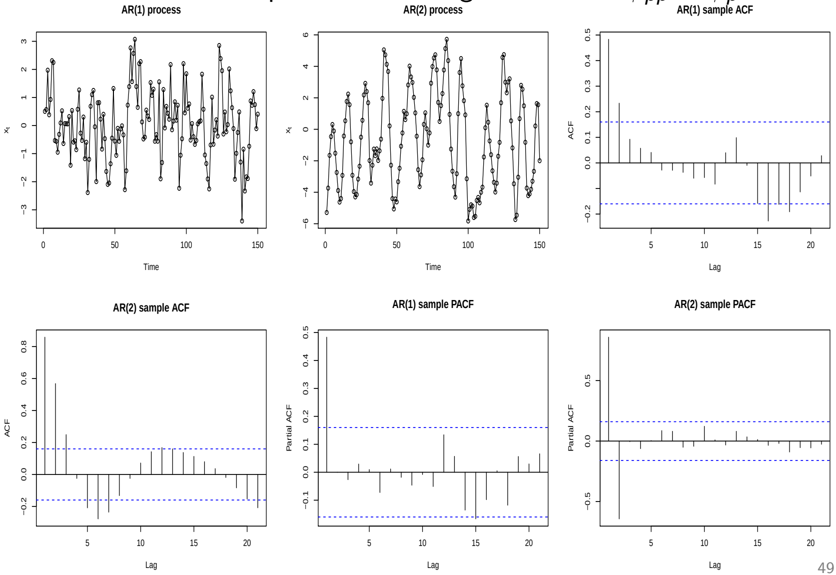
Partial Autocorrelation Function (PACF)

- *E.g.* Let us look at the TRUE PACF's of the following models using R: (i) $x_t = 0.5x_{t-1} + w_t$; (ii) $x_t = 1.5x_{t-1} - 0.75x_{t-2} + w_t$.



Partial Autocorrelation Function (PACF)

- We also generate $n = 150$ observations from the processes above and calculate the sample PACF's using R. Observe $\widehat{\phi_{pp}} \sim \phi_p$.



Partial Autocorrelation Function (PACF)

- Invertible **MA(1)**: $x_t = \theta w_t + w_{t-1}$. The PACF is

$$\phi_{hh} = \frac{(-\theta)^h(1-\theta^2)}{1-\theta^{2(h+1)}}, \qquad h \geq 1.$$

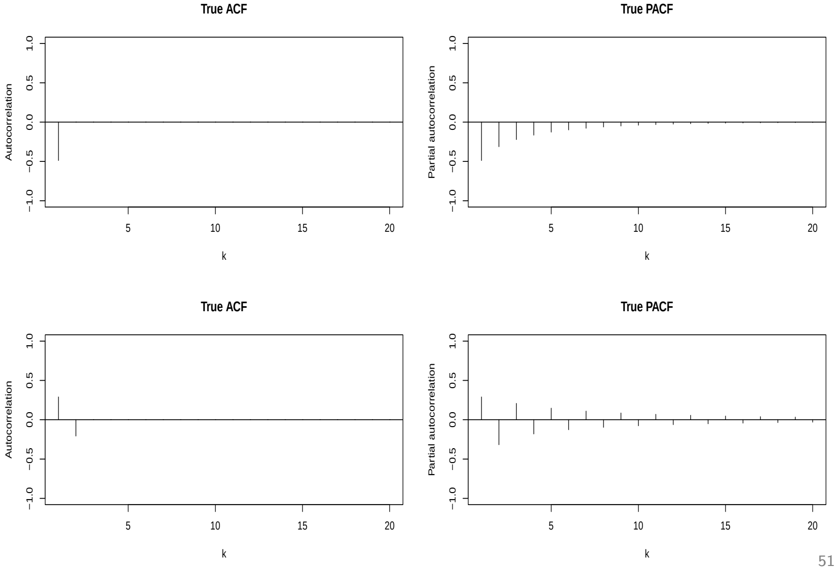
- Clearly,

$$\lim_{h\to\infty} \phi_{hh} = 0 \qquad \text{as} \qquad |\theta| < 1.$$

- The PACF of MA(1) decays to zero as the lag h increases just like the ACF of AR(1).
- Note that Invertibilty of MA(1) suggests that an AR(∞) exists, hence the PACF will never cut off like an AR(p).

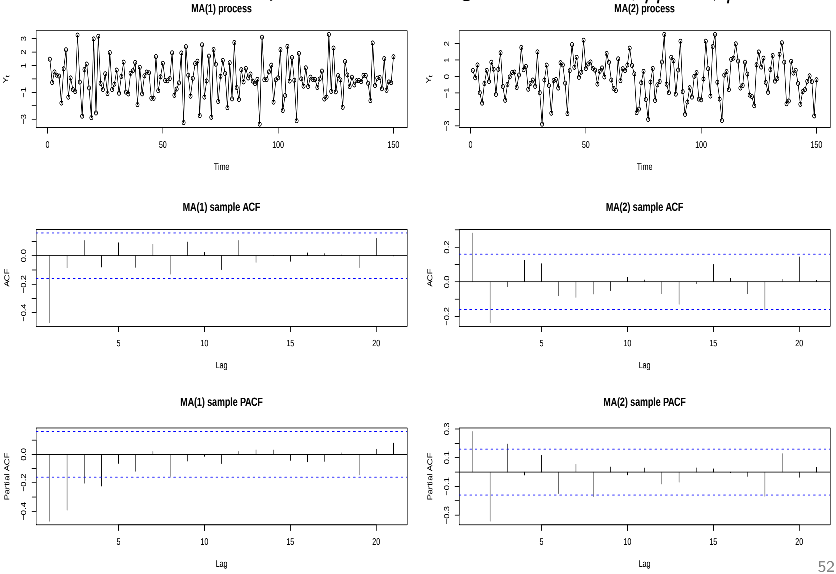
Partial Autocorrelation Function (PACF)

- *E.g.* Let us look at the PACF's of the following models using R:
  (i) $x_t = w_t - 0.8w_{t-1}$; (ii) $x_t = w_t + 0.6w_{t-1} - 0.3w_{t-2}$.



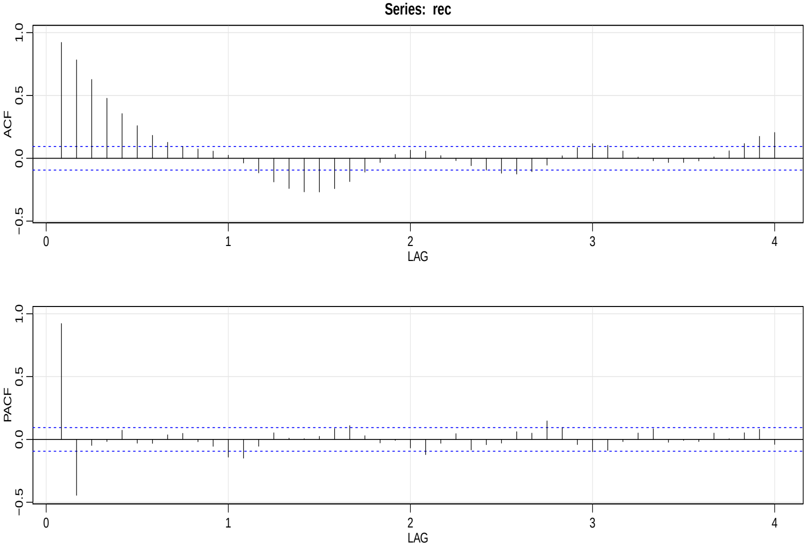
Partial Autocorrelation Function (PACF)

- We also generate $n = 150$ observations from the processes above and calculate the sample PACF's using R. Observe $\widehat{\phi_{pp}} \sim \phi_p$.



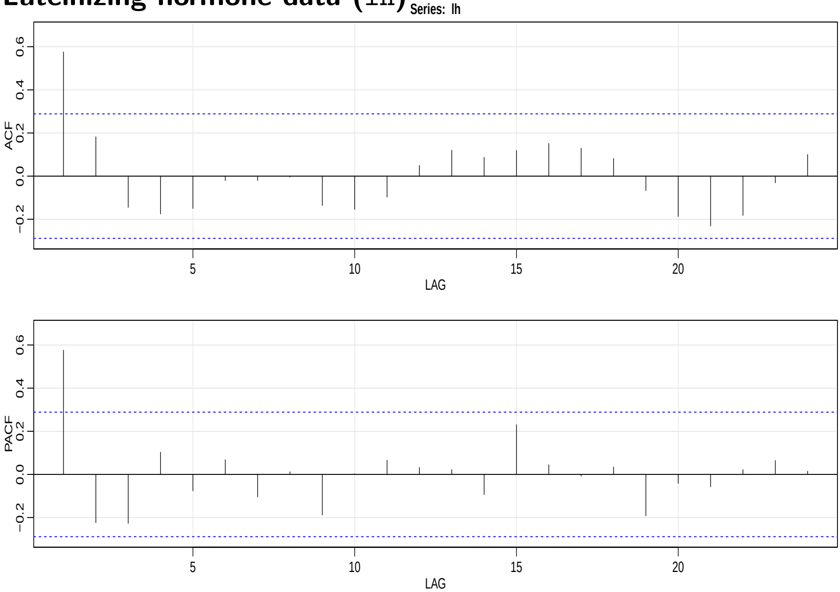
Partial Autocorrelation Function (PACF)

- **El Niño's Recruitment data**: Use `acf2()` from `astsa` package.



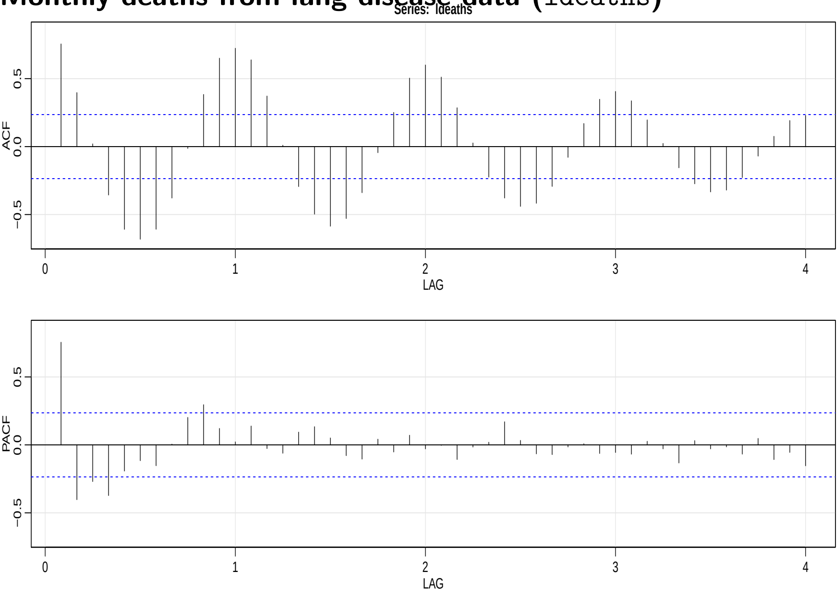
Partial Autocorrelation Function (PACF)

- **Luteinizing hormone data (`lh`)**



Partial Autocorrelation Function (PACF)

- **Monthly deaths from lung disease data (`ldeaths`)**



Partial Autocorrelation Function (PACF)

- How about an ARMA model?
- **Summary:**

	ACF	PACF
AR(p)	tails off	cuts off after lag p
MA(q)	cuts off after lag q	tails off
ARMA(p,q)	tails off	tails off

- Use `acf2` from `astsa` to produce both ACF and PACF plots. Use `ARMAacf` for the true P/ACF.

Lab for Today

- Let $x_t = 0.6x_{t-1} - 0.7x_{t-2} + w_t + w_{t-1} + 0.2w_{t-2}$. Do the following:

  i. Determine if it is invertible or stationary/causal;
  ii. Generate a sample path using $n = 100$ observations, and produce the sample ACF and sample PACF;
  iii. Produce the true ACF and the true PACF and compare with the sample ACF and sample PACF, respectively.

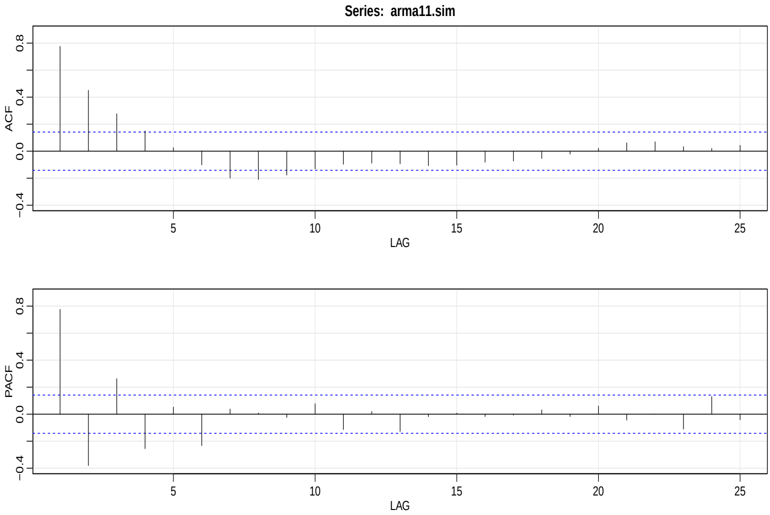
Extended ACF

- How do we choose p and q for ARMA(p, q)?
- The ARMA(p,q) will have a **triangle of zeros**, with the upper left-hand vertex correspond to the orders p and q.
- *E.g.* Ideal sample EACF of ARMA(1,1) in R.

AR\\MA	0	1	2	3	4	5	6	7
0	x	x	x	x	x	x	x	x
1	x	**o**	o	o	o	o	o	o
2	x	x	o	o	o	o	o	o
3	x	x	x	o	o	o	o	o
4	x	x	x	x	o	o	o	o
5	x	x	x	x	x	o	o	o
6	x	x	x	x	x	x	o	o
7	x	x	x	x	x	x	x	*o*

Extended ACF

- *E.g.* The simulated ACF and PACF of a simulated ARMA(1,1) with $\phi = 0.6$ and $\theta = 0.8$.



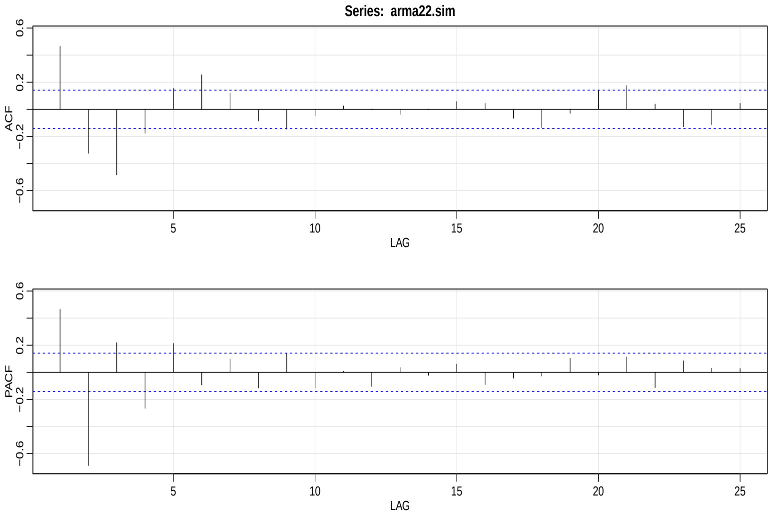
Extended ACF

- *E.g.* The limits of the sample EACF $\widehat{\rho}_j^{(i)}$ of a simulated ARMA($i = 1, j = 1$) with $\phi = 0.6$ and $\theta = 0.8$. Use `eacf` of TSA.

AR\\MA	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	x	x	x	x	o	o	x	x	x	o	o	o	o	o
1	x	**o**	o	o	o	o	x	o	o	o	o	o	o	o
2	x	x	o	o	o	o	x	o	o	o	o	o	o	o
3	x	x	x	o	o	o	o	o	o	o	o	o	o	o
4	x	x	x	o	o	x	o	o	o	o	o	o	o	o
5	x	x	x	o	o	x	x	o	o	o	o	o	o	o
6	x	o	o	x	o	o	o	o	o	o	o	o	o	o
7	x	o	o	o	o	o	o	o	o	o	o	o	o	o

Extended ACF

- *E.g.* The simulated ACF and PACF of a simulated ARMA(2,2) with $\phi = c(0.5, -0.5)$ and $\boldsymbol{\theta} = c(0.8, -0.2)$.



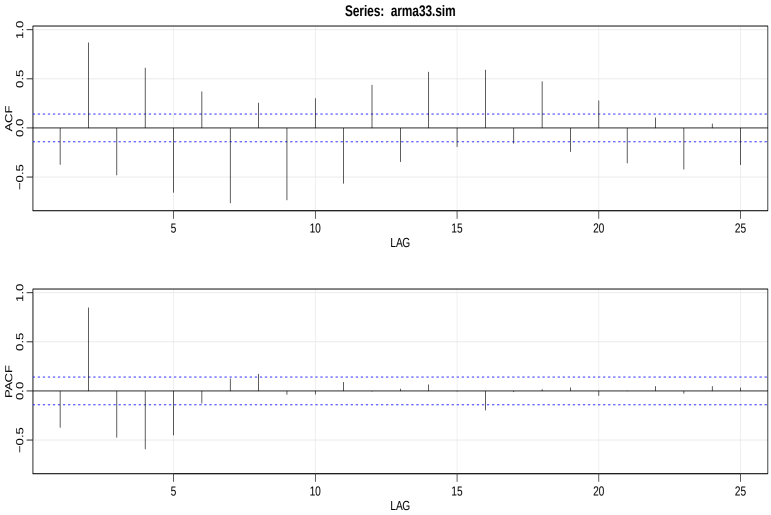
Extended ACF

- *E.g.* The limits of the sample EACF $\widehat{\rho}_j^{(i)}$ of a simulated ARMA($i = 2, j = 2$) with with $\phi = c(0.5, -0.5)$ and $\boldsymbol{\theta} = c(0.8, -0.2)$.

AR\\MA	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	x	x	x	x	x	x	o	o	x	o	o	o	o	o
1	x	x	x	x	x	x	o	o	x	o	o	o	o	o
2	x	x	**o**	o	o	o	o	o	o	o	o	o	o	o
3	x	x	x	o	o	o	o	o	o	o	o	o	o	o
4	x	x	x	o	o	o	o	o	o	o	o	o	o	o
5	x	x	x	o	o	o	o	o	x	o	o	o	o	o
6	x	o	o	x	o	o	o	o	x	o	o	o	o	o
7	x	o	o	o	x	o	o	o	o	o	o	o	o	o

Extended ACF

- *E.g.* The simulated ACF and PACF of a simulated ARMA(3,3) with $\phi = c(0.8, 0.8, -0.9)$ and $\boldsymbol{\theta} = c(-0.9, 0.8, -0.2)$.



Extended ACF

- *E.g.* The limits of the sample EACF $\widehat{\rho}_j^{(i)}$ of a simulated ARMA($i = 3, j = 3$) with with $\phi = c(0.8, 0.8, -0.9)$ and $\boldsymbol{\theta} = c(-0.9, 0.8, -0.2)$.

AR\\MA	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	o	x	x	x	x	o	x	o	x	o	o	o	o	o
1	o	x	x	o	x	o	x	o	x	o	o	o	o	o
2	x	x	x	o	x	x	x	x	x	x	o	o	x	o
3	x	x	x	**o**	o	o	o	o	o	o	o	o	o	o
4	x	x	**o**	o	o	o	o	o	o	o	o	o	x	o
5	x	x	o	o	o	o	o	o	o	o	o	o	x	o
6	x	x	x	o	o	o	o	o	o	o	o	o	o	o
7	x	x	o	o	o	x	o	o	o	o	o	o	x	o

auto.arima()

```
> arma33.sim <- arima.sim(list(order = c(3,0,3), ar = c(0.
> auto.arima(arma33.sim)
Series: arma33.sim
ARIMA(3,0,3) with zero mean

Coefficients:
         ar1     ar2      ar3      ma1    ma2      ma3
      0.8171  0.7823  -0.8958  -1.0431  0.829  -0.2420
s.e.  0.0370  0.0348   0.0334   0.0763  0.089   0.0859

sigma^2 estimated as 1.187:  log likelihood=-301.42
AIC=616.84   AICc=617.42   BIC=639.93
```

- Write down the model. See R for the complete output.

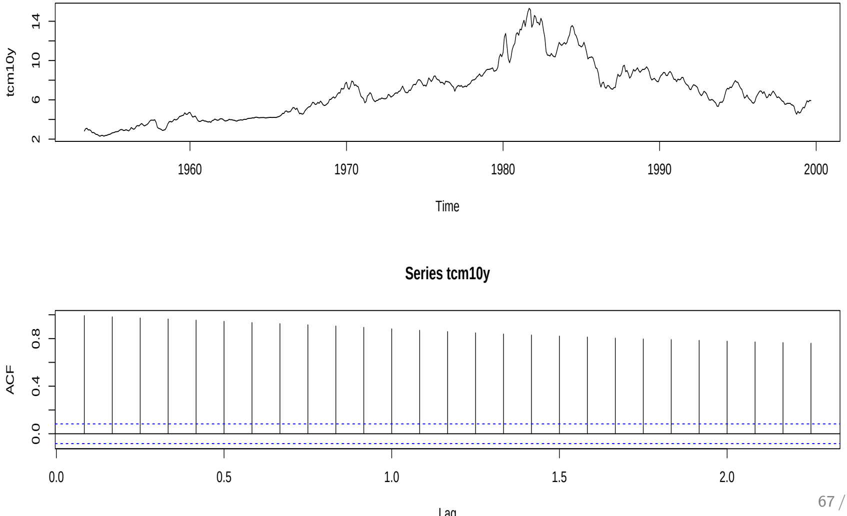
More on auto.arima()

```
> arma33.sim <- arima.sim(list(order = c(3,0,3), ar = c(0.
> auto.arima(arma33.sim)
Series: arma33.sim
ARIMA(5,0,4) with zero mean

Coefficients:
          ar1     ar2     ar3      ar4      ar5     ma1
      -0.7384  1.2000  0.9189  -0.7575  -0.7255  0.6735  0
s.e.   0.1371  0.0885  0.1808   0.0880   0.1018  0.1598  0

sigma^2 estimated as 1.213:  log likelihood=-302.86
AIC=625.72   AICc=626.88   BIC=658.7
```

- Write down the model. See R for the complete output.

ARMA(p,q) Application `diff(tcm10y)`

- Treasury Securities: This data set contains monthly 1 year, 3 year, 5 year, and 10 year yields on treasury securities at constant, fixed maturity from 1953-1999.



ARMA(p,q) Application to `diff(tcm10y)`

- Orders $(p, q) = (0,2), (2,0), (1,0), (0,1), (1,1)$

AR\MA	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	x	x	o	o	x	o	x	o	o	o	x	o	o	o
1	x	x	o	o	o	o	o	x	o	o	o	o	o	o
2	x	o	o	o	o	o	o	x	o	o	x	x	o	o
3	x	o	x	o	o	o	o	o	o	o	x	o	o	o
4	x	x	x	x	o	o	o	o	o	o	o	o	o	o
5	x	x	x	o	x	o	o	o	o	o	o	o	o	o
6	o	x	x	o	o	o	x	o	o	o	o	o	o	o
7	o	x	o	x	o	x	x	o	o	o	o	o	o	o

Box-Jenkins Method

1) Determine the theoretical ACF and PACF for known classes of time series models (e.g., ARMA(p,q)). Use the sample ACF/PACF to match the data to a possible known model.

2) Estimate parameters using a method appropriate to the chosen model and assess the fit, primarily by studying the residuals. The residuals should then "look like" white noise.

3) If the fit is inadequate go back to steps 1 and 2. Otherwise, use the "final" model to forecast.

ARMA(p,q) P/ACF

	ACF	PACF
AR(p)	tails off	cuts off after lag p
MA(q)	cuts off after lag q	tails off
ARMA(p,q)	tails off	tails off

Lab for Today

- Consider the five-year treasury security data `tcm5y`. Identify the possible orders (p, q) for ARMA(p, q).